Topological Gauge Theory,
Cartan Geometry,
and Gravity

by

Derek Keith Wise

B.S. Physics (Abilene Christian University, Texas) 1998
B.A. Mathematics (Abilene Christian University, Texas) 1998
M.S. Mathematics (University of California, Riverside) 2004

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Mathematics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, RIVERSIDE

Committee in charge:
Dr. John C. Baez, Chairperson
Dr. Michel L. Lapidus
Dr. Stefano Vidussi

June 2007

Topological Gauge Theory,
Cartan Geometry,
and Gravity

The dissertation of Derek Keith Wise is approved:

Chair Date

Date

Date

University of California, Riverside

June 2007

Acknowledgments

First, I thank John Baez for his exceptional guidance, mentorship, and patience. I've learned much from John, and in many ways. It is hard to imagine having a better advisor, or getting along better with one's advisor.

I also wish to thank Michel Lapidus and Stefano Vidussi for serving on my thesis committee.

During this work, I was involved in two collaborations of particular relevance to this work. John Baez and Alissa Crans collaborated with me on the exotic statistics project [12] which forms an important part of this thesis. John and I were also involved in a project with Aristide Baratin, Laurent Freidel, and Jeffrey Morton [10]. I am sincerely grateful to each of these collaborators.

I am indebted to Xiao–Song Lin, first because his work on the loop braid group was the impetus for our work on exotic statistics. I also owe a him a great debt of gratitude for all that I learned from him as a student, and for his inspiring example. These are unfortunately debts that can no longer be repaid. I miss you, Xiao–Song.

In the course my graduate studies, I also benefitted greatly from numerous discussions with many others, especially Jeff Morton, Jim Dolan, Josh Willis, Sam Nelson, Alissa Crans, Miguel Carrion Álvarez, Toby Bartels, Aristide Baratin, Prasad Senesi, Blake Winter, Mike Stay, Alex Hoffnung, and Danny Stevenson.

Abstract

Topological Gauge Theory,
Cartan Geometry,
and Gravity

by

Derek Keith Wise
Doctor of Philosophy in Mathematics

University of California, Riverside

Dr. John C. Baez, Chair

We investigate the geometry of general relativity, and of related topological gauge theories, using Cartan geometry. Cartan geometry—an 'infinitesimal' version of Klein's *Erlanger Programm*—allows us to view physical spacetime as tangentially approximated by a homogeneous 'model spacetime', such as de Sitter or anti de Sitter. This idea leads to a common geometric foundation for 3d Chern–Simons gravity, as studied by Witten, and 4d MacDowell–Mansouri gravity. We describe certain topological gauge theories, including BF theory—a natural extension of 3d gravity to arbitrary dimensions—as 'Cartan gauge theories' in which the gauge field is replace by a 'Cartan connection' modeled on some Klein geometry G/H. Cartan-type BF theory has solutions that say spacetime is locally isometric to the G/H itself; in this case Cartan geometry reduces to the theory of 'geometric structures'. This leads to generalizations of 3d gravity based on other 3d Klein geometries, including those in Thurston's classification of 3d Riemannian model geometries.

For BF theory in n-dimensional spacetime, we also describe codimension-2 'branes' as topological defects. These branes—particles in 3d spacetime, strings in 4d, and so on—are shown to be classified by conjugacy classes in the gauge group G of the theory. They also exhibit 'exotic statistics' which are neither Bose–Einstein nor Fermi–Dirac, but are governed by representations of generalizations of the braid group known as 'motion groups'. These representations come from a natural action of the motion group on the moduli space of flat G-bundles on space. We study this in particular detail in the case of strings in 4d BF

theory, where Lin has called the motion group the 'loop braid group', LB_n. This makes 4d BF theory with strings into a 'loop braided quantum field theory'.

We also use ideas from 'higher gauge theory' to study particles as topological defects in 4d BF theory, and find they are classified by adjoint orbits in the Lie algebra of the gauge group. Including both particles and strings in 4d BF theory leads to interesting effects, such as exotic particle/string statistics and a duality between Bohm–Aharonov effects for particles and strings.

Contents

Appendices

Chapter 1

Introduction

One long standing theme in theoretical work on quantum gravity has been to exploit relationships between general relativity and gauge theory. The reason is clear: we know how to quantize the gauge theories of particle physics. But ordinary gauge theories are very different from gravity in an essential way. A typical gauge theory, such as Yang–Mills theory, uses the geometry of spacetime, as encoded in the metric, in its definition. Gravity, on the other hand, is a kind of 'gauge theory' that *determines* the spacetime geometry itself.

Topological gauge theories represent a sort of compromise. On one hand, such theories are formulated in essentially the same language as, say, Yang–Mills theory, and one can try quantizing them using similar methods. On the other hand, they are more similar to gravity in that they do not require any fixed background structure. While being simpler than general relativity, they thus share with it the many of the conceptual issues related to quantizing a generally covariant theory.

What makes a gauge theory 'topological' is a somewhat subjective matter. One possibility is that it should be describable using the functorial definition of topological field theory, or some slight generalization. But this is too strong a requirement to include some of the most interesting examples. A more practical requirement is that all solutions of a 'topological gauge theory' should be locally the same up to gauge transformations. Such theories are more interesting when the topology of spacetime is more interesting, since solutions that look completely trivial on a local scale may yet have interesting global properties. Intuitively, it is natural that topological gauge theories of this sort should be related to topological invariants. This is indeed true: the interplay between pure topology and gauge theories has been enormously fruitful, particularly in work on 3-manifold invariants and knot theory, but more recently also for 4-manifolds.

But one should not be misled into thinking the difference between topological gauge theories and general relativity is too much like the difference between topology and geometry. So called 'topological' theories can actually have a rich *geometric* content. Though they do not have local degrees of freedom, as general relativity has, certain topological gauge theories have field equations that determine the geometry of spacetime in much the same way as in general relativity. The essential idea is to interpret the fiber bundle language of gauge theory not as describing 'internal' degrees of freedom as it does in particle physics, but as describing degrees of freedom in spacetime geometry. This idea in fact has its roots in the work of Élie Cartan, who had a more 'concrete' view of the role of connections and bundles. This thesis is partly a story about geometry, and particularly how Cartan's perspective lets us see the geometry of topological gauge theories transform into the geometry of general relativity. The hope is that a deeper understanding of the geometric content of topological gauge theory will provide insight into the geometry of general relativity itself, and perhaps ultimately its quantization.

In fact, certain topological gauge theories are more than just *analogous* to general relativity. As we shall describe shortly, full-fledged general relativity can be obtained from certain topological theories either by imposing constraints or by symmetry breaking. But perhaps the strongest case, at least initially, for trying to relate general relativity to topological theories is the following fact. In $(2+1)$-dimensional spacetime, general relativity is a topological gauge theory. In fact, 3d general relativity is a special case of one of the most important topological gauge theories for our purposes—a theory called 'BF theory'—so we begin with a description of that.

3d gravity and BF theory

The essential reason that general relativity is topological in 3 dimensions is simply that there are not enough dimensions to admit the variety of curvature possible in 4 or more dimensions. To be precise, in the absense of matter, Einstein's equations imply that the Einstein tensor must vanish. But in 3 dimensions the Einstein tensor vanishes if and only if the full Riemann curvature tensor does, so the field equations imply that 3d spacetime is *flat*. This immediately suggests 3d general relativity is 'topological' in the sense we have described, since flat Levi–Civita connections are all locally the same up to gauge transformations. In fact, 3d general relativity is a special case of 'BF theory', which we now describe more generally.

In n-dimensional spacetime, BF theory with gauge group G—assuming a trivial G-bundle for simplicity—involves two fields: a G-connection A, and a $\mathfrak{g}$-valued $(n-2)$-form E. In the absence of matter, the Lagrangian is simply

$$L = \frac{1}{\kappa}\mathrm{tr}\,(E \wedge F)$$

Here κ plays the role of Newton's constant in the case of 3d gravity, and $F = dA + A \wedge A$ is the curvature of A. The resulting equations of motion:

$$F = 0, \qquad d_A E = 0,$$

simply say that the connection A is flat, and E is covariantly constant—its covariant exterior derivative $d_A E$ vanishes. All flat connections are locally the same up to gauge transformations. In the global setting, covariantly constant E fields are not all related by gauge transformations of the usual sort. However the BF Lagrangian has an additional gauge symmetry given locally by

$$A \mapsto A \qquad E \mapsto E + d_A \eta$$

for any $\mathfrak{g}$-valued $(n-3)$-form η, and all E fields are then locally gauge equivalent in the broader sense. [7]

3d gravity is essentially a BF theory with the 3d Lorentz group $\mathrm{SO}(2,1)$ as gauge group, since this describes the symmetries of a local coordinate frame. In 4 dimensions, general relativity is of course not a BF theory: unlike the 3d case, there is no equation in 4d general relativity that says the Riemann curvature of spacetime vanishes. But 4d BF theory is related to 4d general relativity in important ways. Indeed, the Lagrangian for 4d general relativity may be written as

$$L = \frac{1}{\kappa} tr(e \wedge e \wedge F)$$

where e is a Lie algebra valued 1-form of an appropriate sort. 4d general relativity may thus be viewed as a BF theory subject to the constraint that $E = e \wedge e$ for some choice of e.

In general, since BF theory involves a *flat* connection on a fiber bundle, it is related to 'flat' spacetime geometries—but in a generalized sense where 'flat' really means it looks just locally just like the fiber of a certain bundle of homogeneous spaces. The intuitive idea is actually best to understand in the more general context where solutions are not necessarily 'flat'—namely general relativity itself.

Gravity and Cartan geometry

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 [65]. 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 [45, 46, 74, 94] to background-free quantum gravity [43, 40, 89].

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* [27, 28]. 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' [30], introduced by Cartan's student Charles Ehresmann [33].

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 [86]. 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.

MacDowell–Mansouri gravity

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

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

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

$$\mathfrak{so}(4,1) \cong \mathfrak{so}(3,1) \oplus \mathbb{R}^{3,1}, \tag{1.1}$$

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:

$$S_{\mathrm{MM}} = \frac{-3}{2G\Lambda} \int \mathrm{tr}\,(F \wedge \star \widehat{F}) \tag{1.2}$$

Here $\widehat{F}$ 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 (1.1) 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]

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

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

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

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.

Recently, the basic MacDowell–Mansouri technique has been used with a different action [43, 87, 89], based on BF theory. This work has in turn been applied already in a variety of ways, from cosmology [2] to particle physics [64]. The setup for this theory is much like that of the original MacDowell–Mansouri theory, but in addition to the 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]:

$$S = \int \mathrm{tr}\left(B \wedge F - \frac{G\Lambda}{6} B \wedge \star\widehat{B}\right). \qquad (1.3)$$

We give a treatment of both the original MacDowell–Mansouri action and the Freidel–Starodubtsev reformulation in Section 11.4, 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 $\mathrm{SO}(4,1)$ connection into an $\mathrm{SO}(3,1)$ connection and coframe field? For this it is easiest to first consider a lower-dimensional example, involving $\mathrm{SO}(3)$ and $\mathrm{SO}(2)$. An oriented 2d Riemannian manifold is often thought of in terms of an $\mathrm{SO}(2)$ connection since, in the tangent bundle, parallel

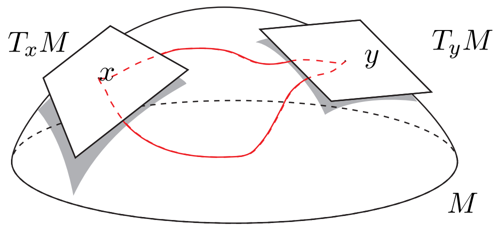
---

case indices running from 0 to 3:

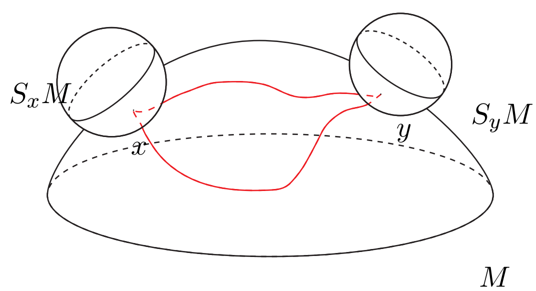
$$I, J, K, \ldots \in \{0, 1, 2, 3, 4\}$$
$$i, j, k, \ldots \in \{0, 1, 2, 3\}.$$

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

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

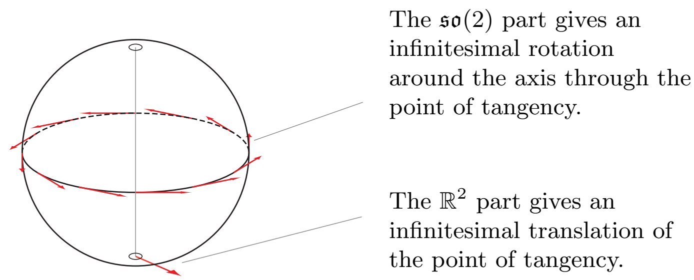
In our example, just as in the extension from the Lorentz group to the de Sitter group, we have an orthogonal splitting of the Lie algebra

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

given in terms of matrix components by

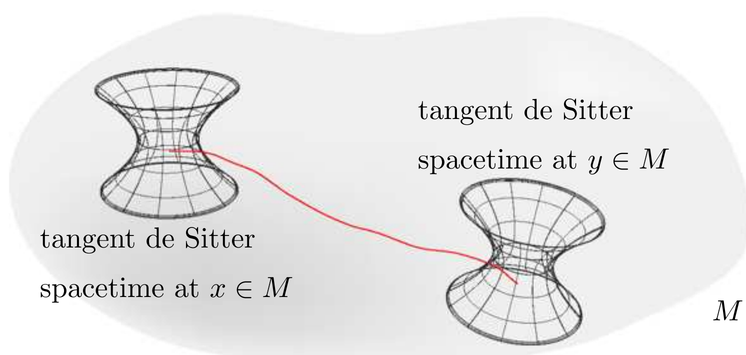
$$\begin{bmatrix} 0 & u & a \\ -u & 0 & b \\ -a & -b & 0 \end{bmatrix} = \begin{bmatrix} 0 & u & \\ -u & 0 & \\ & & 0 \end{bmatrix} + \begin{bmatrix} & & a \\ & & b \\ -a & -b & \end{bmatrix}.$$

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 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 $\mathrm{SO}(4,1)$ 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 [86], Cartan geometry is a common generalization of Riemannian and Klein geometries. The following diagram is an adaptation of one of Sharpe's:

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

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.

Geometric structures and topological gauge theories

We shall describe topological gauge theories as theories involving a Cartan connection, just as in MacDowell–Mansouri gravity. But when a Cartan connection is ‘flat’ there is a great simplification. In this case, the ‘rolling’ described above is essentially trivial, since the spacetime is locally isometric to the Kleinian model spacetime G/H. Cartan geometry then reduces to the theory of ‘geometric structures’ on manifolds, as studied by Thurston [93]. The theory of geometric structures is a major tool in modern geometric toplogy and other areas.

In the body of the thesis, we first describe the theory of geometric structures and how they are obtained as the solutions of topological field theories. Later, when we describe Cartan geometry in general, this leads us to see how 4d general relativity with cosmological constant and 4d BF theory are related in a concrete geometric way.

For now we turn to a different subject: the inclusion of *matter* in topological gauge theory.

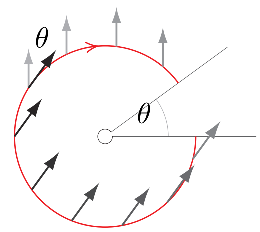
Particles in 3d BF theory and exotic statistics

In 3d BF theory, point particles can be included by considering spacetimes with curves removed: we think of these as the particles' worldlines. Away from these worldlines the BF theory equations still hold, while along the worldlines A becomes singular. To understand the description of matter as topological defects in BF theory, the important point is that solutions of BF theory give flat connections on *space*.

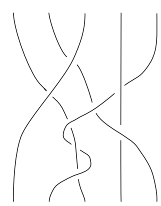
The behavior of a collection of identical particles when they are exchanged goes by the name of 'statistics'. Traditionally, statistics was described using representations of the symmetric group. However, it is well known that in 3d spacetime, 'exotic' statistics are possible, in which the process of exchanging identical particles is described by a representation of the braid group. For example, exchanging two 'abelian anyons' multiplies their wavefunction by a phase, which need not be 1 as it is for bosons, nor -1 as for fermions. This possibility has been investigated in experiments on the fractional quantum Hall effect [23]. Now researchers have begun the search for 'nonabelian anyons', whose statistics are described by more complicated representations of the braid group [20]. Plans are already afoot to use these in quantum computers [37, 58].

Exotic statistics also arise naturally in the context of 3d quantum gravity. As we 'turn on gravity', letting Newton's gravitational constant κ become nonzero, ordinary quantum field theory on 3d Minkowski spacetime deforms into a theory where the Poincaré group goes over to a quantum group called the κ-Poincaré group. Moreover, if we begin with a field theory of bosons, their statistics become exotic as we turn on gravity. For a thorough treatment of these fascinating phenomena, see the papers by Freidel and collaborators [38, 39], the paper by Krasnov [62], and the many references therein.

In fact, the reason for exotic statistics in 3d quantum gravity is very simple. In 3d spacetime, Einstein's equations say that spacetime is *flat* except in regions where matter is present. A point particle at rest bends the nearby space into a cone. This cone is flat everywhere except at its tip, where there is a deficit angle proportional to the particle's mass. If we parallel transport a vector around the particle, it gets rotated by this angle θ:

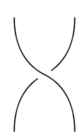


More generally, if we have n particles, space will be flat except for conical singularities at n points. If we exchange these particles by moving them around the plane, they trace out a loop in the space of n-point subsets of the plane. Their energy-momenta will change in a way that depends on this loop—but only on the *homotopy class* of this loop, because they are being parallel transported with respect to a flat connection. A homotopy class of such loops is just an n-strand braid:



So, the group B_n of n-strand braids acts on the Hilbert space of states for n identical particles. In fact, this result holds classically as well: we get an action of B_n on the configuration space for n identical particles.

The holonomy around a loop circling a worldline gives an element of the group G. A collection of n particles in the plane thus gives rise to an n-tuple of elements of G. For simplicity, consider the case $n = 2$. As we exchange two particles by rotating them around each other counterclockwise, they trace out this braid:



As we recall in Section 6, this operation acts as the following map on G^2:

$$(g_1, g_2) \mapsto (g_1 g_2 g_1^{-1}, g_1). \tag{1.4}$$

Applying this map twice does *not* give the identity, so we do not obtain an action of the symmetric group on G^2, but only an action of the braid group. In other words, the particles have exotic statistics!

In the case of 3d gravity, the singularity of the connection along a particle's worldline reflects the fact that the particle's mass creates a conical singularity in the metric. The holonomy around the worldline, an element of $G = \mathrm{SO}(2,1)$, describes the particle's *energy-momentum*. This may seem odd, since we are used to thinking of energy-momentum as a vector in Minkowski spacetime. However, in 3 dimensions Minkowski spacetime is naturally isomorphic to the Lie algebra $\mathfrak{so}(2,1)$, and we can reinterpret Lie algebra elements as group elements via the map:

$$\begin{array}{ccc} \mathfrak{so}(2,1) & \to & \mathrm{SO}(2,1) \\ p & \mapsto & \exp(\kappa p). \end{array}$$

So, we can encode the energy-momentum p of a particle in the holonomy $g = \exp(\kappa p)$ resulting from parallel transport around this particle's worldline.

Thanks to the factor of κ here, the group $\mathrm{SO}(2,1)$ effectively 'flattens out' to $\mathfrak{so}(2,1)$ in the $\kappa \to 0$ limit. For example, multiplication in the group reduces to addition in the Lie algebra plus small corrections:

$$\exp(\kappa p_1)\exp(\kappa p_2) = \exp(\kappa(p_1 + p_2) + \frac{\kappa^2}{2}[p_1, p_2] + \cdots) \tag{1.5}$$

This implies that in terms of $\mathfrak{so}(2,1)$-valued energy-momenta, the braiding in equation (1.4) is given by

$$(p_1, p_2) \mapsto (p_2 + \kappa[p_1, p_2] + \cdots\,,\, p_1)$$

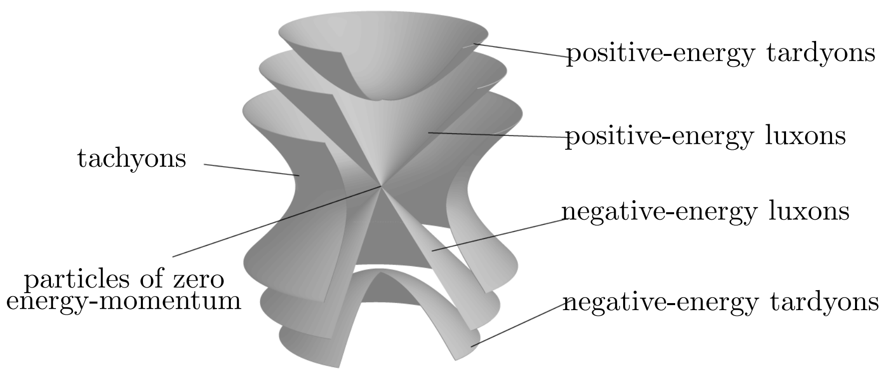
So, the exotic statistics reduce to ordinary bosonic statistics in the limit where Newton's constant goes to zero. They also reduce to bosonic statistics in the limit where the particles are at rest relative to each other, since then p_1 and p_2 become proportional and their commutator vanishes.

The corrections to the usual law for addition of energy-momenta implicit in equation (1.5) are interesting in themselves. Like the exotic statistics, these corrections become negligible in the limit $\kappa \to 0$. Under the name of 'doubly special relativity', modified laws for adding energy-momentum have already been studied by many authors. The paper by Freidel, Kowalski-Glikman and Smolin [39] gives a good account of doubly special relativity in the context of 3d quantum gravity; their paper also explains more of the history of this subject.

Quandle field theory

Besides exotic statistics and corrections to the usual rule for adding energy-momenta, there is yet another surprising consequence of the switch from vector-valued to group-valued energy-momentum as we turn on gravity in 3d physics. The classification of elementary particles changes!

In ordinary quantum field theory on Minkowski spacetime, the Lorentz group acts on the space of possible energy-momenta, and the orbits of this action correspond to different types of spin-zero particles. When spacetime is 3-dimensional, the space of energy-momenta is $\mathfrak{so}(2,1)$, and the orbits look like this:



If we write the energy-momentum as $p = (E, p_x, p_y)$ and let $p \cdot p = E^2 - p_x^2 - p_y^2$, we have six families of orbits, corresponding to six types of spin-zero particles:

1. positive-energy tardyons of mass $m > 0$: $\{p \cdot p = m^2,\ E > 0\}$,
2. negative-energy tardyons of mass $m > 0$: $\{p \cdot p = m^2,\ E < 0\}$,
3. positive-energy luxons: $\{p \cdot p = 0,\ E > 0\}$,
4. negative-energy luxons: $\{p \cdot p = 0,\ E < 0\}$,
5. tachyons of mass im for $m > 0$: $\{p \cdot p = -m^2\}$,
6. particles of vanishing energy-momentum: $\{p = 0\}$.

Given any orbit $Q \subseteq \mathfrak{so}(2,1)$, the Hilbert space for a single particle of type Q is just $L^2(Q)$.

The same philosophy applies when we turn on gravity, but now the space of energy-momenta is not the Lie algebra $\mathfrak{so}(2,1)$ but the Lorentz group itself. This acts on itself by conjugation, and the orbits are conjugacy classes. Types of spin-zero particles now correspond to conjugacy classes in the Lorentz group. Near the identity these conjugacy classes look just like orbits in the Lie algebra, so the classification of particles reduces to the above one in the limit of small energy-momenta. However, there are important differences, which show up for large energy-momenta.

Most notably, under the map

$$p \mapsto \exp(\kappa p)$$

the Lie algebra element $p = (E, 0, 0)$ is mapped to a rotation by the angle κE in the xy plane. So, the holonomy around a stationary particle of energy E is a rotation by the angle κE. This rotation does not change when we add $2\pi/\kappa$ to the particle's energy. Up to factors of order unity, this quantity $2\pi/\kappa$ is just the *Planck energy.* If we call it the Planck energy, then masses in 3d quantum gravity are defined only modulo the Planck mass.

This 'periodicity of mass' affects the classification of tardyons—that is, the most familiar sort of particles, those with timelike energy-momentum. Instead of positive-energy tardyons of arbitrary mass $m > 0$ and negative-energy tardyons of arbitrary mass $m > 0$, we just have tardyons of arbitrary mass $m \in \mathbb{R}/\frac{2\pi}{\kappa}\mathbb{Z}$.

More generally, for any Lie group G, the various allowed types of spin-zero particles in 3d BF theory with gauge group G correspond to conjugacy classes $Q \subseteq G$. Any conjugacy class is closed under the operations

$$g \rhd h = ghg^{-1}, \qquad h \lhd g = g^{-1}hg,$$

and these operations satisfy equations making Q into an algebraic structure called a 'quandle' [56], whose definition we recall in Section 6.1. The Hilbert space for a single particle of type Q is just $L^2(Q)$, defined using a measure on Q that is invariant under these operations. In an easy generalization of 3d BF theory, we can study the exotic statistics of 'particles of type Q' for any quandle Q equipped with an invariant measure. This takes advantage of the well-known relation between quandles and the braid group [35].

Exotic statistics in 4d BF theory

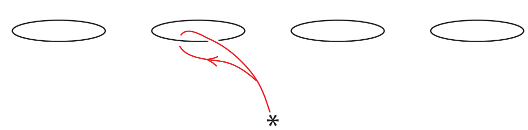
It would be wonderful to generalize all the above results to 4d gravity, but for now all we can handle is a simpler theory: 4d BF theory. This may eventually be relevant to gravity, since one can describe general relativity in 4 dimensions either as the result of *constraining* 4d BF theory with a certain gauge group, or *perturbing around* 4d BF theory with some other gauge group. The first approach goes back to Plebanski [80], and it underlies a great deal of work on spin foam models of quantum gravity [7, 76, 79], especially the Barrett–Crane model. The second approach goes back to MacDowell and Mansouri [65], and has recently been explored by Freidel and Starodubtsev [43]. With a view toward these potential applications, we focus our attention on certain relevant choices of gauge group:

$$\text{Plebanski gravity:} \qquad G = \mathrm{SO}(3,1)$$

$$\text{MacDowell–Mansouri gravity:} \quad \begin{cases} G = \mathrm{SO}(4,1) & \Lambda > 0 \\ G = \mathrm{SO}(3,2) & \Lambda < 0 \end{cases}$$

Our idea is simply to increase the dimension of everything in the previous section by 1. Thus, we consider BF theory on a 4-dimensional spacetime with the worldsheets of several 'closed strings' removed. Really these strings are just unknotted, unlinked circles in space. We call them 'closed strings' for short, even though they behave differently from the closed strings familiar in string theory: the relevant Lagrangian is different. Their dynamics has been studied in a related paper [14]. One of our purposes is to study the exotic statistics exhibited by these strings. Mathematically speaking, this will amout to studying certain representations of a higher-dimensional analogue of the braid group: the 'loop braid group'. Just as the braid group describes the topology of points moving in the plane, the loop braid group describes the topology of circles moving in $\mathbb{R}^3$. In Chapter 7 we describe this group and certain representations of it coming from the moduli space of flat bundles on $\mathbb{R}^3$ with these circles removed.

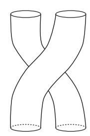
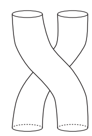
We focus on the case where the manifold representing space is $\mathbb{R}^3 - \Sigma$, where Σ is an 'n-component unlink': a collection of n unknotted unlinked circles. A flat connection on $\mathbb{R}^3 - \Sigma$ gives us a group element for each circle, namely the holonomy of some standard loop going around this circle:



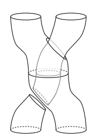
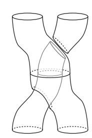
So, just as before, we obtain n-tuples of elements of G. Moreover, any way to exchange the circles in Σ gives a map from G^n to itself.

It is often said that exotic statistics are only possible when space has dimension 2 or less. However, this folklore only applies to point particles. As pointed out by Balanchandran and others [3, 18, 73, 91, 92], exotic statistics are possible for closed strings in 3-dimensional space, since there are topologically nontrivial ways to exchange unknotted unlinked circles in $\mathbb{R}^3$. The statistics of such theories are governed not by the braid group B_n, but by a larger group: the 'loop braid group' LB_n.

Using recent work of Lin [63], we show that this group is isomorphic to the 'braid permutation group' of Fenn, Rimányi and Rourke [34]. This is an apt name, because LB_n has a presentation with generators s_i that describe two strings trading places without passing through each other, just as if they were point particles:

 $=$ 

but also generators σ_i that describe one string passing through another:

 $\neq$ 

So, this group is a kind of 'hybrid' of the symmetric group and the braid group. Indeed, the elements s_i generate a copy of the symmetric group S_n in LB_n, while the elements σ_i generate a copy of the braid group B_n.

In a one-dimensional unitary representation of the loop braid group, the permutation generators s_i all act as ± 1, while the braid generators σ_i all act as an arbitrary phase $q \in \mathrm{U}(1)$. We could call particles that transform in this way 'abelian bose-anyons' and 'abelian fermi-anyons', respectively. They act like bosons or fermions when we switch them using the generators s_i, but like abelian anyons when we switch them using the generators σ_i.

BF theory gives us more interesting unitary representations of the loop braid group: whenever the group G is unimodular, we obtain a unitary representation of LB_n on $L^2(G^n)$. All the groups listed above are unimodular, so we get an interesting variety of exotic statistics for closed strings in 4d BF theory.

We can also restrict attention to a specific conjugacy class $Q \subseteq G$ and get a unitary representation of the loop braid group on $L^2(Q^n)$, as long as Q is equipped with a measure invariant under conjugation. As already mentioned, in the case of 3d gravity a choice of conjugacy class in $G = \mathrm{SO}(2,1)$ essentially amounts to choosing a specific *mass* for our point particles, which is a very natural thing to do. In the case of 4d BF theory with $G = \mathrm{SO}(3,1)$, choosing a conjugacy class essentially amounts to choosing a specific *mass density* for our closed strings.

Higher gauge theory and particles in 4d BF theory

4d BF theory is actually a topological 'higher gauge theory': it involves not just an ordinary connection, but a '2-connection', which has a 1-form part A and a 2-form part E [15]. The equations of BF theory say this 2-connection (A, E) is 'flat'. So, we get both particles and strings in 4d BF theory in a purely topological way. 'Strings' appear in 4d BF theory because integrating the 1-form A along a loop enclosing a string-shaped hole in space gives a 'Wilson loop' observable. Similarly, 'particles' appear in 4d BF theory because integrating the 2-form E over a surface enclosing a point puncture in space gives a 'Wilson surface' observable:

We shall study these particles and strings, which exhibit interesting collective behavior, such as a kind of combined particle/string exotic statistics.

Part I

Geometric Structures and Topological Gauge Theory

Chapter 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'. The importance of Klein geometry for our purposes is that it explains the geometry of the most basic conceptions of spacetime, including the spacetimes of Galilean and Einsteinian special relativity, but also of important generalizations such as de Sitter spacetime. While homogeneous geometry by itself is inadequate to describe theories with less symmetric geometry, such as general relativity, the Kleinian perspective is essential to understanding C artan geometry. so we review it here in some detail.

2.1 Klein geometry

A **homogeneous space** (G, X) is an abstract space[1] 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 $H \subset G$

[1] 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.

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:

$$H_x = gH_{x'}g^{-1}$$

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:

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

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

$$\begin{array}{c} G \\ \downarrow \\ G/H \end{array}$$

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 [86]

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

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 inner product[2] 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 Survey of homogeneous spacetimes

Let us now describe some of the standard homogeneous models of spacetime, from the Kleinian perspective. Here we give only a brief description of the relevant symmetry groups. For a more thorough treatment of many aspects of homogeneous spacetime from the perspective of symmetry groups, we refer the reader to Penrose and Rindler [78]. For introductions the de Sitter and anti de Sitter spacetimes (in 3+1 dimensions) see, for example, Hawking and Ellis [53].

[2]Unless otherwise indicated, by 'inner product' we always mean a possibly indefinite inner product. We also use the term 'metric' on a vector space interchangeably, especially when the innner product in question is naturally part of a metric on a vector bundle.

2.3.1 3d spacetimes

Let us first consider 3-dimensional (more properly, (2+1)-dimensional) homogeneous spacetimes. Though these are slightly easier to deal with than 4d spacetimes, and much easier to visualize, 3d geometry is sufficiently rich to provide an accurate view the features of homogeneous spacetime geometry. Moreover, the most important examples in 4 dimensions are immediate analogs of 3d ones we shall consider.

The three basic Lorentzian models we discuss in this section are listed below, along with the their symmetry groups G and point stabilizer subgroups H. They correspond to the maximally symmetric Lorentzian spacetimes with the three possible signs of the cosmological constant Λ.

			G	H
1.	3d de Sitter spacetime	$\Lambda > 0$	$\mathrm{SO}(3,1)$	$\mathrm{SO}(2,1)$
2.	3d Minkowski spacetime	$\Lambda = 0$	$\mathrm{ISO}_0(2,1)$	$\mathrm{SO}(2,1)$
3.	3d anti-de Sitter spacetime	$\Lambda < 0$	$\mathrm{SO}(2,2)$	$\mathrm{SO}(2,1)$

We begin with the most familiar, Minkowski spacetime.

3d Minkowski spacetime

The symmetries of 3-dimensional Minkowski spacetime are given by the 3d Poincaré group $\mathrm{ISO}(2,1) = \mathrm{SO}(2,1) \ltimes \mathbb{R}^3$, whose Lie algebra is

$$\mathfrak{iso}(2,1) = \begin{pmatrix} \mathfrak{so}(2,1) & \mathbb{R}^3 \\ 0 & 0 \end{pmatrix} = \left\{ \left(\begin{array}{cccc} 0 & -c & b & x \\ -c & 0 & a & y \\ b & -a & 0 & z \\ 0 & 0 & 0 & 0 \end{array} \right) \middle| \; a,b,c,x,y,z \in \mathbb{R} \right\}$$

It is often convenient to use the fact that the connected 2d Lorentz group $\mathrm{SO}_0(2,1)$ is double covered by $\mathrm{SL}(2,\mathbb{R})$. In other words, $\mathrm{SL}(2,\mathbb{R})$ is the group of 'spin transformations' of $(2+1)$-dimensional Minkowski vector space. Let us review how this double cover description works. The essential point is that the Lie algebra $\mathfrak{sl}(2,\mathbb{R})$ is isomorphic to Minkowski vector space $\mathbb{R}^{2,1}$. To see this, note that $\mathfrak{sl}(2,\mathbb{R})$ consists of traceless matrices, which can be written in the form

$$\begin{pmatrix} x & y+t \\ y-t & x \end{pmatrix}$$

and the Killing form $\mathrm{tr}\,(\mathrm{ad}(\cdot)\mathrm{ad}(\cdot))$ on such matrices is proportional to the Minkowski metric on the standard coordinates (t, x, y). The Lie bracket amounts to the Minkowskian analog of the cross product of vectors in $\mathbb{R}^3$.

Now the adjoint action of $\mathrm{SL}(2, \mathbb{R})$ on its Lie algebra:

$$\begin{aligned} \mathrm{Ad}\colon \mathrm{SL}(2,\mathbb{R}) \times \mathfrak{sl}(2,\mathbb{R}) &\to \mathfrak{sl}(2,\mathbb{R}) \\ (g,p) &\mapsto gpg^{-1} \end{aligned}$$

preserves the Killing form, so acts as symmetries of Minkowski vector space, that is, as Lorentz transformations. Since $\mathrm{SL}(2, \mathbb{R})$ is connected, a Lorentz transformation coming from an element of $\mathrm{SL}(2, \mathbb{R})$ in this way must live in the connected part of the Lorentz group. We thus get a map:

$$\mathrm{SL}(2,\mathbb{R}) \to \mathrm{SO}_0(2,1).$$

by writing any given transformation as a 3×3 matrix acting on the coordinates (t, x, y). Clearly this map is at least two–to–one, since $g, -g \in \mathrm{SL}(2, \mathbb{R})$ induce the same Lorentz transformation. In fact, the map is exactly two-to-one and onto. So we get an isomorphism

$$\mathrm{SO}_0(2,1) \cong \mathrm{PSL}(2,\mathbb{R}) = \mathrm{SL}(2,\mathbb{R})/\{\pm 1\}$$

Using the adjoint action of $\mathrm{SL}(2, \mathbb{R})$, we can construct the semidirect product $\mathrm{SL}(2, \mathbb{R}) \ltimes \mathfrak{sl}(2, \mathbb{R})$, which is the double cover of the connected Poincaré group for $(2 + 1)$-dimensional Minkowski spacetime.

The symmetry groups are be summarized in the following diagram:

$$\begin{tikzcd} \mathrm{SL}(2,\mathbb{R}) \arrow[r] \arrow[d, "\text{2-1}"] & \mathrm{SL}(2,\mathbb{R}) \ltimes \mathfrak{sl}(2,\mathbb{R}) \arrow[d, "\text{2-1}"] \\ \mathrm{SO}_0(2,1) \arrow[r] & \mathrm{SO}_0(2,1) \ltimes \mathbb{R}^3 \end{tikzcd}$$

As a Klein geometry, we may regard 3d Minkowski spacetime either as

$$\mathbb{R}^{2,1} \cong (\mathrm{SL}(2,\mathbb{R}) \ltimes \mathfrak{sl}(2,\mathbb{R}))/\mathrm{SL}(2,\mathbb{R})$$

or

$$\mathbb{R}^{2,1} \cong (\mathrm{SO}_0(2,1) \ltimes \mathbb{R}^{2,1})/\mathrm{SO}_0(2,1).$$

The two descriptions give isometric Lorentzian affine spaces, but the but of redundancy in the symmetry groups in the first description accounts for spin.

3d de Sitter spacetime

The 3-dimensional analog of de Sitter spacetime with cosmological constant $\Lambda > 0$ is the hyperboloid

$$\left\{(t,x,y,z) \in \mathbb{R}^{3,1} | -t^2 + x^2 + y^2 + z^2 = \frac{1}{\Lambda}\right\}$$

in $(3+1)$-dimensional Minkowski vector space. The Lorentz group acts on $\mathbb{R}^{3,1}$ preserving the metric $-t^2 + x^2 + y^2 + z^2$, so the full symmetry group of de Sitter spacetime is simply $\mathrm{SO}(3,1)$. In physics, we generally restrict to the connected component of this group, $G = \mathrm{SO}_0(3,1)$, which preserves orientation and time orientation. The stabilizer of a point in de Sitter spacetime is the stabilizer of a spacelike vector in the ambient Minkowski vector space, namely $H = \mathrm{SO}_0(2,1)$.

We often have occasion to use a matrix representation of the Lie algebra, which is given as follows:

$$\mathfrak{so}(3,1) = \left\{ \left. \begin{pmatrix} 0 & x & y & a \\ x & 0 & z & b \\ y & -z & 0 & c \\ a & -b & -c & 0 \end{pmatrix} \right| x,y,z,a,b,c \in \mathbb{R} \right\}$$

We also make frequent use of an alternate description of deSitter spacetime, based on the double cover of $\mathrm{SO}_0(3,1)$, namely $\mathrm{SL}(2,\mathbb{C})$. To see how this double cover works, we first note that $(3+1)$-dimensional Minkowski vector space is isomorphic to the space $\mathcal{H}$ of all 2×2 Hermitian matrices:

$$\mathcal{H} := \left\{ \begin{pmatrix} t+z & x+iy \\ x-iy & t-z \end{pmatrix} \ : \ t,x,y,z \in \mathbb{R} \right\}$$

The Minkowski metric is conveniently given by the determinant:

$$\det \begin{pmatrix} t+z & x+iy \\ x-iy & t-z \end{pmatrix} = t^2 - x^2 - y^2 - z^2$$

$\mathrm{SL}(2,\mathbb{C})$ acts on $\mathcal{H}$ by

$$\begin{array}{ccc} \mathrm{SL}(2,\mathbb{C}) \times \mathcal{H} & \to & \mathcal{H} \\ (g,X) & \mapsto & gXg^\dagger \end{array}$$

preserving the metric, so just as with $\mathrm{SL}(2,\mathbb{R})$ and $\mathrm{SO}_0(2,1)$ in the previous section, we get a map

$$\mathrm{SL}(2,\mathbb{C}) \to \mathrm{SO}_0(3,1)$$

which again is two–to–one and onto. De Sitter spacetime is the det $= -1$ hypersurface in $\mathcal{H}$, so $\mathrm{SO}(3,1)$ is the analog of the Poincaré group for 3d deSitter spacetime. The stabilizer of a point is $\mathrm{SO}(2,1)$, which lifts to a subgroup isomorphic to $\mathrm{SL}(2,\mathbb{R})$ in the double cover. As with the Minkowski case, we can summarize the symmetry groups in the following diagram:

$$\begin{tikzcd} \mathrm{SL}(2,\mathbb{R}) \arrow[r] \arrow[d, "2\text{-}1"] & \mathrm{SL}(2,\mathbb{C}) \arrow[d, "2\text{-}1"] \\ \mathrm{SO}_0(2,1) \arrow[r] & \mathrm{SO}_0(3,1) \end{tikzcd}$$

3d Anti-de Sitter spacetime and $\mathrm{SL}(2,\mathbb{R})$ geometry

Finally, we come to the 3d spacetime with cosmological constant $\Lambda < 0$. Like deSitter spacetime, this space can be seen as a hyperboloid in a vector space with an inner product, this time of signature $(-++-)$:hyperboloid

$$\left\{(t,x,y,z) \in \mathbb{R}^{2,2} \mid -t^2 + x^2 + y^2 - z^2 = \frac{1}{\Lambda}\right\}$$

Actually, it is standard practice to take 3d anti de Sitter space to be the universal cover of this hyperboloid, since the hyperboloid itself has closed timelike loops. However, for most of our work, the important thing is the local geometry of anti de Sitter space, so there is generally no need to pass to the universal cover. The symmetry group of the hyperboloid is $\mathrm{SO}(2,2)$, and as with the previous cases we will often consider only the connected component of the identity, $\mathrm{SO}_0(2,2)$.

A convenient matrix representation of the Lie algebra is given by

$$\mathfrak{so}(2,2) = \left\{ \left. \begin{pmatrix} 0 & x & y & a \\ x & 0 & z & b \\ y & -z & 0 & c \\ -a & b & c & 0 \end{pmatrix} \right| x,y,z,a,b,c \in \mathbb{R} \right\}$$

We have seen that $\mathrm{SL}(2,\mathbb{R})$ is the double cover of the 3d Lorentz group $\mathrm{SO}_0(2,1)$, while $\mathrm{SL}(2,\mathbb{C})$ is the double cover of $\mathrm{SO}_0(3,1)$. Similarly, the double cover of $\mathrm{SO}_0(2,2)$ is $\mathrm{SL}(2,\mathbb{R}) \times \mathrm{SL}(2,\mathbb{R})$. This description is particularly nice, since the 3d anti de Sitter space can be seen as the group $\mathrm{SL}(2,R)$ itself. This is made clear by considering the following representation of $\mathrm{SL}(2,\mathbb{R})$ as a matrix Lie group:

$$\mathrm{SL}(2,\mathbb{R}) = \left\{ \frac{1}{\ell} \begin{pmatrix} t+x & y+z \\ y-z & t-x \end{pmatrix} \;:\; t,x,y,z \in \mathbb{R},\ -t^2 + x^2 + y^2 - z^2 = -\ell^2 \right\}.$$

where we think of t, x, y, z as having dimensions of length and choose the normalizing length scale ℓ such that $-\ell^2 = 1/\Lambda$.

$\mathrm{SL}(2,\mathbb{R})$ acts on itself by both left and right translation. So we get a right action of $\mathrm{SL}(2,\mathbb{R})^2$ on 3d anti de Sitter spacetime $\mathrm{SL}(2,\mathbb{R})$ given by

$$\begin{array}{rcl} \mathrm{SL}(2,\mathbb{R})^2/\{\pm 1\} \times \mathrm{SL}(2,\mathbb{R}) & \to & \mathrm{SL}(2,\mathbb{R}) \\ ((g_1, g_2), h) & \mapsto & g_1 h g_2^{-1} \end{array}$$

The only elements that act trivially are elements (g, g) with g in the center of $\mathrm{SL}(2,\mathbb{R})$, namely $g = \pm 1$. In fact, we get an isomorphism

$$SO_0(2,2) \cong (\mathrm{SL}(2,\mathbb{R}) \times \mathrm{SL}(2,\mathbb{R})/\{\pm 1\}$$

and $\mathrm{SL}(2,\mathbb{R})^2$ is the group of spin transformations of 3d anti de Sitter spacetime. The stabilizer of the identity element is clearly the diagonal subgroup $H \cong \mathrm{SL}(2,\mathbb{R})$ consisting of all elements of the form $(g, g) \in \mathrm{SL}(2,\mathbb{R})^2$.

As with the previous two spacetimes, we summarize the groups of symmetries in a diagram:

$$\begin{tikzcd} \mathrm{SL}(2,\mathbb{R}) \arrow[r] \arrow[d, "2\text{-}1"] & \mathrm{SL}(2,\mathbb{R}) \times \mathrm{SL}(2,\mathbb{R}) \arrow[d, "2\text{-}1"] \\ \mathrm{SO}_0(2,1) \arrow[r] & \mathrm{SO}_0(2,2) \end{tikzcd}$$

2.3.2 Contractions and Wick rotations

In this section we describe two ways of getting new homogeneous spaces from old ones. These are based on two well-known process for groups: 'contractions' and 'Wick rotations'. From a physical point of view, 'contractions' can be thought of as 'limits' of Lie groups as some parameter approaches a specified value. The easiest example is what might be called the 'Columbus contraction', in which the parameter of interest is the radius of a spherical Earth. For any value of the radius, the group of symmetries is the rotation group SO(3), but if radius becomes infinite, the group suddenly becomes the Euclidean group of the plane, ISO(2).

Wick rotation is also a sort of limiting process, but in a slightly different sense. The classic example in physics is the Wick rotation from the Lorentz group $\mathrm{SO}(3,1)$ to the orthogonal group SO(4). The idea is to allow the speed of light c to take complex values and then rotate in the complex plane from c to ic, so that the Minkowski metric

$-c^2t^2 + x^2 + y^2 + z^2$ switches sign in the time direction. More rigorously, we think of $\mathrm{SO}(3,1)$ and $\mathrm{SO}(4)$ both as sitting inside the complex Lie group $\mathrm{SO}(4,\mathbb{C})$ and consider a 1-parameter family of subgroups that interpolates between them.

Starting with the three models of homogeneous spacetime we have already discussed, and using contractions and Wick rotations, we construct a family of 9 homogeneous 3d models of spacetime. In fact, we could just start with one spacetime, say the 3d de Sitter model. Anti de Sitter may then be obtained by a Wick rotation, not with respect to c, but with respect to the cosmological constant Λ. Minkowski spacetime is a limit of both the de Sitter and anti de Sitter models as $\Lambda \to 0$. Wick rotations of these three models give the three 3d Riemannian symmetric spaces, the Hyperbolic, Euclidean, and Spherical 3d geometries. But there is another pair of group contractions that is equally important: the flat Lorentzian and Riemannian spacetimes have the spacetime of Galilean relativity as a common limit, the limit as $c^2 \to \infty$. We can also consider what happens if we take the $c \to \infty$ limit *before* the $\Lambda \to 0$ limit. These give analogs of Galilean spacetime with positive and negative spatial curvature. If we include these "Galilean spacetimes with cosmological constant" we get one type of homogeneous spacetime for each of the nine combinations of sign choices for Λ and $1/c^2$. We can describe the relationships between these homogeneous spacetimes in the following diagram:

$$
\begin{tikzcd}
\begin{array}{c}\frac{1}{c^2} > 0 \\ \text{Lorentzian}\end{array} & \begin{array}{c}\text{3d Anti-de Sitter} \\ \mathrm{SO}(2,2)/\mathrm{SO}(2,1)\end{array} \arrow[r, "\text{contraction}"] \arrow[d, "\text{contraction}"] \arrow[rr, bend left, leftrightarrow, "\text{Wick rotation}"] & \begin{array}{c}\text{3d Minkowski} \\ \mathrm{ISO}(2,1)/\mathrm{SO}(2,1)\end{array} \arrow[d, "\text{contraction}"] & \begin{array}{c}\text{3d de Sitter} \\ \mathrm{SO}(3,1)/\mathrm{SO}(2,1)\end{array} \arrow[l, "\text{contraction}"'] \arrow[d, "\text{contraction}"] \arrow[dd, bend left, leftrightarrow, "\text{Wick rotation}"] \\
\begin{array}{c}\frac{1}{c^2} = 0 \\ \text{Galilean}\end{array} & \begin{array}{c}\text{3d hyperbolic Galilean} \\ \mathrm{ISO}(2,1)/\mathrm{ISO}(2)\end{array} \arrow[r, "\text{contraction}"] & \begin{array}{c}\text{3d Galilean} \\ (\mathrm{ISO}(2) \ltimes \mathbb{R}^3)/\mathrm{ISO}(2)\end{array} & \begin{array}{c}\text{3d spherical Galilean} \\ \mathrm{ISO}(3)/\mathrm{ISO}(2)\end{array} \arrow[l, "\text{contraction}"'] \\
\begin{array}{c}\frac{1}{c^2} < 0 \\ \text{Riemannian}\end{array} & \begin{array}{c}\text{3d hyperbolic} \\ \mathrm{SO}(3,1)/\mathrm{SO}(3)\end{array} \arrow[r, "\text{contraction}"] \arrow[u, "\text{contraction}"] & \begin{array}{c}\text{3d Euclidean} \\ \mathrm{ISO}(3)/\mathrm{SO}(3)\end{array} \arrow[u, "\text{contraction}"] & \begin{array}{c}\text{3d spherical} \\ \mathrm{SO}(4)/\mathrm{SO}(3)\end{array} \arrow[l, "\text{contraction}"'] \arrow[u, "\text{contraction}"] \\
 & \Lambda < 0 & \Lambda = 0 & \Lambda > 0
\end{tikzcd}
$$

In what follows, we describe contractions and Wick rotations in general, but also the six spacetimes above that we have not yet considered.

Wick rotation

To understand how the spacetimes in the above diagrams are related by Wick rotations, we need to complexify the symmetry groups. The complexification of any $\mathrm{SO}(p,q)$ is $\mathrm{SO}(n,\mathbb{C})$ where $n = p+q$. This is the group of linear transformations that preserves a nondegenerate symmetric bilinear form on $\mathbb{C}^n$.[3] $\mathrm{SO}(n,\mathbb{C})$ acts on $\mathbb{C}^n$ in the obvious way, and their semidirect product is the complexification of any $\mathrm{ISO}(p,q)$ with $n = p+q$. This semidirect product deserves to be called $\mathrm{ISO}(n,\mathbb{C})$:

$$\mathrm{ISO}(n,\mathbb{C}) := \mathrm{SO}(n,\mathbb{C}) \ltimes \mathbb{C}^n$$

Now, when we complexify all of the groups in diagram of homogeneous spacetimes, we get only four distinct complex homogeneous spaces:

$$\begin{tikzcd}
\mathrm{SO}(4,\mathbb{C})/\mathrm{SO}(3,\mathbb{C}) \arrow[r] \arrow[d] & \mathrm{ISO}(3,\mathbb{C})/\mathrm{SO}(3,\mathbb{C}) \arrow[d] & \mathrm{SO}(4,\mathbb{C})/\mathrm{SO}(3,\mathbb{C}) \arrow[l] \arrow[d] \\
\mathrm{ISO}(3,\mathbb{C})/\mathrm{ISO}(2,\mathbb{C}) \arrow[r] & (\mathrm{ISO}(2,\mathbb{C}) \ltimes \mathbb{C}^3)/\mathrm{ISO}(2,\mathbb{C}) & \mathrm{ISO}(3,\mathbb{C})/\mathrm{ISO}(2,\mathbb{C}) \arrow[l] \\
\mathrm{SO}(4,\mathbb{C})/\mathrm{SO}(3,\mathbb{C}) \arrow[r] \arrow[u] & \mathrm{ISO}(3,\mathbb{C})/\mathrm{SO}(3,\mathbb{C}) \arrow[u] & \mathrm{SO}(4,\mathbb{C})/\mathrm{SO}(3,\mathbb{C}) \arrow[l] \arrow[u]
\end{tikzcd}$$

This complexified diagram is symmetric under vertical and horizontal reflections. In particular, the four corners are the same, as are the left/right and top/bottom pairs. There is no distinction between Lorentzian and Riemannian, nor between positive and negative cosmological constant in complex spacetime. There *is* a distinction between Galilean and non-Galilean, as between zero and nonzero cosmological constant.

Let us be a bit more explicit about how the Wick rotations work. Since all nondegenerate symmetric bilinear forms on a complex vector space are equivalent, we can define $\mathrm{SO}(n,\mathbb{C})$ to be the group of isometries of $\mathbb{C}^n$ with the standard dot product:

$$z \cdot w = \delta^{ij} z_i z_j$$

The real forms of $\mathrm{SO}(n,\mathbb{C})$ can all be described as subgroups

$$\mathrm{SO}(n,\mathbb{C})_\sigma := \{g \in \mathrm{SO}(n,\mathbb{C}) \mid g\sigma = \sigma g\}$$

[3] Note that this is not a complex inner product on $\mathbb{C}^n$, which would be *sesquilinear*. Sesquilinear inner products on $\mathbb{C}^n$ come in various signatures, giving groups $\mathrm{SU}(p,q)$. All *symmetric* inner products on $\mathbb{C}^n$ are isomorphic, so there's just one $\mathrm{SO}(n,\mathbb{C})$.

consisting of all elements commuting with a given conjugate linear involution $\sigma\colon \mathbb{C}^n \to \mathbb{C}^n$. In particular, if V is a real inner product space of signature $\varepsilon = (\varepsilon_1 \cdots \varepsilon_n)$, $\varepsilon_i = \pm 1$, then

$$\mathrm{SO}(V) \cong \mathrm{SO}(n, \mathbb{C})_\sigma$$

where

$$\begin{aligned} \sigma\colon \mathbb{C}^n &\to \mathbb{C}^n \\ z_k &\mapsto \varepsilon_k \bar{z}_k \end{aligned}$$

To understand Wick rotation from $\mathrm{SO}(p,q)$ to $\mathrm{SO}(p',q')$, where $p+q = p'+q'$, it suffices to consider the simplest example, which contains all of the features of the general case. For the group $\mathrm{SO}(2,\mathbb{C})$, for each $\phi \in [0,\pi]$ define a conjugate linear involution of $\mathbb{C}^2$ by

$$\sigma(\phi)\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} e^{i\phi}\bar{z}_1 \\ \bar{z}_2 \end{pmatrix}.$$

This gives a 1-parameter family of subgroups $\mathrm{SO}(2,\mathbb{C})_{\sigma(\phi)} \subseteq \mathrm{SO}(2,\mathbb{C})$. When $\phi = 0$, the subgroup is may be identified with $\mathrm{SO}(2)$; when $\phi = \pi$, it is $\mathrm{SO}(1,1)$. This is Wick rotation.

the 3-sphere

Let us now describe the three Riemannian spacetimes obtained by Wick rotating the de Sitter, Minkowski and anti de Sitter models. Since these are quite similar to their Lorentzian cousins, but even simpler, we keep these descriptions terse. The 3-sphere can be described as a Klein geometry with $(G,H) = (\mathrm{SO}(4), \mathrm{SO}(3))$. The Lie algebra has a matrix representation given by

$$\mathfrak{so}(4) = \left\{ \begin{pmatrix} 0 & z & y & a \\ -z & 0 & x & b \\ -y & -x & 0 & c \\ -a & -b & -c & 0 \end{pmatrix} \middle| \, x,y,z,a,b,c \in \mathbb{R} \right\}$$

where as in the previous cases the stabilizer subalgebra consists of those matrices in with $a,b,c = 0$. The double cover description works in almost exactly the same way as in the anti-de Sitter case, since $S^3 \cong SU(2)$, and the group of isometries of $\mathrm{SU}(2)$ is $G = \mathrm{SU}(2) \times \mathrm{SU}(2)/\{\pm 1\}$ acting by left and right multiplication:

$$\begin{aligned} G \times \mathrm{SU}(2) &\to \mathrm{SU}(2) \\ (\pm(h_1,h_2), g) &\mapsto h_1 g h_2^{-1} \end{aligned}$$

As in the Lorentzian case, the stabilizer of 1 is the diagonal subgroup.

We thus summarize the effective and spin symmetry groups of S^3 as follows:

$$\begin{tikzcd} \mathrm{SU}(2) \arrow[r] \arrow[d, "2\text{-}1"] & \mathrm{SU}(2) \times \mathrm{SU}(2) \arrow[d, "2\text{-}1"] \\ \mathrm{SO}(3) \arrow[r] & \mathrm{SO}(4) \end{tikzcd}$$

Euclidean 3-space

3-dimensional Euclidean geometry is described by $(G, H) = (\mathrm{ISO}(3), \mathrm{SO}(3)$, where $\mathrm{ISO}(3) = \mathrm{SO}(3) \ltimes \mathbb{R}^3$ is the Eudliean group with Lie algebra

$$\mathfrak{iso}(3) = \begin{pmatrix} \mathfrak{so}(3) & \mathbb{R}^3 \\ 0 & 0 \end{pmatrix} = \left\{ \begin{pmatrix} 0 & x & y & a \\ -x & 0 & z & b \\ -y & -z & 0 & c \\ 0 & 0 & 0 & 0 \end{pmatrix} \middle| \, x, y, z, a, b, c \in \mathbb{R} \right\}$$

As in the 3d Minkowski case, we have an SU(2)-equivariant isomorphism $\mathbb{R}^3 \cong \mathfrak{su}(2)$, where the actions of SU(2) on $\mathbb{R}^3$ comes from the obvious action of SO(3) and the covering map $\mathrm{SU}(2) \to SO(3)$, and the action on $\mathfrak{su}(2)$ is the adjoint action. We thus have the following description of the symmetry groups and their double covers:

$$\begin{tikzcd} \mathrm{SU}(2) \arrow[r] \arrow[d, "2\text{-}1"] & \mathrm{SU}(2) \ltimes \mathfrak{su}(2) \arrow[d, "2\text{-}1"] \\ \mathrm{SO}(3) \arrow[r] & \mathrm{ISO}(3) \end{tikzcd}$$

3d Hyperbolic space

Finally, we come to 3d hyperbolic space. This is the most interesting and important of the three Riemannian geometries we are considering here. Its symmetries are given by the Lorentz group $\mathrm{SO}(3,1)$. This is easy to see if we think of hyperbolic space as the velocity space in special relativity. That is, H^3 is the hypersurface of all unit future-pointing timelike vectors in Minkowski spacetime. The stabilizer of a point is simply the subgroup $\mathrm{SO}(3) \subseteq \mathrm{SO}(3,1)$ consisting of spatial rotations around the corresponding velocity vector. Though we have seen these groups before, we again write down a matrix representation of the Lie algebra:

$$\mathfrak{so}(3,1) = \left\{ \begin{pmatrix} 0 & w & -v & x \\ -w & 0 & u & y \\ v & -u & 0 & z \\ x & y & z & 0 \end{pmatrix} \middle| \, x, y, z, a, b, c \in \mathbb{R} \right\}$$

This differs from the description in the section on de Sitter spacetime by a change of basis. Here we've chosen the signature $(+++-)$, so that the stabilizer subalgebra is still the upper left 3×3 block matrices with zeros elsewhere.

We have already described the double cover of $\mathrm{SO}_0(3,1)$, since it is also the symmetry group of 3d de Sitter spacetime. We've also described the double cover of the stabilizer $\mathrm{SO}(3)$, so we get the following diagram for the symmetry groups and double covers:

$$\begin{tikzcd}
\mathrm{SU}(2) \arrow[r] \arrow[d, two heads, "\text{2-1}"] & \mathrm{SL}(2,\mathbb{C}) \arrow[d, two heads, "\text{2-1}"] \\
\mathrm{SO}(3) \arrow[r] & \mathrm{SO}_0(3,1)
\end{tikzcd}$$

Contractions

Besides Wick rotations, our homogeneous spacetimes are related by 'contractions' or 'limits' of their symmetry groups. There are actually several possibilities for defining contractions. We shall use the following definition, adapted from Hermann [54, p. 87][4]:

Definition 3 *Let G be a Lie group with Lie subgroups H, H'. Then H' is called a contraction of H within G if there is a sequence $g_1, g_2, \ldots \in G$ such that:*

1. *for every sequence $h_1, h_2, \ldots \in H$ such that $\mathrm{AD}(g_1)h_1, \mathrm{AD}(g_2)h_2, \ldots$ converges, the limit is an element of H';*

2. *every element $h' \in H'$ can be written as*

$$h' = \lim_{k \to \infty} \mathrm{AD}(g_k) h_k$$

for some sequence h_k.

When H' is a contraction of H by the sequence $g_1, g_2, \ldots \in G$, we write

$$H' = \lim_{k \to \infty} \mathrm{AD}(g_k) H$$

Hermann also offers a Lie algebra version of this definition, which is naturally easier to work with in most cases:

Definition 4 *Let $\mathfrak{g}$ be the Lie algebra of the Lie group G, and let $\mathfrak{h}, \mathfrak{h}' \subseteq \mathfrak{g}$ be Lie subalgebras. Then $\mathfrak{h}'$ is called a contraction of $\mathfrak{h}$ within $\mathfrak{g}$ if there is a sequence $g_1, g_2, \ldots \in G$ such that:*

[4] Hermann actually leaves out the second clause in the definition, though it is obviously needed, since otherwise any group containing H (or any contraction of it) would be called a contraction of H.

1. *for every sequence* $h_1, h_2, \ldots \in \mathfrak{h}$ *such that* $\lim_{k\to\infty} \mathrm{Ad}(g_k)h_k$ *exists, the limit is an element of* $\mathfrak{h}'$

2. *every element* $h' \in \mathfrak{h}'$ *can be written as*

$$\lim_{k\to\infty} \mathrm{Ad}(g_k)h_k$$

for some sequence h_k*.*

When $\mathfrak{h}'$ *is a contraction of* $\mathfrak{h}$ *by the sequence* $g_1, g_2, \ldots \in G$*, we write*

$$\mathfrak{h}' = \lim_{k\to\infty} \mathrm{Ad}(g_k)\mathfrak{h}$$

Galilean Spacetimes

Let us describe the Galilean spacetimes in some detail, since they may be the least familiar from the perspective of symmetry groups. In ordinary Galilean relativity, space is Euclidean and the only measurements one can make are the distance between simultaneous events and the time interval between events. After making an initial choice of coordinate system, a Galilean transformation consists of some combination of a translation T, Galilei boost B, and rotation R (see, e.g. [4, p. 6]):

$$T_{s_o,\vec{s}}(t,\vec{x}) = (t + s_0, \vec{x} + \vec{s}) \qquad B_{\vec{v}}(t,\vec{x}) = (t, \vec{x} + t\vec{v}) \qquad R_\Theta(t,\vec{x}) = (t, \Theta\vec{x})$$

where $(s_0, \vec{s})$ is a displacement of the origin, $\vec{v}$ is a velocity, and Θ is a rotation matrix. Composing these shows that the rotation group acts both on boosts and the spatial parts of translations via the defining representation of the rotation group, while boosts act on translations as follows:

$$[B(\vec{v}) \circ T(s_o, \vec{s}) \circ B(\vec{v})^{-1}](t,\vec{x}) = (t + s_0, \vec{x} + \vec{s} + s\vec{v}) = T(s_0, \vec{s} + s_0\vec{v})(t,\vec{x}).$$

This gives the **Galilei group** the structure of a nested semidirect product:

$$\mathrm{Gali}(n+1) = (\mathrm{SO}(n) \ltimes \mathbb{R}^n) \ltimes \mathbb{R}^{n+1}$$

where n denotes the dimension of Galilean 'space'; for present purposes, $n = 2$. Since rotations act on boosts, while both rotations and boosts act on translations, the simplest factorization of a general Galilean transformation is of the form $T \circ B \circ R$:

$$[T_{s_o,\vec{s}} \circ B_{\vec{v}} \circ R_\Theta](t,\vec{x}) = (t + s_0, \Theta\vec{x} + t\vec{v} + \vec{s})$$

Rewriting this as a matrix equation:

$$\begin{pmatrix} 1 & 0 & s_0 \\ \vec{v} & \Theta & \vec{s} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} t \\ \vec{x} \\ 1 \end{pmatrix} = \begin{pmatrix} t + s_0 \\ \Theta\vec{x} + t\vec{v} + \vec{s} \\ 1 \end{pmatrix}$$

suggests the following matrix representation of the Galilei group:

$$\mathrm{Gali}(n+1) := \left\{ \begin{pmatrix} 1 & 0 & s_0 \\ \vec{v} & \Theta & \vec{s} \\ 0 & 0 & 1 \end{pmatrix} \middle| \, s_0 \in \mathbb{R}, \vec{s}, \vec{v} \in \mathbb{R}^n, \Theta \in \mathrm{SO}(n) \right\} \subseteq \mathrm{GL}(n+2, \mathbb{R})$$

The subgroup stabilizing the origin is the $\mathrm{ISO}(n)$ subgroup generated by boosts and rotations. These are of course the elements in the upper left $(n+1) \times (n+1)$ block.

For the case $n = 3$, we get the following matrix representation of the Lie algebra

$$\mathfrak{gali}(3) = \left\{ \begin{pmatrix} 0 & 0 & 0 & u \\ b & 0 & a & v \\ c & -a & 0 & w \\ 0 & 0 & 0 & 0 \end{pmatrix} \middle| \, u, v, w, a, b, c \in \mathbb{R} \right\}$$

analogous to the Lie algebra representations we have written down for the other spacetimes.

We wish to see Galilean spacetime as the Newtonian limit $c \to \infty$ of Minkowski spacetime. Consider the point stabilizer group for 3d Minkowski spacetime: the Lorentz group $\mathrm{SO}_0(2,1)$. This group by definition preserves the inner product on $\mathbb{R}^n$ given by

$$\langle x, y \rangle_c = x^T \eta(c) y$$

where ^{T} denotes the matrix transpose and the matrix of the inner product is

$$\eta(c) = \begin{pmatrix} -c^2 & 0 \\ 0 & I \end{pmatrix}.$$

For understanding contractions, it is useful to denote the Lorentz group preserving this inner product by $\mathrm{SO}(2,1)_c$—or more generally by $\mathrm{SO}(n,1)_c$—as a reminder that the group depends on the value of c. These groups are all isomorphic for $0 < c < \infty$, but we wish to see what happens when we let c become infinite.

Following the calculation by Hermann [54], suppose we now adjust the speed of light from c to some other c'. Then note that

$$\eta(c') = P^T \eta(c) P$$

where

$$P = \begin{pmatrix} c'/c & 0 \\ 0 & I \end{pmatrix}.$$

If $\mathcal{O} \in \mathrm{SO}(n,1)'_c$, then

$$\mathcal{O}^T \eta(c')\mathcal{O} = \eta(c')$$

or equivalently,

$$(P\mathcal{O}P^{-1})^T \eta(c)(P\mathcal{O}P^{-1}) = \eta(c).$$

Therefore, $\mathcal{O} \in \mathrm{SO}(2,1)_{c'}$ if and only if $P\mathcal{O}P^{-1} \in \mathrm{SO}(n,1)_c$. In particular, the Lorentz groups for different values of c are all conjugate subgroups of $\mathrm{GL}(3,\mathbb{R})$. This of course implies that the Lie algebras $\mathfrak{so}(2,1)_c$ are all conjugate within $\mathfrak{gl}(3,\mathbb{R})$. If we increment the value of c successively, we thus get a sequence of elements $g_1, g_2, \ldots \in \mathrm{GL}(3,\mathbb{R})$ relating the corresponding Lie algebras, as in Definition 4.

The form of a general element ξ of the Lie algebra $\mathfrak{so}(2,1)_c$ can be found using the fact that $\xi^T\eta(c) + \eta(c)\xi = 0$. One obtains

$$\xi = \begin{pmatrix} 0 & -w & v \\ -c^2 w & 0 & u \\ c^2 v & -u & 0 \end{pmatrix}$$

for arbitrary u, v, w. Now we consider u, v, w as functions of the value of the speed of light c. If the above matrix is to converge as $c \to \infty$, then we must have

$$\lim_{c\to\infty} w(c) = 0$$
$$\lim_{c\to\infty} v(c) = 0$$

but of course we may arrange for $c^2 w(c)$ and $c^2 v(c)$ to have any limit we wish. So, in the limit as $c \to \infty$, if an $\mathfrak{so}(2,1)$-valued function of c is to converge, it must converge to an element of the form

$$\begin{pmatrix} 0 & 0 & 0 \\ -w & 0 & u \\ v & -u & 0 \end{pmatrix}$$

But this is just an element of $\mathfrak{iso}(2)$, and this is the point stabilizer subalgebra for $\mathfrak{gali}(3)$. We say that $\mathfrak{so}(2,1)$ contracts to $\mathfrak{iso}(2)$ as $c \to 0$.

If we do the above contraction process not just for the Lorenz group, but for the full Poincareǵroup, we see that $\mathfrak{iso}(2,1)$ contracts to $\mathfrak{gali}(3)$. A typical element of $\mathfrak{iso}(2,1)$

has the form

$$\begin{pmatrix} 0 & -w & v & x \\ -w & 0 & u & y \\ v & -u & 0 & z \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

In the $c \to 0$ limit, any convergent matrix-valued function of the speed of light converges to an element of the form

$$\begin{pmatrix} 0 & -w & v & x \\ -w & 0 & u & y \\ v & -u & 0 & z \\ 0 & 0 & 0 & 0 \end{pmatrix} \in \mathfrak{gali}(3).$$

It is clear that there is nothing special here about 3 dimensions, and in fact

$$\lim_{c\to\infty} \mathfrak{so}(n,1) = \mathfrak{iso}(n)$$

$$\lim_{c\to\infty} \mathfrak{iso}(n,1) = \mathfrak{gali}(n).$$

We have done the above contractions using the speed of light c as a parameter, but it is just as easy to use the cosmological constant Λ, or perhaps more appropriate in the Riemannian cases, the radius. Doing this, we get a precise sense in which the five of the nine spacetimes in our diagram on p. 29 are limiting cases of others, either as $c \to \infty$ or as $\Lambda \to 0$. It is worth rewriting this diagram including the matrices for typical Lie algebra elements, since it is instructive to see how the matrix entries change when we perform Wick rotations or contractions. In each case in the diagram below, the stabilizer subalgebra consists of those matrices that are zero outside of the upper 3×3 block.

$$
\begin{tikzcd}
\begin{array}{c}\frac{1}{c^2}>0\\ \text{Lorentzian}\end{array} &
\begin{array}{c}\text{3d Anti-de Sitter}\\ \mathrm{SO}(2,2)/\mathrm{SO}(2,1)\\ (-{+}{+}-)\\ \begin{pmatrix} 0 & -w & v & x\\ -w & 0 & u & y\\ v & -u & 0 & z\\ -x & y & z & 0\end{pmatrix}\end{array}
\arrow[r, "c\to\infty"] \arrow[d, "\Lambda\to 0"] \arrow[rr, bend left, leftrightarrow, "\text{Wick rotation}"] &
\begin{array}{c}\text{3d Minkowski}\\ \mathrm{ISO}(2,1)/\mathrm{SO}(2,1)\\ (-{+}{+}0)\\ \begin{pmatrix} 0 & -w & v & x\\ -w & 0 & u & y\\ v & -u & 0 & z\\ 0 & 0 & 0 & 0\end{pmatrix}\end{array}
\arrow[d, "\Lambda\to 0"] &
\begin{array}{c}\text{3d de Sitter}\\ \mathrm{SO}(3,1)/\mathrm{SO}(2,1)\\ (-{+}{+}{+})\\ \begin{pmatrix} 0 & -w & v & x\\ -w & 0 & u & y\\ v & -u & 0 & z\\ x & -y & -z & 0\end{pmatrix}\end{array}
\arrow[l, "c\to\infty"'] \arrow[d, "\Lambda\to 0"] \arrow[dd, bend left, leftrightarrow, "\text{Wick rotation}"] \\
\begin{array}{c}\frac{1}{c^2}=0\\ \text{Galilean}\end{array} &
\begin{array}{c}\text{3d hyperbolic Galilean}\\ \mathrm{ISO}(2,1)/\mathrm{ISO}(2)\\ (0{+}{+}-)\\ \begin{pmatrix} 0 & 0 & 0 & x\\ -c & 0 & a & y\\ b & -a & 0 & z\\ 0 & y & z & 0\end{pmatrix}\end{array}
\arrow[r, "c\to\infty"] &
\begin{array}{c}\text{3d Galilean}\\ (\mathrm{ISO}(2)\ltimes\mathbb{R}^3)/\mathrm{ISO}(2)\\ (0{+}{+}0)\\ \begin{pmatrix} 0 & 0 & 0 & x\\ -w & 0 & u & y\\ v & -u & 0 & z\\ 0 & 0 & 0 & 0\end{pmatrix}\end{array} &
\begin{array}{c}\text{3d spherical Galilean}\\ \mathrm{ISO}(3)/\mathrm{ISO}(2)\\ (0{+}{+}{+})\\ \begin{pmatrix} 0 & w & -v & x\\ 0 & 0 & u & y\\ 0 & -u & 0 & z\\ 0 & -y & -z & 0\end{pmatrix}\end{array}
\arrow[l, "c\to\infty"'] \\
\begin{array}{c}\frac{1}{c^2}<0\\ \text{Riemannian}\end{array} &
\begin{array}{c}\text{3d hyperbolic}\\ \mathrm{SO}(3,1)/\mathrm{SO}(3)\\ ({+}{+}{+}-)\\ \begin{pmatrix} 0 & w & -v & x\\ -w & 0 & u & y\\ v & -u & 0 & z\\ x & y & z & 0\end{pmatrix}\end{array}
\arrow[r, "c\to\infty"] \arrow[u, "\Lambda\to 0"'] &
\begin{array}{c}\text{3d Euclidean}\\ \mathrm{ISO}(3)/\mathrm{SO}(3)\\ ({+}{+}{+}0)\\ \begin{pmatrix} 0 & w & -v & x\\ -w & 0 & u & y\\ v & -u & 0 & z\\ 0 & 0 & 0 & 0\end{pmatrix}\end{array}
\arrow[u, "\Lambda\to 0"'] &
\begin{array}{c}\text{3d spherical}\\ \mathrm{SO}(4)/\mathrm{SO}(3)\\ ({+}{+}{+}{+})\\ \begin{pmatrix} 0 & w & -v & x\\ -w & 0 & u & y\\ v & -u & 0 & z\\ -x & -y & -z & 0\end{pmatrix}\end{array}
\arrow[l, "c\to\infty"'] \arrow[u, "\Lambda\to 0"'] \\
& \Lambda<0 & \Lambda=0 & \Lambda>0
\end{tikzcd}
$$

2.3.3 4d spacetimes

The generalization to 4d spacetimes is immediate, the only real difference being that in 4 dimensions we do not have all of the same nice descriptions of the double covers of symmetry groups, since there are fewer coincidences of Lie groups as we move up to higher dimensions. We still have the same nine types of homogeneous spacetime, which can be given descriptions in terms of Klein geometry G/H in an obvious way:

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

Chapter 3

Geometric structures and flat bundles

Thanks to Thurston's famous geometrization conjecture— recently proven by Perelman —we have a solid understanding of what Riemannian 3-manifolds look like. Roughly, Thurston's conjecture states that the interior of any compact 3-manifold can be chopped up into pieces which look locally like one of eight model Klein geometries, listed below. The full conjecture says much more about what these pieces can look like, and in fact implies the Poincaré conjecture: every compact connected and simply connected 3-manifold is a homeomorph of S^3. We shall not concern ourselves with the full power of Thurston's conjecture, instead referring the reader to [71] for a very readable account, and to Thurston's book [93] for much of the background in 3d geometry.

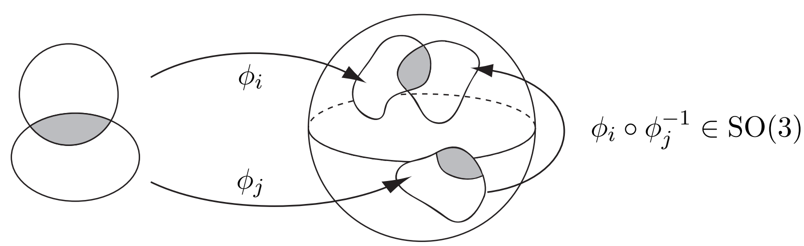
For our purposes, Thurston's idea of a geometric manifold provides a useful way of looking at solutions of topological gauge theories, like 3d general relativity. As we will see in the next chapter, solutions of 3d general relativity look locally like Klein geometries, but not necessarily globally. Geometric structures on manifolds make this idea precise.

In fact, it has been known for some time that the theory of geometric structures is related to 3d gravity. This point of view has been beautifully explained by Carlip [24, 25]. But as we will see, geometric structures may be just as important for understanding 4d gravity. This will become apparent when finally study Cartan geometries, the geometries that show up in MacDowell-Manouri-like approaches to 4d gravity: Cartan geometries are just curved versions of geometric structures. But first we must understanding the flat case in detail and, in the next chapter, its relation to topological gauge theories.

3.1 Geometric manifolds

There are many variations of the idea of a manifold, generally obtained by modifying the model space and the properties which a transition function must satisfy. For example, one defines a manifold with boundary by choosing as the model space, rather than simply $\mathbb{R}^n$, a closed half-space of $\mathbb{R}^n$, say the region where the first coordinate is non-negative. Likewise, a Banach manifold has some Banach space as the model space. One defines versions of manifolds which are topological, smooth, real analytic, and so on, by demanding that the transition functions are maps preserving the corresponding structure in the model space.

Geometric manifolds are an example of such generalization which is quite rigid. The model space is any homogeneous space, while the transition functions are required to be *isometries*. In a geometric manifold M modeled on a sphere S^2 of radius r, for example, coordinate maps are continuous functions from open sets of M to the sphere, such that two different sets of local coordinates around a point agree up to a rotation of the model sphere:



Formally, we have the following definition.

Definition 5 *Let $X = G/H$ be a Klein Geometry. A (G,X)–manifold M is a (smooth or analytic) manifold M equipped with an open cover $\{U_i\}$ and a family of homeomorphic embeddings $\phi_i\colon U_i \to X$ such that the transition functions $\phi_i \circ \phi_j^{-1}$ live in G.*

In the context of this definition, we say M is **geometrically modeled on** the homogeneous space G/H. Borrowing the usual terminology for manifolds, we call the maps $\phi_i\colon U_i \to X$ **(coordinate) charts** for the geometric manifold.

Though we work globally as often as possible, it will sometimes be convenient to adopt the following notation for local coordinate charts. Denote intersections of sets in the open cover by concatenating their substripts:

$$
\begin{aligned}
U_{ij} &:= U_i \cap U_j \\
U_{ijk} &:= U_i \cap U_j \cap U_j \\
&\vdots
\end{aligned}
$$

Also, denote the transition functions as follows:

$$\phi_{ij} := \phi_i \circ \phi_j^{-1} \colon \phi_j(U_{ij}) \to \phi_i(U_{ij})$$

We will also use the customary level of sloppiness about domains and ranges of coordinate charts and transition functions for manifolds. In particular, we also denote by ϕ_{ij} the transformation of X induced by the transition function.

After defining geometric manifolds, we should define the appropriate sort of maps between them. We adopt the following definition essentially from Goldman [48].

Definition 6 *Given two (G, X)–manifolds M and M', a (G, X)–map $f \colon M \to M'$ is a map such that for every every coordinate chart (U, ϕ) for M and every coordinate chart (V, ψ) for M' with $f(U) \cap V \neq 0$, $\psi \circ f \circ \phi^{-1}$ extends to a transformation of X in G.*

3.2 The flat bundle associated to a geometric structure

Ultimately, our goal is to relate geometric structures to gauge theory. Since gauge theory is written in the language of connections on fiber bundles, we need a description of geometric structures in this language.

A naive guess at the relationship between bundles and geometric structures on manifolds would be that geometric structures correspond to *flat* fiber bundles with $X = G/H$ as fiber. This makes sense—moving around in a geometric manifold seems to give trivial parallel transport in some bundle of Klein geometries—and in fact it is almost correct. The difference between flat (G, X)-bundles and (G, X) structures has to with the idea of 'symmetry breaking', which will be a theme for us throughout this thesis.[1] We begin to see this symmetry breaking in the following standard result on reduction of bundles, which we shall need for our characterization of geometric structures:

[1]For the relationship between reduction of bundles and symmetry breaking in particle physics, *c.f.* Choquet–Bruhat, *et. al.*, vol. 2 [30]

Proposition 1 *Suppose H is a closed subgroup of a Lie group G and that $Q \to M$ is a principal G bundle. Then Q admits a reduction to structure group H if and only if the associated bundle $Q \times_G G/H$ admits a global section.*

Proof: See, for example, Thurston [93, p. 162], or Choquet–Bruhat *et. al.* [30, p. 385]. Here we will simply point out that the section $s\colon M \to Q \times_G G/H$ and the corresponding principal H bundle $P \to M$ reducing Q are related via the pullback:

$$\begin{array}{ccc} P & \longrightarrow & Q \\ \downarrow & & \downarrow \\ M & \xrightarrow{s} & Q \times_G G/H \end{array}$$

where Q is viewed as a principal H-bundle over $Q \times_G G/H$. □

We are now ready to prove the following correspondence between geometric manifolds and bundles with certain extra structure:

Theorem 2 *Let $X = G/H$ be an Klein geometry. Then there is a canonical one-to-one correspondence between the following types of structures that can be put on a manifold M of the same dimension as X:*

(i) a (G, X)-structure on M;

(ii) a bundle over M with standard fiber X and structure group G, equipped with both a flat connection and a global section transverse to the connection;

(iii) a principal H-bundle $P \to M$ together with a flat connection on the associated principal G bundle $Q = P \times_H G$, induced by inclusion of H in G, such that the sub-bundle P is everywhere transverse to the connection.

Proof: Most of the elements of the proof are contained—explicitly or implicitly—in Thurston's book [93, §3.6]. It seems worthwhile, though, to assemble these ingredients into a complete proof, so that the explicit correspondence between these descriptions may be evident. We proceed by establishing the equivalence of (i) and (ii), then of (ii) and (iii).

(i) $\Longrightarrow$ (ii) Given a (G, X) structure on M with charts $\phi_i\colon U_i \to X$, we obtain a flat (G, X)-bundle E with transverse section s as follows. First describe E locally over each U_i as a product bundle

$$\begin{tikzcd} U_i \times X \arrow[d] \\ U_i \end{tikzcd}$$

This bundle has its standard flat connection as a trivial bundle, whose parallel transport is trivial in the second factor. It also has an obvious[2] section transverse to the connection, namely the map

$$\begin{aligned} s_i \colon U_i &\to U_i \times X \\ u &\mapsto (u, \phi_i(u)). \end{aligned}$$

The product bundles over U_i can be assembled into a bundle E over M with fiber X, using the ϕ_{ij} to identify fibers. Explicitly, E is the disjoint union of the bundles $U_i \times X$, modulo identification of $(u, x) \in U_j \times X$ with $(u, \phi_{ij}(x)) \in U_i \times X$. The ϕ_{ij} clearly satisfy the cocycle condition that $\phi_{ij}\phi_{jk}\phi_{ki}$ is the identity on U_{ijk}, so that we really get a fiber bundle. It is clear that the standard flat connections agree on overlaps U_{ij}. Moreover, the local sections s_i assemble into a global section $s \colon M \to E$, since if $u \in U_{ij}$ then $s_j(u) = (u, \phi_j(u))$ gets identified with $(u, \phi_{ij}(\phi_j(u))) = (u, \phi_i(u)) = s_i(u)$ in E. We have thus constructed a flat bundle $E \to M$ with standard fiber X and a section that is everywhere transverse to the connection.

(ii) $\Longrightarrow$ (i) Now suppose we have a flat (G, X)-bundle $p \colon E \to M$, equipped with a section s transverse to the connection. Pick local trivializations $\psi_i \colon p^{-1}(U_i) \to U_i \times X$ of E, where without loss of generality we may assume each of the open sets $U_i \subseteq M$ is contractible. Since E is flat, the connection gives a horizontal foliation of E, and since the U_i are contractible we may choose the local trivializations ψ_i in such a way that the leaves of this foliation are locally the horizontal slices $U_i \times \{x\}$. Define a (G, X) structure on M

[2]This section is canonical: it is the pullback along ϕ_i of the diagonal section $\Delta \colon X \to X \times X$ of the product bundle $X \times X \to X$:

$$\begin{tikzcd} U_i \times X \arrow[r, "\phi_i \times 1_X"] \arrow[d] & X \times X \arrow[d] \\ U_i \arrow[r, "\phi_i"] & X \end{tikzcd}$$

using charts ϕ_i given by the following composite

$$\begin{tikzcd} & p^{-1}(U_i) \arrow[r, "\psi_i"] & U_i \times X \arrow[dr, "p_2"] & \\ U_i \arrow[ur, "s|U_i"] \arrow[rrr, "\phi_i"'] & & & X \end{tikzcd}$$

That s is transverse means the differential of $\psi_i \circ s|_{U_i}$ has an image complementary to the tangent space of the horizontal leaf $U_i \times \{x\}$. Thus nearby points in U_i get mapped to distinct leaves so that, choosing the U_i small enough, the composite $\phi_i = p_2 \circ \psi_i \circ s|_{U_i} \colon U_i \to X$ is injective. The local trivializations ψ_i give gluing maps γ_{ij} defined by $\psi_{ij}(u, x) = (u, \gamma_{ij}(x))$. Since the bundle has structure group G, the gluing maps live in G, and hence so do the transition functions ϕ_{ij}. Therefore, the collection of charts (U_i, ϕ_i) defines a (G, X)-structure on M.

(ii) $\Longrightarrow$ (iii) Given a flat (G, X)-bundle $p \colon E \to M$, equipped with a transverse section s, let Q be the bundle of generalized 'frames' $f \colon X \to E_m$. Then Q is a principal G-bundle. Since $E \cong Q \times_G X$ has a section, Q reduces in structure to a principal H-bundle, say P, by Proposition 1. We then have $Q \cong P \times_H G$, and we have only to show that the subbundle P is transverse to the flat connection on Q. P is the a pullback along the section s:

$$\begin{tikzcd} P \arrow[r, "\iota"] \arrow[d] & Q \arrow[d, "\pi"] \\ M \arrow[r, "s"] & Q \times_G G/H \end{tikzcd}$$

Locally, say in a trivialization over $U \subseteq M$, the bundle on the right is simply the canonical projection $U \times G \to U \times G/H$, where the flat connection on each is simply given by the horizontal foliation. If we differentiate all of the maps in the above commuting square and note that the vertical maps are submersions, we see that the images of $d\iota$ and ds are simultaneously transverse to the connections on Q and $Q \times_G G/H$, respectively.

(iii) $\Longrightarrow$ (ii) Finally, suppose we have a principal H bundle P with a flat connection on $Q = P \times_H G$, such that the image of the inclusion $P \to Q$ is a submanifold transverse to the foliation defined by the flat connection. The bundle $E = P \times_H G/H$

manifestly has a section, namely

$$\begin{aligned} s\colon M &\to P \times_H G/H \\ u &\mapsto [u, H]. \end{aligned}$$

We now show that E can be viewed as an associated bundle of Q, so that E inherits the flat connection from Q. In fact, given the reduction to P, we have a canonical isomorphism of fiber bundles

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

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. Thus E inherits the flat connection from Q. $\square$

3.3 Moduli space of geometric structures

In light of the previous section, we can describe geometric structures by flat bundles. It follows that we can describe geometric structures up to isomorphism by specifying the corresponding flat bundles up to gauge transformations. This gets us closer to relating geometric structures to the gauge theory.

In general, given a fiber bundle E with structure group G, let $\mathcal{A}(E)$ denote the space of connections on P, $\mathcal{A}_0(E)$ the subspace of flat connections, and $\mathcal{G}(E)$ the group of gauge transformations of the bundle. When the bundle is clear from the context, we write simply $\mathcal{A}, \mathcal{A}_0$, or $\mathcal{G}$. Gauge equivalence classes of flat connections on P thus live in $\mathcal{A}_0/\mathcal{G}$, the **moduli space of flat connections** on P.

The space $\mathcal{A}_0/\mathcal{G}$ is a bit difficult to handle. It is often more convenient to start by fixing a basepoint $* \in X$ and working with $\mathcal{A}_0/\mathcal{G}_0$, where

$$\mathcal{G}_0 = \{g \in \mathcal{G} \colon\ g(*) = 1\}.$$

This amounts to choosing a **frame** at $*$—an identification of the fiber over $*$ with the standard fiber, G itself in the case of a principal bundle. The group $\mathcal{G}/\mathcal{G}_0 \cong G$ acts on $\mathcal{A}_0/\mathcal{G}_0$ in a natural way. This lets us form $\mathcal{A}_0/\mathcal{G}$ as the quotient of the bigger space $\mathcal{A}_0/\mathcal{G}_0$ by this action of G.

The advantage of the space $\mathcal{A}_0/\mathcal{G}_0$ for a principal G bundle P is that any point $[A]$ in this space gives a homomorphism

$$\mathrm{hol}([A]) \colon \pi_1(X) \to G$$

which sends any homotopy class of loops $[\gamma]$ to the holonomy of A around γ. This gives a map

$$\mathrm{hol} \colon \mathcal{A}_0/\mathcal{G}_0 \to \hom(\pi_1(X), G)$$

which is known to be one-to-one. Note that G acts on $\hom(\pi_1(X), G)$ by conjugation:

$$(gf)(\gamma) = gf(\gamma)g^{-1}$$

where $f \colon \pi_1(X) \to G$ is any homomorphism. Moreover, the map hol is compatible with this group action:

$$\mathrm{hol}([gA]) = g\,\mathrm{hol}([A]).$$

So far we have fixed a principal G-bundle P. But, in gauge theory it is often better to treat this bundle as variable—part of the physical field along with the connection A. For example, path integrals in quantum chromodynamics involve a sum over bundles, which represent instantons. The mathematical advantage of treating P as variable is that all points of $\hom(\pi_1(X), G)$ are in the image of hol if we allow ourselves to vary P [59]. In fact, we have the following more general result [93]:

Proposition 3 *Let G be a Lie group, M a connected manifold with a basepoint, and X a G-manifold such that the action of G is effective. Then there is a one-to-one correspondence between:*

1. *homomorphisms $\pi_1(M) \to G$, and*

2. *gauge-equivalence classes of flat bundles over M with fiber X and structure group G, equipped with a frame at the basepoint.*

A point in this space represents a 'G-bundle with flat connection over X, mod gauge transformations that equal the identity at the basepoint'. Modding out by the rest of the gauge transformations we get a space known as the **moduli space of flat bundles**, $\hom(\pi_1(X), G)/G$. In the context of the theory of geometric structures, this same space has been called the **deformation space of geometric structures** [49].

3.4 Model geometries and exotic spacetimes

In Chapter 2 we have examined the geometry of several 3d homogeneous spacetimes. But there are other interesting 3d geometries as well. In fact, in the Riemannian case, Thurston has classified the homogeneous spaces which serve as model geometries for geometric 3-manifolds. He defines a 'model geometry' to be a Klein geometry $X = G/H$ such that [93]:

1. G/H is connected and simply connected;

2. H is compact;

3. G is maximal in the sense that it is not contained in any larger group of diffeomorphisms of X with compact point stabilizers;

4. there exists at least one compact manifold with the local geometry of (G/H), in the sense of Definition 6.

Some explanation of these criteria is in order. The first requirement, that the model geometry be connected and simply connected, keeps one from counting geometries as distinct when they are locally isometric. For example, it would be absurd to consider the disjoint union of two copies of the plane as a model geometry distinct from a single connected copy. Similarly, the Klein geometry $\mathrm{SO}(2,2)/\mathrm{SO}(2,1)$ and its universal cover, the 3d anti de Sitter spacetime, should not be considered separate model geometries. The canonical choice is to always let the universal cover represent the model geometry. Equivalently, we could define model geometries as equivalence classes of Klein geometries. Since the main purpose here is not the classification of geometries but the application to physics, we shall be somewhat relaxed about such issues, feeling free to refer to geometries as model geometries even when they do not satisfy this requirement.

The second requirement, that the point stabilizer subgroup be compact, is necessary to guarantee that the model be a Riemannian manifold. In fact, stabilizers of points in homogeneous Riemannian n-manifolds must be subgroups of $\mathrm{O}(n)$, which is compact. This requirement is of course too restrictive for more general 'model geometries' for Lorentzian physics: the Lorentz groups are noncompact.

The third requirement says we are not allowed to construct new model geometries by crippling the symmetry groups of old geometries. The symmetry groups must be maximal. The spacetime geometries as described in Chapter 2 also do not meet this requirement. In particular, since $\mathrm{SO}(p,q) \subset \mathrm{O}(p,q)$, we have not considered the maximal symmetry groups for these spacetimes. Technically, then, the model geometries on which these spacetimes are based are the unoriented versions, with larger symmetry groups.

The fourth requirement is obviously sensible if we wish to model compact manifolds on such geometries. For spacetime physics, this requirement is less obviously important, but it has the advantage of simplifying the classification considerably.

For two dimensions, there are just 3 model geometries satisfying properties 1-4: the sphere, the Euclidean plane, and the hyperbolic plane:

		G	H	$X = G/H$
1.	2-sphere	SO(3)	SO(2)	S^2
2.	Euclidean plane	ISO(2)	SO(2)	$\mathbb{R}^2$
3.	hyperbolic plane	SO(2, 1)	SO(2)	H^2

In three dimensions, there are eight, including the obvious analogs of the 2d geometries, products of 2d geometries with the unique 1d geometry, and three more exotic ones:

		G	H	$X = G/H$
1.	3-sphere	SO(4)	SO(3)	S^3
2.	Euclidean 3-space	ISO(3)	SO(3)	$\mathbb{R}^3$
3.	hyperbolic 3-space	SO(3, 1)	SO(3)	H^3
4.	spherical cylinder	$\mathrm{SO}(3) \times \mathbb{R}$	SO(2)	$S^2 \times \mathbb{R}$
5.	hyperbolic cylinder	$\mathrm{SO}(2,1) \times \mathbb{R}$	SO(2)	$\mathrm{H}^2 \times \mathbb{R}$
6.	'nilgeometry'	$\mathrm{Nil}^3 \rtimes \mathrm{SO}(2)$	SO(2)	Nil^3
7.	3d Lorentz group	$\mathrm{SL}(2,\mathbb{R}) \times \mathrm{SO}(2)$	SO(2)	$\mathrm{SL}(2,\mathbb{R})$
8.	'solvgeometry'	$\mathbb{R}^* \ltimes \mathbb{R}^2$	1	$\mathbb{R}^* \ltimes \mathbb{R}^2$

We emphasize that in the above charts, $(G, G/H$ does not necessarily satisfy axioms 1 and 3 in Thurston's definition of model geometry. In particular, we have ignored orientation-reversing isometries and not necessarily described the simply connected member of a given family of locally isometric geometries. We have simply given the simplest way to construct these homogeneous manifolds as coset spaces with the correct stabilizer dimension.

We might consider these as potential models for Riemannian 3d spacetimes. The first three are just the totally isotropic Riemannian spacetimes in the bottom row of the chart on p. 29, which have 3-dimensional point stabilizer groups. Four additional ones have 1-dimensional point stabilizers:

3d hyperbolic cylinder $(\mathrm{SO}(2,1) \times \mathbb{R})/\mathrm{SO}(2)$	3d spherical cylinder $(\mathrm{SO}(3) \times \mathbb{R})/\mathrm{SO}(2)$
nilgeometry $(\mathrm{Nil}^3 \rtimes \mathrm{SO}(2))/\mathrm{SO}(2)$	$\mathrm{SL}(2,\mathbb{R})$ geometry $P(\mathrm{SL}(2,\mathbb{R}) \times \mathrm{SO}(2))/\mathrm{SO}(2)$

The first two of these, the spherical and hyperbolic cylinders, describe nonrelativistic spacetimes in the strictest sense. They have a notion of 'absolute space', as in Newtonian gravity.[3] The time translation subgroup $\mathbb{R}$ gives a canonical way of moving from one slice of space to a later one. The time translations $\mathbb{R}$ form a normal subgroup, and factoring these out gives the symmetry group of absolute space: $\mathrm{SO}(3)$ in the sperical case, or $\mathrm{SO}(2,1)$ in the hyperbolic. Of course, these spacetimes have a common contraction limit to one with symmetry group $\mathrm{ISO}(2)\times\mathbb{R}$ and $\mathrm{SO}(2)$ as point stabilizer. The reason this is not considered a separate geometry in Thurston's classification is that it violates property 3: the resulting manifold is isometric with the Euclidean model $\mathrm{ISO}(3)/\mathrm{SO}(3)$, even though it has by convention a reduced group of symmetries. Nonetheless, for purposes of modeling spacetime we can certainly consider Euclidean space with a preferred time direction, obtaining the contraction family:

$$\begin{tikzcd}
\begin{array}{c}\textsf{3d hyperbolic cylinder}\\ (\mathrm{SO}(2,1)\times\mathbb{R})/\mathrm{SO}(2)\end{array} \arrow[r, "\text{contraction}"] \arrow[rr, bend left, leftrightarrow, "\text{Wick rotation}"] & \begin{array}{c}\textsf{3d Euclidean cylinder}\\ \mathrm{ISO}(2)\times\mathbb{R}/\mathrm{SO}(2)\end{array} & \begin{array}{c}\textsf{3d spherical cylinder}\\ (\mathrm{SO}(3)\times\mathbb{R})/\mathrm{SO}(2)\end{array} \arrow[l, "\text{contraction}"'] \\
\Lambda < 0 & \Lambda = 0 & \Lambda > 0
\end{tikzcd}$$

The next two geometries are slightly more exotic, but in involve Lie groups that are more or less familiar from physics. We have already seen that $\mathrm{SL}(2,\mathbb{R})$ is the double cover of the connected 3d Lorentz group $\mathrm{SO}_0(2,1)$, and can also be viewed as 3d anti de Sitter spacetime. But here we consider $\mathrm{SL}(2,\mathbb{R})$ not as Lorentzian, but as a Riemannian manifold. We explain this shortly. 'Nilgeometry' is the geometry of the 3d 'Heisenberg group', a Lie group whose Lie algebra satisfies relations equivalent to the canonical commutation relations of quantum mechanics.

Finally, there is an even more exotic model geometry in Thurston's classification, the so called 'solvgeometry'. The connected isometry group of this geometry has trivial point stabilizer, so in a spacetime modeled on this geometry, there are not even isometric rotations of 'space'. In the next section we describe these geometries in more detail.

3.4.1 3d model geometries

We have already discussed the three symmetric geometries, and the two cylinder geometries are need little clarification. In this section we describe the three exotic Reimannian model geometries.

[3] Rovelli has written a fascinating account, in his book [81], of Newtonian absolute space, as it stands in relation to more modern conceptions of spacetime.

geometry of $\widetilde{SL}(2,\mathbb{R})$

We have seen that $\widetilde{SL}(2,\mathbb{R})$ is isomorphic as a Lorentzian manifold to 3d anti de Sitter spacetime. But $\widetilde{SL}(2,\mathbb{R})$ also serves as an important homogeneous model for 3d *Riemannian* geometry. How is this? Even though $\widetilde{\mathrm{SL}}(2,\mathbb{R})$ is most naturally a Lorentzian manifold, any Lorentzian manifold M that admits a nonvanishing timelike vector field ∂_t can be turned into a Riemannian manifold as follows. We partially diagonalize the metric g at each point, writing

$$g = -dt_2 + g_S$$

where g_S is the positive definite metric defined pointwise on the orthogonal complements of ∂_t. We define a new positive definite metric on all of M by flipping the sign in the ∂_t direction:

$$\tilde{g} = dt^2 + g_S.$$

Let us apply this procedure to the group $\mathrm{SL}(2,\mathbb{R})$, and describe the groups of symmetries of the resulting homogeneous Riemannian manifold. Recall the following representation of the group:

$$\mathrm{SL}(2,\mathbb{R}) = \left\{ \begin{pmatrix} a+b & c+d \\ c-d & a-b \end{pmatrix} \;:\; a,b,c,d \in \mathbb{R},\; a^2-b^2-c^2+d^2=1 \right\}$$

The corresponding Lie algebra representation consists of all 2×2 traceless matrices:

$$\mathfrak{sl}(2,\mathbb{R}) = \left\{ \begin{pmatrix} p_1 & p_2+p_0 \\ p_2-p_0 & -p_1 \end{pmatrix} \;:\; p_0,p_1,p_2 \in \mathbb{R} \right\}$$

Using these matrix representations it is straightforward to compute the stabilizer of a unit timelike vector field. The obvious timelike vector at the identity, with respect to the Killing form, is

$$v_1 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in \mathfrak{sl}(2,\mathbb{R}).$$

We can push this vector field forward by left translation to get a vector field on all of $\mathrm{SL}(2,\mathbb{R})$,

$$v_h = hv_1 \in T_h\mathrm{SL}(2,\mathbb{R}).$$

where we are taking advantage of the embedding of $\mathrm{SL}(2,\mathbb{R})$ in $\mathrm{Mat}_{2\times 2}(\mathbb{R}) \cong \mathbb{R}^4$ to identify each tangent space with a vector subspace of $\mathbb{R}^4$. Explicitly:

$$v_1 = \begin{pmatrix} -c-d & a+b \\ -a+b & c-d \end{pmatrix} \in T_h\mathrm{SL}(2,\mathbb{R}) \qquad \text{where } h = \begin{pmatrix} a+b & c+d \\ c-d & a-b \end{pmatrix}$$

Recall that the full symmetry group of $\mathrm{SL}(2,\mathbb{R})$ as a Lorentzian manifold is $\mathrm{SL}(2,\mathbb{R}) \times \mathrm{SL}(2,\mathbb{R})/\{\pm 1\}$, where

$$(g, g') \in \mathrm{SL}(2,\mathbb{R}) \times \mathrm{SL}(2,\mathbb{R})$$

acts on $\mathrm{SL}(2,\mathbb{R})$ by

$$\begin{array}{cccc} (g,g')\colon & \mathrm{SL}(2,\mathbb{R}) & \to & \mathrm{SL}(2,\mathbb{R}) \\ & h & \mapsto & ghg'^{-1} \end{array}$$

Such a transformation of $\mathrm{SL}(2,\mathbb{R})$ preserves the *Riemannian* metric if and only if it preserves the vector field v. That is, $(g, g') \in \mathrm{SL}(2,\mathbb{R}) \times \mathrm{SL}(2,\mathbb{R})$ is a symmetry of Riemannian $\mathrm{SL}(2,\mathbb{R})$ if and only if

$$g v_h g'^{-1} = v_{ghg'^{-1}}$$

or equivalently,

$$g' v_1 = v_1 g'.$$

A direct computation shows that the subgroup of $\mathrm{SL}(2,\mathbb{R})$ commuting with the matrix v_1 is

$$\left\{ \begin{pmatrix} a & d \\ -d & a \end{pmatrix} \,:\, a, d \in \mathbb{R},\ a^2 + d^2 = 1 \right\} \cong \mathrm{SO}(2)$$

Thus, as a Riemannian homogeneous space, $\mathrm{SL}(2,\mathbb{R})$ may be described as

$$\mathrm{SL}(2,\mathbb{R}) = G/H$$

where

$$G = \mathrm{SL}(2,\mathbb{R}) \times SO(2)$$

and

$$H = \{(g,h) \in \mathrm{SL}(2,\mathbb{R}) \times \mathrm{SO}(2) : g = h\} \cong \mathrm{SO}(2).$$

It is worth mentioning another well-known way of getting the Riemannian structure on $\widetilde{\mathrm{SL}}(2,\mathbb{R})$ (see, for example, Scott [85]). The 3d Lorentz group $\mathrm{SO}_0(2,1) \cong \mathrm{PSL}(2,\mathbb{R})$ is isomorphic to the unit tangent bundle of velocity space, which is just the hyperbolic plane, H^2. Since H^2 is a Riemannian manifold, so is its unit tangent bundle $\mathrm{PSL}(2,\mathbb{R})$, and its metric may be pulled back to any covering group, in particular to the universal cover $\widetilde{\mathrm{SL}}(2,\mathbb{R})$. By taking quotients of the hyperbolic plane by discrete groups, the unit tangent bundle of any surface of genus at least two has a geometric structure modeled on $\widetilde{\mathrm{SL}}(2,\mathbb{R})$ [69].

nilgeometry

Nilgeometry is the geometry, as a Riemannian manifold, of the unique 3d connected and simply connected Lie group which is nilpotent but not abelian [93, p. 185]:

$$\mathrm{Nil}^3 := \left\{ \begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} \,:\, a, b, c \in \mathbb{R} \right\}$$

The operations of multiplication and inversion in Nil^3 are simple enough to be worth writing out in general:

$$\begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & a' & c' \\ 0 & 1 & b' \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & a + a' & c + ab' + c' \\ 0 & 1 & b + b' \\ 0 & 0 & 1 \end{pmatrix}$$

and

$$\begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -a & ab - c \\ 0 & 1 & -b \\ 0 & 0 & 1 \end{pmatrix}$$

From the multiplication rule, it is easy to see that the center of the Heisenberg group is the 1-parameter subgroup defined by $a = b = 0$:

$$Z(\mathrm{Nil}^3) = \left\{ \begin{pmatrix} 1 & 0 & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \,:\, c \in \mathbb{R} \right\}$$

The group Nil^3 is also called the Heisenberg group, since its Lie algebra $\mathfrak{nil}^3$ can be presented as a set of generators:

$$p = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad q = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \qquad h = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

satisfying relations which are formally the canonical commutation relations of quantum mechanics:

$$[p, q] = h \qquad [p, h] = [q, h] = 0$$

To get a metric on Nil^3, we choose an inner product on $\mathfrak{nil}^3$ and left translate to each tangent space. In particular, if we choose the inner product such that the matrices $\{p, q, h\}$ given above form an orthonormal basis of $\mathfrak{nil}^3$, and if

$$g := \begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} \in \mathrm{Nil}^3,$$

then an orthonormal basis for the tangent space $T_g\mathrm{Nil}^3$ is $\{p_g, q_g, h_g\}$ where

$$p_g := gp = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = p$$

$$q_g := gq = \begin{pmatrix} 0 & 0 & a \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} = q + ah$$

$$h_g := gh = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = h$$

This metric is invariant by construction under left translation by group elements. A direct calculation shows that the only right translations preserving the metric are translations by central elements, so these don't give any new symmetries. In fact the rest of the symmetries come from a non-obvious circle's worth of outer automorphisms of Nil^3. For any $\theta \in \mathbb{R}/2\pi\mathbb{Z}$ we get an automorphism described by the transformation [85, p. 468]

$$\begin{array}{ccc} a & \mapsto & a\cos\theta - b\sin\theta \\ b & \mapsto & a\sin\theta + b\cos\theta \\ c & \mapsto & c + \frac{1}{2}\sin\theta[(a^2 - b^2)\cos\theta - 2ab\sin\theta] \end{array}$$

where a, b, c are the matrix entries of an arbitrary element

$$g := \begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} \in \mathrm{Nil}^3.$$

This action of $\mathrm{SO}(2)$ on Nil preserves the metric, and stabilizes the identity, since it acts as automorphisms. The only other isometries for the metric come from the automorphism

$$\begin{array}{ccc} a & \mapsto & -a \\ b & \mapsto & -b \\ c & \mapsto & c \end{array}$$

So in fact we have $\mathrm{O}(2) \cong \mathrm{SO}(2) \rtimes \mathbb{Z}/2$, so the full isometry group of Nil^3 is the group generated by Nil^3 itself acting as left translations, together with the $\mathrm{O}(2)$ group of automorphisms. We can describe Nil^3 as a Klein geometry as:

$$\mathrm{Nil}^3 \cong (\mathrm{Nil}^3 \rtimes \mathrm{O}(2))/\mathrm{O}(2),$$

or if we want only the oriented geometry,

$$\mathrm{Nil}^3 \cong (\mathrm{Nil}^3 \rtimes \mathrm{SO}(2))/\mathrm{SO}(2),$$

solvgeometry

Solvgeometry is the geometry of the connected Poincaré group for $1+1$ dimensional spacetime, $\mathrm{ISO}_0(1,1)$, considered as a Riemannian manifold [69]. Since $\mathrm{SO}_0(1,1)$ is isomorphic to the real line, the 2d Poincaré group is really

$$\mathrm{ISO}_0(1,1) \cong \mathbb{R} \ltimes \mathbb{R}^2,$$

where a boost of rapidity $\rho \in \mathbb{R}$ acts on $\mathbb{R}^2$ as matrix multiplication by

$$\begin{pmatrix} \cosh\rho & \sinh\rho \\ \sinh\rho & \cosh\rho \end{pmatrix}.$$

Transforming to a lightlike basis diagonalizes this matrix so that the action of $\mathbb{R}$ on $\mathbb{R}^2$ becomes

$$(u,v) \mapsto (e^{\rho}u, e^{-\rho}v)$$

where $u = (t-x)/\sqrt{2}$ and $v = (t+x)/\sqrt{2}$ are lightlike coordinates on $\mathbb{R}^2$. To get a Riemannian metric on the group $\mathbb{R} \ltimes \mathbb{R}^2$, we simply pick one at the origin and left translate by group elements. This metric is obviously preserved by left translations, but it is not preserved by any right translations. In fact, the geometry is homogeneous but maximally anisotropic: the connected component of the isometry group is Nil^3 itself, hence has trivial point stabilizer subgroup. The full isometry group G has eight components, with $G/G_0 \cong D_4$, the dihedral group of the square. [85].

We ignore the discrete isometries and consider nilgeometry as the Klein geometry $\mathrm{Nil}^3/1$. The product in $\mathbb{R} \ltimes \mathbb{R}^2$ is given by the formula

$$(\rho,u,v)(\rho',u',v') = (\rho+\rho', u + e^{\rho}u', v + e^{-rho}v')$$

while the inverse of an element ρ, u, v is

$$(\rho,u,v)^{-1} = (-\rho, -e^{-\rho}u, -e^{\rho}v).$$

It is sometimes convenient to use the matrix representation:

$$\mathbb{R}^* \ltimes \mathbb{R}^2 \cong \left\{ \begin{pmatrix} 1 & 0 & 0 \\ u & e^{\rho} & 0 \\ v & 0 & e^{-\rho} \end{pmatrix} \;\middle|\; \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2, t \in \mathbb{R}^* \right\}$$

The Lie Algebra:

$$\mathfrak{g} = \left\{ \begin{pmatrix} 0 & 0 & 0 \\ p_x & E & 0 \\ p_y & 0 & -E \end{pmatrix} \Big| x, y, t \in \mathbb{R} \right\}$$

Generators:

$$X = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \qquad Y = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad T = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

$$[X, Y] = 0 \qquad [Y, T] = Y \qquad [T, X] = X$$

so $[\mathfrak{g}, \mathfrak{g}]$ is the abelian Lie algebra with basis $\{X, Y\}$. This implies solvability, since a Lie algebra is solvable if and only if the Killing form vanishes on $[\mathfrak{g}, \mathfrak{g}] \subseteq \mathfrak{g}$. The adjoint action sends

$$\mathrm{Ad}((e^t, x, y))(E, p_x, p_y) = (E, e^t p_x - xE, e^{-t} p_y + yE)$$

It is not hard to check that any Ad-invariant quadratic form on the Lie algebra must be proportional to E^2.

3.4.2 Exotic Lorentzian spacetimes

So far, we have said nothing of exotic 3d *Lorentzian* spacetimes, though there are bound to be interesting examples to consider here as well. Let us just mention one possibility. We saw that the Riemannian geometry of $\mathrm{SL}(2, \mathbb{R})$ is obtained by starting with the 3d anti de Sitter spacetime, picking a timelike vector field, and flipping the sign of the metric in that direction. We could do exactly the opposite of this procedure using the Wick rotatated version of $\mathrm{SL}(2, \mathbb{R})$, namely SU(2). As a Riemannian manifold SU(2) is just the 3-sphere geometry we have considered. We get a nonvanishing vector field on SU(2) by differentiating the circle group action preserving a Hopf fibration. If we flip the sign of the Riemannian metric in the direction of this field, we obtain a 3d Lorentzian spacetime where every point lies on a closed timelike loop. Unlike anti de Sitter spacetime, however, these timelike loops cannot be removed by passing to a universal cover, since S^3 is already simply connected. 'Time' in this spacetime is a circle; 'space' is locally S^2 but there is no foliation into spacelike slices, since the Hopf fibration admits no global section. This rather bizarre spacetime is reminiscent of Gödel's famous solutions of 4d general relativity [47].

The Riemannian $\mathrm{SL}(2,\mathbb{R})$ geometry and this Hopf spacetime are evidently related by a Wick rotation:

$$\begin{tikzcd} \begin{array}{c}\mathrm{SL}(2,\mathbb{R})\ \textsf{geometry}\\ P(\mathrm{SL}(2,\mathbb{R})\times \mathrm{SO}(2))/\mathrm{SO}(2)\end{array} \arrow[rr, bend left, leftrightarrow, "\text{Wick rotation}"] & & \begin{array}{c}\textsf{`Hopf spacetime'}\\ P(\mathrm{SU}(2)\times \mathrm{SO}(2))/\mathrm{SO}(2)\end{array} \end{tikzcd}$$

An interesting question to consider is what the common contraction limit of these two geometries is. One obvious candidate to try is nilgeometry, since it fibers over E^2, while $\mathrm{SL}(2,\mathbb{R})$ and $\mathrm{SU}(2)$ fiber respectively over H^2 and S^2, but we shall not pursue this idea further here.

Chapter 4

Topological gauge theories

4.1 From general relativity to BF theory

General relativity in 3d spacetime is drastically simpler than 4d general relativity, both classically and quantum mechanically [24, 32, 99]. In fact, 3d gravity is a 'topological field theory' in the sense that all solutions of Einstein's equations in 3 spacetime dimensions are locally the same up to gauge transformations. BF theory may be viewed as an alternative generalization of 3d general relativity, which retains its topological character in any dimension.

In arbitrary spacetime dimension n, general relativity may be described in the so-called 'Palatini formalism' as follows. Spacetime is represented by an n-dimensional oriented smooth manifold M. From the Palatini perspective, M does not come equipped with a metric, but acquires one by pulling back along a map of vector bundles called the **coframe field** e:

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

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 is also equipped with a *fixed* metric η of signature (p, q), where $p + q = n$.[1] This idea shifts the focus in general relativity from

[1] The case of greatest physical interest is of course the case $n = 4$ of Lorentzian signature $(3, 1)$. However, there is also much theoretical interest in other dimensions, and in the case of Riemannian signature $(n, 0)$.

the metric to the coframe field, which now becomes the key dynamical variable. This might seem artificial at first sight, but this perspective is more suitable for describing gravity as a gauge theory (see the book by Ashtekar [5] for a discussion). In fact, as explained by Rovelli [81], from a modern perspective it is the coframe field that deserves to be called the 'gravitational field' in general relativity.

Besides the coframe field, the Palatini formalism uses one other physical field: a connection A on the vector bundle $\mathcal{T}$. There is no *a priori* assumption that this connection is compatible with the metric on $\mathcal{T}$. The Palatini Lagrangian for general relativity, without source terms, is then

$$\mathrm{tr}\left(\underbrace{e \wedge \cdots \wedge e}_{n-2} \wedge F\right) \tag{4.1}$$

where F is the curvature of A. Here the wedge product applies both to differential form parts and to the vector bundles in which the forms take values. That is, the wedge product is the obvious bilinear map

$$\wedge \colon \Omega^k(M, \Lambda^s\mathcal{T}) \otimes \Omega^\ell(M, \Lambda^t\mathcal{T}) \to \Omega^{k+\ell}(M, \Lambda^{s+t}\mathcal{T})$$

given by antisymmetrizing both spacetime and internal indices. Thus, since e is a $\mathcal{T}$-valued 1-form, the $(n-2)$-fold wedge product $e \wedge \cdots \wedge e$ is an $(n-2)$-form with values in the vector bundle $\Lambda^{n-2}\mathcal{T}$. The curvature F is a 2-form with values in $\Lambda^2\mathcal{T}$. Hence, the expression in parentheses in (4.1) is a $\Lambda^n\mathcal{T}$-valued n-form on M, and the 'trace' is really a map that turns such a form into an ordinary real-valued form:

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

by contraction with the volume form on the internal space $\mathcal{T}$, which exists since $\mathcal{T}$ has a metric and orientation.

The classical field equations deriving from the Palatini Lagrangian (4.1) are

$$\underbrace{e \wedge \cdots \wedge e}_{n-3} \wedge F = 0 \qquad \underbrace{e \wedge \cdots \wedge e}_{n-3} \wedge d_A e = 0$$

We postpone a detailed discussion of the Palatini action, including why this really is the action for general relativity, until Chapter 11. Our interest here is simply to make explicit the relationship between the Palatini Lagrangian and the Lagrangian for the topological gauge theory 'BF theory'.

To pass to BF theory, we begin by recasting our fields e and A in the language of principal bundles. Let P be the bundle of orthonormal frames $\mathcal{T}_x \to \mathbb{R}^{p,q}$. This is a principal $\mathrm{SO}(p,q)$-bundle over M, which is isomorphic via pullback along e to the frame bundle FM. Thus $\mathcal{T}$ is the bundle associated to P via the defining representation of $\mathrm{SO}(p,q)$ on the vector space $\mathbb{R}^{p,q}$:

$$\mathcal{T} = P \times_{SO(p,q)} \mathbb{R}^{p,q}.$$

Now our connection A may be viewed as a connection on P itself rather than on $\mathcal{T}$. In BF theory, we will simply take A to be a connection on a principal bundle.

To understand what takes the place of the coframe field in BF theory, first define $\mathrm{Ad}(P)$ to be the vector bundle associated to P via the adjoint action of $\mathrm{SO}(p,q)$ on its Lie algebra.[2] We then have the following result.

Lemma 4 *Given the framework described above for the Palatini formalism, the (n-2)nd exterior power of the coframe field*

$$\underbrace{e \wedge \cdots \wedge e}_{n-2}$$

corresponds canonically to an $(n-2)$*-form on* M *with values in* $\mathrm{Ad}(P)$.

Proof: For the coframe field, note that since e transforms at each point $x \in M$ according to the defining representation of $\mathrm{SO}(p,q)$ on $\mathcal{T}_x \cong \mathbb{R}^{p,q}$, the field

$$\underbrace{e \wedge \cdots \wedge e}_{n-2}$$

transforms under the induced representation ρ on $\Lambda^{n-2}\mathbb{R}^{p,q}$. But this induced representation is equivalent to the adjoint representation of $\mathrm{SO}(p,q)$ on its Lie algebra. That is, for each $g \in \mathrm{SO}(p,q)$, the diagram:

$$\begin{tikzcd}
\Lambda^{n-2}\mathbb{R}^{p,q} \arrow[r, "\rho(g)"] \arrow[d, "\alpha"'] & \Lambda^{n-2}\mathbb{R}^{p,q} \arrow[d, "\alpha"] \\
\mathfrak{so}(p,q) \arrow[r, "\mathrm{Ad}(g)"'] & \mathfrak{so}(p,q)
\end{tikzcd}$$

[2]In general, if P is a principal G-bundle, $\mathrm{Ad}(P)$ is the vector bundle whose standard fiber is the Lie algebra $\mathfrak{g}$ of G, associated to P via the adjoint action:

$$\mathrm{Ad}(P) = P \times_{\mathrm{Ad}} \mathfrak{g} := \frac{P \times \mathfrak{g}}{(p,v) \sim (pg^{-1}, Ad(g)v)\ \forall g \in G}$$

commutes, where the intertwining operator $\alpha\colon \Lambda^{n-2}\mathbb{R}^{p,q} \to \mathfrak{so}(p,q)$ is actually the composite of the Hodge star operator (defined using the chosen orientation) on $\Lambda\mathbb{R}^{p,q}$ and the natural correspondence between bivectors and infinitesimal pseudorotations:

$$\begin{tikzcd} \Lambda^{p+q-2}\mathbb{R}^{p,q} \arrow[r, "\star"] \arrow[rr, bend left, "\alpha"] & \Lambda^2\mathbb{R}^{p,q} \arrow[r, "\sim"] & \mathfrak{so}(p,q) \end{tikzcd}$$

□

Using this lemma, we pass from the Palatini Lagrangian (4.1) to the Lagrangian

$$\mathrm{tr}\ (E \wedge F) \tag{4.2}$$

where E is an $\mathrm{Ad}(P)$-valued $(n-2)$-form which takes the place of $e \wedge \cdots \wedge e$, and the trace now comes from the Killing form on $\mathrm{SO}(p,q)$. This is the Lagrangian for **$\mathbf{SO}(p,q)$ BF theory**. At first, it may seem like simply the Lagrangian for general relativity written in a slightly more sophisticated form. However, we now allow E to be an *arbitrary* $(n-2)$ form, not necessarily of the form $e\wedge\cdots\wedge e$. Perhaps more to the point, in deriving the equations of motion for this new Lagrangian, we minimize the action with respect to *arbitrary* variations of E, not just variations coming from the variation of e. General relativity may thus be viewed as $\mathrm{SO}(p,q)$ BF theory subject to the constraint $E = e \wedge \cdots \wedge e$. Notice that this constraint is no constraint at all when $n = p + q = 3$: *(2+1)-dimensional gravity is an* $\mathrm{SO}(2,1)$ *BF theory.* It has been shown that this constraint may be imposed in a natural way to recover general relativity in any dimension from BF theory, by adding a term to the Lagrangian [42].

From a mathematical perspective, one advantage of the BF Lagrangian over the Lagrangian for full-fledged general relativity is the way it generalizes to gauge groups other than $\mathrm{SO}(p,q)$. In general we define BF theory as follows [7]. Let G be any Lie group whose Lie algebra $\mathfrak{g}$ is equipped with a nondegenerate Ad-invariant bilinear form

$$\langle \cdot, \cdot \rangle \colon \mathfrak{g} \times \mathfrak{g} \to \mathbb{R}$$

Let 'spacetime' be described by an n-dimensional manifold M and choose a principal G-bundle $P \to M$. The fields of BF theory are:

$$\begin{aligned} A &\quad \text{— a connection on } P \to M \\ E &\quad \text{— an } \mathrm{ad}(P)\text{-valued } (n-2)\text{-form} \end{aligned}$$

The Lagrangian of BF theory is then

$$L = \mathrm{tr}\,(E \wedge F)$$

where the curvature $F = dA + \frac{1}{2}[A, A]$ is an $\mathrm{Ad}(P)$-valued 2-form, and

$$\mathrm{tr}\,(\cdot \wedge \cdot)\colon \Omega^{n-2}(M, \mathrm{Ad}(P)) \times \Omega^2(M, \mathrm{Ad}(P)) \to \Omega^n(M, \mathbb{R})$$

denotes the operation that wedges differential form parts and, at each point, applies the bilinear form $\langle \cdot, \cdot \rangle$ to Lie algebra parts. The notation 'tr ' is retained since this bilinear form is often the Killing form.

The equations of motion are derived by taking the critical points of the action[3]

$$S = \int_M \mathrm{tr}\ (E \wedge F)\,.$$

Taking the variation of the action gives:

$$\begin{aligned}
\delta S &= \int_M \mathrm{tr}\ (\delta E \wedge F + E \wedge d_A \delta A) \\
&= (-1)^n \int_{\partial M} \mathrm{tr}\ (E \wedge \delta A) + \int_M \mathrm{tr}\ (\delta E \wedge F - (-1)^n d_A E \wedge \delta A)
\end{aligned}$$

Here we have used the identity[4]

$$\delta F = d_A \delta A \tag{4.3}$$

and an integration by parts. So, when the boundary term vanishes, the field equations for BF theory just say

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

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

In 3 and 4 dimensions, where the degree of E divides the dimension of spacetime, it is possible to add a 'cosmological constant term' to the BF Lagrangian. In 3 spacetime dimensions, the BF action with cosmological term takes the form

$$S = \int_M \mathrm{tr}\ \left(E \wedge F + \frac{\Lambda}{3} E \wedge E \wedge E\right).$$

[3]In some sense, the 'action' here is a purely formal entity—it is not generally a convergent integral. However, convergence of S is not what is strictly needed, at least in the classical theory. What we need is for the variation δS to be a convergent integral for all compactly supported smooth variations $\delta A, \delta E$.

[4]The proof of this identity is straightforward: $\delta F = \delta(dA + \frac{1}{2}[A, A]) = d\delta A + \frac{1}{2}[\delta A, A] + \frac{1}{2}[A, \delta A] = d\delta A + [A, \delta A] = d_A \delta A$, using the superbracket of $\mathfrak{g}$-valued forms.

Varying this action gives

$$\begin{aligned}
\delta S &= \int_M \mathrm{tr}\ (E \wedge d_A \delta A + \delta E \wedge (F + \Lambda E \wedge E)) \\
&= -\int_{\partial M} \mathrm{tr}\ (E \wedge \delta A) + \int_M \mathrm{tr}\ (d_A E \wedge \delta A + \delta E \wedge (F + \Lambda E \wedge E))
\end{aligned}$$

where we have used the identity $\delta F = d_A(\delta A)$ and hence, when the boundary term vanishes, the equations of motion

$$F + \Lambda E \wedge E = 0 \qquad d_A E = 0.$$

In 4 dimensions, since the E field is a 2-form, we may add to the action a cosmological term of the form $E \wedge E$:

$$S = \int_M \mathrm{tr}\ \left(E \wedge F + \frac{\alpha}{2} E \wedge E\right)$$

In this case, we get the equations of motion:

$$\begin{aligned}
F - \alpha E &= 0 \\
d_A E &= 0
\end{aligned}$$

but the second equation is implied by the first, provided $\alpha \neq 0$, by the Bianchi identity $d_A F = 0$. We thus reduce to a single equation of motion:

$$F = \alpha E$$

The connection A may thus be chosen arbitrarily, with $E = \frac{1}{\alpha} F[A]$ a consequence of the choice of A. This brings out a sharp contrast between BF theory with nonzero cosmological constant term and ordinary BF theory. When $\alpha = 0$ the classical theory involves a flat connections; for $\alpha \neq 0$ any connection will do.

4.2 BF theory and geometric structures

The classical field equations for BF theory are

$$F = 0 \qquad d_A E = 0$$

In light of the relationship between flat bundles and geometric structures (Theorem 2, p. 43), the equation $F = 0$ immediately suggests one way solutions of BF theory are related to geometric structures. If spacetime is n-dimensional, choose an n-dimensional model Klein geometry G/H, and consider BF theory with gauge group G, assuming we are working over a principal bundle that admits a reduction of structure to the subgroup H. Theorem 2 implies a solution of the equation $F = 0$ from BF theory will then give a geometric structure on spacetime, modeled on G/H, provided we can choose a reduction to an H-bundle that is everywhere transverse to the flat connection. Even if we can't pick a reduction that is *everywhere* transverse, we may be able to do it over an open dense subset of spacetime. We can interpret this as describing a spacetime that is locally isometric to the model Klein geometry *except* at certain 'singularities'.

In fact, this way of getting geometric structures out of BF theory is related to the MacDowell–Mansouri approach to gravity. As noted in the Introduction, this theory involves a connection with gauge group $G = \mathrm{SO}(4,1)$. It may be written as a BF theory, but with an additional term that breaks the symmetry down to $H = \mathrm{SO}(3,1)$. This theory is not a topological gauge theory: it has local degrees of freedom. Nonetheless, it has "flat" solutions (in the generalized sense of flatness relative to a Klein model) which are nothing but geometric structures modeled on de Sitter spacetime $\mathrm{SO}(4,1)/\mathrm{SO}(3,1)$. The more generic solutions of MacDowell–Mansouri gravity are just "wavy" versions of these geometric structures, or in other words Cartan geometries. But this subject will have to wait until Part III.

In 3d spacetime, there is a more well–known way in which solutions of BF theory correspond to geometric structures. The basic example is 3d general relativity, viewed as BF theory with gauge group $H = \mathrm{SO}_0(2,1)$ [24]. For this application, let us go back to calling the connection and coframe field by their more standard names in general relativity, ω and e. If P is our principal $\mathrm{SO}_0(2,1)$-bundle, then we have the fields

$$\begin{array}{ll} \omega & \text{— a connection on } P \to M \\ e & \text{— an } \mathrm{ad}(P)\text{-valued 1-form} \end{array}$$

The equations of BF theory are then the usual equations of general relativity in 3 dimensions:

$$R[\omega] = 0 \qquad d_\omega e = 0 \tag{4.4}$$

Locally, both of these fields are $\mathfrak{so}(2,1)$-valued 1-forms. The trick is combining them into a single connection, not for the group $\mathrm{SO}_0(2,1)$, but for the Poincaré group

$\mathrm{ISO}_0(2,1)$. This means using the isomorphism

$$\mathfrak{so}(2,1) \cong \mathbb{R}^{2,1},$$

which gives

$$\mathrm{ISO}_0(2,1) \cong \mathrm{SO}_0(2,1) \ltimes \mathfrak{so}(2,1).$$

This lets us split the Lie algebra into vector subspaces

$$\mathfrak{iso}(2,1) \cong \mathfrak{so}(2,1) \oplus \mathfrak{so}(2,1)_{\mathrm{ab}}$$

in an $\mathrm{AD}(\mathrm{SO}_0(2,1))$-invariant way. We then think of ω and e as taking values respectively in these two subspaces.

The curvature F of this connection $A = \omega + E$ is given by

$$\begin{aligned} F[A] &= dA + \tfrac{1}{2}[A,A] \\ &= \big(d\omega + \tfrac{1}{2}[\omega,\omega] + \tfrac{1}{2}[e,e]\big) + (de + [\omega, e]) \\ &= R[\omega] + d_\omega e \\ &= 0 \end{aligned}$$

where we have used (4.4) and the fact that e takes values in an abelian Lie algebra, $\mathbb{R}^3$ or equivalently the Lie algebra of the additive group $\mathfrak{so}(2,1)$. Thus the fields ω and e assemble to give a *flat* $\mathrm{ISO}_0(2,1)$-connnection. The local connection 1-forms and curvature 2-forms can be summarized diagrammatically as follows:

$$\begin{tikzcd} & & \mathfrak{so}(2,1) \\ TU \arrow[urr, bend left, "\omega"] \arrow[r, "A"] \arrow[drr, bend right, "e"'] & \mathfrak{iso}(2,1) \arrow[ur] \arrow[dr] & \\ & & \mathfrak{so}(2,1) \end{tikzcd} \qquad \begin{tikzcd} & & \mathfrak{so}(2,1) \\ \Lambda^2(TU) \arrow[urr, bend left, "R"] \arrow[r, "F"] \arrow[drr, bend right, "d_\omega e"'] & \mathfrak{iso}(2,1) \arrow[ur] \arrow[dr] & \\ & & \mathfrak{so}(2,1) \end{tikzcd}$$

Flat $\mathrm{ISO}(2,1)$ connections A constructed in this way already live on a bundle that has been reduced to the stabilizer $\mathrm{SO}(2,1)$, since we started with a BF theory with this gauge group. When A is transverse (perhaps just away from certain singular points), we get a geometric structure modeled on 3d Minkowski spacetime $\mathrm{ISO}_0(2,1)/\mathrm{SO}_0(2,1)$. More generally, using the BF formulation of gravity with cosmological constant term, one can again subsume the ω and e fields into a connection A, but now giving geometric structures modeled on 3d de Sitter $\mathrm{SO}_0(3,1)/\mathrm{SO}_0(2,1)$ or anti de Sitter $\mathrm{SO}_0(2,2)/\mathrm{SO}_0(2,1)$ models, depending on the sign of the cosmological constant. Geometric structures of all of these kinds have been studied rather extensively as solutions of 3d Einstein gravity; see the book by Carlip [24] for some nice examples and further references. We will see more about the $\Lambda \neq 0$ cases soon, when we review the Chern–Simons formulation of 3d general relativity, which is really the most natural setting for these extended connections. For now, we turn to higher dimensional generalization.

We consider BF theory in n-dimensional spacetime, with gauge group $H = \mathrm{SO}(p,q)$, such that $p+q=n$. Let us write the connection as ω, so we have

ω — a connection on a principal $\mathrm{SO}(p,q)$ bundle $P \to M$

E — an $\mathrm{Ad}(P)$-valued $n-2$-form

satisfying the equations:

$$R[\omega] = 0 \qquad d_\omega E = 0 \tag{4.5}$$

In general, there is no way of combining the fields ω and E as we did in the 3d case, since E is not a 1-form. However, we can describe an important class of solutions in a related way. To do this, choose a 1-form e on spacetime with values in the vector bundle associated to P by the defining representation of $\mathrm{SO}(p,q)$:

$$e \in \Omega^1(M) \otimes (P \times_{\mathrm{SO}(p,q)} \mathbb{R}^{p,q})$$

and assume e is covariantly constant:

$$d_\omega e = 0.$$

We then define

$$E = \underbrace{e \wedge \cdots \wedge e}_{n-2}$$

This is an $(n-2)$-form with values in the induced vector bundle whose standard fiber is $\Lambda^{n-2}\mathbb{R}^{p,q}$:

$$E \in \Omega^{n-2}(M) \otimes (P \times_{\mathrm{SO}(p,q)} \Lambda^{n-2}\mathbb{R}^{p,q}).$$

But this vector bundle is equivalent, by Hodge duality, to the associated bundle with standard fiber $\Lambda^2\mathbb{R}^{p,q} \cong \mathfrak{so}(p,q)$, namely the bundle $\mathrm{Ad}(P)$. Since the equation $d_\omega E = 0$ follows from $d_\omega e = 0$, we thus have a solution of BF theory. Now everything goes through as before:

we can combine ω and e into an $\mathfrak{iso}(p,q)$ connection A using the the invariant splitting, as vector spaces,

$$\mathfrak{iso}(p,q) \cong \mathfrak{so}(p,q) \oplus \mathbb{R}^{p,q}.$$

Whenever A is transverse to P, we get a geometric structure on spacetime, modeled on the pseudo-Riemannian space $\mathrm{ISO}(p,q)/\mathrm{SO}(p,q)$ of signature (p,q).

4.3 Generalized 3d gravities

We have seen that the trick to writing 3d general relativity as a BF theory lies in the equivalence between the adjoint representation of $\mathrm{SO}(2,1)$ and its obvious representation on the 3d Minkowski vector space $\mathbb{R}^{2,1}$. Indeed, it is the isomorphism $\mathfrak{so}(2,1) \cong \mathbb{R}^{2,1}$ that lets us write the coframe field, which is locally $\mathbb{R}^{2,1}$-valued, as an $\mathfrak{so}(2,1)$-valued 1-form. But this is not strictly what is necessary to define an action like that of 3d general relativity. What *is* needed is a pairing

$$\langle e, F \rangle$$

between an appropriate generalization of the coframe field and the curvature of a connection. From the perspective of Kleinian geometry, the coframe field should take values in the infinitesimal geometry $\mathfrak{g}/\mathfrak{h}$. This idea leads to a more general theory that includes 3d gravity but also some interesting related theories, each of which corresponds to a different kind of geometric structure.

The types of model spacetimes we shall be interested in are Riemannian or Lorentzian 'symmetric spaces'. In fact, for our purposes, the following infinitesimal version of symmetric spaces will suffice [60]:

Definition 7 *A* **symmetric lie algebra** $\mathfrak{g}$ *is a Lie algebra equipped with a spitting (as vector spaces):*

$$\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{p}$$

such that

$$[\mathfrak{h},\mathfrak{h}] \subseteq \mathfrak{h}$$
$$[\mathfrak{h},\mathfrak{p}] \subseteq \mathfrak{p}$$
$$[\mathfrak{p},\mathfrak{p}] \subseteq \mathfrak{h}$$

The elements of $\mathfrak{p}$ *are called* **transvections**.

The conditions on the bracket in this definition say that $\mathfrak{h}$ is a subalgebra, $\mathfrak{p}$ is $\mathfrak{h}$-invariant, and the bracket of transvections lives in the subalgebra $\mathfrak{h}$. If in this definition we required further that the Killing form be negative definite on $\mathfrak{h}$ and positive definite on $\mathfrak{p}$, we would have the definition of a Cartan decomposition of the Lie algebra $\mathfrak{g}$, but in the cases we are most interested in this does not hold. For our purposes, a **symmetric space** will simply be a homogeneous space G/H, such that $(\mathfrak{g}, \mathfrak{h}, \mathfrak{p})$ is a symmetric Lie algebra for a suitable choice of $\mathfrak{p}$.

In fact all of the homogeneous spacetimes in the diagram on p. 29 are symmetric spaces. One easy way to see this is using the following more general result.

Proposition 5 *Given a matrix Lie group $G \subseteq \mathrm{GL}(n, \mathbb{C})$, let $H \subseteq G$ be a closed subgroup such that the Lie algebra $\mathfrak{h}$ consists of all block-diagonal matrices in $\mathfrak{g}$ of the form:*

$$\begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix} \in \mathfrak{g} \qquad X \in \mathfrak{gl}(m, \mathbb{C}), Y \in \mathfrak{gl}(n-m, \mathbb{C})$$

for fixed $m \leq n$. Then the homogeneous space G/H is a symmetric space.

Proof: The proof is an easy calculation. Define the complement $\mathfrak{p}$ of $\mathfrak{h}$ to consist of all elements of $\mathfrak{g}$ of the form

$$\begin{pmatrix} 0 & A \\ B & 0 \end{pmatrix}.$$

The bracket of two generic elements of $\mathfrak{g}$ is then:

$$\left[\begin{pmatrix} X & A \\ B & Y \end{pmatrix}, \begin{pmatrix} X' & A' \\ B' & Y' \end{pmatrix}\right] = \begin{pmatrix} [X, X'] + AB' - A'B & XA' - X'B + AY' - A'Y \\ BX' - B'X + YB' - Y'B & [D, D'] + BA' - B'A \end{pmatrix}$$

From this formula it is easy to see that $[\mathfrak{h}, \mathfrak{h}] \subseteq \mathfrak{h}$, $[\mathfrak{h}, \mathfrak{p}] \subseteq \mathfrak{p}$, and $[\mathfrak{p}, \mathfrak{p}] \subseteq \mathfrak{h}$. $\square$

This proposition gives us a wide variety of examples of symmetric spaces, including:

1. the 9 basic homogeneous spacetimes we have discussed in Section 2.3.1:
   - anti de Sitter, Minkowski, de Sitter
   - hyperbolic Galilean, Galilean, spherical Galilean
   - hyperbolic, Euclidean, spherical
2. The hyperbolic and spherical cylinders $S^n \times \mathrm{H}^1 = (\mathrm{SO}(n+1) \times \mathbb{R})/\mathrm{SO}(n)$

3. Real and complex projective spaces $\mathbb{R}P^n = SO(n+1)/O(n)$, $\mathbb{C}P^n = \mathrm{SU}(n+1)/\mathrm{U}(n)$

These are all symmetric spaces.

For generalizing 3d gravity, the nice feature of using a symmetric model geometry becomes apparent when we write out the equation of the curvature. Suppose G/H is a symmetric space, and we have a G-connection $A = \omega + e$, where ω takes values in $\mathfrak{h}$, e in $\mathfrak{p}$. The curvature F of this connection is:

$$\begin{aligned} F[A] &= dA + \tfrac{1}{2}[A, A] \\ &= \big(d\omega + \tfrac{1}{2}[\omega, \omega] + \tfrac{1}{2}[e, e]\big) + (de + [\omega, e]) \\ &= (R[\omega] + \tfrac{1}{2}[e, e]) + d_\omega e \end{aligned}$$

where $R[\omega]$ is the curvature of ω. Using the definition of symmetric Lie algebra, we see that the $\mathfrak{h}$-valued part of this curvature is $R[\omega] + \frac{1}{2}[e, e]$, while the $\mathfrak{p}$-valued part is the torsion $d_\omega e$.

Now suppose we have a bilinear form

$$\langle \cdot, \cdot \rangle \colon \mathfrak{p} \otimes \mathfrak{h} \to \mathbb{R}$$

which is invariant under the adjoint action of H, and satisfies the following cyclic property:

$$\langle x, [y, z] \rangle = \langle y, [z, x] \rangle \qquad \forall x, y, z \in \mathfrak{p}. \tag{4.6}$$

Note that this property is satisfied if the bilinear form in question is the restriction to $\mathfrak{p} \otimes \mathfrak{h}$ of an invariant bilinear form $b_\mathfrak{g}$ defined on $\mathfrak{g} \otimes \mathfrak{g}$, since differentiating the invariance equation

$$b_\mathfrak{g}(\mathrm{Ad}(g)x, \mathrm{Ad}(g)y) = b_\mathfrak{g}(x, y)$$

with $g \colon I \to G$ a path with $g(0) = 1$, $g'(0) = z$, yields

$$b_\mathfrak{g}([z, x], y) + b_\mathfrak{g}(x, [z, y]) = 0$$

which by symmetry of $k_\mathfrak{g}$ and antisymmetry of the bracket is equivalent to

$$b_\mathfrak{g}(x, [y, z]) = b_\mathfrak{g}(y, [z, x]).$$

But we wish this equation to hold even if our bilinear form does not extend to all of $\mathfrak{g}$.

Given these data, we can write down the action

$$\int \langle e, R + \frac{1}{6}[e, e] \rangle$$

Before deriving the equations of motion, we prove a lemma:

Lemma 6 *Let $\langle \cdot, \cdot \rangle \colon \mathfrak{p} \otimes \mathfrak{h} \to \mathbb{R}$ be a bilinear pairing of between the two parts of a symmetric Lie algebra $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{p}$ and assume $\langle \cdot, \cdot \rangle$ satisfies the invariance property (4.6). Then the induced pairing between $\mathfrak{p}$- and $\mathfrak{h}$-valued 1-forms X, Y, and Z satisfies*

$$\langle X, [Y, Z] \rangle = \langle Y, [Z, X] \rangle.$$

Proof: We can describe the $\mathfrak{p}$-valued 1-forms in local coordinates as

$$X = X_\mu \otimes dx^\mu \qquad Y = Y_\mu \otimes dx^\mu \qquad Z = Z_\mu \otimes dx^\mu.$$

We then have

$$\begin{aligned} \langle X, [Y, Z] \rangle &= \langle X_\mu, [Y_\nu, Z_\sigma] \rangle \otimes dx^\mu \wedge dx^\nu \wedge dx^\sigma \\ &= \langle Y_\nu, [Z_\sigma, X_\mu] \rangle \otimes dx^\nu \wedge dx^\sigma \wedge dx^\mu \\ &= \langle Y, [Z, X] \rangle, \end{aligned}$$

using the property (4.6) of $\langle \cdot, \cdot \rangle$ and the graded commutativity of the wedge product. $\square$

Now to derive the equations of motion, we take the variation of the action:

$$\begin{aligned} \delta S &= \int \langle e, \delta R + \frac{1}{3}[\delta e, e] \rangle + \langle \delta e, R + \frac{1}{6}[e, e] \rangle \\ &= \int \langle e, d_\omega \delta\omega \rangle + \langle \delta e, R + \frac{1}{2}[\delta e, e] \rangle \\ &= \int d\langle e, \delta\omega \rangle - \langle d_\omega e, \delta\omega) \rangle + \langle \delta e, R + \frac{1}{2}[e, e] \rangle, \end{aligned}$$

where in the second inequality we have used the above Lemma along with the identity $\delta R = d_\omega \delta\omega$, and in the third an integration by parts. When the boundary term vanishes, we thus get the equations of motion:

$$\begin{aligned} R + \frac{1}{2}[e, e] &= 0 \\ d_\omega e &= 0 \end{aligned}$$

which are direct analogs of the equations for 3d general relativity.

In the next section, we present the Chern–Simons formulation of 3d general relativity, which will turn out to be related to this generalization of 3d gravity we have presented here.

4.4 Chern-Simons theory and 3d gravity

The relationship between Chern–Simons theory and 3d gravity has been well explored [24, 99], including aspects of the relationship between Chern–Simons and geometric structures. In this section, we simply wish to demonstrate one such relationship, in the particular case where the model is a symmetric space.

Let us first recall a few basic facts about Chern–Simons theory. The Chern-Simons action is defined by:

$$S_{\mathrm{CS}}(A) = \int \mathrm{tr}\,(A \wedge dA + \frac{2}{3} A \wedge A \wedge A)$$

where A is a connection on some principal G-bundle Q over the 3d spacetime manifold M. To work out the equations of motion, we take the variation of the action corresponding to an arbitrary variation of A:

$$\begin{aligned}
\delta S_{\mathrm{CS}} &= \int_M \mathrm{tr}\,(\delta A \wedge dA + A \wedge d(\delta A) + 2\delta A \wedge A \wedge A) \\
&= \int_M \mathrm{tr}\,(2\delta A \wedge dA + 2\delta A \wedge A \wedge A) - \int_{\partial M} A \wedge \delta A.
\end{aligned}$$

Recognizing $dA + A \wedge A$ as the curvature F of A, we thus have the equation of motion

$$F = 0,$$

provided that the boundary term vanishes.

Now suppose that G/H is a symmetric space, with canonical decomposition

$$\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{p}$$

and that the structure group of Q can be reduced to H. Then we can write A as the sum of a connection ω on a principal H-bundle and a 1-form e on P with values in $\mathfrak{p}$:

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

The important point for our purposes is that, since $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{p}$ is a symmetric Lie algebra,

$$A' = \omega - e$$

is also a connection, whose curvature is:

$$F' = R + \tfrac{1}{2}[e, e] - d_\omega e.$$

A surprising fact is that if we take Chern–Simons theories for both A and A', with the proper normalization, their difference is precisely the generalization of 3d gravity discussed in the previous section.

Proposition 7

$$\frac{1}{\sqrt{\Lambda}}\left(S_{\mathrm{CS}}(\omega+\sqrt{\Lambda}e)-S_{\mathrm{CS}}(\omega-\sqrt{\Lambda}e)\right)=4\int \mathrm{tr}\,\left(e\wedge F+\frac{\Lambda}{3}e\wedge e\wedge e\right)$$

Proof: To reduce notational clutter, set $\Lambda=1$. We then calculate:

$$\begin{aligned}
&S_{\mathrm{CS}}(\omega+e)-S_{\mathrm{CS}}(\omega-e)\\
&\quad=\int \mathrm{tr}\,\left((\omega+e)\wedge d(\omega+e)+\frac{2}{3}(\omega+e)\wedge(\omega+e)\wedge(\omega+e)\right)\\
&\quad\quad-\int \mathrm{tr}\,\left((\omega-e)\wedge d(\omega-e)+\frac{2}{3}(\omega-e)\wedge(\omega-e)\wedge(\omega-e)\right)\\
&\quad=\int \mathrm{tr}\,\left(2\omega\wedge de+2e\wedge d\omega+\frac{4}{3}(\omega\wedge\omega\wedge e+\omega\wedge e\wedge\omega+e\wedge\omega\wedge\omega+e\wedge e\wedge e)\right)
\end{aligned}$$

The trace is graded cyclic— $\mathrm{tr}\,(\mu\wedge\nu)=(-1)^{pq}\mathrm{tr}\,(\nu\wedge\mu)$ for any p-form μ and q-form ν. So, since e and ω are each 1-forms, we have

$$\mathrm{tr}\ \omega\wedge\omega\wedge e=\mathrm{tr}\ \omega\wedge e\wedge\omega=\mathrm{tr}\ e\wedge\omega\wedge\omega.$$

Also, provided the boundary term vanishes, integration by parts gives us

$$\int \mathrm{tr}\ \omega\wedge de=\int \mathrm{tr}\ de\wedge\omega=\int \mathrm{tr}\ e\wedge d\omega$$

So, we get:

$$\begin{aligned}
S_{\mathrm{CS}}(\omega+e)-S_{\mathrm{CS}}(\omega-e)&=\int \mathrm{tr}\,\left(4e\wedge d\omega+4e\wedge\omega\wedge\omega+\frac{4}{3}e\wedge e\wedge e\right)\\
&=4\int \mathrm{tr}\,\left(e\wedge d\omega+e\wedge\omega\wedge\omega+\frac{1}{3}e\wedge e\wedge e\right)\\
&=4\int \mathrm{tr}\,\left(e\wedge(d\omega+\omega\wedge\omega)+\frac{1}{3}e\wedge e\wedge e\right)\\
&=4\int \mathrm{tr}\,\left(e\wedge F+\frac{1}{3}e\wedge e\wedge e\right)
\end{aligned}$$

which is four times the action for 3d general relativity with cosmological constant $\Lambda=1$. Restoring the cosmological constant Λ by a simple change of variables yields the desired equality. $\square$

Hence general relativity in 3 dimensions with nonzero cosmological constant splits into two noninteracting Chern-Simons theories. But what about 3d general relativity with $\Lambda = 0$?

The Lie algebra $\mathfrak{iso}(2,1)$ is not semisimple, as is easily checked by the Cartan criterion: its Killing form is degenerate. However, as pointed out by Witten [99], thanks to a coincidence of three dimensions, the Lie algebra $\mathfrak{iso}(2,1)$ has a nondegenerate symmetric bilinear form which is invariant under the adjoint action of $\mathrm{ISO}(2,1)$. This bilinear form is given as follows. First, an element of $\mathfrak{so}(2,1)$ can be thought of as an element of 3d Minkowski vector space $\mathbb{R}^{2,1}$ by exploiting the sequence of isomorphisms:

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

We shall also call the composite map $\star$. Explicitly, if $X \in \mathfrak{so}(2,1)$ has matrix components $X^i{}_j$, then $\star X$ has components

$$\star X^k = \frac{1}{2!}\epsilon^{ijk}\eta_{i\ell}X^\ell{}_j$$

or, written in terms of matrices:

$$\left[\begin{array}{rrr} 0 & w & v \\ -w & 0 & u \\ v & -u & 0 \end{array}\right] \overset{\eta}{\longleftrightarrow} \left[\begin{array}{rrr} 0 & w & -v \\ -w & 0 & u \\ v & -u & 0 \end{array}\right] \overset{\star}{\longleftrightarrow} \left[\begin{array}{r} u \\ v \\ w \end{array}\right]$$

Now define a quadratic form that takes any $(X, v) \in \mathfrak{so}(2,1) \ltimes \mathbb{R}^{2,1} = \mathfrak{iso}(2,1)$ to $\eta(\star X, v)$, where η denotes the metric on $\mathbb{R}^{2,1}$, which is invariant under the action of the Poincaré group. Polarizing this quadratic form gives a bilinear form

$$\langle (X,v),(X',v')\rangle = \frac{1}{2}\left(\eta(\star X, v') + \eta(\star X', v)\right)$$

which can be used to define the action in $\mathrm{ISO}(2,1)$ gauge theory.

Consider a Chern–Simons theory based on a symmetric space G/H such that $\mathfrak{g} = \mathfrak{h} + \mathfrak{p}$, and suppose that $\mathfrak{h}$ and $\mathfrak{p}$ are both null subspaces with respect to the inner product on $\mathfrak{g}$. That is, suppose $\mathrm{tr}\,(X,Y)$ vanishes whenever X and Y are either both in $\mathfrak{h}$ or both in $\mathfrak{p}$. Then

$$S_{\mathrm{CS}}(\omega + e) = S_{\mathrm{BF}+\Lambda}(\omega, e).$$

That is:

$$\frac{1}{\sqrt{\Lambda}}\int \mathrm{tr}\,\left(A \wedge \left(dA + \frac{1}{3}[A,A]\right)\right) = 2\int \mathrm{tr}\,\left(e \wedge \left(d\omega + \frac{1}{2}[\omega,\omega] + \frac{2\Lambda}{3}[e,e]\right)\right)$$

$$
\begin{aligned}
S_{\mathrm{CS}}(\omega+e) &= \int \mathrm{tr}\left((\omega+e)\wedge\left(d\omega+de+\frac{1}{3}[\omega+e,\omega+e]\right)\right)\\
&= \int \mathrm{tr}\left((\omega+e)\wedge\left(d\omega+de+\frac{1}{3}[\omega,\omega]+\frac{2}{3}[\omega,e]+\frac{1}{3}[e,e]\right)\right)\\
&= \int \mathrm{tr}\left(\omega\wedge\left(de+\frac{2}{3}[\omega,e]\right)+e\wedge\left(d\omega+\frac{1}{3}[\omega,\omega]+\frac{1}{3}[e,e]\right)\right)
\end{aligned}
$$

where we have used the fact that $\mathfrak{p}$ and $\mathfrak{h}$ are null subspaces with respect to the inner product to eliminate half of the terms. Performing an integration by parts and using the invariance of the inner product to combine terms, we obtain

$$
\frac{1}{\sqrt{\Lambda}}S_{\mathrm{CS}}(\omega+\sqrt{\Lambda}e) = -\int \mathrm{tr}\left(d(\omega+e)\right)+2\int \mathrm{tr}\left(e\wedge\left(d\omega+\frac{1}{2}[\omega,\omega]+\frac{2\Lambda}{3}[e,e]\right)\right).
$$

4.5 3d Galilean general relativity and the Newtonian limit of 3d gravity

What is the limit of $(2{+}1)d$ general relativity as the speed of light tends to infinity? Perhaps the most interesting aspect of this question is that the answer is *not* Newtonian gravity! To see this, consider that Newtonian gravity admits closed orbits, even in low-dimensional space. Indeed, when space is n-dimensional, a small mass mass m may orbit a large one M at fixed distance r in the x_1-x_2 plane as follows:

$$
x_1 = r\cos\left(\sqrt{\frac{kM}{r^n}}t\right) \quad x_2 = r\sin\left(\sqrt{\frac{kM}{r^n}}t\right) \quad x_i = 0,\ i\neq 1,2.
$$

On the other hand, general relativity in 2d space does not admit circular orbits around point masses. This is obvious when the total deficit angle around the large mass is less than π, since in this case no geodesic motion ever passes through the same point in space twice. But even if the total deficit angle is greater than π, we cannot construct a closed orbit. To see this, cut space from the mass out to infinity, and lay it flat so that geodesics are straight lines in the Euclidean plane. A particle in geodesic motion near the singularity may come back to its starting point provided the deficit angle ϕ is greater than π, but its velocity vector will have rotated by ϕ upon its return.

So, what is the 'Newtonian limit' of $(2{+}1)$ gravity, if not Newtonian gravity? One sensible way to try answering this question is to think of $(2{+}1)$ gravity as a gauge theory as in the previous sections, and try performing an Inönü-Wigner contraction of the Poincaré group down to the Galilei group. However, there is one apparent obstacle to doing this: the Galilei Lie algebra

$$\mathfrak{gali} = (\mathfrak{so}(2) \ltimes \mathbb{R}^2) \ltimes \mathbb{R}$$

does not have a nondegenerate invariant inner product. In fact, it does not even have an inner product invariant under the point stabilizer $\mathrm{ISO}_0(2)$ consisting of boosts and rotations. Consequently, we cannot write down a nondegenerate Lagrangian patterned after the Lagrangian for 3d general relativity.

We could simply forget about the Lagrangian, taking the viewpoint that a classical theory is directly determined by specifying its equations of motion. If we do so, the equation of motion

$$F = dA + \tfrac{1}{2}[A, A] = 0$$

makes as much sense for a Galilei group connection as for a Poincaré group connection. In fact, we can still write

$$A = \omega + e$$

with ω an ISO(2)-connection and e a coframe field. Since Galilean spacetime is a symmetric space, this gives

$$F = (R + \tfrac{1}{2}[e, e]) + d_\omega e$$

and get geometric structures modeled on 3d Galilean spacetime.

But we can do better than this, at least in the case $\Lambda \neq 0$. To see this, recall that the spherical and hyperbolic versions of Galilean spacetime are specified by the following symmetry groups:

$$G = \mathrm{ISO}(2,1) \quad H = \mathrm{ISO}(2) \qquad \text{hyperbolic}$$

$$G = \mathrm{ISO}(3) \quad H = \mathrm{ISO}(2) \qquad \text{spherical}$$

We have already used Witten's observation that $\mathrm{ISO}(2,1)$ has a nondegenerate invariant inner product, so we can try to describe the $c \to \infty$ limit of 3d general relativity with $\Lambda < 0$ as an $\mathrm{ISO}(2,1)$ Chern–Simons theory. Similarly, we can describe the $c \to \infty$ limit in the $\Lambda > 0$ case, using ISO(3) Chern–Simons theory with the analogous inner product. In fact, this inner product has the same interpretation as it does in the flat Lorentzian case. To see this, look again at a typical matrix element of $\mathfrak{iso}(3)$:

$$\begin{pmatrix} 0 & w & -v & x \\ -w & 0 & u & y \\ v & -u & 0 & z \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

In the Euclidean case, the stabilizer subalgebra is the upper left 3×3 block, while in the 'spherical Galilean' case, the stabilizer is the lower right 3×3 block. Thinking in terms of the Euclidean geometry, the quadratic form on this Lie algebra element simply takes the dot product of its transvection (or vector) part and its stabilizer (or axial vector) part:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \cdot \begin{pmatrix} u \\ v \\ w \end{pmatrix}$$

But this can also be interpreted as a product of the stabilizer and transvection parts in the spherical Galilean picture:

$$\begin{pmatrix} x \\ -v \\ w \end{pmatrix} \cdot \begin{pmatrix} u \\ -y \\ z \end{pmatrix}$$

4.6 Transforms between Chern–Simons theories

The previous section raises a subtle issue: we have already seen that ISO$(2,1)$ Chern–Simons theory is equivalent to $(2+1)$ general relativity with vanishing cosmological constant. Are we suggesting that the Lorentzian $\Lambda = 0$ case is somehow equivalent to the Galilean $\Lambda < 0$ case? The answer is, "not quite." The essential point is whether the bundle is reducible to the appropriate stabilizer, and whether we get a (generically) transverse section in the associated bundle with the Klein model as standard fiber. If the bundle is reducible in both ways, we may be able to 'transform' between these two types of Chern–Simons theories.

In fact, this idea shows up without resorting to these strange non-flat Galilean models: Looking at the chart on p. 29, we see that an SO$(3,1)$ Chern–Simons theory may be interpreted as giving geometric structures that are either hyperbolic or de Sitter, depending on our choice of stabilizer, SO(3) or SO(2). There is again the possibility of transforms between these two theories, based on choice bundle reduction.

Part II

Particles and Strings

Chapter 5

Statistics and Motion Groups

5.1 The Hamiltonian picture of BF theory

So far we have described BF theory in the Lagrangian formulation. We now describe the corresponding covariant Hamiltonian picture. This will allow us to discuss 'matter' in BF theory, which arises in a purely topological way.

Consider BF theory on a spacetime M diffeomorphic—though not canonically—to $\mathbb{R} \times S$ for some $(n-1)$-manifold S. Picking a specific diffeomorphism

$$\phi \colon M \to \mathbb{R} \times S$$

gives a way of splitting the spacetime M into 'time' and 'space'. More precisely, ϕ gives a foliation of M whose leaves are the hypersurfaces

$$S_t := \phi^{-1}(\{t\} \times S) \subset M \qquad t \in \mathbb{R}$$

where we think of S_t as 'space at time t'. Using this time-parameterized slicing of spacetime M, we wish to reformulate the equations BF theory as predictive (in fact, also *retro*dictive) equations, showing how initial data on S_0 evolve as t varies to give solutions of BF theory on all of M.

To do this, we put the connection A in temporal gauge. The diffeomorphism ϕ induces a global time coordinate on M and hence a nonvanishing vector field ∂_t in the time direction. We say that the connection A is in **temporal gauge** if

$$A(\partial_t) = 0$$

A standard tool of gauge theory is the fact that any connection is gauge equivalent to one in temporal gauge. We review the proof here:

Lemma 8 *If A is a principal connection on a manifold $M \cong \mathbb{R} \times S$, then there is a gauge transformation $A \mapsto A'$ such that $A'(\partial_t) = 0$. Moreover, A' is unique up to gauge transformation on space S.*

Proof: Starting with an arbitrary connection A and using the global time coordinate on M, we can write

$$A = A_0\, dt + A_S$$

where $A_0 = A(\partial_t)$ and A_S is a t-dependent connection on space S. We wish to find a gauge transformation so that the time component A_0 of the the connection vanishes. A gauge transformation is a section $g\colon M \to P \times_{\mathrm{Ad}} G$ which acts on the connection A by

$$A \mapsto A' = gAg^{-1} + gdg^{-1}.$$

The time component of A' is:

$$A_0' = gA_0g^{-1} + g\partial_t g^{-1}$$

This vanishes if and only if g^{-1} is a solution of the first-order differential equation

$$\partial_t g^{-1} = -A_0 g^{-1}.$$

The basic existence theorem for solutions of linear differential equations guarantees such a g exists and is uniquely determined by initial conditions on $g|_{S_0}$, that is, a gauge transformation on space S. $\square$

Using this lemma, we now assume our connection A has been put in temporal gauge. The only gauge transformations we can do that keep A in temporal gauge are gauge transformations on space. Thus A is now effectively a time-dependent connection on the restricted bundle $P|_{S_t} \to S_t$, and the curvature is given by

$$\begin{aligned} F &= dA + [A, A] \\ &= dt \wedge \partial_t A_S + d_S A_S + [A_S, A_S] \end{aligned}$$

where d_S is the differential on space. The momentum canonically conjugate to A_S is found by differentiating the Lagrangian

$$L = \mathrm{tr}\,(E \wedge (dt \wedge \partial_t A_S + d_S A_S + [A_S, A_S]))$$

with respect to $\partial_t A_S$:

$$\frac{\delta L}{\delta(\partial_t A_S)} = \mathrm{tr}\,(E \wedge dt \wedge -)\,.$$

The usual practice in quantum field theory is to identify this linear functional with E itself, so we say the momentum conjugate to the connection A_S is E. Indeed there is an obvious dual pairing, given as follows. First E is an $\mathrm{Ad}(P)$-valued $(n-2)$-form. Connections, on the other hand, form an affine space modeled on the vector space of $\mathrm{Ad}(P)$-valued 1-forms. Thus a tangent vector δA to the space of connections $\mathcal{A}$ is an $\mathrm{Ad}(P)$-valued 1-form. Since space is $(n-1)$-dimensional, we have a dual pairing:

$$\langle E, \delta A\rangle = \int_S \mathrm{tr}\,(E \wedge \delta A)$$

We should mention that if space S is not compact, there are issues of converge to deal with here. We ignore this issue for now.

5.2 Canonical quantization and motion group statistics

In this section we recall Dahm's [31] action of the 'motion group' $\mathrm{Mo}(S, \Sigma)$ on the fundamental group of $S - \Sigma$ and describe how this gives a unitary representation of the motion group on a certain Hilbert space of states for BF theory on $S - \Sigma$. The general idea of a 'motion group' goes back at least to Dahm's 1962 thesis [31], which unfortunately was never published. In the 1970's and 80's, some papers by Wattenberg [96] and Goldsmith [51, 52] clarified and expanded on Dahm's work.

Quite generally, suppose that S is a smooth oriented manifold and $\Sigma \subseteq S$ is a smooth oriented submanifold. Let $\mathrm{Diff}(S)$ be the group of orientation-preserving diffeomorphisms of S. Let $\mathrm{Diff}(S, \Sigma)$ be the subgroup of $\mathrm{Diff}(S)$ maps that restrict to give orientation-preserving diffeomorphisms of Σ.

We define a **motion** of Σ in S to be a smooth map $f\colon [0,1] \times S \to S$, which we write as $f_t\colon S \to S$ $(t \in [0,1])$, with the following properties:

- for all t, f_t lies in $\mathrm{Diff}(S)$;

- for all t sufficiently close to 0, f_t is the identity;

- for all t sufficiently close to 1, f_t is independent of t and lies in $\mathrm{Diff}(S, \Sigma)$.

Intuitively, a motion is a way of moving Σ through S so that it comes back to itself—not pointwise, but as a set—at $t = 1$. This suggests that one can 'multiply' motions by doing one after the other, and indeed this is true. Given motions f and g, one can define a motion $f \cdot g$ called their **product** as follows:

$$(f \cdot g)_t = \begin{cases} f_{2t} & \text{for} \quad 0 \leq t \leq \frac{1}{2} \\ g_{2t-1} \circ f_1 & \text{for} \quad \frac{1}{2} \leq t \leq 1 \end{cases}$$

Given a motion f we can also define a motion called its **reverse**, denoted $\bar{f}$, by:

$$\bar{f}_t = f_{1-t} \circ f_1^{-1}.$$

We say two motions f and g are **equivalent** if $\bar{f} \cdot g$ is smoothly homotopic, as a path in $\mathrm{Diff}(S)$ with fixed endpoints, to a path that lies entirely in $\mathrm{Diff}(S, \Sigma)$. One can check that this is indeed an equivalence relation and that the operations of product and reverse make equivalence classes of motions into a group. This is called the **motion group** $\mathrm{Mo}(S, \Sigma)$.

Next we turn to examples:

- When $\Sigma \subset \mathbb{R}^d$ is a collection of n points and $d > 2$, $\mathrm{Mo}(\mathbb{R}^d, \Sigma)$ is the symmetric group S_n.

- When $\Sigma \subset \mathbb{R}^2$ is a collection of n points, $\mathrm{Mo}(\mathbb{R}^2, \Sigma)$ is the braid group B_n.

- When $\Sigma \subset \mathbb{R}^3$ is a collection of n unknotted and unlinked oriented circles, we call $\mathrm{Mo}(\mathbb{R}^3, \Sigma)$ the **loop braid group** LB_n.

In general, the motion group $\mathrm{Mo}(S, \Sigma)$ acts in a natural way on $\pi_1(X)$. The idea goes back to Dahm's original work on the motion group [31], and it has been nicely explained by Goldsmith [51]. The idea is simple: elements of the motion group $\mathrm{Mo}(S, \Sigma)$ give equivalence classes of diffeomorphisms of $X = S - \Sigma$, and these act on homotopy classes of loops in X. The only problem is that the fundamental group is defined using *based* loops, and the diffeomorphisms used in the definition of the motion group need not preserve the basepoint in X. Luckily, Wattenberg [96] has shown that we can use compactly supported diffeomorphisms in the definition of the motion group without changing this group. In the examples above, we can assume without loss of generality that these diffeomorphisms are supported in a fixed large ball containing Σ. So, if we choose a basepoint $* \in S$ that is sufficiently far from Σ, we can assume this basepoint is preserved by all the diffeomorphisms in the definition of the motion group. This makes it easy to check that $\mathrm{Mo}(S, \Sigma)$ acts as automorphisms of $\pi_1(X)$.

Now let us return to BF theory in n-dimensional spacetime. We take 'space' to be of the form $X = S - \Sigma$, where S is an oriented manifold of dimension $n-1$, and $\Sigma \subset S$ is an oriented submanifold. We let G be a Lie group and let $P \to X$ be a principal G-bundle. As seen in the previous section, the fields A and E in BF theory are canonically conjugate, so the 'naive configuration space' of BF theory is just $\mathcal{A}_0/\mathcal{G}$, where $\mathcal{A}_0$ is the space of flat connections on P and $\mathcal{G}$ is the group of gauge transformations. By 'naive' we mean that we are ignoring boundary conditions. One could study only examples where there are no boundary conditions worry about, such as when X is just a compact manifold, but we shall mainly be interested in two examples that do not fall into this category:

1. X is $\mathbb{R}^2$ with a finite set of points removed (describing *point particles*):

$$X = S - \Sigma, \qquad S = \mathbb{R}^2, \qquad \Sigma = \{z_1, \ldots, z_n\}.$$

2. X is $\mathbb{R}^3$ with a finite set of unlinked unknotted circles removed (describing what one might call *closed strings*):

$$X = S - \Sigma, \qquad S = \mathbb{R}^3, \qquad \Sigma = \ell_1 \cup \cdots \cup \ell_n.$$

A rigorous study of BF theory may require that we impose boundary conditions at Σ. We ignore this issue now, leaving it for future research.

Recall from Section 3.3 that the union of the spaces $\mathcal{A}_0(P)/\mathcal{G}(P)$ over principal bundles $P \to X$ is the moduli space of flat bundles

$$\hom(\pi_1(X), G)/G.$$

This is the naive configuration space for BF theory where we treat the bundle P as variable. Applying Schrödinger quantization to this configuration space, we obtain the (naive) Hilbert space for BF theory:

$$L^2(\hom(\pi_1(X), G)/G)$$

Of course, defining this L^2 space requires that we choose a measure on the moduli space of flat bundles. Alternatively, we can try to form a Hilbert space

$$L^2(\hom(\pi_1(X), G))$$

on which G acts as follows:

$$(g\psi)(f) = \psi(g^{-1}f).$$

Again, this requires choosing a measure on $\hom(\pi_1(X), G)$. Moreover, G will only have a unitary representation on $L^2(\hom(\pi_1(X), G)$ if this measure is G-invariant.

In Chapters 6 and 7 we will show that for the two examples above, there is a 'natural' choice of G-invariant measure on $\hom(\pi_1(X), G)$. In both these examples the motion group $\mathrm{Mo}(S, \Sigma)$ acts on $\pi_1(X)$ and thus on $\hom(\pi_1(X), G)$. By saying a measure on $\hom(\pi_1(X), G)$ is 'natural', we simply mean that it is preserved by this action.

Using such a natural measure to define the Hilbert space $L^2(\hom(\pi_1(X), G))$, and using the action of $\mathrm{Mo}(S, \Sigma)$ on $\pi_1(X)$, we obtain a unitary representation of the motion group on this Hilbert space. This representation describes the statistics of point particles or closed strings in BF theory. In the first example the motion group is the braid group B_n, while in the 4d case it is the loop braid group LB_n. So, we obtain 'exotic statistics' in both cases.

In the next two chapters, we apply these ideas in some detail to the cases of 3d and 4d BF theory, particularly with the gauge groups most relevant to 3d and 4d gravity.

Chapter 6

Point particles in 3d BF theory

Now let us apply the general ideas of the previous chapter to the case of a plane with n punctures:

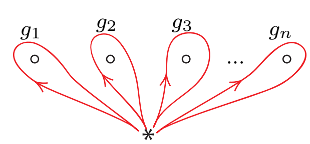
$$X = S - \Sigma, \qquad S = \mathbb{R}^2, \qquad \Sigma = \{z_1, \ldots, z_n\}$$

If we interpret these punctures as 'particles', we shall see that a state of 3d BF theory on this space describes a collection of identical point particles with exotic statistics governed by the braid group.

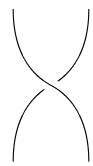
The fundamental group of X is the free group on n generators, so we have

$$\hom(\pi_1(X), G) = G^n$$

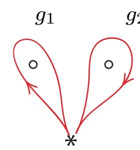
The n group elements here are nothing but the holonomies of a flat connection around based loops going clockwise around the particles:



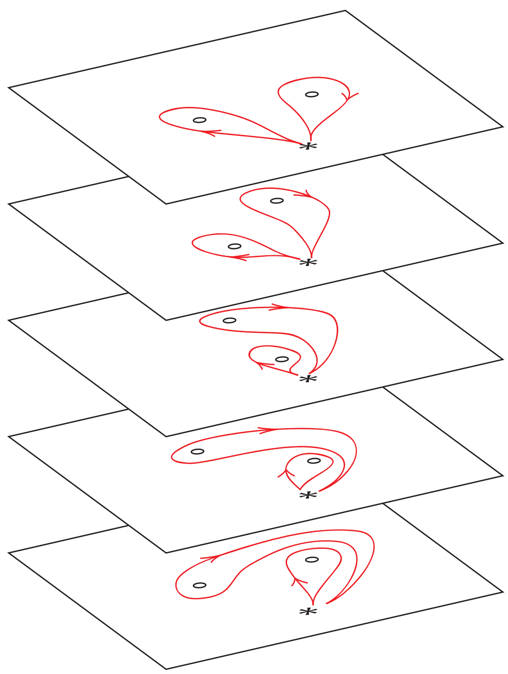
Having described particles as punctures in this theory, let us now consider what sort of statistics such particles obey. The previous section shows that the interchange of identical particles is described by an action of the n-strand braid group B_n on G^n, but we would like to work it out explicitly. For simplicity, consider the case $n = 2$ and consider what happens when the two particles switch places. As remarked earlier, there are infinitely many topologically distinct ways for the particles to move around each other, but they are all powers of the braid group generator σ_1:



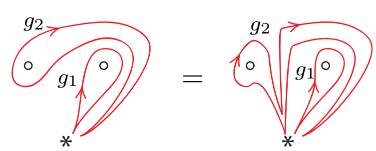
If the holonomies around the two particles are g_1 and g_2:



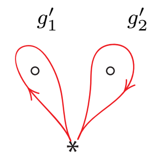
switching them via σ_1 induces a diffeomorphism of the plane which deforms the loops around which the holonomies are taken:



To see how the system changes in this process, compare the final frame in this 'movie' to the first frame. Given that $(g_1, g_2) \in G^2$ describes the holonomies initially, a slight deformation of the loops in the final frame:

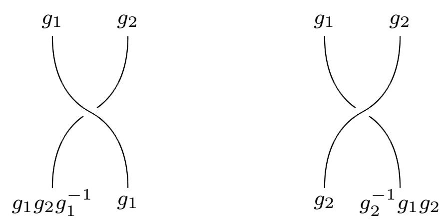


makes it clear that the corresponding holonomies around *these* loops in the final configuration:



are $(g_1', g_2') = (g_1 g_2 g_1{}^{-1}, g_1)$. Thus the effect of switching the two particles via σ_1 is to send (g_1, g_2) to $(g_1 g_2 g_1{}^{-1}, g_1)$.

We can work out the action of σ_1^{-1} in the same way, or simply derive it algebraically from the fact that it must undo the effect of σ_1. The easiest way to remember the results is with this picture:



More generally, we have a right action of the braid group B_n on G^n given as follows:

$$(g_1, \ldots, g_i, g_{i+1}, \ldots, g_n)\sigma_i = (g_1, \ldots, g_i g_{i+1} g_i^{-1}, g_i, \ldots, g_n).$$

As mentioned in the previous section, we also have a left action of G on G^n via gauge transformations at the basepoint $*$. This works as follows:

$$g(g_1, \ldots, g_n) = (g g_1 g^{-1}, \ldots, g g_n g^{-1}).$$

We would like a measure on G^n that is invariant under both these group actions, so that the braid group and gauge transformations act as unitary operators on $L^2(G^n)$. Such a measure exists whenever G is **unimodular**, meaning that its left-invariant Haar measure is also right-invariant. A Lie group is automatically unimodular if it is compact, or abelian, or semisimple. In particular, the groups $\mathrm{SO}(p, q)$ are all unimodular. Since these groups act on Minkowski spacetime in a way that preserves its Lebesgue measure, the Poincaré groups $\mathrm{ISO}(p, q)$ are also unimodular. Also, the identity component of a unimodular group is unimodular, as is any covering space of a unimodular group.

From this we see that the 3d Lorentz group $\mathrm{SO}(2, 1)$ is unimodular, as are its identity component $\mathrm{SO}_0(2, 1)$ and the double cover of its identity component, namely $\mathrm{SL}(2, \mathbb{R})$. All these are reasonable choices of gauge group when treating 3-dimensional—or more properly, (2+1)-dimensional—Lorentzian gravity as a BF theory.

Given a unimodular Lie group, Haar measure is typically not the only measure invariant under conjugation: we can multiply Haar measure by any function that only depends on the conjugacy class. As an extreme example, we can even try to multiply Haar measure by a 'delta function' supported on one conjugacy class. More precisely, we can look for a conjugation-invariant measure supported on a single conjugacy class of G. In this case we might as well be working not with G but with just the conjugacy class. It turns out that in the case of 3d quantum gravity, this amounts to studying identical particles *of a specified mass.* This leads us to our next subject: quandle field theory.

6.1 Quandle field theory

In the previous section we considered BF theory in 3 dimensions, and were led to a natural action of the braid group B_n on the space G^n for any group G. Notice that we did not actually need the multiplication in G to define this action; we only needed the operation of *conjugation.* This suggests that we can work more generally, replacing the group G by some algebraic structure that captures the properties of conjugation. Such a thing is called a 'quandle'.

More precisely, a **quandle** is a nonempty set Q equipped with two binary operations $\triangleright \colon Q \times Q \to Q$ and $\triangleleft \colon Q \times Q \to Q$ called **left** and **right conjugation**, which satisfy:

(i) left idempotence: $x \triangleright x = x$

(i′) right idempotence: $x \triangleleft x = x$

(ii) left inverse law: $x \triangleright (y \triangleleft x) = y$

(ii′) right inverse law: $(x \triangleright y) \triangleleft x = y$

(iii) left distributive law: $x \triangleright (y \triangleright z) = (x \triangleright y) \triangleright (x \triangleright z)$

(iii′) right distributive law: $(x \triangleleft y) \triangleleft z = (x \triangleleft z) \triangleleft (y \triangleleft z)$

for all $x, y, z \in Q$. In general, the operations of left and right conjugation in a quandle are neither associative nor commutative.

Quandles were first introduced as a source of knot invariants by David Joyce [56] in 1982. Many examples of quandles can be found in the work of Fenn and Rourke [35] and other authors [21, 56, 57]. For us, the most important examples come from taking a group G, letting Q be any union of conjugacy classes of G, and making Q into a quandle with

$$g \rhd h = ghg^{-1}, \qquad h \lhd g = g^{-1}hg.$$

We are especially interested in the case where Q is either the whole group G or a single conjugacy class.

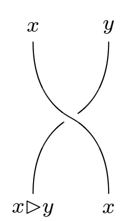
We can do some of the same things with quandles as with groups. For example, we can define a **topological quandle** to be a topological space that is also a quandle in such a way that the quandle operations $\rhd$ and $\lhd$ are continuous [83]. If G is a Lie group and $Q \subseteq G$ is a conjugacy class, Q becomes a topological quandle with the induced topology.

Given a topological quandle Q, we define an **invariant measure** on Q to be a Borel measure that is invariant under left conjugation by any element of Q—or equivalently, invariant under right conjugation by any element of Q. This implies that

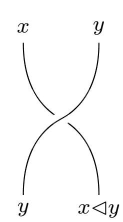
$$\begin{aligned} \int f(x)\, d\mu(x) &= \int f(q \rhd x)\, d\mu(x) \\ &= \int f(x \lhd q)\, d\mu(x) \end{aligned}$$

for any $q \in Q$ and any integrable function f on Q. As noted earlier, invariant measures on quandles are far from unique in general. In particular, we may multiply an invariant measure on a Lie group by any class function and obtain a new invariant measure.

In the previous section, we saw that the n-strand braid group B_n acts on G^n for any group G. But, since our argument relied only on properties of conjugation, it works just as well for a quandle. The idea is that we can braid two elements of a quandle past each other using left conjugation:



The inverse braiding uses right conjugation:



It is well known that with these rules, the braid group relations follow from the quandle axioms. So, generalizing our result from the previous section, we easily obtain:

Theorem 9. *Suppose Q is a topological quandle equipped with an invariant measure. Then there is a unitary representation ρ of the braid group B_n on $L^2(Q^n)$ given by*

$$(\rho(\sigma)\psi)(q_1,\ldots,q_n) = \psi((q_1,\ldots,q_n)\sigma)$$

for all $\sigma \in B_n$, where B_n has a right action on Q^n given by:

$$(q_1,\ldots,q_i,q_{i+1},\ldots,q_n)\sigma_i = (q_1,\ldots,q_i \triangleright q_{i+1}, q_i,\ldots,q_n).$$

There is also a unitary operator $U(q)$ on $L^2(Q^n)$ for each element $q \in Q$, given by

$$(U(q)\psi)(q_1,\ldots,q_n) = \psi(q \triangleright q_1,\ldots,q \triangleright q_n).$$

The operators $U(q)$ represent gauge transformations when Q is a group, so we can think of them as representing some sort of 'gauge transformation' even when Q is a quandle. Of course, if Q is a conjugacy class in a group G, there will be gauge transformations even for elements of G that do not lie in Q.

It is instructive to work out the details in the case of (2+1)-dimensional quantum gravity. This theory can be viewed as a BF theory with G being the connected Lorentz group $\mathrm{SO}_0(2,1)$, or perhaps better, its double cover $\mathrm{SL}(2,\mathbb{R})$. In either case we shall see that different conjugacy classes Q describe different types of spinless particles. The Hilbert space for n particles of this type is $L^2(Q^n)$, and Theorem 9 describes the exotic statistics and gauge invariance of this n-particle system.

In quantum field theory without gravity on 3d Minkowski spacetime, we can describe the energy-momentum of a particle by an element $p \in \mathfrak{sl}(2,\mathbb{R})$:

$$p = \begin{pmatrix} p_x & p_y + E \\ p_y - E & -p_x \end{pmatrix}$$

Note that

$$\det p = E^2 - p_x^2 - p_y^2.$$

The adjoint action of $\mathrm{SL}(2,\mathbb{R})$ on its Lie algebra:

$$\begin{aligned} \mathrm{SL}(2,\mathbb{R}) \times \mathfrak{sl}(2,\mathbb{R}) &\to \mathfrak{sl}(2,\mathbb{R}) \\ (g,p) &\mapsto gpg^{-1} \end{aligned}$$

preserves the determinant of p. So, the adjoint action gives an action of $\mathrm{SL}(2,\mathbb{R})$ as Lorentz transformations on the space of energy-momenta. As explained in the Introduction, an orbit of this action is just a type of spin-zero particle.

When we turn on gravity, we must describe energy-momenta not by elements of the Lie algebra $\mathfrak{sl}(2,\mathbb{R})$ but by elements of the group $\mathrm{SL}(2,\mathbb{R})$. Particle types are then described not by adjoint orbits but by conjugacy classes $Q \subseteq \mathrm{SL}(2,\mathbb{R})$. However, this new description is compatible with the old one, at least for energy-momenta that are small compared to the Planck energy $2\pi/\kappa$. The reason is that we can identify group elements near the identity with Lie algebra elements via the map

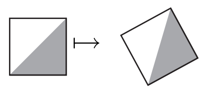
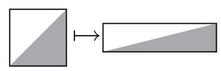
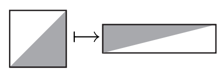
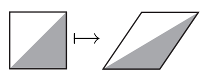
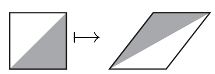
$$\begin{array}{ccc} \mathfrak{sl}(2,\mathbb{R}) & \to & \mathrm{SL}(2,\mathbb{R}) \\ p & \mapsto & \exp(\kappa p) \end{array}$$

This maps any adjoint orbit of $\mathfrak{sl}(2,\mathbb{R})$ into a conjugacy class of $\mathrm{SL}(2,\mathbb{R})$. Indeed, it gives a one-to-one correspondence between the set of adjoint orbits close to $0 \in \mathfrak{sl}(2,\mathbb{R})$ and the set of conjugacy classes close to $1 \in \mathrm{SL}(2,\mathbb{R})$. But, as mentioned in the Introduction, important differences show up for large energy-momenta.

To understand the conjugacy classes in $\mathrm{SL}(2,\mathbb{R})$, it is handy to use the representation

$$\mathrm{SL}(2,\mathbb{R}) = \left\{ \left(\begin{array}{cc} a+b & c+d \\ c-d & a-b \end{array} \right) \ : \ a,b,c,d \in \mathbb{R}, \ a^2 - b^2 - c^2 + d^2 = 1 \right\}$$

which says $\mathrm{SL}(2,\mathbb{R})$ is geometrically a 'unit hyperboloid' in a space of signature $(+--+)$. Since conjugate matrices have the same eigenvalues, the *trace* and thus the number a is an invariant of conjugacy classes. It is not a complete invariant, but it is except for matrices with $\operatorname{tr} g = \pm 2$. Every matrix in $\mathrm{SL}(2, \mathbb{R})$ is conjugate to one of these five kinds:

		conjugate to...	trace
rotations		$\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$	$-2 \le \operatorname{tr} g \le 2$
boosts		$\begin{pmatrix} e^{\alpha} & 0 \\ 0 & e^{-\alpha} \end{pmatrix}$	$\operatorname{tr} g \ge 2$
antiboosts		$\begin{pmatrix} -e^{\alpha} & 0 \\ 0 & -e^{-\alpha} \end{pmatrix}$	$\operatorname{tr} g \le -2$
shears		$\begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix}$	$\operatorname{tr} g = 2$
antishears		$\begin{pmatrix} -1 & \alpha \\ 0 & -1 \end{pmatrix}$	$\operatorname{tr} g = -2.$

Some explanation of this table is in order. Every 'rotation' maps to a rotation in the connected Lorentz group $\mathrm{SO}_0(2, 1)$: in other words, a transformation that preserves a timelike vector in 3d Minkowski spacetime. Similarly, every 'boost' maps to a transformation that preserves a spacelike vector, and every 'shear' maps to a transformation that preserves a lightlike vector. Since the two-to-one map from $\mathrm{SL}(2, \mathbb{R})$ to $\mathrm{SO}_0(2, 1)$ maps the matrix -1 to the identity, 'antiboosts' get mapped to the same elements as boosts, and 'antishears' get mapped to the same elements as shears. (An 'antirotation' would be just another rotation.)

The above chart counts certain conjugacy classes more than once. First of all, there is an overlap at $\operatorname{tr} g = 2$, since the identity rotation is also the identity shear and identity boost. Similarly, there is an overlap at $\operatorname{tr} g = -2$, since a rotation by π is also an antishear and an antiboost. Finally, all shears (resp. antishears) with $\alpha > 0$ are conjugate to each other, and all shears (resp. antishears) with $\alpha < 0$ are conjugate to each other. These are all the redundancies.

Knowing this, we can list all the conjugacy classes in $\mathrm{SL}(2, \mathbb{R})$ without any redundancies. However, it is less tiresome to list the conjugacy classes in $\mathrm{SO}_0(2, 1)$, since the elements $\pm g \in \mathrm{SL}(2, \mathbb{R})$ get identified in $\mathrm{SO}_0(2, 1)$, so we do not need to worry about 'antiboosts' and 'antishears'.

Here are all the conjugacy classes in $\mathrm{SO}_0(2,1)$, and the corresponding five types of spin-zero particles:

1. For any $0 < m < 2\pi/\kappa$ there is a conjugacy class containing the image of
$$\begin{pmatrix} \cos \kappa m/2 & -\sin \kappa m/2 \\ \sin \kappa m/2 & \cos \kappa m/2 \end{pmatrix} \in \mathrm{SL}(2, \mathbb{R}).$$
This corresponds to a tardyon of mass m.

2. For any $0 < m < \infty$ there is a conjugacy class containing the image of
$$\begin{pmatrix} e^{\kappa m/2} & 0 \\ 0 & e^{-\kappa m/2} \end{pmatrix} \in \mathrm{SL}(2, \mathbb{R}).$$
This corresponds to a tachyon of mass im.

3. There is a conjugacy class containing the image of
$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \mathrm{SL}(2, \mathbb{R}).$$
This corresponds to a positive-energy luxon.

4. There is a conjugacy class containing the image of
$$\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \in \mathrm{SL}(2, \mathbb{R}).$$
This corresponds to a negative-energy luxon.

5. There is a conjugacy class containing the image of
$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \in \mathrm{SL}(2, \mathbb{R}).$$
This corresponds to a particle of vanishing energy-momentum.

The factors of $1/2$ here arise from the double cover $\mathrm{SL}(2, \mathbb{R}) \to \mathrm{SO}_0(2,1)$. As explained in the Introduction, masses of tardyons really take values in the circle $\mathbb{R}/\frac{2\pi}{\kappa}\mathbb{Z}$. These conjugacy classes are also worked out in Goldman's doctoral thesis, which also features a drawing [48].

Each conjugacy class $Q \subseteq \mathrm{SO}_0(2,1)$ admits an invariant measure which is unique up to an overall scale. So, Theorem 9 applies: we can form a Hilbert space $L^2(Q)$ for particles of type Q, and more generally an n-particle Hilbert space $L^2(Q^n)$, on which the braid group and $\mathrm{SO}_0(2,1)$ gauge transformations act as unitary transformations.

We can also consider the corresponding Riemannian case. The procedure goes through in the same manner, since the double cover of $SO(3)$ is $\mathrm{SU}(2)$, which is just a Wick rotated version of $\mathrm{SL}(2,\mathbb{R})$:

$$\mathrm{SU}(2) = \left\{ \begin{pmatrix} a+ib & c+id \\ -c+id & a-ib \end{pmatrix} \;:\; a,b,c,d \in \mathbb{R},\; a^2+b^2+c^2+d^2 = 1 \right\}$$

In fact, the classification here turns out to be much simpler since there is no distinction between timelike and spacelike directions: two rotations are conjugate if and only if they are rotations by the same angle. Since a rotation by an angle θ is a rotation by angle $2\pi - \theta$ around the opposite axis, we need only consider $0 \le \theta \le \pi$. Thus types of spin-0 particles in Riemannian 3d gravity are simply classified by a mass $m \in [0, \pi]$.

6.2 $\mathrm{ISO}(2,1)$ Chern–Simons gravity and spin

The preceding analysis gave us a classification of the spin-0 particle types in 3d general relativity, by viewing this theory as a BF theory with $G = \mathrm{SO}_0(2,1)$. But this raises the question of how spin should be incorporated into the classification. The key lies in using the the trick of combining the connection and coframe field into a single connection for the Poincaré group $\mathrm{ISO}_0(2,1)$. In other words, the trick is viewing 3d general relativity (without cosmological constant) as a theory whose solutions are geometric structures modeled on Minkowski spacetime. In Chapter 4, we recalled how this idea works, and how it leads to the Chern–Simons formulation of 3d gravity.

In fact, using the formulation of 3d general relativity as an $\mathrm{ISO}(2,1)$ Chern–Simons theory, Philipp de Sousa Gerbert [88] calculates the holonomy of the $\mathrm{ISO}(2,1)$ connection around a tardyon with mass M and spin S, and finds it is conjugate to

$$\begin{pmatrix} 1 & 0 & 0 & -2\pi S \\ 0 & \cos 2\pi M & -\sin 2\pi M & 0 \\ 0 & \sin 2\pi M & \cos 2\pi M & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{6.1}$$

That is, parallel transport around a tardyon gives a rotation proportional to the mass and a timelike translation proportional to the spin. Note that for the spin-0 case this is consistent with our results above: parallel transport simply gives a rotation corresponding to the angle defecit caused by the mass.

To fully understand the classification of spinning particles in 3d gravity, one should of course work out the conjugacy classes in the 3d Poincaré group. These conjugacy classes should give modified versions of each of the particle types in the previous section, just as the class of the matrix above is a modification of the spin-0 tardyon type found above. We do not pursue this further in this context, but instead turn to a somewhat more interesting problem: the classification of particles with spin in 3d gravity with a positive cosmological constant.

6.3 $\mathrm{SO}(3,1)$ Chern–Simons gravity

In Section 6.1 we recalled the classification of conjugacy classes in $\mathrm{SO}_0(2,1)$ and its double cover $\mathrm{SL}(2,\mathbb{R})$. The classification for $\mathrm{SO}_0(3,1)$ and its double cover $\mathrm{SL}(2,\mathbb{C})$ is very similar, but simpler, because every complex number has a square root. It is also more familiar, since any element of

$$\mathrm{SL}(2,\mathbb{C}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \; : \; a,b,c,d \in \mathbb{C}, \; ad - bc = 1 \right\}$$

gives a fractional linear transformation

$$z \mapsto \frac{az+b}{cz+d}.$$

Such transformations are precisely the conformal transformations of the Riemann sphere. Note that both 1 and -1 in $\mathrm{SL}(2,\mathbb{C})$ map to the identity fractional linear transformation, so the conformal group of the Riemann sphere is

$$\mathrm{SL}(2,\mathbb{C})/\{\pm 1\} \cong \mathrm{SO}_0(3,1).$$

Indeed, Lorentz transformations can be thought of as conformal transformations of the 'celestial sphere': the set of light rays through an observer at the origin [78]. A list of conjugacy classes in $\mathrm{SO}_0(3,1)$ can thus be read off from the well-known classification of conformal transformations of the Riemann sphere [72]. But in fact, it is easy enough to construct this list from first principles.

Every element of $\mathrm{SO}_0(3,1)$ is either conjugate to the image of the shear

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \mathrm{SL}(2,\mathbb{C})$$

or conjugate to the image of

$$\begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix} \in \mathrm{SL}(2, \mathbb{C})$$

for some $\lambda \neq 0$. The conjugacy class of the latter element is unchanged if we make the replacement $\lambda \mapsto 1/\lambda$, and its image in $\mathrm{SO}_0(3,1)$ is unchanged if we make the replacement $\lambda \mapsto -\lambda$. These replacements (and their composite) are the only ways we can change λ without changing the conjugacy class of the corresponding element of $\mathrm{SO}_0(3,1)$. Using this, we can see there are five types of conjugacy classes in $\mathrm{SO}_0(3,1)$:

1. For any real m with $0 < m \leq \pi/\kappa$ there is a conjugacy class containing the image of

$$\begin{pmatrix} e^{i\kappa m/2} & 0 \\ 0 & e^{-i\kappa m/2} \end{pmatrix} \in \mathrm{SL}(2, \mathbb{C}).$$

   An element conjugate to one of this form is called **elliptic**.

2. For any purely imaginary m with $0 < \mathrm{Im}(m) < \infty$ there is a conjugacy class containing the image of

$$\begin{pmatrix} e^{i\kappa m/2} & 0 \\ 0 & e^{-i\kappa m/2} \end{pmatrix} \in \mathrm{SL}(2, \mathbb{C}).$$

   An element conjugate to one of this form is called **hyperbolic**.

3. For any $m \in \mathbb{C}$ with $0 < \mathrm{Re}(m) < 2\pi/\kappa$ and $0 < \mathrm{Im}(m) < \infty$ there is a conjugacy class containing the image of

$$\begin{pmatrix} e^{i\kappa m/2} & 0 \\ 0 & e^{-i\kappa m/2} \end{pmatrix} \in \mathrm{SL}(2, \mathbb{C}).$$

   An element conjugate to one of this form is called **loxodromic**.

4. There is a conjugacy class containing the image of

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \mathrm{SL}(2, \mathbb{C}).$$

   An element conjugate to one of this form is called **parabolic**.

5. There is a conjugacy class containing the image of

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \in \mathrm{SL}(2, \mathbb{C}).$$

   This class contains only the identity element.

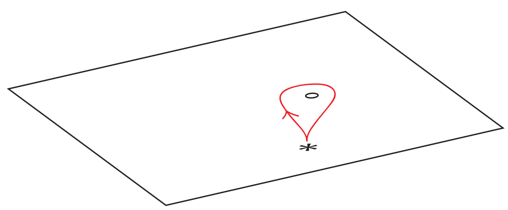
Let us now interpret these conjugacy classes as particle types in 3d gravity with positive cosmological constant. It will be convenient to work at the level of the double cover $\mathrm{SL}(2,\mathbb{C})$. First, recall how $\mathrm{SL}(2,\mathbb{C})$ acts on 4d Minkowski spacetime. Minkowski spacetime is conveniently represented as the vector space $\mathcal{H}$ of all 2×2 Hermitian matrices:

$$X = \begin{pmatrix} t+z & x+iy \\ x-iy & t-z \end{pmatrix}$$

where the Minkowski metric is given by the determinant. $\mathrm{SL}(2,\mathbb{C})$ acts on $\mathcal{H}$ by

$$\begin{array}{ccc} \mathrm{SL}(2,\mathbb{C}) \times \mathcal{H} & \to & \mathcal{H} \\ (g, X) & \mapsto & gXg^\dagger \end{array}$$

preserving the metric. De Sitter spacetime is the $\det X = 1$ hypersurface in the space of 2×2 Hermitian matrices. Suppose we have a single particle in Chern–Simons gravity with positive cosmological constant, and suppose the holonomy of the connection around the particle is $g \in \mathrm{SL}(2,\mathbb{C})$:



1. If g is elliptic, it acts on Minkowski spacetime as a spatial rotation in some reference frame. Explicitly, picking the reference frame where

$$g = \begin{pmatrix} e^{i\kappa m/2} & 0 \\ 0 & e^{-i\kappa m/2} \end{pmatrix} \qquad 0 < m \leq \pi/\kappa$$

one finds that the transformation induced on the coordinates t, x, y, z is given by

$$\begin{array}{ccc} t & \mapsto & t \\ x & \mapsto & x \cos \kappa m - y \sin \kappa m \\ y & \mapsto & x \sin \kappa m + y \cos \kappa m \\ z & \mapsto & z \end{array}$$

Thinking of this as a transformation of de Sitter spacetime, with x, y as spatial coordinates, we see that parallel translation around the particle gives a spatial rotation by angle κm. This is consistent with the first type of spin-0 particle found in on page 93, so it is appropriate to call the particle a **spin 0 tardyon**.

2. If g is hyperbolic, it acts on Minkowski spacetime as a boost in some reference frame. Explicitly, picking the reference frame where

$$\begin{pmatrix} e^{i\kappa m/2} & 0 \\ 0 & e^{-i\kappa m/2} \end{pmatrix} \qquad m \text{ imaginary, } 0 < \mathrm{Im}(m) < \infty$$

we get the transformation

$$\begin{array}{rcc} t & \mapsto & t\cosh\kappa m - z\sinh\kappa m \\ x & \mapsto & x \\ y & \mapsto & y \\ z & \mapsto & -t\sinh\kappa m + z\cosh\kappa m \end{array}$$

This is analogous to the timelike translation proportional to spin, as in the holonomy matrix (6.1) on p. 94. There is no global meaning of 'translation' in the curved de Sitter spacetime, but this boost in the 'z direction' in the ambient 4d Minkowski spacetime appears as a local time translation to an on observer with our chosen frame. Evidently, such a particle deserves to be called a **massless spinning particle**. While massless, it is not a luxon, since the spin-0 case does not reduce to the holonomy around a spinless luxon, but to the holonomy around a particle with no energy–momentum. We can say this particle type carries no linear momentum, but only angular momentum.

But there is really a second type of particle corresponding to this conjugacy class. As far as Minkowski spacetime is concerned, there is nothing special about the 'z' direction: the hyperbolic conjugacy class also corresponds to transformations like

$$\begin{array}{rcc} t & \mapsto & t\cosh\kappa m - x\sinh\kappa m \\ x & \mapsto & -t\sinh\kappa m + x\cosh\kappa m \\ y & \mapsto & y \\ z & \mapsto & z \end{array}$$

This kind of transformation is conjugate to the previous one within $\mathrm{SO}_0(3,1)$, *but it is not conjugate by an element of the stabilizer* $\mathrm{SO}_0(2,1)$! The point is that in the geometric structure picture, there is an assumed reduction of structure to the stabilizer subgroup, and it is only gauge transformations of the reduced bundle that leave the geometric structure invariant. In this case, the holonomy around the particle is a boost in the 3d de Sitter spacetime, so it corresponds to a **massive tachyon**.

3. If g is loxodromic, it acts on Minkowski spacetime as a combined rotation and boost about the same axis in some reference frame. Since the elliptic and hyperbolic transformations above are degenerate cases of loxodromic transformations, it comes as no

surprise that we get the transformation

$$\begin{array}{rcl} t & \mapsto & t\cosh \kappa m_2 - z \sinh \kappa m_2 \\ x & \mapsto & x \cos \kappa m_1 - y \sin \kappa m_1 \\ y & \mapsto & x \sin \kappa m_1 + y \cos \kappa m_1 \\ z & \mapsto & -t \sinh \kappa m_2 + z \cosh \kappa m_2 \end{array}$$

where

$$m_1 = \mathrm{Re}(m), \qquad m_2 = \mathrm{Im}(m).$$

We can still think of m_1 and m_2 as the mass and spin of **tardyon**, where $0 < m_1 < 2\pi/\kappa$ and $0 < m_2 < \infty$.

4. If g is parabolic, it acts on Minkowski spacetime as a Lorentz transformation fixing a single null vector and a single spacelike vector. Explicitly, picking the frame where

$$g = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

the transformation we get corresponds to the matrix

$$\begin{pmatrix} 3/2 & 1 & 0 & -1/2 \\ 1 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 1/2 & 1 & 0 & 1/2 \end{pmatrix} \in \mathrm{SO}_0(3,1).$$

This matrix has a unique eigenvalue $\lambda = 1$ of algebraic multiplicity 4, with only a 2-dimensional eigenspace, spanned by the eigenvectors

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

5. If g is the identity, we can say the particle carries no momentum and no angular momentum.

6.4 $\mathrm{SO}(2,2)$ Chern–Simons gravity

We have, in fact, already done the hardest part of the work necessary to describe spinning particles in 3d general relativity with negative cosmological constant. This is because the double cover of $\mathrm{SO}_0(2,2)$ is simply the product $\mathrm{SL}(2,\mathbb{R})\times\mathrm{SL}(2,\mathbb{R})$. A conjugacy class in $\mathrm{SL}(2,\mathbb{R})^2$ is simply a pair of conjugacy classes in $\mathrm{SL}(2,\mathbb{R})$. There are 5 basic types of conjugacy classes of $\mathrm{SL}(2,\mathbb{R})$, so there are 25 types of conjugacy classes to deal with; the situation is more complicated than in the $\mathrm{SO}(3,1)$ case since there are now two 'time' directions. Moreover, these conjugacy classes generally bifurcate to give two distinct particle types since, as we saw for the $\mathrm{SL}(2,\mathbb{C})$ case, it is really conjugacy under the stabilizer subgroup that counts. This gives ~ 50 different particle types. We make no attempt here to describe every particle type here. However it is appropriate to point out at least one of the most relevant examples.

- When g in $\mathrm{SL}(2,\mathbb{R})\times\mathrm{SL}(2,\mathbb{R})$ is a pair of rotations:

$$\left(\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}\right)$$

we find that the corresponding transformation of the coordinates (t,x,y,z) is given by

$$\begin{pmatrix} \cos(\theta-\phi) & 0 & 0 & \sin(\theta-\phi) \\ 0 & \cos(\theta+\phi) & -\sin(\theta+\phi) & 0 \\ 0 & \sin(\theta+\phi) & \cos(\theta+\phi) & 0 \\ -\sin(\theta-\phi) & 0 & 0 & \cos(\theta-\phi) \end{pmatrix}.$$

In particular, if $\theta=\phi=\kappa m/2$, we get a spatial rotation corresponding to the deficit angle around a **tardyon** of mass m. On the other hand, if $\theta=-\phi=S/2$ we get the 3d anti de Sitter analog of a 'time translation' in the Poincaré group, corresponding to a particle with vanishing mass but with 'spin parameter' S. More generally, if we define

$$2\pi M = \theta+\phi;$$
$$-2\pi S = \theta-\phi$$

Then we get

$$\begin{pmatrix} \cos 2\pi S & 0 & 0 & -\sin 2\pi S \\ 0 & \cos 2\pi M & -\sin 2\pi M & 0 \\ 0 & \sin 2\pi M & \cos 2\pi M & 0 \\ \sin 2\pi(S) & 0 & 0 & \cos 2\pi S \end{pmatrix}$$

as the analog of the holonomy (6.1) in the $\Lambda=0$ case. Notice that for small angular momentum this is approximately

$$\begin{pmatrix} 1 & 0 & 0 & -2\pi S \\ 0 & \cos 2\pi M & -\sin 2\pi M & 0 \\ 0 & \sin 2\pi M & \cos 2\pi M & 0 \\ -2\pi S & 0 & 0 & 1 \end{pmatrix}$$

which makes the resemblance to the Poincaré case (6.1) more plain. In the contraction limit as $\Lambda \to 0$, we recover the former case exactly.

Also note that this example shows the two $\mathrm{SL}(2,\mathbb{R})$ factors in the gauge group seem to be related by 'chirality': if we switch the angles θ and ϕ, then the mass M is unaffected, while the spin S switches sign.

6.5 Particle types and contractions

We might ask what becomes of the (spin-0) particle types in certain limits, particularly, the $c \to 0$ limit. The important observation is that the definition of contraction for Lie groups does not use group multiplication directly, but only conjugation. Thus, Definition 3 on p. 33 generalizes immediately to any smooth quandle:

Definition 8 *Let U be a smooth quandle with subquandles Q, Q'. Then Q' is called a contraction of Q within U if there is a sequence $u_1, u_2, \ldots \in U$ such that:*

1. *for every sequence $q_1, q_2, \ldots \in Q$ such that the sequence $u_1 \triangleright q_1, u_2 \triangleright q_2, \ldots \in U$ converges, the limit is an element of Q';*
2. *every element $q' \in Q'$ can be written as*

$$q' = \lim_{k\to\infty} u_k \triangleright h_k$$

for some sequence h_k.

When Q' is a contraction of Q by the sequence $g_1, g_2, \ldots \in U$, we write

$$Q' = \lim_{k\to\infty} u_k \triangleright Q.$$

Notice in particular that when U, Q, Q' are the underlying quandles of a group and two subgroups, this definition reduces to the one for contraction of Lie groups. But the slightly more general case most relevant to our purposes is the one where U is the quandle of a group, and Q, Q' are conjugacy classes.

Lemma 10 *Suppose H, H' are subgroups of a group G, and suppose H' is a contraction of H in the sense of Definition 3:*

$$H' = \lim_{k\to\infty} g_k \triangleright H$$

for some sequence $g_k \in G$. Let $Q \subseteq H$ be a conjugacy class in H, and consider the subset $Q' \subseteq G'$ consisting of all elements of G' that are limits of elements of Q:

$$Q' = \lim_{k\to\infty} g_k \triangleright Q$$

Then for every $q' \in Q'$ and every $h' \in H'$, we have $h' \triangleright q' \in Q'$. In other words, Q' is a union of conjugacy classes.

Proof: Since $q' \in Q'$ and $h' \in H'$, there exist sequences $q_k \in Q$ and $h_k \in H$ such that

$$q' = \lim_{k\to\infty} g_k \triangleright q_k$$
$$h' = \lim_{k\to\infty} g_k \triangleright h_k$$

Then

$$\begin{aligned}
\lim_{k\to\infty} g_k \triangleright (h_k \triangleright q_k) &= \lim_{k\to\infty} ((g_k \triangleright h_k) \triangleright (g_k \triangleright q_k)) \\
&= \lim_{k\to\infty} (g_k \triangleright h_k) \triangleright \lim_{k\to\infty} (g_k \triangleright q_k) \\
&= h' \triangleright q'.
\end{aligned}$$

so in particular, $h' \triangleright q' \in Q'$. □

A single conjugacy class (or particle type) may split into multiple types in the contraction limit. For a simple example, consider the contraction of SO(3) to ISO(2). In SO(3), two rotations are conjugate if and only if they are rotations by the same angle. But in the ISO(2) limit, some rotations become pure rotations, while other become pure translations, and these can never be conjugate. The opposite behavior is also possible: conjugacy classes may get identified during a contraction. An example is the Galilean limit ($c \to \infty$) of special relativity, where for example the past and future light cones get identified.

Chapter 7

Strings in 4d BF Theory

All the work in the previous two sections generalizes nicely from 3 to 4 dimensions, using the 'loop braid group' as a substitute for the braid group. In this chapter we describe the loop braid group and then study how it governs the exotic statistics of strings in 4d BF theory.

7.1 The loop braid group

The loop braid group LB_n consists of all ways a collection of oriented, unknotted, unlinked circles can move around in $\mathbb{R}^3$ and come back to their original positions, perhaps trading places. More precisely, it consists of 'isotopy classes' of such motions. This group thus plays the same role in describing the interchange of closed strings in $\mathbb{R}^3$ that the symmetric group S_n plays for point particles in $\mathbb{R}^3$, and the braid group plays for point particles in $\mathbb{R}^2$. In the notation of Section 5.2, the Loop braid group is the motion group:

$$LB_n := \mathrm{Mo}(\mathbb{R}^3, \Sigma)$$

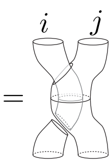
where $\Sigma \subset \mathbb{R}^3$ is a collection of n unknotted and unlinked oriented circles. In this section we use the work of Lin [63] to obtain two presentations of the loop braid group. McCool [67] and Rubinsztein [82] have also studied the motion group for unknotted and unlinked circles in $\mathbb{R}^3$. Surya has also given a description of the loop braid group as an iterated semidirect product [91]. Much of this work considers the motion of unoriented circles. Since we use oriented circles, we obtain a smaller motion group, which lacks the 'circle-flipping' operations that reverse orientations.

We shall use the work of Lin [63] to give two presentations of LB_n. First note that there is a homomorphism

$$p\colon LB_n \to S_n$$

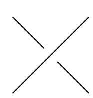
which simply forgets the details of the braiding, remembering only how the circles get permuted in the process. The image of p is all of S_n. We call the kernel of p the **pure loop braid group** PLB_n.

Suppose, just to be specific, that $\Sigma = \ell_1 \cup \cdots \cup \ell_n$ where $\ell_1, \ldots, \ell_n$ are disjoint unit circles in the xy plane, lined up from left to right with their centers on the x axis. Lin proves that PLB_n has a presentation with generators σ_{ij} for $i, j \in \{1, \ldots, n\}$ with $i \neq j$. The generator σ_{ij} describes a motion in which the ith circle floats up and over the jth circle, shrinks slightly and passes down through the jth circle, expands to its original size, and then moves straight back to its starting position. We draw this as follows:

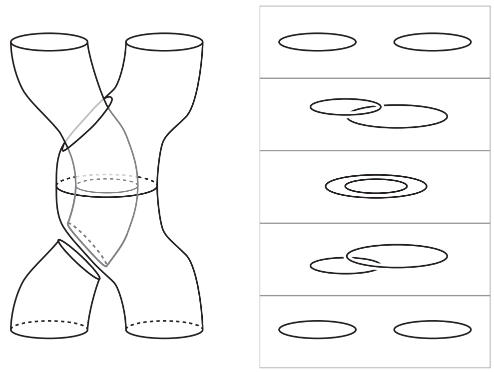
$\sigma_{ij} =$ 

where for purely artistic reasons we let the jth circle move a bit to the left in the process.

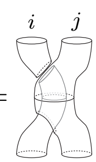
Here we are using a drawing style adapted from Carter and Saito's work on surfaces in 4 dimensions [29]. Crossings in a braid or knot are usually drawn with an artificial 'break' in one of the strands to indicate that it lies under the other:



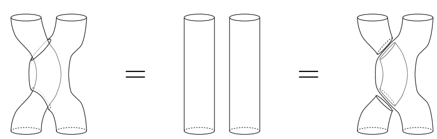
Similarly, Carter and Saito draw 3d projections of knotted surfaces in 4 dimensions, indicating by a broken surface which one passes 'under' the other in the suppressed fourth dimension. In our context, we take this suppressed dimension to be one of the spatial dimensions, in order to make room for *time*, which we decree to flow downward in all our diagrams. The broken surfaces in σ_{ij} indicate whether one circle is above or below the other in the suppressed spatial dimension, so that the following diagram and 'movie' illustrate the same process:



The inverse of σ_{ij} is of course obtained by running the movie backwards, which in diagrammatic notation becomes:

$$\sigma_{ij}{}^{-1} =$$ 

One advantage of this drawing style is that it immediately suggests Reidemeister–like moves for loop braids, such as this:



We shall study the loop braid group algebraically, relying on such diagrams for our intuition.

Given Lin's presentation of PLB_n, we can obtain a presentation of LB_n using the short exact sequence

$$1 \to PLB_n \xrightarrow{i} LB_n \xrightarrow{p} S_n \to 1.$$

First, note that there is a homomorphism

$$j\colon S_n \to LB_n$$

which takes a given permutation to what Lin calls a 'permutation path' in the motion group: a loop braid in which circles trade places without any circle passing through another in a topologically nontrivial way. For example, we can have them trade places while remaining on the xy plane. This map j is well-defined since all such permutation paths are homotopic. Moreover, the composite $p \circ j\colon S_n \to S_n$ is the identity homomorphism on S_n, so j is a splitting of the short exact sequence above.

Since j is one-to-one, we may identify elements of S_n with their images in LB_n. Since PLB_n is a normal subgroup, elements of S_n act on PLB_n via conjugation. This allows us to define the semidirect product $S_n \ltimes PLB_n$, and thanks to our split exact sequence, we get an isomorphism

$$\begin{aligned} f\colon LB_n &\to S_n \ltimes PLB_n \\ g &\mapsto (p(g), j(p(g))^{-1}g) \end{aligned}$$

with inverse

$$\begin{aligned} f^{-1}\colon S_n \ltimes PLB_n &\to LB_n \\ (s,\sigma) &\mapsto s\sigma \end{aligned}$$

Writing the loop braid group as a semidirect product in this way, we easily obtain a presentation for it:

Theorem 11. *The loop braid group LB_n has a presentation with generators s_i for $1 \le i \le n-1$ and σ_{ij} for $1 \le i, j \le n$ with $i \ne j$, together with the following relations:*

(a) the relations for the standard generators s_i of S_n:

$$s_i s_j = s_j s_i \qquad \text{for } |i-j| > 1 \tag{7.1}$$

$$s_i s_{i+1} s_i = s_{i+1} s_i s_{i+1} \qquad \text{for } 1 \le i \le n-2 \tag{7.2}$$

$$s_i^2 = 1 \qquad \text{for } 1 \le i \le n-1 \tag{7.3}$$

(b) Lin's relations for the generators σ_{ij} of PLB_n:

$$\sigma_{ij}\sigma_{k\ell} = \sigma_{k\ell}\sigma_{ij} \qquad \text{for } i,j,k,\ell \text{ distinct} \tag{7.4}$$

$$\sigma_{ik}\sigma_{jk} = \sigma_{jk}\sigma_{ik} \qquad \text{for } i,j,k \text{ distinct} \tag{7.5}$$

$$\sigma_{ij}\sigma_{kj}\sigma_{ik} = \sigma_{ik}\sigma_{kj}\sigma_{ij} \qquad \text{for } i,j,k \text{ distinct} \tag{7.6}$$

(c) relations expressing the action of S_n on PLB_n:

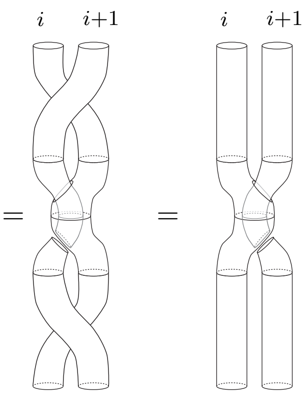
$$s_i\sigma_{i(i+1)} = \sigma_{(i+1)i}s_i \qquad \text{for } 1 \le i \le n-1 \tag{7.7}$$

$$s_k\sigma_{ij} = \sigma_{ij}s_k \qquad \text{for } i,j,k,k+1 \text{ distinct} \tag{7.8}$$

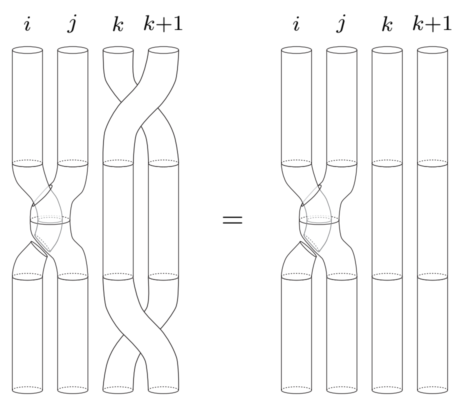
$$s_j\sigma_{ij} = \sigma_{i(j+1)}s_j \qquad \text{for } i,j,j+1 \text{ distinct} \tag{7.9}$$

$$s_i\sigma_{ij} = \sigma_{(i+1)j}s_i \qquad \text{for } i,i+1,j \text{ distinct} \tag{7.10}$$

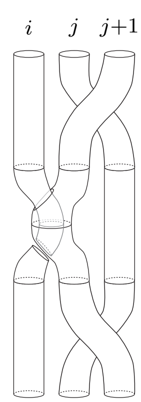
Proof: Since the presentation (a) of S_n is well-known, and Lin [63] proved that PLB_n has the presentation (b), to present their semidirect product LB_n it suffices to add relations that express the result of conjugating any of Lin's generators σ_{ij} by the symmetric group generators s_k. For $1 \leq i \leq n-1$ we have:

$s_i\sigma_{i(i+1)}s_i^{-1}$ =  $= \sigma_{(i+1)i}$

For i, j, k and $k+1$ all distinct, we have:

$s_k\sigma_{ij}s_k^{-1}$ =  $= \sigma_{ij}$

For i, j and $j+1$ distinct, we have:

$s_j\sigma_{ij}{s_j}^{-1}$ =  $= \sigma_{i(j+1)}$

and using a similar picture we see that for $i, i+1$ and j distinct, $s_i\sigma_{ij}{s_i}^{-1} = \sigma_{(i+1)j}$. The reader may notice that we have not included all possible conjugations of generators of PLB_n

by generators of S_n—we would naively expect two additional such classes, yielding two more relations:

$$s_{j-1}\sigma_{ij} = \sigma_{i(j-1)}s_{j-1} \qquad \text{for } i, j-1, j \text{ distinct} \tag{7.11}$$

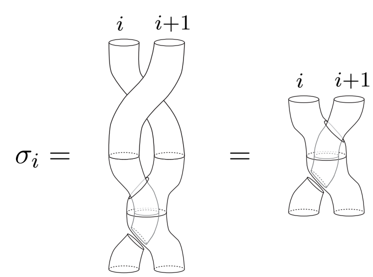
$$s_{i-1}\sigma_{ij} = \sigma_{(i-1)j}s_{i-1} \qquad \text{for } i-1, i, j \text{ distinct} \tag{7.12}$$

but these follow, respectively, from (7.9) and (7.10) combined with (7.3). So, we have precisely the relations in part (c), as desired. □

From this presentation of the loop braid group we now derive a presentation with fewer generators. We keep all the generators s_i, but replace the σ_{ij} with new generators defined as follows:

$$\sigma_i = s_i\sigma_{i(i+1)}$$

for $1 \leq i \leq n-1$. We can draw these as follows:



where we twist the picture a bit in the second step. To see that the generators s_i and σ_i indeed give a new presentation, note that we can express the old generators σ_{ij} in terms of these new ones as follows. First, repeatedly applying (7.9) we obtain:

$$\sigma_{ij} = s_{j-1}s_{j-2}\cdots s_{i+1}\sigma_{i(i+1)}s_{i+1}s_{i+2}\cdots s_{j-2}s_{j-1} \quad \text{for } \ i<j.$$

If instead of (7.9) we use its equivalent form (7.11), we obtain:

$$\sigma_{ij} = s_j s_{j+1}\cdots s_{i-2}\sigma_{i(i-1)}s_{i-2}\cdots s_{j+1}s_j \quad \text{for } \ i>j.$$

Rewriting these in terms of the new generators σ_i, and in the second case using relation (7.7), we obtain a way to write σ_{ij} in terms of the new generators:

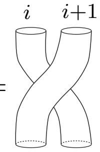
$$\sigma_{ij} = \begin{cases} s_{j-1}s_{j-2}\cdots s_i\sigma_i s_{i+1}s_{i+2}\cdots s_{j-2}s_{j-1} & \text{for } \ i<j \\ s_j s_{j+1}\cdots s_{i-2}\sigma_{i-1}s_{i-1}s_{i-2}\cdots s_{j+1}s_j & \text{for } \ i>j \end{cases} \tag{7.13}$$

Sometimes it is more convenient to use an alternate formula, obtained by applying (7.10), its equivalent form (7.12), and (7.7) again:

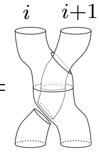
$$\sigma_{ij} = \begin{cases} s_i s_{i+1} \cdots s_{j-1} \sigma_{j-1} s_{j-2} \cdots s_{i+1} s_i & \text{for } i < j \\ s_{i-1} s_{i-2} \cdots s_{j+1} \sigma_j s_j s_{j+1} \cdots s_{i-2} s_{i-1} & \text{for } i > j. \end{cases} \tag{7.14}$$

What these formulas say is that when $j \neq i+1$ we can construct the loop braid σ_{ij} by permuting either the ith circle or the jth until they are adjacent, braiding one through the other, and then permuting the circles back to where they started.

The nice thing about using s_i and σ_i as generators of the loop braid group is that s_i describes how two neighboring circles can trade places by going around each other:

$s_i =$ 

while σ_i describes how two neighboring circles can trade places with the right one passing over and then down through the left one:

$\sigma_i =$ 

As a result, the generators s_i generate a subgroup of LB_n isomorphic to the symmetric group S_n, while the σ_i generate a subgroup isomorphic to the braid group B_n. There are also 'mixed relations' involving generators of both kinds:

Theorem 12. *The loop braid group LB_n has a presentation with generators s_i and σ_i for $1 \leq i \leq n-1$ together with the following relations:*

(a) relations for the standard generators s_i of S_n:

$$s_i s_j = s_j s_i \qquad \text{for } |i-j| > 1 \tag{7.15}$$

$$s_i s_{i+1} s_i = s_{i+1} s_i s_{i+1} \qquad \text{for } 1 \leq i \leq n-2 \tag{7.16}$$

$$s_i^2 = 1 \qquad \text{for } 1 \leq i \leq n-1 \tag{7.17}$$

(b′) relations for the standard generators σ_i of B_n:

$$\sigma_i \sigma_j = \sigma_j \sigma_i \qquad \text{for } |i-j| > 1 \tag{7.18}$$

$$\sigma_i \sigma_{i+1} \sigma_i = \sigma_{i+1} \sigma_i \sigma_{i+1} \qquad \text{for } 1 \leq i \leq n-2 \tag{7.19}$$

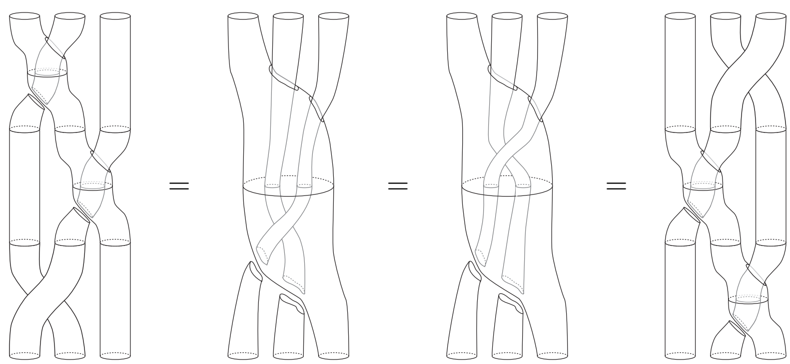
(c′) the following **mixed relations***:*

$$s_i\sigma_j = \sigma_j s_i \qquad \text{for } |i-j|>1 \tag{7.20}$$

$$s_i s_{i+1}\sigma_i = \sigma_{i+1} s_i s_{i+1} \qquad \text{for } 1 \le i \le n-2 \tag{7.21}$$

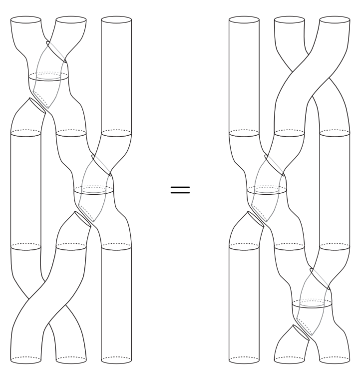
$$\sigma_i\sigma_{i+1} s_i = s_{i+1}\sigma_i\sigma_{i+1} \qquad \text{for } 1 \le i \le n-2 \tag{7.22}$$

Proof: The proof is somewhat lengthy, so we defer it to the Appendix. It is, however, simple to convince oneself using pictures that the given relations express topologically allowed moves for loop braids. Perhaps the least obvious of these is (7.22), for which we supply a visual proof below:



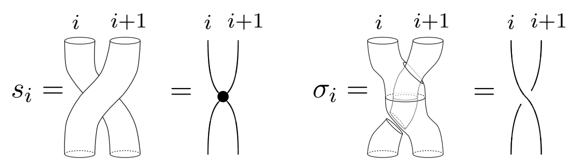
□

If we omit relations (7.22) we obtain the 'virtual braid group' VB_n of Vershinin [95]. This plays a role in virtual knot theory analogous to that of the usual braid group in ordinary knot theory. If we include these relations, which say:

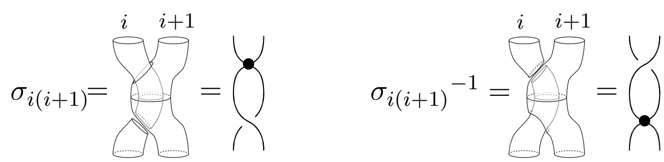


then we obtain precisely the 'braid permutation group' BP_n of Fenn, Rimányi and Rourke [34]. So, the loop braid group is isomorphic to the braid permutation group.

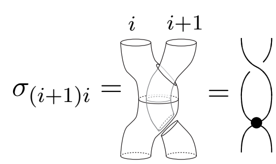
The isomorphism $LB_n \cong BP_n$ yields a simplified diagrammatic way of working with loop braids, which is in fact the method used by Fenn, Rimányi and Rourke in their original paper on BP_n. In the theory of 'welded braids', the generators σ_i in BP_n correspond to the kind of crossings found in ordinary braids: , while the s_i describe 'welded crossings', drawn like this: . These crossings are called 'welded' because one imagines that the two strands have been 'welded down' at the crossing. The point is that elements of the abstract group presented in Theorem 16 can be represented either as loop braid diagrams or as welded braid diagrams, as follows:



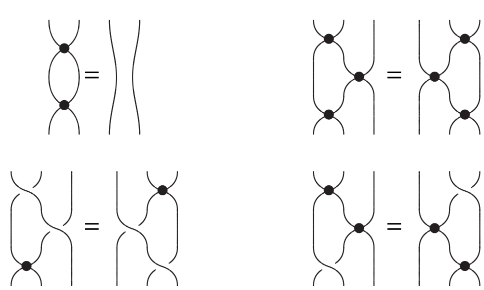
For the *pure* loop braid group PLB_n, the above correspondence implies the following welded braid pictures of the generators $\sigma_{i(i+1)}$ and their inverses:



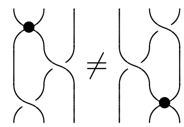
The other generators σ_{ij} can be obtained from these by conjugation, using (7.13) or (7.14). For example:



Diagrammatic calcuations with welded braids—and hence with loop braids—can be carried out by using the usual Reidemeister moves for real crossings, along with 'welded Reidemeister moves':



which are of course simply graphical restatements of the relations in (a) and (c'). The nonexistence of the following move:



is the rationale for the term 'welded braid'—we are not allowed to pass a strand under the weld.

It is easy from the presentation in Theorem 16 to work out the 1-dimensional unitary representations of the loop braid group. If $\rho\colon LB_n \to \mathrm{U}(1)$ is such a representation, we must have

$$\rho(s_i) = \pm 1$$

and

$$\rho(\sigma_i) = q$$

for all $1 \leq i < n$, where $q \in \mathrm{U}(1)$ is a fixed phase. We call the representations with $\rho(s_i) = 1$ **bose-anyons**, and the representations with $\rho(s_i) = -1$ **fermi-anyons**. These have been studied in physics at least since the work of Balachandran [18], and recently Niemi has shown how they arise in the dynamics of vortices in a quantum fluid [73].

In Section 7 we describe more interesting unitary representations of the loop braid group, using the action of the motion group on the moduli space of flat bundles. In related work, Szabo [92] has obtained a different class of representations using BF theory with abelian gauge group. Surya [91] has also studied representations of the loop braid group.

7.2 Loop braid statistics and representations

Let space be $\mathbb{R}^3$ with n unknotted and unlinked circles removed:

$$X = S - \Sigma, \qquad S = \mathbb{R}^3, \qquad \Sigma = \ell_1 \cup \cdots \cup \ell_n.$$

The fundamental group of X is the free group on n generators, so for any Lie group G we have

$$\hom(\pi_1(X), G) = G^n.$$

As explained in Section 3.3, a point in this space represents a G-bundle with flat connection over X, mod gauge transformations that equal the identity at a chosen basepoint. The n elements of G describing this point are just the holonomies around the circles $\ell_1, \ldots, \ell_n$.

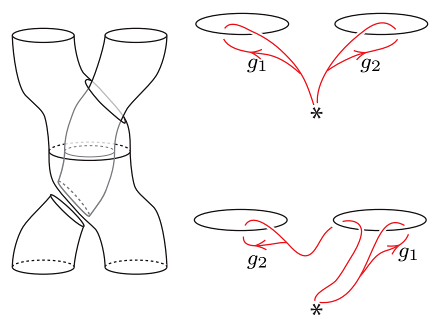
Physically, we think of these circles as string-like 'topological defects' where the flat connection on space becomes singular.

We explained quite generally in Section 5.2 how the motion group $\mathrm{Mo}(S, \Sigma)$ acts on $\hom(\pi_1(X), G)$. In the present case the motion group is just the loop braid group LB_n, and its generators act on $\hom(\pi_1(X), G) = G^n$ as follows:

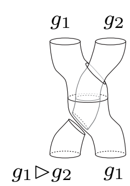
$$(g_1, \ldots, g_i, g_{i+1}, \ldots, g_n)s_i = (g_1, \ldots, g_{i+1}, g_i, \ldots, g_n),$$

$$(g_1, \ldots, g_i, g_{i+1}, \ldots, g_n)\sigma_i = (g_1, \ldots, g_i g_{i+1} g_i^{-1}, g_i, \ldots, g_n).$$

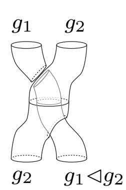
This is easy to see using pictures. For example, the generator σ_1 has the following effect:



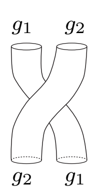
By an argument like the one we made in Section 6 for the ordinary braid group action in 3d BF theory, it follows that σ_1 acts on the holonomies g_1, g_2 by switching them while left conjugating g_2 by g_1:



Similarly, the inverse of σ_1 acts to switch the group elements while right conjugating g_1 by g_2:



The generator s_1 simply switches the holonomies g_1 and g_2:



It is easy to see that if G is unimodular, this action of the loop braid group on G^n gives rise to a unitary representation of the loop braid group on $L^2(G^n)$. And, just as in 3 dimensions, we can generalize this result to the case of a quandle:

Theorem 13. *Suppose Q is a topological quandle equipped with an invariant measure. Then there is a unitary representation ρ of the loop braid group LB_n on $L^2(Q^n)$ given by*

$$(\rho(\sigma)\psi)(q_1,\ldots,q_n) = \psi((q_1,\ldots,q_n)\sigma)$$

for all $\sigma \in LB_n$, where LB_n has a right action on Q^n given by:

$$\begin{array}{rcl} (q_1,\ldots,q_i,q_{i+1},\ldots,q_n)s_i & = & (q_1,\ldots,q_{i+1},q_i,\ldots,q_n) \\ (q_1,\ldots,q_i,q_{i+1},\ldots,q_n)\sigma_i & = & (q_1,\ldots,q_i \rhd q_{i+1},q_i,\ldots,q_n) \end{array}$$

There is also a unitary operator $U(q)$ on $L^2(Q^n)$ for each element $q \in Q$, given by

$$(U(q)\psi)(q_1,\ldots,q_n) = \psi(q \rhd q_1,\ldots,q \rhd q_n).$$

Proof: While the proof is straightforward, it is worth comparing Theorem 5.1 of Fenn, Rimányi and Rourke [34]. This says that the braid permutation group BP_n is the group of automorphisms of the free quandle on n generators. Since BP_n is isomorphic to the loop braid group LB_n, it follows that LB_n acts on Q^n for any quandle Q. The action is precisely as above. $\square$

Let us illustrate these ideas in the case where the gauge group is the connected Lorentz group $\mathrm{SO}_0(3,1)$ or its double cover $\mathrm{SL}(2,\mathbb{C})$. With either of these gauge groups, BF theory in 4 dimensions is sometimes called 'topological gravity'. We take space to be $\mathbb{R}^3$ and remove a collection of unknotted unlinked circles $\ell_1,\ldots,\ell_n$. For brevity let us call these circles 'closed strings'. A flat connection on space will have some holonomy $g_i \in \mathrm{SO}_0(3,1)$ around the ith string. We have already listed the conjugacy classes for these groups, in Section 6.3: they correspond to the various types of Möbius transformations of the Riemann sphere. Here we can simply reinterpret these conjugacy classes to list possible 'types' of strings, just as we used conjugacy classes in $\mathrm{SO}_0(2,1)$ to list types of spin-0 point particles in 3d gravity. This list is the analog for strings of the particle classification on p. 93

1. If g_i is elliptic, it acts on Minkowski spacetime as a spatial rotation in some reference frame. In this reference frame, parallel transport around the string ℓ_i is a spatial rotation by some angle $0 < \theta \leq \pi$ about some axis. (A rotation by an angle $\theta > \pi$ is a rotation by $\theta - \pi$ about the opposite axis.) This angle θ is proportional to the real number m which appears in item 1 of the above list, as follows:

   $$\theta = \kappa m.$$

   By analogy to 3d gravity, we could call the string a **tardyon** in this case, and call the number m its 'mass density'. The number m is real and takes values $0 < m \leq \pi/\kappa$.

2. If g_i is hyperbolic, it acts on Minkowski spacetime as a boost in some reference frame. In this reference frame, parallel transport around the string ℓ_i is a boost with rapidity $0 < \rho < \infty$ along some axis. The rapidity ρ is proportional to the imaginary number m which appears in item 2 of the above list, as follows:

   $$\rho = \kappa \mathrm{Im}(m).$$

   By analogy to 3d gravity, we could call the string a **tachyon** in this case, and call the number m its 'mass density'. The number m is purely imaginary and takes values in the upper half of the imaginary axis: $0 < \mathrm{Im}(m) < \infty$.

3. If g_i is loxodromic, it acts on Minkowski spacetime as a combined rotation and boost about the same axis in some reference frame. In this reference frame, parallel transport around the string ℓ_i is a combination of a rotation by an angle $0 < \theta < 2\pi$ and a boost with rapidity $0 < \rho < \infty$ about the same axis, where

   $$\theta = \kappa \mathrm{Re}(m), \qquad \rho = \kappa \mathrm{Im}(m).$$

   This case has no analogue in 3d gravity. We can still think of m as some sort of mass density, but it is complex, with $0 < \mathrm{Re}(m) < 2\pi/\kappa$ and $0 < \mathrm{Im}(m) < \infty$.

4. If g_i is parabolic, it acts on Minkowski spacetime as a Lorentz transformation fixing a single null vector. By analogy to 3d gravity, we could call the string a **luxon** in this case, and say $m = 0$.

5. If g_i is the identity, we can say the string carries no energy-momentum, and again say $m = 0$.

Each of these conjugacy classes $Q \subseteq \mathrm{SO}_0(3,1)$ is a quandle. The question then arises which of these quandles admits an invariant measure, and whether this measure is unique up to scale. One can work this out on a case-by-case basis.

One important case is when Q is the conjugacy class containing all rotations by some fixed angle $0 < \theta < \pi$. This conjugacy class corresponds to a 'tardyonic' closed string with a given mass density $0 < m \leq \pi/\kappa$. It is easy to see that this conjugacy class Q indeed admits an invariant measure. To see this, note that to specify a rotation by the angle θ one must first pick a future-pointing unit timelike vector $u \in \mathbb{R}^4$, to split Minkowski spacetime into space and time, and then pick a unit spacelike vector v orthogonal to u, to serve as the axis of rotation. The allowed choices of u lie in the hyperboloid

$$H = \{(t, x, y, z) \colon \; t^2 - x^2 - y^2 - z^2 = 1, \; t > 0\}.$$

This hyperboloid H is a Riemannian submanifold of $\mathbb{R}^4$. An allowed choice of u together with v amounts to a point in SH, the unit sphere bundle of H. So, we have $Q \cong SH$. Since the unit sphere bundle of a Riemannian manifold is itself a Riemannian manifold in a natural way, we get a well-defined Lebesgue measure on SH and thus Q, which is invariant under $\mathrm{SO}_0(3,1)$, since our construction respected the Lorentz group symmetry.

Given an invariant measure on Q, we obtain a Hilbert space $L^2(Q^n)$ for n strings of type Q. Note that we do not try to 'symmetrize' the states in this Hilbert space. Instead we describe the statistics using a representation of the loop braid group, following Theorem 13. Of course, one should work out the details explicitly, but we leave this for future research.

Chapter 8

Higher gauge theory and particles in 4d BF theory

8.1 The idea of a p-connection

We have seen that particles in 3d BF theory and strings in 4d BF theory have 'group-valued energy-momentum' given by the holonomy of the connection. More generally, we've seen the connection assigns group-valued momentum to any 'brane' of codimension 2. The question arises whether codimension-2 branes are the only possibility for generalizing the inclusion of matter in 3d general relativity to arbitrary BF theories. Can we include particles?

Of course, we cannot take the holonomy of a flat connection 'around a particle' unless spacetime is 3-dimensional, since the fundamental group is oblivious to obstructions with any codimension but 2. This would kill our hopes of including particles in 4d BF theory in a purely topological way if the connection A were the only field in the theory. However, we also have the B field, which in n-dimensional spacetime is a $\mathfrak{g}$-valued $(n-2)$-form. The dimension makes B perfect for integrating not over a loop, but over an $(n-2)$-sphere—just the right dimension of sphere to enclose a *particle*, regardless of n! In particular,

$$\pi_{d-1}(\mathbb{R}^d - \{*\}) \cong \pi_{d-1}(S^{d-1}) \cong \mathbb{Z},$$

where $d = n - 1$ is the dimension of space. When $d = 3$, we can draw a generator of the

2nd homotopy group:

In any dimension, we have hope of being able to detect particle-like defects using the B field.

In fact, at least in the case $n = 4$, one can perform such an integral and get a group element in a gauge invariant way. This has been shown in the context of 'higher gauge theory' by Baez and Schreiber [16], which is the appropriate mathematical setting for describing such higher dimensional notions of holonomy. Higher gauge theory is a rich subject, to which this chapter should by no means be considered an adequate introduction. As a proper discussion of the relationship between BF theory and higher gauge theory would present too great a distraction from our purposes, we refer the interested reader to the references [8, 15, 16, 19]. Here we shall simply describe enough of the ingredients of higher gauge theory to develop a picture of point particles as topological defects in 4d BF theory.

Higher gauge theory is a sort of hybrid of ordinary gauge theory and higher category theory. In higher gauge theory, all of the familiar gadgets from gauge theory get generalized to categorical analogs:

- bundles become 'p-bundles' (or $(p-1)$-gerbes),
- connections become 'p-connections',
- gauge groups become 'gauge p-groups',

and so on, where p is a positive integer, $p = 1$ being the base case. In fact, except in certain 'abelian' cases, these types of generalizations have so far mainly been developed for $p = 2$. The nontrivial process of generalizing 'things' to '2-things' (and beyond) is known as 'categorification' [13], since it involves replacing *sets* with *categories*. We will not describe the general idea categorification further, but we will give one concrete example that is immediately applicable to our problem of describing particles in 4d BF theory—the passage from groups to 2-groups.

8.2 From groups to 2-groups

To describe holonomies along both paths and surfaces, it turns out one should use not just an ordinary group but a '2-group'. In this section we define 2-groups; in the next we see how certain 2-groups show up naturally in 4d BF theory.

8.2.1 2-groups as 2-categories

Any group G can be thought of as a category with a single object $\star$, morphisms labeled by elements of G, and composition defined by multiplication in G:

$$\star \xrightarrow{g_1} \star \xrightarrow{g_2} \star \quad = \quad \star \xrightarrow{g_2 g_1} \star$$

In fact, one can define a group to be a category with a single object and all morphisms invertible. The object $\star$ can be thought of as an object whose symmetry group is G.

In a 2-group, we add an additional layer of structure to this picture, and capture the idea of *symmetries between symmetries*. In addition to having a single object $\star$ and its automorphisms, we have isomorphisms *between* automorphisms of $\star$:

$$\begin{tikzcd} \star \arrow[r, bend left, "g"] \arrow[r, bend right, "g'"'] \arrow[r, phantom, "\Downarrow h"] & \star \end{tikzcd}$$

In other words, a 2-group is a '2-category' with one object, all morphisms invertible, and all 2-morphisms invertible.

To understand 2-groups, it is thus helpful to recall the more general idea of a 2-category. A 2-category consists of

- objects: $X, Y, Z, \ldots$
- morphisms between objects: $X \xrightarrow{f} Y$
- 2-morphisms between morphisms: $\begin{tikzcd} X \arrow[r, bend left, "f"] \arrow[r, bend right, "f'"'] \arrow[r, phantom, "\Downarrow\alpha"] & Y \end{tikzcd}$

Morphisms can be composed as in a category, and 2-morphisms can be composed in two distinct ways, either vertically:

$$
\begin{tikzcd}
X \arrow[r, bend left=50, "f"{name=A}] \arrow[r, "f'"{description, name=B}] \arrow[r, bend right=50, "f''"'{name=C}] \arrow[from=A, to=B, phantom, "\Downarrow\alpha"] \arrow[from=B, to=C, phantom, "\Downarrow\alpha'"] & Y
\end{tikzcd}
\quad = \quad
\begin{tikzcd}
X \arrow[r, bend left=50, "f"{name=A}] \arrow[r, bend right=50, "f''"'{name=C}] \arrow[from=A, to=C, phantom, "\Downarrow\alpha'\cdot\alpha"] & Y
\end{tikzcd}
$$

or horizontally:

$$
\begin{tikzcd}
X \arrow[r, bend left=40, "f_1"{name=A}] \arrow[r, bend right=40, "f_1'"'{name=B}] \arrow[from=A, to=B, phantom, "\Downarrow\alpha_1"] & Y \arrow[r, bend left=40, "f_2"{name=C}] \arrow[r, bend right=40, "f_2'"'{name=D}] \arrow[from=C, to=D, phantom, "\Downarrow\alpha_2"] & Z
\end{tikzcd}
\quad = \quad
\begin{tikzcd}
X \arrow[r, bend left=40, "f_2f_1"{name=A}] \arrow[r, bend right=40, "f_2'f_1'"'{name=B}] \arrow[from=A, to=B, phantom, "\Downarrow\alpha_2\circ\alpha_1"] & Y
\end{tikzcd}
$$

There are certain axioms that must hold for this to be a 2-category. The axioms give, for example, the expected behavior of identity morphisms and 2-morphisms under composition. We shall not be concerned with listing these axioms here, deferring instead to standard references on category theory [66]. The critical axiom is that the vertical composition "$\cdot$" and horizontal composition "$\circ$"of 2-morphisms satisfy the **exchange law**:

$$(\beta' \cdot \beta) \circ (\alpha' \cdot \alpha) = (\beta' \circ \beta) \cdot (\alpha' \circ \alpha) \tag{8.1}$$

so that diagrams of the form

$$
\begin{tikzcd}
X \arrow[r, bend left=50, "f_1"{name=A}] \arrow[r, "f_1'"{description, name=B}] \arrow[r, bend right=50, "f_1''"'{name=C}] \arrow[from=A, to=B, phantom, "\Downarrow\alpha"] \arrow[from=B, to=C, phantom, "\Downarrow\alpha'"] & Y \arrow[r, bend left=50, "f_2"{name=D}] \arrow[r, "f_2'"{description, name=E}] \arrow[r, bend right=50, "f_2''"'{name=F}] \arrow[from=D, to=E, phantom, "\Downarrow\beta"] \arrow[from=E, to=F, phantom, "\Downarrow\beta'"] & Z
\end{tikzcd}
$$

are unambiguous.

When a 2-category has a unique object, and all 1-morphisms and 2-morphisms are invertible, it is a **2-group**.

8.2.2 Constructing 2-groups

We can construct a 2-group from the following data: a pair (G, H) of groups, equipped with an action $\rhd$ of G as automorphisms of H, *i.e*

$$g \rhd (h_1h_2) = (g \rhd h_1)(g \rhd h_2) \tag{8.2}$$

$$g \rhd 1 = 1 \tag{8.3}$$

for all $g \in G$ and $h_1, h_2 \in H$, and a group homomorphism $\delta : H \to G$ compatible with $\rhd$ in the following sense:

$$\begin{aligned} \delta(g \rhd h) &= g\delta(h)g^{-1} && (8.4) \\ \delta(h) \rhd h' &= hh'h^{-1}. && (8.5) \end{aligned}$$

Such a system $(G, H, \rhd, \delta)$ is called a **crossed module**, and in fact it is a theorem that every 2-group arises from some crossed module in the way we now describe [36].

Given a crossed module $(G, H, \rhd, \delta)$, we construct a 2-group with

- object $\star$
- elements of G as 1-morphisms $\star \xrightarrow{g} \star$
- pairs $u = (g, h) \in G \times H$ as 2-morphisms, where (g, h) is a 2-morphism from g to $\delta(h)g$. We draw this as: $u = \begin{tikzcd} \star \arrow[r, bend left, "g"] \arrow[r, bend right, "g'"'] \arrow[r, phantom, "\Downarrow h"] & \star \end{tikzcd}$ where $g' = \delta(h)g$.

The composition of 1-morphisms and vertical composition of 2-morphisms are induced by multiplication in G and H respectively:

$$\star \xrightarrow{g_1} \star \xrightarrow{g_2} \star \quad = \quad \star \xrightarrow{g_2 g_1} \star$$

and

$$\begin{tikzcd} \star \arrow[r, bend left=60, "g"] \arrow[r, "g'" near start] \arrow[r, bend right=60, "g''"'] \arrow[r, phantom, shift left=3ex, "\Downarrow h"] \arrow[r, phantom, shift right=3ex, "\Downarrow h'"] & \star \end{tikzcd} \quad = \quad \begin{tikzcd} \star \arrow[r, bend left, "g"] \arrow[r, bend right, "g''"'] \arrow[r, phantom, "\Downarrow h'h"] & \star \end{tikzcd}$$

with $g' = \delta(h)g$ and $g'' = \delta(h')\delta(h)g = \delta(h' \cdot h)g$. In other words, writing the vertical composition with a dot $\cdot$, we have

$$u' \cdot u = (g', h') \cdot (g, h) = (g, h'h) \tag{8.6}$$

defined for 2-morphisms $u = (g, h)$ and $u' = (g', h')$ such that $g' = \delta(h)g$. The horizontal composition "$\circ$" of two 2-morphisms

$$\begin{tikzcd} \star \arrow[r, bend left, "g_1"] \arrow[r, bend right, "g_1'"'] \arrow[r, phantom, "\Downarrow h_1"] & \star \arrow[r, bend left, "g_2"] \arrow[r, bend right, "g_2'"'] \arrow[r, phantom, "\Downarrow h_2"] & \star \end{tikzcd} \quad = \quad \begin{tikzcd} \star \arrow[r, bend left, "g_2 g_1"] \arrow[r, bend right, "g_2' g_1'"'] \arrow[r, phantom, "\Downarrow h_2 \cdot (g_2 \rhd h_1)"] & \star \end{tikzcd}$$

promotes the pair (G, H) to a semidirect product $G \ltimes H$ with multiplication:

$$(g_2, h_2) \circ (g_1, h_1) \equiv (g_2 g_1, h_2(g_2 \rhd h_1)). \tag{8.7}$$

One can check that the exchange law

$$(u_2' \cdot u_2) \circ (u_1' \cdot u_1) = (u_2' \circ u_2) \cdot (u_1' \circ u_1) \tag{8.8}$$

holds for 2-morphisms $u_i = (g_i, h_i)$ and $u_i' = (g_i', h_i')$, so that the diagram

$$\begin{tikzcd}
\star \arrow[r, bend left=50, "g_1"] \arrow[r, "g_1'" description] \arrow[r, bend right=50, "g_1''"'] \arrow[r, phantom, shift left=3, "\Downarrow h_1"] \arrow[r, phantom, shift right=3, "\Downarrow h_1'"] & \star \arrow[r, bend left=50, "g_2"] \arrow[r, "g_2'" description] \arrow[r, bend right=50, "g_2''"'] \arrow[r, phantom, shift left=3, "\Downarrow h_2"] \arrow[r, phantom, shift right=3, "\Downarrow h_2'"] & \star
\end{tikzcd}$$

is well defined.

Example: 2-groups from semidirect products

Given any semidirect of groups $G \ltimes H$, we can construct a 2-group as above by letting $\delta \colon H \to G$ be the trivial homomorphism. For example, the 'Poincaré 2-groups', come from the ordinary Poincaré groups $\mathrm{SO}(p, q) \ltimes \mathbb{R}^{p+q}$ in precisely this way.

In this special case where δ is trivial, a 2-morphism u given by the pair (g, h) has g as its source morphism and $\delta(h)g = g$ as its target morphism. Thus, the 2-group has only 2-*auto*morphisms, and each morphism g has precisely one automorphism for each element of H:

$$\begin{tikzcd}
\star \arrow[r, bend left=50, "g"] \arrow[r, bend right=50, "g"'] \arrow[r, phantom, "\Downarrow h"] & \star
\end{tikzcd}$$

8.3 2-connections and 2-groups for 4d BF theory

In 4d spacetime BF theory has an interpretation as a 'higher gauge theory'—a theory of a '2-connection' [16] on a principal '2-bundle' [19] with some 2-group as its 'gauge 2-group'. We only give a very naive description of this subject, since anything more would take us too far afield of our immediate goal: describing particles in 4d BF theory.

Roughly, though, the data for a 2-connection on a trivial 2-bundle with gauge 2-group $\mathcal{G} = (G, H, t, \rhd)$ are:

- A $\mathfrak{g}$-valued 1-form A, and

- An $\mathfrak{h}$-valued 2-form E.

This is analogous to ordinary connections, which are locally Lie algebra valued 1-forms. In fact, the pair (A, E) should be thought of as taking values in the 'Lie 2-algebra' [11] $(\mathfrak{g}, \mathfrak{h})$, which amounts to a 'differential' version of the gauge 2-group described by (G, H).

The essential point for our purposes is that a 2-connection allows us to assign elements of H to surfaces, as well as elements of G to paths. Recall first how this works for ordinary connections. In local coordinates, a G-connection on spacetime M is simply a Lie algebra valued 1-form:

$$A\colon TM \to \mathfrak{g}.$$

In this description, calculating the holonomy of A around a based loop $\gamma\colon I \to M$ amounts to "integrating" the Lie group elements $exp(A(\gamma'(t)))$ around the loop. Of course, this is an oversimplification—when G is nonabelian we must really use the path-ordered exponential [9]

$$P\exp\left(-\int_0^1 A(\gamma'(t))dt\right) = \sum_{n=0}^{\infty}(-1)^n\int_{1\geq t_1\geq\cdots\geq t_n\geq 0} A(\gamma'(t_1))\cdots A(\gamma'(t_n))dt_n\cdots dt_1$$

to maintain gauge invariance. Similarly, if we want to assign group elements to surfaces using a 2-connection, we can try integrating the 2-form E over the surface to get an element of H.

In fact, Baez and Schreiber have shown that one *can* perform a kind of 'integral' of the E field over a surface. This calculation assigns well-defined gauge-invariant '2-holonomy' to the surface, but only if A and E satisfy an additional compatibility condition called **fake flatness**:

$$F + dt \circ E = 0 \tag{8.9}$$

Here $F = dA + \frac{1}{2}[A, A]$ is the curvature of A, and

$$dt\colon \mathfrak{h} \to \mathfrak{g}$$

is the differential of the homomorphism $t\colon H \to G$ in the crossed module description of our 2-group. The computation of 2-holonomies is really the analog of the path ordered exponential used to calculate the holonomy of a connection, and deserves to be called the 'surface-ordered exponential'. It can also be calculated as an ordinary path ordered exponential in a suitable space of paths in spacetime. [16]

Now let us turn to 4d BF theory, with trivial principal G bundle $P = M \times G$. The connection A is a $\mathfrak{g}$-valued 1-form, while the E field is a $\mathfrak{g}$-valued 2-form, so we seem to have the right ingredient for a 2-connection, provided the 2-group (G, H) has both $\mathfrak{g} = \mathfrak{h}$. At first, there might seem to be just one sensible choice. However, we only get well-defined holonomies over surfaces if we can choose the 2-group in such a way that the field equations imply the fake flatness condition. These considerations lead to different 2-groups for BF theory, depending on whether we include a cosmological term in the Lagrangian [8]:

Lagrangian	field equations	2-group
$\mathrm{tr}\,(E \wedge F)$	$F = 0 \quad d_A E = 0$	$(G, \mathfrak{g}, \mathrm{Ad}, 0)$
$\mathrm{tr}\,(E \wedge F + \frac{\Lambda}{2} E \wedge E)$	$F = \Lambda E$	$(G, G, \mathrm{AD}, 1)$

Let us first discuss the case without cosmological term. As explained in the previous section, we get a 2-group from any semidirect product of groups. The 2-group $(G, \mathfrak{g}, \mathrm{Ad}, 0)$ comes from the semidirect product $G \ltimes \mathfrak{g}$, where G acts on the additive group of its Lie algebra via the adjoint action. The field equations for the theory,

$$F = 0 \qquad d_A E = 0$$

clearly imply the fake flatness condition $F + dt \circ B = 0$, since $F = 0$ and the target homomorphism $t\colon \mathfrak{g} \to G$ is trivial. So with this choice of 2-group we get a 2-connection with well defined surface holonomies.

The field equations for BF theory put additional constraints on the 2-connection. The equation $F = 0$, implies as usual that the holonomy along a path is invariant under homotopy (preserving the endpoints). This is nothing new, of course: it is what we have been using all along to describe ordinary flat connections. In particular, this gives us a homomorphism

$$\mathrm{hol}\colon \pi_1(M) \to G.$$

The second equation of BF theory says that the **curvature 3-form** $d_A E$ also vanishes. This is believed to imply that surface holonomies are also homotopy invariant. In particular, we get a homomorphism

$$2\mathrm{hol}\colon \pi_2(M) \to H.$$

We have seen that BF theory has a much different character when we include a cosmological term. In this case the Bianchi identity reduces the equations of motion to a single equation:

$$F = \Lambda E$$

The 2-group used above for the $\Lambda = 0$ therory will not do in this case: the equations of motion would not imply fake flatness. However, it is easy to guess a 2-group that will work: $(G, G, \mathrm{AD}, 1_G)$. This 2-group has G as its group of morphisms, $G \ltimes G$ as 2-morphisms, with the action of G on itself by conjugation.

Since one of our big themes has been the behavior of gauge theories under contractions of groups, it is interesting to note that the 2-groups for BF theory with and without cosmological constant seem to be related by a kind of '2-group contraction'

$$(G, G, \mathrm{AD}, 1) \xrightarrow{\Lambda \to 0} (G, \mathfrak{g}, \mathrm{Ad}, 0)$$

We make no attempt here to describe contractions for Lie 2-groups in any rigorous way, but we hope the reader can see this is an appealing idea which deserves further attention. The idea is that the group G 'flattens out' to its Lie algebra $\mathfrak{g}$ when the cosmological term becomes zero, while the adjoint action of G on itself becomes the adjoint action on $\mathfrak{g}$.

8.4 Particles

We now return to 4d BF theory without cosmological constant term, and with gauge group G. As we have seen, this can be viewed as a higher gauge theory with gauge 2-group $(G, \mathfrak{g}, 1_G, \mathrm{AD})$. When spacetime is of the form $M = X \times \mathbb{R}$ for some 3-manifold X representing 'space', we can restrict the fields A and E to get a 2-connection on X. This 2-connection is again flat in the sense that the curvature 2-form F and the curvature 3-form $d_A E$ both vanish. We warn the reader that this section is rather less precise than the analogous cases we have considered using ordinary gauge theory. Our goal is to provide a sketch, leaving the hard work in this area to future research.

By analogy with what we did for particles in 3d BF theory in Chapter 6, and for strings in 4d BF theory in Chapter 7, let us consider 4d BF theory where space is $\mathbb{R}^3$ with n punctures:

$$X = S - \Sigma, \qquad S = \mathbb{R}^3, \qquad \Sigma = \{z_1, \ldots, z_n\}$$

and interpret these punctures as 'particles'. In this case the motion group is

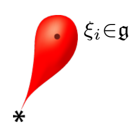
$$\mathrm{Mo}(S, \Sigma) \cong S_n$$

so only ordinary bosonic or fermionic statistics are possible.

The fundamental group of X is trivial, so there are no interesting 1-holonomies. The second homotopy group $\pi_2(X)$ however is the free abelian group on n generators, so we get 2-holonomies living in

$$\hom(\pi_2(X), \mathfrak{g}) = \mathfrak{g}^n$$

Of course, we have only described 2-connections on a *trivial* principal 2-bundle. In the case of ordinary connections, we needed to let the bundle be variable in order to get all homomorphisms $\pi_1 \to G$ as holonomies. But let us assume an analogous result holds in the present case, so that the 2-holonomy can be specified by picking n Lie algebra elements arbitrarily. Each of these n Lie algebra elements, say $\xi_i \in \mathfrak{g}$, represents the holonomy over a chosen generator of π_2, enclosing the ith particle:



As in the lower-dimensional cases, we expect these holonomies to describe certain properties of the particle, such as analogs of energy-momentum and spin. After factoring out gauge transformations, we expect to get

$$\hom(\pi_2(X), \mathfrak{g})/G$$

as the space of 2-holonomies, where $g \in G$ acts as gauge transformations by conjugating the 2-holonomies:

$$(\xi_1, \ldots, \xi_n) \to (\mathrm{Ad}(g)\xi_1, \ldots, \mathrm{Ad}(g)\xi_n)$$

This seems to indicate that particles in 4d BF theory are classified not by conjugacy classes in the gauge group but by adjoint orbits in its Lie algebra.

Actually, saying that adjoint orbits give 'particle types' in 4d BF theory is perhaps too strong of a statement. In the 3d case, we classified a particle type according to the conjugacy class of the holonomy around it. But the flat connection A corresponds to a point in the configuration space for the theory; the field E does not—it is the momentum conjugate to A. So in Schrödinger quantization of the theory, one does not describe time evolution of the E field. However, one important point is that BF theory is not ultimately what we are interested in. As we review in see in the final part of this thesis, BF theory can be deformed into general relativity, via Freidel and Starodubtsev's reformulation of MacDowell–Mansouri theory. In this deformation, the symplectic structure changes, and we are not yet certain how these 'particle type' data contained in the 2-holonomy will transform in the deformation.

In fact, there are clues that adjoint orbits may relate to types of particles in 4d quantum gravity. In a paper with Baez, Baratin, Freidel, and Morton [10], we classify the '2-representations' of the Poincaré 2-groups mentioned in Section 8.2.2. The results of this paper can be immediately extended to see that irreducible representations of the tangent 2-group $(G, \mathfrak{g}, 0, \mathrm{Ad})$ are classified by the adjoint orbits in $\mathfrak{g}$. This work is related to the Baratin–Freidel spin foam model, which essentially reproduces ordinary quantum field theory in 4d spacetime

In any case, at least at the classical level, the 2-holonomy really is some invariant thing that we can calculate about a particle. If we do BF theory based on some 4d Kleinian geometry, such as the deSitter Model $\mathrm{SO}(4,1)/SO(3,1)$, we expect that this holonomy tells us how the particle affects the geometry of the space around it, just as the holonomy around a particle in 3d gravity tells us the angle deficit introduced by its energy-momentum. These holonomies also represent 'conserved quantities', since if two particle with 2-holonomies ξ_1, ξ_2 coalesce in an interaction, homotopy invariance implies the 2-holonomy around the resulting particle must be $\xi_1 + \xi_2$.

8.5 Particle/string statistics and Bohm–Aharonov duality

The most interesting possibilities happen when we allow both particles and strings in 4d BF theory. In fact, even when there are just strings around, one gets holonomies over both paths and surfaces. For example, suppose we just have a single string, so space is

$$X = \mathbb{R}^3 - \{\text{unknot}\}.$$

Then X has a deformation retraction onto the union of a sphere and a line through its origin, which shows X has the homotopy type

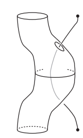
$$S^1 \vee S^2,$$

hence has nontrivial π_1 and π_2. More generally, space is $\mathbb{R}^3$ minus a collection of n strings and m particles, its homotopy type is

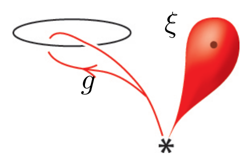
$$\underbrace{S^2 \vee \cdots \vee S^2}_{n+m} \vee \underbrace{S^1 \vee \cdots \vee S^1}_{n}$$

That is, we get a 1-holonomy for each string, a 2-holonomy for each particle *and* each string.

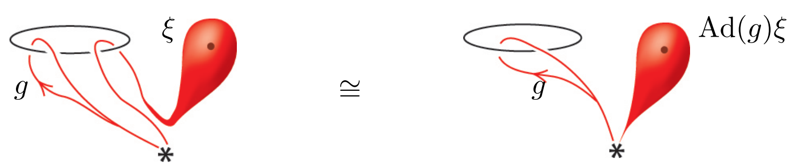
To the reader who has studied our diagrams for loop braids, the meaning of the following diagram should be clear:



This represents a particle passing over and then down through a string, and then the two returning to their original places. The string is labeled by its holonomy $g \in G$, while the particle is labeled by its 2-holonomy $\xi \in \mathfrak{g}$:



When we perform the operation depicted above, passing the particle down through the string, nothing happens to the holonomy around the string, but the 2-holonomy around the particle gets conjugated:



This 'conjugation' of the 2-holonomy by the 1-holonomy just corresponds to the usual action of π_1 on π_2. At the level of the 2-groups $(G, \mathfrak{g}, \mathrm{Ad}, 0)$, what is happening here can be represented by the diagram

$$\begin{tikzcd} \star \arrow[r, "g^{-1}"] & \star \arrow[r, bend left, "1"] \arrow[r, bend right, "1"'] \arrow[r, phantom, "\Downarrow\xi"] & \star \arrow[r, "g"] & \star \end{tikzcd}$$

which represents 'conjugation' of the 2-morphism $\xi \colon 1 \to 1$ by the morphism g. This sort of operation is called 'whiskering' a 2-morphism, for obvious reasons. To make sense of it, we interpret the morphism g as really representing the identity 2-morphism $1_g\colon g \to g$, and similarly for g^{-1}. This is of course the same as the trick for getting the action of π_1 on π_2, we think of a path as a degenerate surface. Noting that the identity morphism 1_g just corresponds to the identity $0 \in \mathfrak{g}$, and applying the rule for horizontal composition of 2-morphisms we get

$$\begin{tikzcd} \star \arrow[r, bend left, "g^{-1}"] \arrow[r, bend right, "g^{-1}"'] \arrow[r, phantom, "\Downarrow 0"] & \star \arrow[r, bend left, "1"] \arrow[r, bend right, "1"'] \arrow[r, phantom, "\Downarrow \xi"] & \star \arrow[r, bend left, "g"] \arrow[r, bend right, "g"'] \arrow[r, phantom, "\Downarrow 0"] & \star \end{tikzcd} \; = \; \begin{tikzcd} \star \arrow[r, bend left, "1"] \arrow[r, bend right, "1"'] \arrow[r, phantom, "\Downarrow Ad(g)\xi"] & \star \end{tikzcd}$$

That is, the 1-holonomy acts on the 2-holonomy via the adjoint action, just as in the above illustration.

These ideas lead to a kind of 'exotic particle/string statistics'. If we allow the definition of manifold to include the possibility that different connected components have different dimensions, then the definition of the motion group $\mathrm{Mo}(S, \Sigma)$ in Section 5.2 carries over immediately to the case where space is $\mathbb{R}^3$ with a finite set of unlinked unknotted circles removed (describing *closed strings*), *and* with a finite set of points removed (describing *point particles*):

$$X = S - \Sigma, \qquad S = \mathbb{R}^3, \qquad \Sigma = \ell_1 \cup \cdots \cup \ell_n. \cup \{z_1, \ldots, z_m\}.$$

The motion group of n strings and m particles should have a presentation with all the generators of LB_n and S_m, together with generators passing a particle through a string, as in the above pictures.

There is another interesting way to interpret these effects. We can think of the points $z_1, \ldots, z_m$ as particles, but think of the circles $\ell_1 \ldots, \ell_n$ simply as 'obstacles', which our particles may or may not pass through. At the quantum level, this gives something like the Bohm–Aharonov effect. But dually, we can think of the points as obstacles, and the circles as string–like matter that can either pass around the obstacles or not. This gives something like the the 2-form Bohm–Aharonov effect that shows up in 2-form electromagnetism [98]. So, we get a kind of duality between Bohm–Aharonov effects for particles scattering off of strings on one hand, and strings scattering off of particles on the other.

8.6 Adjoint orbits

In the previous section we saw that particles in 4d BF theory are classified by adjoint orbits in the Lie algebra of the gauge group. Let us describe the adjoint orbits in the Lie algebras

$$G = \mathfrak{so}(p,q),$$

using the results of Burgoyne and Cushman [22].

Burgoyne and Cushman classify adjoint orbits by classifying the Jordan canonical forms (over $\mathbb{C}$) of their representatives. Let us introduce the notation $J_k(\zeta)$ for a $k \times k$ elementary Jordan block with ζ down the diagonal:

$$J_k(\zeta) := \underbrace{\begin{pmatrix} \zeta & 1 & & \\ & \zeta & \ddots & \\ & & \ddots & 1 \\ & & & \zeta \end{pmatrix}}_{k}$$

In particular $J_1(\zeta)$ denotes a 1×1 matrix whose entry is ζ. If $X \in \mathfrak{g}$, then X is conjugate to a direct sum (i.e. a block diagonal matrix) with blocks of the following types.

type	Jordan form		n_+	n_-
(a)	$\begin{pmatrix} J_k(\zeta) & & & \\ & J_k(-\zeta) & & \\ & & J_k(\bar\zeta) & \\ & & & J_k(-\bar\zeta) \end{pmatrix}$	$\zeta \neq \pm\bar\zeta$	$2k$	$2k$
(b)	$\begin{pmatrix} J_k(\zeta) & 0 \\ 0 & J_k(-\zeta) \end{pmatrix}$	$\zeta \neq 0$ real	k	k
(c)	$\begin{pmatrix} J_k(\zeta) & 0 \\ 0 & J_k(-\zeta) \end{pmatrix}$	$\zeta \neq 0$ imaginary k odd	$k \pm 1$	$k \mp 1$
(d)	$\begin{pmatrix} J_k(\zeta) & 0 \\ 0 & J_k(-\zeta) \end{pmatrix}$	ζ imaginary or 0 k even	k	k
(e)	$J_k(0)$	k odd	$\frac{k\pm1}{2}$	$\frac{k\mp1}{2}$

Let us work out a few examples. The easiest case is $G = \mathrm{SO}(n)$. Since there are no minus signs in the metric, each block must have $n_- = 0$. Inspecting the table, we see that cases (c) and (e) with $k = 1$ are the only possibilities. Thus adjoint orbits in $\mathfrak{so}(n)$ are represented by block matrices whose nonzero blocks have Jordan canonical form

$$\begin{pmatrix} i\xi & 0 \\ 0 & -i\xi \end{pmatrix}$$

for some $\xi \in \mathbb{R}$. This matrix is conjugate to

$$\begin{pmatrix} 0 & \xi \\ -\xi & 0 \end{pmatrix} \in \mathfrak{so}(2).$$

The most general adjoint orbit in $\mathfrak{so}(n)$ thus has a block diagonal representative with $\lfloor n/2 \rfloor$ blocks of this form.

The case $G = \mathrm{SO}(n,1)$ is more interesting. Cases (a) and (d) are still not possible here, but case (b) now works, with $k = 1$. For example, for the de Sitter Lie algebra $\mathfrak{so}(4,1)$, we get the following types of blocks:

type	Jordan form	$\mathfrak{so}(4,1)$ block	n_+	n_-
(b)	$\begin{pmatrix} u & 0 \\ 0 & -u \end{pmatrix}$	$\begin{pmatrix} 0 & u \\ u & 0 \end{pmatrix}$	1	1
(c)	$\begin{pmatrix} iu & 0 \\ 0 & -iu \end{pmatrix}$	$\begin{pmatrix} 0 & u \\ -u & 0 \end{pmatrix}$	2	0
(e)	(0)	(0)	1 0	0 1
	$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}$	2	1

We can assemble a representative of an adjoint orbit by combining the $\mathfrak{so}(4,1)$ blocks into 5×5 matrices, making sure that the total of the signatures (n_+, n_-) is $(4,1)$. Doing this, we see that any adjoint orbit has an representative that looks like one of the following:

1. $$\begin{pmatrix} 0 & u & & & \\ u & 0 & & & \\ & & 0 & v & \\ & & -u & 0 & \\ & & & & 0 \end{pmatrix}$$

2. $$\begin{pmatrix} 0 & & & & \\ & 0 & u & & \\ & -u & 0 & & \\ & & & 0 & v \\ & & & -u & 0 \end{pmatrix}$$

3. $$\begin{pmatrix} 0 & 1 & 0 & & \\ 1 & 0 & 1 & & \\ 0 & -1 & 0 & & \\ & & & 0 & u \\ & & & -u & 0 \end{pmatrix}$$

Part III

Cartan Geometry and Gravity

Chapter 9

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 to 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, we refer the reader to the references for further details on this very rich subject. In particular, the book by Sharpe [86] and the article by Alekseevsky and Michor [1] are helpful resourses, and serve as the major references for our explanation here.

We 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. Our purpose in reviewing this definition is merely to easily contrast it with the definition of a 'Cartan connection', to be given in Section 9.2.

9.1 Ehresmann connections

Before giving the definition of Cartan connection we review the more familiar notion of an Ehresmann connection on a principal bundle. In 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 [30]. The definition we shall use is the following one.

Definition 9 *An* **Ehresmann connection** *on a principal H bundle*

$$\begin{tikzcd} P \arrow[d, "\pi"] \\ M \end{tikzcd}$$

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.

9.2 Definition of Cartan geometry

We are ready to state the formal definition of Cartan geometry, essentially as given by Sharpe [86].

Definition 10 *A* **Cartan geometry** $(\pi\colon P \to M, A)$ *modeled on the Klein Geometry* (G, H) *is a principal right H bundle*

$$\begin{tikzcd} P \arrow[d, "\pi"] \\ M \end{tikzcd}$$

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 9.4.

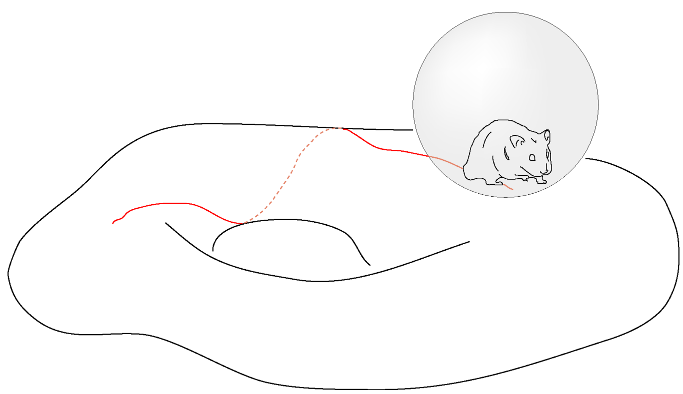
9.3 Geometric interpretation: rolling Klein geometries

In Section 1, we 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.

We 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.

9.4 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:

$$\begin{tikzcd}
 & & \mathfrak{h} \\
TP \arrow[rru, bend left, "\omega"] \arrow[r, "A"] \arrow[rrd, bend right, "e"'] & \mathfrak{g} \arrow[ru] \arrow[rd] & \\
 & & \mathfrak{g}/\mathfrak{h}
\end{tikzcd}$$

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:

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

The $\mathfrak{g}/\mathfrak{h}$ part T is the **torsion**. The $\mathfrak{h}$ part $\widehat{F}$ is related to the curvature of the Ehresmann connection ω, but there is an important difference: *The 'curvature' $\widehat{F}$ 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,

$$d\omega_G = -\frac{1}{2}[\omega_G, \omega_G], \tag{9.1}$$

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 4d Kleinian model spacetimes discussed in Chapter 2—de Sitter, Minkowski, anti de Sitter, and their Riemannian analogs—are all reductive.

	$\Lambda < 0$	$\Lambda = 0$	$\Lambda > 0$
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[1]

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

where the value of ϵ depends on the algebra:

$$\epsilon = \begin{cases} 1 & \mathfrak{g} = \mathfrak{so}(4,1) \\ 0 & \mathfrak{g} = \mathfrak{iso}(3,1) \\ -1 & \mathfrak{g} = \mathfrak{so}(3,2) \end{cases} \tag{9.2}$$

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

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[2]

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

which is invariant under $\mathrm{Ad}(G)$, hence under $\mathrm{Ad}(H)$. For the $\epsilon = 0$ case, we do not have a G-invariant metric. However, we really only require a metric invariant 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.

In any of these cases, with respect to this appropriate 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

$$\lambda = \frac{3\epsilon}{\ell^2} \tag{9.3}$$

[2]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.

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 9.4, in the context of Cartan 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.

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 (9.2). 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 (9.3) of the model homogeneous spacetime. Indeed, the Maurer–Cartan form ω_G is a Cartan connection for the model spacetime, and the structural equation (9.1) 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 modeled 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:

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

Chapter 10

Cartan-type gauge theory

Part of the case we 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, we 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.

10.1 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

$$\iota\colon P \to P \times_H G$$
$$p \mapsto [p, 1_G]$$

is a canonical inclusion of H bundles. We 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 10, 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 $\ker\tilde{A} \cap \iota_*(TP) = 0$ [86].

10.2 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 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 it is helpful 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 isomorphic as fiber bundles, the isomorphism is not an isomorphism of associated bundles (in the sense described by Isham [55]), 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 the 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

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

where

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

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.

10.3 Holonomy and development

Just as a Cartan geometry has two notions of parallel translation, it also has two notions of holonomy, taking values in 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 we 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 [86].

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

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

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 $\boldsymbol{\gamma}$ on $\boldsymbol{G/H}$** starting at gH. This path is independent of the lifting γ, depending only on the path in the base manifold M. [86]

In the SO(3)/SO(2) example of Section 9.3, the development is the path traced out on the ball itself by the point of tangency on the surface, as the ball rolls.

10.4 Cartan-type $\boldsymbol{BF}$ Theory

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' [7]. Special cases of such a theory have already been considered by Freidel and Starodubtsev, in connection with MacDowell–Mansouri gravity [43], 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

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

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' [93], which have been used to study a particular low-dimensional case of BF theory, namely, 3d quantum gravity [25].

Let us work out a more explicit example: Cartan-type BF theory based on one of the Lorentzian reductive models discussed in Sections 2.3.3 and 9.4. 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 $b = \widehat{B}$ 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 (9.2). 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.

Chapter 11

From Palatini to MacDowell–Mansouri

In this chapter we 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. This in turn leads to generalizations of MacDowell–Mansouri theory for alternative model geometries, just as we explored generalizations of 3d gravity in terms of geometric structures. But first, we review the Palatini approach in some detail.

11.1 The Palatini formulation of general relativity

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 may thus be viewed as an extension of general relativity to degenerate metrics. While classical solutions of general relativity do not have degenerate metrics, imposing a nondegeneracy constraint is particularly troublesome when we try path-integral quantization.

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

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

where $D_\mu := D_{\partial_\mu}$ denotes the covariant derivative in the μth coordinate direction. 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:

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

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

$$\begin{aligned}
(\nabla_\mu w)^\alpha &= e^\alpha_a (D_\mu e_\beta w^\beta)^a \\
&= e^\alpha_a \left(\partial_\mu (e^a_\beta w^\beta) + A^a_{\mu b} e^b_\beta w^\beta\right) \\
&= e^\alpha_a \left(e^a_\beta \partial_\mu w^\beta + (\partial_\mu e^a_\beta) w^\beta + A^a_{\mu b} e^b_\beta w^\beta\right) \\
&= \partial_\mu w^\alpha + (e^\alpha_a \partial_\mu e^a_\beta + e^\alpha_a A^a_{\mu b} e^b_\beta) w^\beta
\end{aligned}$$

Hence,

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

where

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

The curvature of course transforms by a simpler formula:

$$R^\alpha_{\mu\nu\beta} = e^\alpha_a F^a_{\mu\nu b} e^b_\beta.$$

as can be shown by a direct, though rather lengthy calculation in the index notation. The simpler way to see that this formula is true is to compare to the way the curvature in gauge theory transforms under gauge transformations.

The Palatini action is

$$S_{\mathrm{Pal}}(\omega, e) = \frac{1}{2G}\int_M \mathrm{tr}\left(e\wedge e\wedge R + \frac{\Lambda}{6}\, e\wedge e\wedge e\wedge e\right). \tag{11.1}$$

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 4-form on M, and the 'trace' is really a map that turns such a form into an ordinary 4-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:

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

Taking the variation of the action gives

$$\begin{aligned}
\delta S &= \int \mathrm{tr}\,(2\delta e\wedge e\wedge R + e\wedge e\wedge \delta R + \frac{2\Lambda}{3}\,\delta e\wedge e\wedge e\wedge e)\\
&= \int \mathrm{tr}\,(2\delta e\wedge (e\wedge R + \frac{\Lambda}{3}\, e\wedge e\wedge e) + e\wedge e\wedge d_\omega\delta\omega)\\
&= \int \mathrm{tr}\,(2\delta e\wedge (e\wedge R + \frac{\Lambda}{3}\, e\wedge e\wedge e) \pm d_\omega(e\wedge e)\wedge\delta\omega)
\end{aligned}$$

where we used the identity $\delta R = d_\omega\delta\omega$ and performed an integration by parts. The variations of ω and e give us the respective equations of motion

$$d_\omega(e\wedge e) = 0 \tag{11.2}$$

$$e\wedge R + \frac{\Lambda}{3}\, e\wedge e\wedge e = 0. \tag{11.3}$$

In Section 11.2, we review the correspondence between these equations and the equations of general relativity in the perhaps more familiar tensor notation. Briefly, 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.

11.2 Equivalence of formulations

There are many equivalent ways of writing general relativity. In this section we provide the reader with a translation between the Palatini formulation and the tensor formulation that is usually presented first in any general relativity text. We restrict to the sourceless case.

Proposition 14 *At each point where the coframe field is an isomorphism $e\colon T_xM \xrightarrow{\sim} \mathcal{T}_x$, we have the following relationships between the internal ($\star$) and spacetime ($*$) Hodge duals of wedge powers of e:*

$$p!\,\star(\underbrace{e\wedge\cdots\wedge e}_{p}) \quad = \quad (n-p)!\,*(\underbrace{e\wedge\cdots\wedge e}_{n-p})$$

Proof: While the proof of the equation is straightforward to carry out in general, we avoid much notational clutter by considering a particular example. It is easy to see the same proof holds in general. Consider the case of 4 dimensions, supposing that $e\colon T_xM \xrightarrow{\sim} \mathbb{R}^4$ is an isomorphism, and let us prove that

$$1!\,\star e = 3!\,*(e\wedge e\wedge e). \tag{11.4}$$

First note that at each point $x \in M$ we have

$$e \in \mathbb{R}^4 \otimes T^*_xM \qquad \text{and} \qquad (e\wedge e\wedge e) \in \Lambda^3\mathbb{R}^4 \otimes \Lambda^3 T^*_xM$$

Applying the internal and spacetime Hodge stars we see that $\star e$ and $*(e\wedge e\wedge e)$ both live in

$$\Lambda^3\mathbb{R}^4 \otimes T^*_xM.$$

The two elements in question have components given by (see Appendix **??**)

$$\star e^{ijk}_{\lambda} = \varepsilon^{ijk}{}_{\ell} e^{\ell}_{\lambda}$$

and

$$\begin{aligned} *(e \wedge e \wedge e)^{ijk}_{\lambda} &= \frac{1}{3!} \varepsilon^{\mu\nu\rho}{}_{\lambda} e^{[i}_{\mu} e^{j}_{\nu} e^{k]}_{\rho} \\ &= \frac{1}{3!} \varepsilon^{\mu\nu\rho}{}_{\lambda} e^{i}_{\mu} e^{j}_{\nu} e^{k}_{\rho} \end{aligned}$$

where the second equality follows from antisymmetry of $\varepsilon^{\mu\nu\rho}{}_{\lambda}$ in μ, ν, ρ, and permuting the order of the factors e^{a}_{α}. The epsilon tensors in these two expressions are related by the isomorphism e. In particular, indices of $\varepsilon^{ijk}{}_{\ell}$ are raised using the metric η, while indices of $\varepsilon^{\mu\nu\rho}{}_{\lambda}$ are raised using the pullback metric g. Using the definition (**??**) of g we see that the components of $\star e$ and $3! * (e \wedge e \wedge e)$ are both equal to the components of the 'mixed Levi–Civita tensor'

$$\varepsilon^{ijk}{}_{\lambda} = \varepsilon^{ijk}{}_{\ell} e^{\ell}_{\lambda} = \varepsilon^{\mu\nu\rho}{}_{\lambda} e^{i}_{\mu} e^{j}_{\nu} e^{k}_{\rho}$$

so the equality is established. $\square$

It is worth writing down the general expression for one of the mixed Levi–Civita tensors coming up in the proof.

$$\varepsilon^{i_1 \cdots i_{n-p}}{}_{\mu_1 \cdots \mu_p} = \varepsilon^{i_1 \cdots i_{n-p}}{}_{j_1 \cdots j_p} e^{j_1}_{\mu_1} \cdots e^{j_p}_{\mu_p} = \varepsilon^{\nu_1 \cdots \nu_{n-p}}{}_{\mu_1 \cdots \mu_p} e^{j_1}_{\nu_1} \cdots e^{j_{n-p}}_{\nu_{n-p}}.$$

We also have the following:

Proposition 15 *Suppose $e \colon V \to W$ is an isomorphism of N-dimensional vector spaces. Then*

$$e \wedge - \colon \Lambda^p V^* \otimes \Lambda^q W \to \Lambda^{p+1} V^* \otimes \Lambda^{q+1} W$$

is injective if and only if $p + q \leq N - 1$.

Applying this result to the case where $V = T_x M$, $W = \mathcal{T}_x$, we see that $e \wedge d_A e = 0$ if and only if $d_A e = 0$, since

$$d_A(e \wedge e) = (d_A e) \wedge e - e \wedge (d_A e) = 2(d_A e) \wedge e$$

and $- \wedge e$ is injective.

The equation $d_A e = 0$ says precisely that ∇ is torsion free. To see this, it is easiest to work locally. We have

$$\begin{aligned} 0 = (d_A e)^a_{\mu\nu} &= (de)^a_{\mu\nu} + (A \wedge e)^a_{\mu\nu} \\ &= \partial_\mu e^a_\nu - \partial_\nu e^a_\mu + A^a_{\mu b} e^b_\nu - A^a_{\nu b} e^b_\mu \end{aligned}$$

Applying e^α_a to this we get

$$e^\alpha_a \partial_\mu e^a_\nu + e^\alpha_a A^a_{\mu b} e^b_\nu = e^\alpha_a \partial_\nu e^a_\mu + e^\alpha_a A^a_{\nu b} e^b_\mu$$

or

$$\Gamma^\alpha_{\mu\nu} = \Gamma^\alpha_{\nu\mu},$$

which is the usual index-based way of saying that the connection ∇ with Christoffel symbols Γ is torsion free. In fact, ∇ is also metric preserving, since D is, so that ∇ is the Levi-Civita connection on spacetime.

The other equation of motion is Einstein's equation in disguise. In index notation, this equation may be written

$$\epsilon_{ijkl}(e^i_\lambda \wedge R^{jk}{}_{\mu\nu} - \frac{2\Lambda}{3} e^i_\lambda \wedge e^j_\mu \wedge e^k_\nu) = 0.$$

Since the curvature tensor is antisymmetric in spacetime indices and internal indices, we may write it as

$$R^{jk}{}_{\mu\nu} = R^{jk}{}_{mn} e^m_\mu \wedge e^n_\nu$$

Using this and applying the Hodge star operator on spacetime indices we get

$$\begin{aligned} 0 &= \varepsilon^{\lambda\mu\nu}{}_\pi \varepsilon_{ijk\ell}(R^{jk}{}_{mn} e^i_\lambda \wedge e^m_\mu \wedge e^n_\nu - \frac{2\Lambda}{3} e^i_\lambda \wedge e^j_\mu \wedge e^k_\nu) \\ &= \varepsilon_{ijk\ell}(R^{jk}{}_{mn} \varepsilon^{imn}{}_p - \frac{2\Lambda}{3} \varepsilon^{ijk}{}_p) e^p_\pi \\ &= \varepsilon_{ijk\ell}(R^{jk}{}_{mn} \varepsilon^{imnq} - \frac{2\Lambda}{3} \varepsilon^{ijkq}) \eta_{pq} e^p_\pi \\ &= (-3!\, \delta^m_{[j} \delta^n_k \delta^q_{\ell]} R^{jk}{}_{mn} + 4\Lambda\, \delta^{q\ell}) \eta_{pq} e^p_\pi \end{aligned}$$

where in the second equality we have used Proposition 14, and in the fourth we've used standard contraction identities for Levi–Civita symbols (see Appendix **??**). The antisymmetrized δ's in the first term serve to contract the curvature into the internal version of the

Einstein tensor, as follows:

$$\begin{aligned}
3!\, \delta^m_{[j}\delta^n_k\delta^q_{\ell]} R^{jk}{}_{mn} =& \delta^m_j\delta^n_k\delta^q_\ell R^{jk}{}_{mn} + \delta^m_k\delta^n_\ell\delta^q_j R^{jk}{}_{mn} + \delta^m_\ell\delta^n_j\delta^q_k R^{jk}{}_{mn} \\
& - \delta^m_k\delta^n_j\delta^q_\ell R^{jk}{}_{mn} - \delta^m_j\delta^n_\ell\delta^q_k R^{jk}{}_{mn} - \delta^m_\ell\delta^n_k\delta^q_j R^{jk}{}_{mn} \\
=& R^{mn}{}_{mn}\delta^q_\ell + R^{qm}{}_{mn}\delta^n_\ell + R^{nq}{}_{mn}\delta^m_\ell \\
& - R^{nm}{}_{mn}\delta^q_\ell - R^{mq}{}_{mn}\delta^n_\ell - R^{qn}{}_{mn}\delta^m_\ell \\
=& 2R\delta^q_\ell - 4R^q{}_\ell = -4G^q{}_\ell
\end{aligned}$$

where the **internal Einstein tensor** is given by

$$G_{mn} = R_{mn} - \tfrac{1}{2}R\eta_{mn}.$$

Using this result we get

$$G_{mn} + \Lambda\eta_{mn} = 0,$$

or, applying the coframe field to turn internal indices to spacetime indices:

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 0.$$

11.3 The coframe field

We now begin our transition from the Palatini approach to the more intrinsically Cartan–geometric approach of MacDowell and Mansouri. 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 explained 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 we 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 10.1.

11.4 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 (11.1) using the internal Hodge star operator as

$$S_{\mathrm{Pal}} = \frac{-1}{G}\int \mathrm{tr}\left((e \wedge e \wedge \star R - \frac{\Lambda}{6} e \wedge e \wedge \star(e \wedge e)\right). \tag{11.5}$$

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 $\mathfrak{so}(3,1)$-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 2.3.3, 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

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

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

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

The $R \wedge \star R$ term here is a topological invariant, having vanishing variation due to the Bianchi identity. The classical theory is unaffected if we simply discard this term. If we also recognize that $\widehat{F} \wedge \star \widehat{F} = F \wedge \star \widehat{F}$, we obtain the **MacDowell–Mansouri action** (1.2) as we presented it in Section 1:

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

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

$$S = \int \mathrm{tr}\, \left(B \wedge F - \frac{\alpha}{2} B \wedge \star \widehat{B} \right). \tag{11.6}$$

where

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

Calculating the variation, we get:

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

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,

$$F = \alpha \star \widehat{B} \tag{11.7}$$

$$d_A B = 0 \tag{11.8}$$

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, noting that $\star^2 = -1$, we obtain

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

which is precisely the MacDowell–Mansouri action. However, it is instructive to see Einstein's equations coming directly from the equations of motion (11.7) and (11.8).

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 10.4. Indeed, if we take $G = 0$, they are identical. This is good, because it says turning off Newton's gravitational constant 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 (11.5):

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

namely, Einstein's equation.

11.5 Lagrangians for gravity and topological gauge theories

As mentioned in the previous section, beginning with the Freidel–Starodubtsev Lagrangian 11.6 and substituting the equation of motion 11.8, one obtains the original MacDowell–Mansouri Lagrangian. In fact, one can do the analogous trick to an ordinary 4d BF theory with cosmological term and obtain the 2nd Chern form [9]. We thus get a commutative diagram of theories:

$$\begin{tikzcd}
\text{topological Lagrangians} & \text{Lagrangians for general relativity} \\
\mathrm{tr}\left(B \wedge F + \frac{\alpha}{2} B \wedge B\right) \arrow[r, mapsto, "\text{break symmetry}"] \arrow[d, mapsto, "\substack{\text{substitute} \\ B = -\frac{1}{\alpha}F}"'] & \mathrm{tr}\left(B \wedge F + \frac{\alpha}{2} B \wedge \star\widehat{B}\right) \arrow[d, mapsto, "\substack{\text{substitute} \\ B = -\frac{1}{\alpha}F}"] \\
\frac{-1}{2\alpha}\mathrm{tr}\,(F \wedge F) \arrow[r, mapsto, "\text{break symmetry}"] & \frac{-1}{2\alpha}\mathrm{tr}\left(F \wedge \star\widehat{F}\right)
\end{tikzcd}$$

11.6 Generalized MacDowell–Mansouri theory

Just as we generalized 3d general relativity to geometric structures modeled on alternative 3d symmetric spaces, so we can generalize the MacDowell–Mansouri theory to other types of Cartan geometries. For simplicity, let us restrict attention to the original MacDowell–Mansouri action

$$S = -\frac{1}{2\alpha}\int \mathrm{tr}\,(\widehat{F} \wedge \widehat{F})$$

The question we wish to answer in this section is precisely what data are necessary to generalize this action to other Cartan geometries. In the previous sections, we did not derive the equations of motion directly from this action. We will do so using the more general setup we now describe.

Since the curvature is a 2-form, the action only makes sense for 4-dimensional Cartan geometry. So, suppose $P \to M$ is a principal H-bundle over a 4-manifold M, and G/H is a 4-dimensional Klein geometry. Let

$$A = \omega + \frac{1}{\ell} e$$

be a Cartan connection on P, where ω is the $\mathfrak{h}$-valued part, and e is the $\mathfrak{g}/\mathfrak{h}$-valued part. The curvature of A is then

$$F = R + \frac{1}{2\ell^2}[e, e] + \frac{1}{\ell}\left(de + [\omega, e]\right)$$

The similarity to general relativity is clearest when G/H is a symmetric space, so that $\mathfrak{h}$ and $\mathfrak{p} = \mathfrak{g}/\mathfrak{h}$ satisfy $[\mathfrak{h}, \mathfrak{p}] \subseteq \mathfrak{p}$, $[\mathfrak{p}, \mathfrak{p}] \subseteq \mathfrak{h}$. In this case, we have

$$\widehat{F} = R + \frac{1}{2\ell^2}[e, e]$$

as the $\mathfrak{h}$-valued part of F, and

$$T = \frac{1}{\ell}\left(de + [\omega, e]\right)$$

as the $\mathfrak{g}/\mathfrak{h}$-valued part.

For the action, we still use

$$S = -\frac{1}{2\alpha} \int \mathrm{tr}\,(\widehat{F} \wedge \widehat{F}).$$

where ‘tr’, as usual, denotes a nondegenerate invariant innner product on the Lie algebra. Note that this inner product only needs to be H-invariant, not fully G-invariant. Taking the variation of this action gives

$$\begin{aligned} \delta S &= -\frac{1}{\alpha} \int \mathrm{tr}\,(\delta\widehat{F} \wedge \widehat{F}) \\ &= -\frac{1}{\alpha} \int \mathrm{tr}\, \left((\delta R + \frac{1}{\ell^2}[\delta e, e]) \wedge \widehat{F} \right) \\ &= -\frac{1}{\alpha} \int \mathrm{tr}\, \left(d_\omega \delta\omega \wedge \widehat{F} + \frac{1}{\ell^2}[\delta e, e] \wedge \widehat{F} \right) \end{aligned}$$

The variation of ω, after an integration by parts, gives us

$$d_\omega \widehat{F} = 0$$

which, thanks to the Bianchi identity $d_\omega R = 0$, reduces to

$$\begin{aligned} & d_\omega [e, e] = 0 \\ \text{or} \quad & [d_\omega e, e] = 0 \end{aligned}$$

which is equivalent to

$$d_\omega e = 0$$

when $[\cdot, e]$ is injective.

The second equation of motion is a bit more subtle. It comes from the equation

$$\int \mathrm{tr}\, \left([\delta e, e] \wedge \widehat{F}\right) = 0 \qquad \forall\, \delta e.$$

Although we only needed H-invariance of our bilinear form $\mathrm{tr}\,(\cdot \wedge \cdot)$, it is tricky to derive an equation of motion from the vanishing of this variation unless we also have G-invariance. When this is the case, differentiating the invariance equation

$$\mathrm{tr}\,(\mathrm{Ad}(g)\delta e \wedge \mathrm{Ad}(g)\widehat{F}) = \mathrm{tr}\,(\delta e \wedge \widehat{F})$$

for a path with $g'(0) = e$ gives

$$\mathrm{tr}\,([e, \delta e] \wedge \widehat{F} + \delta e \wedge [e, \widehat{F}]) = 0$$

or

$$\mathrm{tr}\,([\delta e, e] \wedge \widehat{F}) = -\mathrm{tr}\,(\delta e \wedge [e, \widehat{F}]).$$

We thus conclude $[e, \widehat{F}] = 0$, or

$$[e, R] + \frac{1}{2\ell^2}[e, [e, e]] = 0$$

When $\mathfrak{g}$ is a matrix Lie algebra, so we can define the wedge product of $\mathfrak{g}$-valued forms using matrix multiplication, this becomes

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

where we define the 'cosmological constant' $\Lambda = 3/\ell^2$ by analogy with the physical case where $G = \mathrm{SO}(4,1)$.

Appendix A

Presentations of the loop braid group

Here we present a proof of Theorem 16 on p. 169, which we repeat here for the reader's convenience:

Theorem 16. *The loop braid group LB_n has a presentation with generators s_i and σ_i for $1 \leq i \leq n-1$ together with the following relations:*

(a) relations for the standard generators s_i of S_n:

$$s_i s_j = s_j s_i \qquad \text{for } |i-j| > 1 \tag{A.1}$$

$$s_i s_{i+1} s_i = s_{i+1} s_i s_{i+1} \qquad \text{for } 1 \leq i \leq n-2 \tag{A.2}$$

$$s_i^2 = 1 \qquad \text{for } 1 \leq i \leq n-1 \tag{A.3}$$

(b′) relations for the standard generators σ_i of B_n:

$$\sigma_i \sigma_j = \sigma_j \sigma_i \qquad \text{for } |i-j| > 1 \tag{A.4}$$

$$\sigma_i \sigma_{i+1} \sigma_i = \sigma_{i+1} \sigma_i \sigma_{i+1} \qquad \text{for } 1 \leq i \leq n-2 \tag{A.5}$$

(c′) the following **mixed relations***:*

$$s_i \sigma_j = \sigma_j s_i \qquad \text{for } |i-j| > 1 \tag{A.6}$$

$$s_i s_{i+1} \sigma_i = \sigma_{i+1} s_i s_{i+1} \qquad \text{for } 1 \leq i \leq n-2 \tag{A.7}$$

$$\sigma_i \sigma_{i+1} s_i = s_{i+1} \sigma_i \sigma_{i+1} \qquad \text{for } 1 \leq i \leq n-2 \tag{A.8}$$

Proof: Recall that we proved in Theorem 11 that the loop braid group LB_n has a presentation with generators s_i for $1 \leq i \leq n-1$ and σ_{ij} for $1 \leq i, j \leq n$ with $i \neq j$, together with the following relations:

(a) the relations for the standard generators s_i of S_n:

$$s_i s_j = s_j s_i \qquad \text{for } |i-j| > 1 \tag{A.9}$$

$$s_i s_{i+1} s_i = s_{i+1} s_i s_{i+1} \qquad \text{for } 1 \leq i \leq n-2 \tag{A.10}$$

$$s_i^2 = 1 \qquad \text{for } 1 \leq i \leq n-1 \tag{A.11}$$

(b) Lin's relations for the generators σ_{ij} of PLB_n:

$$\sigma_{ij}\sigma_{k\ell} = \sigma_{k\ell}\sigma_{ij} \qquad \text{for } i, j, k, \ell \text{ distinct} \tag{A.12}$$

$$\sigma_{ik}\sigma_{jk} = \sigma_{jk}\sigma_{ik} \qquad \text{for } i, j, k \text{ distinct} \tag{A.13}$$

$$\sigma_{ij}\sigma_{kj}\sigma_{ik} = \sigma_{ik}\sigma_{kj}\sigma_{ij} \qquad \text{for } i, j, k \text{ distinct} \tag{A.14}$$

(c) relations expressing the action of S_n on PLB_n:

$$s_i \sigma_{i(i+1)} = \sigma_{(i+1)i} s_i \qquad \text{for } 1 \leq i \leq n-1 \tag{A.15}$$

$$s_k \sigma_{ij} = \sigma_{ij} s_k \qquad \text{for } i, j, k, k+1 \text{ distinct} \tag{A.16}$$

$$s_j \sigma_{ij} = \sigma_{i(j+1)} s_j \qquad \text{for } i, j, j+1 \text{ distinct} \tag{A.17}$$

$$s_i \sigma_{ij} = \sigma_{(i+1)j} s_i \qquad \text{for } i, i+1, j \text{ distinct} \tag{A.18}$$

We begin by demonstrating that the relations in the statement of Theorem 16 follow from those given in Theorem 11. It clearly suffices to show that the relations in (b') and (c') follow from the relations in $(a), (b)$ and (c).

In what follows, we make frequent use of the correspondence between generators σ_{ij} of PLB_n and generators σ_i of LB_n as given in (7.13) and (7.14). In fact, since these follow from different relations in the presentation of Theorem 11, it suffices for our purposes to take one expression from each of these, say

$$\sigma_{ij} = \begin{cases} s_i s_{i+1} \cdots s_{j-1} \sigma_{j-1} s_{j-2} \cdots s_{i+1} s_i & \text{for } \ i < j \\ s_j s_{j+1} \cdots s_{i-2} \sigma_{i-1} s_{i-1} \cdots s_{j+1} s_j & \text{for } \ i > j \end{cases} \tag{A.19}$$

These representations of σ_{ij} follow directly from the definition of σ_i along with the relations (A.15), (A.17), and (A.18).

• Relation (A.6): We wish to show that $s_j\sigma_i = \sigma_i s_j$ for $|i-j|>1$. To check this, we begin with relation (A.16) in the form:

$$s_j\sigma_{i(i+1)} = \sigma_{i(i+1)}s_j,$$

where $|i-j|>1$. Using (A.19) above, this becomes:

$$s_js_i\sigma_i = s_i\sigma_is_j.$$

Applying relation (A.9) of to the left-hand side and then cancelling s_i from each side gives $s_j\sigma_i = \sigma_i s_j$ when $|i-j|>1$, which is (A.6).

• Relation (A.7): We wish to show that $s_is_{i+1}\sigma_i = \sigma_{i+1}s_is_{i+1}$ for $1 \le i \le n-2$. Beginning with relation (A.17) with $j = i+1$, we obtain:

$$s_{i+1}\sigma_{i(i+1)} = \sigma_{i(i+2)}s_{i+1}.$$

By (A.19) this gives:

$$s_{i+1}s_i\sigma_i = s_is_{i+1}\sigma_{i+1}s_is_{i+1}.$$

Multiplying on the right by $s_{i+1}s_i$ and on the left by s_is_{i+1}, we have:

$$\begin{aligned}\sigma_is_{i+1}s_i &= s_is_{i+1}s_is_{i+1}\sigma_{i+1} \\ &= s_is_is_{i+1}s_i\sigma_{i+1} && \text{by (A.10)} \\ &= s_{i+1}s_i\sigma_{i+1} && \text{by (A.11)}\end{aligned}$$

This can be rewritten as $s_is_{i+1}\sigma_i = \sigma_{i+1}s_is_{i+1}$, which is (A.7).

• Relation (A.8): We wish to show that $\sigma_i\sigma_{i+1}s_i = s_{i+1}\sigma_i\sigma_{i+1}$ for $1 \le i \le n-2$. To verify this we use relation (A.13) with $i, i+1$ and $i+2$, which gives:

$$\sigma_{i(i+2)}\sigma_{(i+1)(i+2)} = \sigma_{(i+1)(i+2)}\sigma_{i(i+2)}.$$

By (A.19) this becomes:

$$(s_is_{i+1}\sigma_{i+1}s_i)(s_{i+1}\sigma_{i+1}) = (s_{i+1}\sigma_{i+1})(s_is_{i+1}\sigma_{i+1}s_i).$$

Applying relation (A.7) on the left hand side gives:

$$s_is_{i+1}s_is_{i+1}\sigma_i\sigma_{i+1} = (s_{i+1}\sigma_{i+1})(s_is_{i+1}\sigma_{i+1}s_i).$$

Multiplying by $s_i s_{i+1} s_i$ on the left produces:

$$\begin{aligned} s_{i+1}\sigma_i\sigma_{i+1} &= s_i s_{i+1} s_i s_{i+1}\sigma_{i+1} s_i s_{i+1}\sigma_{i+1} s_i \\ &= s_{i+1} s_i \sigma_{i+1} s_i s_{i+1}\sigma_{i+1} s_i && \text{by (A.2)} \\ &= \sigma_i\sigma_{i+1} s_i && \text{by (A.7)} \end{aligned}$$

which is (A.8).

• Relation (A.4): We wish to show that $\sigma_i\sigma_j = \sigma_j\sigma_i$ for $|i-j| > 1$. To do so, we use relation (A.12) with $i, i+1, j, j+1$, which are clearly all distinct for $|i-j| > 1$. We therefore have:

$$\sigma_{i(i+1)}\sigma_{j(j+1)} = \sigma_{j(j+1)}\sigma_{i(i+1)},$$

which, by (A.19), becomes:

$$s_i\sigma_i s_j\sigma_j = s_j\sigma_j s_i\sigma_i.$$

Applying (A.6) to both sides of this equation, followed by relation (A.9), we obtain:

$$\sigma_i\sigma_j = \sigma_j\sigma_i$$

with $|i-j| > 1$, which is (A.4).

• Relation (A.5): We wish to show that $\sigma_i\sigma_{i+1}\sigma_i = \sigma_{i+1}\sigma_i\sigma_{i+1}$ for $1 \leq i \leq n-2$. To check this we start with relation (A.14) with $i, i+1$, and $i+2$, which are clearly all distinct. Thus, we have:

$$\sigma_{i(i+1)}\sigma_{(i+2)(i+1)}\sigma_{i(i+2)} = \sigma_{i(i+2)}\sigma_{(i+2)(i+1)}\sigma_{i(i+1)}.$$

Using the correspondence given in (A.19) and cancelling s_i from both sides, we obtain:

$$\begin{aligned} s_{i+1}\sigma_{i+1} s_i\sigma_{i+1} s_{i+1} s_i\sigma_i &= \sigma_i\sigma_{i+1} s_{i+1} s_i s_{i+1}\sigma_{i+1} s_i \\ &= \sigma_i\sigma_{i+1} s_i s_{i+1} s_i \sigma_{i+1} s_i && \text{by (A.10)} \\ &= \sigma_i\sigma_{i+1} s_i \sigma_i s_{i+1} && \text{by (A.7), (A.11)} \\ &= s_{i+1}\sigma_i\sigma_{i+1} s_{i+1} && \text{by (A.8).} \end{aligned}$$

Cancelling s_{i+1} on the left and multiplying by s_{i+1} on the right produces:

$$\begin{aligned} \sigma_i\sigma_{i+1}\sigma_i &= \sigma_{i+1} s_i\sigma_{i+1} s_{i+1} s_i\sigma_i s_{i+1} \\ &= \sigma_{i+1}\sigma_i\sigma_{i+1} \end{aligned}$$

where in the last step we used (A.13) in the form $s_i\sigma_i\sigma_{i+1}s_{i+1} = \sigma_{i+1}s_{i+1}s_i\sigma_i$. This is (A.5).

The loop braid group thus has generators that satisfy all of the relations of the braid permutation group. It remains to show that these relations are sufficient, which we do by demonstrating that the relations in the statement of Theorem 11 follow from those given in Theorem 16. In this direction of the proof it is convenient to use both of the equivalent expressions (7.13) and (7.14) as the correspondence between generators σ_i and σ_{ij}.

• Relation (A.15): This relation simply says $s_i\sigma_{i(i+1)} = \sigma_{(i+1)i}s_i$, which is immediate from (A.19) since both sides are equal to σ_i.

• Relation (A.16): We wish to show $s_k\sigma_{ij} = \sigma_{ij}s_k$, whenever $i, j, k, k+1$ are distinct. When either $k+1 < i < j$ or $i < j < k$, s_k commutes with each of the factors in the expansion

$$\sigma_{ij} = s_i s_{i+1} \cdots s_{j-1}\sigma_{j-1}s_{j-2} \cdots s_{i+1}s_i$$

by (A.1) and (A.6). Similarly, when $k+1 < j < i$ or $j < i < k$, s_k commutes with each factor in

$$\sigma_{ij} = s_j s_{j+1} \cdots s_{i-2}\sigma_{i-1}s_{i-1} \cdots s_{j+1}s_j.$$

When $i < k < k+1 < j$ we also need two applications of (A.2):

$$\begin{aligned}
s_k\sigma_{ij} &= s_k s_i \cdots s_{j-1}\sigma_{j-1}s_{j-2} \cdots s_i \\
&= s_i \cdots s_{k-2}s_k s_{k-1}s_k s_{k+1} \cdots s_{j-1}\sigma_{j-1}s_{j-2} \cdots s_i && \text{by (A.1)} \\
&= s_i \cdots s_{k-2}s_{k-1}s_k s_{k-1}s_{k+1} \cdots s_{j-1}\sigma_{j-1}s_{j-2} \cdots s_i && \text{by (A.2)} \\
&= s_i \cdots s_{k-2}s_{k-1}s_k s_{k+1} \cdots s_{k-1}s_{j-1}\sigma_{j-1}s_{j-2} \cdots s_i && \text{by (A.1)} \\
&= s_i \cdots s_{j-1}\sigma_{j-1}s_{j-2} \cdots s_{k+1}s_{k-1}s_k s_{k-1}s_{k-2} \cdots s_i && \text{by (A.1), (A.6)} \\
&= s_i \cdots s_{j-1}\sigma_{j-1}s_{j-2} \cdots s_{k+1}s_k s_{k-1}s_k s_{k-2} \cdots s_i && \text{by (A.2)} \\
&= \sigma_{ij}s_k && \text{by (A.1)}
\end{aligned}$$

The only remaining case is $j < k < k+1 < i$, which is handled similarly.

• Relation (A.17): We wish to show that $s_j\sigma_{ij} = \sigma_{i(j+1)}s_j$ whenever $i \neq j+1$. When $i < j$ we have:

$$\begin{aligned}
s_j\sigma_{ij} &= s_is_{i+1}\cdots s_js_{j-1}\sigma_{j-1}s_{j-2}\cdots s_{i+1}s_i && \text{by (A.1)}\\
&= s_is_{i+1}\cdots s_{j-1}s_js_{j-1}s_j\sigma_{j-1}s_{j-2}\cdots s_{i+1}s_i && \text{by (A.2)}\\
&= s_is_{i+1}\cdots s_{j-1}s_j\sigma_js_{j-1}s_js_{j-2}\cdots s_{i+1}s_i && \text{by (A.6)}\\
&= \sigma_{i(j+1)}s_j && \text{by (A.1), (A.19)}
\end{aligned}$$

and the case $i > j+1$ is similar.

• Relation (A.18): The proof that $s_i\sigma_{ij} = \sigma_{(i+1)j}s_i$ is essentially the same as the proof of (A.17) above.

• Relation (A.12): We wish to show $\sigma_{ij}\sigma_{k\ell} = \sigma_{k\ell}\sigma_{ij}$, whenever i, j, k, and ℓ are distinct. Naively there are 4! orderings of i, j, k, ℓ to consider, but symmetry of the relation implies only 8 are independent. All cases are proved similarly; we demonstrate only the case $i < j < k < \ell$:

$$\begin{aligned}
\sigma_{ij}\sigma_{k\ell} &= (s_i\cdots s_{j-1}\sigma_{j-1}s_{j-2}\cdots s_i)(s_k\cdots s_{\ell-1}\sigma_{\ell-1}s_{\ell-2}\cdots s_k)\\
&= s_k\cdots s_{\ell-1}(s_i\cdots s_{j-1}\sigma_{j-1}s_{j-2}\cdots s_i)(\sigma_{\ell-1}s_{\ell-2}\cdots s_k) && \text{by (A.1), (A.6)}\\
&= s_k\cdots s_{\ell-1}\sigma_{\ell-1}(s_i\cdots s_{j-1}\sigma_{j-1}s_{j-2}\cdots s_i)(s_{\ell-2}\cdots s_k) && \text{by (A.6), (A.4)}\\
&= (s_k\cdots s_{\ell-1}\sigma_{\ell-1}s_{\ell-2}\cdots s_k)(s_i\cdots s_{j-1}\sigma_{j-1}s_{j-2}\cdots s_i) && \text{by (A.1), (A.6)}\\
&= \sigma_{k\ell}\sigma_{ij}.
\end{aligned}$$

• Relation (A.13): We wish to show that $\sigma_{ik}\sigma_{jk} = \sigma_{jk}\sigma_{ik}$ when i, j, k are distinct. We have three independent cases: $i < j < k$, $i < k < j$, and $k < i < j$. In the case $i < j < k$, we first note that if $j \neq i+1$, then by (A.16) and (A.17) we have:

$$\begin{aligned}
\sigma_{ik}\sigma_{jk} &= s_{j-1}(\sigma_{ik}\sigma_{(j-1)k})s_{j-1}\\
\text{and}\quad \sigma_{jk}\sigma_{ik} &= s_{j-1}(\sigma_{(j-1)k}\sigma_{ik})s_{j-1}.
\end{aligned}$$

By repeated application of these facts, it suffices to consider the subcase where $j = i+1$. Similarly, if $k \neq j+1$, we can use (A.16) and (A.18) to reduce to the case where $k = j+1$.

Thus it suffices to consider only the cases where i, j, k are consecutive:

$$
\begin{aligned}
\sigma_{i(i+2)}\sigma_{(i+1)(i+2)} &= (s_i s_{i+1}\sigma_{i+1}s_i)(s_{i+1}\sigma_{i+1}) \\
&= s_i s_{i+1} s_i s_{i+1}\sigma_i\sigma_{i+1} && \text{by (A.7)} \\
&= s_{i+1} s_i \sigma_i \sigma_{i+1} && \text{by (A.2)} \\
&= s_{i+1} s_i s_{i+1}\sigma_i\sigma_{i+1}s_i && \text{by (A.8)} \\
&= s_{i+1}\sigma_{i+1}s_i s_{i+1}\sigma_{i+1}s_i && \text{by (A.7)} \\
&= \sigma_{(i+1)(i+2)}\sigma_{i(i+2)}.
\end{aligned}
$$

This proves the case $i < j < k$. The remaining two cases are similar.

- Relation (A.14): We wish to show that $\sigma_{ij}\sigma_{kj}\sigma_{ik} = \sigma_{ik}\sigma_{kj}\sigma_{ij}$ when i, j, k are distinct. In light of (A.13) this equation is symmetric under the interchange of i and k, and this symmetry reduces the number of independent cases to 3: $i < j < k$, $i < k < j$, and $j < i < k$. In the case $i < j < k$, we first note that if $j \neq i + 1$, then by (A.16) and (A.17) we have

$$
\begin{aligned}
\sigma_{ij}\sigma_{kj}\sigma_{ik} &= s_{j-1}(\sigma_{i(j-1)}\sigma_{k(j-1)}\sigma_{ik})s_{j-1} \\
\text{and}\quad \sigma_{ik}\sigma_{kj}\sigma_{ij} &= s_{j-1}(\sigma_{ik}\sigma_{k(j-1)}\sigma_{i(j-1)})s_{j-1}
\end{aligned}
$$

By repeated application of these facts, it suffices to consider the subcase where $j = i + 1$. Similarly, if $k \neq j + 1$, we can use (A.16) and (A.18) to reduce to the case where $k = j + 1$. Thus it suffices to consider only the cases where i, j, k are consecutive:

$$
\begin{aligned}
\sigma_{i(i+1)}\sigma_{(i+2)(i+1)}\sigma_{i(i+2)} &= (s_i\sigma_i)(\sigma_{i+1}s_{i+1})(s_i s_{i+1}\sigma_{i+1}s_i) \\
&= s_i\sigma_i\sigma_{i+1}s_i s_{i+1}s_i\sigma_{i+1}s_i && \text{by (A.2)} \\
&= s_i\sigma_i\sigma_{i+1}s_i\sigma_i s_{i+1} && \text{by (A.7)} \\
&= s_i s_{i+1}\sigma_i\sigma_{i+1}\sigma_i s_{i+1} && \text{by (A.8)} \\
&= s_i s_{i+1}\sigma_{i+1}\sigma_i\sigma_{i+1}s_{i+1} && \text{by (A.5)} \\
&= \sigma_{i(i+2)}\sigma_{i(i+1)}\sigma_{(i+2)(i+1)} \\
&= \sigma_{i(i+2)}\sigma_{(i+2)(i+1)}\sigma_{i(i+1)} && \text{by (A.13)}
\end{aligned}
$$

This proves the case of $i < j < k$. The other two independent cases are similar.

Thus, the relations of Theorem 16 imply those of Theorem 11. □

As pointed out by Blake Winter, one can also prove Theorem 16 as follows. Fenn, Rimányi, and Rourke [34] show that the braid permutation group BP_n is isomorphic to the subgroup of $\mathrm{Aut}(F_n)$ generated by all permutations of basis elements, together with all operations of conjugating one basis element by another. Let X be $\mathbb{R}^3$ with unlinked unknotted circles $\ell_1, \ldots, \ell_n$ removed. As we have seen, $\pi_1(X) = F_n$, the free group on n generators, so by the work of Dahm, the loop braid group acts as automorphisms of F_n. Let $D\colon LB_n \to \mathrm{Aut}(F_n)$ be the resulting homomorphism. Goldsmith [51] shows that the image of D is precisely the above subgroup of $\mathrm{Aut}(F_n)$ and that, moreover, D is one-to-one. It follows that LB_n and BP_n are isomorphic. Since Fenn, Rimányi and Rourke prove that BP_n has the presentation given in Theorem 16, it follows that LB_n also has this presentation.

Appendix B

The Lie algebras $\mathfrak{so}(p,q)$ and $\mathfrak{iso}(p,q)$

We use the convention that $O(p,q)$ denotes the group of transformations of $\mathbb{R}^{p+q}$ preserving the generalized Minkowski metric of signature $p-q$:

$$\eta = \begin{bmatrix} -I_q & 0 \\ 0 & I_p \end{bmatrix}$$

Equivalently, as matrices, $g \in O(p,q)$ satisfies $g^\dagger = g^{-1}$ where the adjoint is the transpose conjugated by the metric: $g^\dagger = \eta^{-1} g^T \eta$. Write an element $X \in \mathfrak{so}(p,q)$ as $X = \gamma'(0)$ where $\gamma\colon (-\epsilon,\epsilon) \to \mathrm{SO}(p,q)$ is a path in the group with $\gamma(0) = 1$. In matrix form, the metric may be written as

$$\langle v, w \rangle = v^T \eta w$$

where the superscript T denotes the transpose. Differentiating the equation

$$\langle \gamma(t)v, \gamma(t)w \rangle = \langle v, w \rangle$$

at $t = 0$ gives

$$v^T (X^T \eta + \eta X) w = 0.$$

Since this must be true for any $v, w \in \mathbb{R}^{p+q}$, we conclude

$$X^T = -\eta X \eta^{-1}$$

(Note this actually makes sense for any $\mathrm{SO}(V)$ with symmetric matrix η as the matrix for a metric in a given basis.) A matrix in $\mathfrak{so}(p,q)$ may therefore be written

$$\begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix} \quad \text{with } {A_1}^T = -A_1,\ {A_2}^T = A_3,\ {A_4}^T = -A_4$$

Appendix C

Hodge duality in Lie algebras

C.1 Hodge duality for inner product spaces.

We first recall some basic facts about the Hodge dual in ΛV, for an n-dimensional vector space V with inner product η of signature $(n-s,s)$ [9, p. 91]. If $\{\xi_i \mid i=1,\ldots,n\}$ is an ordered basis for V (defining an orientation), and

$$\omega = \frac{1}{p!}\omega_{i_1\cdots i_p}\xi^{i_1}\wedge\cdots\wedge\xi^{i_p}$$

is an element of $\Lambda^p V$, then the Hodge dual of ω is the element of $\Lambda^{n-p}V$ given by

$$\star\omega = \frac{1}{(n-p)!}\star\omega_{j_1\cdots j_{n-p}}\xi^{j_1}\wedge\cdots\wedge\xi^{j_{n-p}}$$

with the components given by

$$\star\omega_{j_1\cdots j_{n-p}} = \frac{1}{p!}\varepsilon^{i_1\cdots i_p}{}_{j_1\cdots j_{n-p}}\omega_{i_1\cdots i_p}$$

The Hodge star operator satisfies

$$\star^2 = (-1)^{p(n-p)+s}.$$

C.2 $\mathfrak{so}(4)$, $\mathfrak{so}(3,1)$, and $\mathfrak{so}(2,2)$

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:

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

Also note that $\star$ gives an isomorphism of Lie algebras

$$\begin{aligned} \star\colon \mathfrak{so}(3,1) &\to \star\mathfrak{so}(3,1) \\ X &\mapsto \star X \end{aligned}$$

where we define $\star\mathfrak{so}(3,1)$ to be the Lie algebra obtained from $\mathfrak{so}(3,1)$ by changing the bracket $[\cdot,\cdot]$ to $[\cdot,\cdot]_\star := -\star[\cdot,\cdot]$. It is easy to check that the Jacobi identity holds for $[\cdot,\cdot]_\star$, and

$$[\star X, \star X']_\star = -\star[\star X, \star X'] = -\star(\star)^2[X, X'] = \star[X, X']$$

shows $\star$ is a Lie algebra homomorphism, hence an isomorphism since it is bijective.

Appendix D

Identities for forms and tensors

D.1 Permutation symbols

The Levi–Civita tensor $\varepsilon_{i_1\cdots i_n}$ corresponding to a given metric of signature $(s, n-s)$ satisfies the contraction relations [97, p. 433]:

$$\varepsilon_{i_1\cdots i_p i_{p+1}\cdots i_n}\varepsilon^{i_1\cdots i_p j_{p+1}\cdots j_n} = (-1)^s p!(n-p)!\delta^{i_{p+1}}_{[j_{p+1}}\cdots\delta^{i_n}_{j_n]}$$

for $p = 0, 1, \ldots n$. It is worth writing out these relations explicitly in the most physical example: 4d Lorentzian spacetime of signature $(-+++)$:

$$\begin{aligned}
\varepsilon_{ijkl}\varepsilon^{mnpq} &= -0!4!\,\delta^m_{[i}\delta^n_j\delta^p_k\delta^q_{\ell]}\\
\varepsilon_{ijk\ell}\varepsilon^{imnp} &= -1!3!\,\delta^m_{[j}\delta^n_k\delta^p_{\ell]}\\
\varepsilon_{ijk\ell}\varepsilon^{ijmn} &= -2!2!\,\delta^m_{[k}\delta^n_{\ell]}\\
\varepsilon_{ijk\ell}\varepsilon^{ijkm} &= -3!1!\,\delta^m_\ell\\
\varepsilon_{ijk\ell}\varepsilon^{ijk\ell} &= -4!0!
\end{aligned}$$

Here we use the convention for antisymmetrization over indices which includes the factor $(1/p!)$:

$$X_{[i_1 i_2\cdots i_p]} := \frac{1}{p!}\sum_{\sigma\in S_p}\mathrm{sgn}(\sigma)X_{i_{\sigma(1)}i_{\sigma(2)}\cdots i_{\sigma(p)}}$$

D.2 Lie algebra-valued differential forms

(See AMP for most of this stuff) A differential form on M with values in the Lie algebra $\mathfrak{g}$ is an element of

$$\Omega^p(M,\mathfrak{g}) := \Omega^p(M)\otimes\mathfrak{g}.$$

The differential

$$d : \Omega^p(M, \mathfrak{g}) \to \Omega^{p+1}(M, \mathfrak{g})$$

just acts on the form parts via the ordinary differential, letting the Lie algebra parts go along along for the ride:

$$d(\omega) = d(\omega^\alpha \otimes v_\alpha) := d\omega^\alpha \otimes v_\alpha$$

where $\{v_\alpha\}$ is a basis of $\mathfrak{g}$.

We define the bracket of $\mathfrak{g}$-valued forms by using the wedge product on form parts and the Lie bracket on Lie algebra parts:

$$[\omega, \mu] := (\omega^\alpha \wedge \mu^\beta) \otimes [v_\alpha, v_\beta].$$

If ω is a p-form, μ a q-form, then switching ω and μ produces a factor $(-1)^{pq}$ from the graded commutativity of the wedge product, and an additional (-1) from the anticommutativity of the Lie bracket. Hence

$$[\omega, \mu] = (-1)^{pq+1}[\mu, \omega] \tag{D.1}$$

It is easy to see that the differential d on $\mathfrak{g}$-valued forms is a graded derivation with respect to the bracket:

$$\begin{aligned} d[\omega, \mu] &= d(\omega^\alpha \wedge \mu^\beta) \otimes [v_\alpha, v_\beta] \\ &= (d\omega^\alpha \wedge \mu + (-1)^p \omega^\alpha \wedge d\mu^\beta \\ &= [d\omega, \mu] + (-1)^p [\omega, d\mu] \end{aligned}$$

In local coordinates, a connection on a principal bundle is itself a Lie algebra-valued 1-form, and the exterior covariant derivative is given by:

$$d_A\omega = d\omega + [A, \omega]$$

The facts that the differential d of is a graded derivation on $\Omega(M)$, and the Lie bracket is a derivation on $\mathfrak{g}$ (the Jacobi identity), imply that d_A is a graded derivation on $\Omega(M, \mathfrak{g})$:

$$\begin{aligned}
d_A[\omega, \mu] &= d[\omega, \mu] + [A, [\omega, \mu]] \\
&= (d\omega^\alpha \wedge \mu^\beta + (-1)^p \omega^\alpha \wedge d\mu^\beta) \otimes [v_\alpha, v_\beta] + A^\gamma \wedge \omega^\alpha \wedge \mu^\beta \otimes [v_\gamma, [v_\alpha, v_\beta]] \\
&= (d\omega^\alpha \wedge \mu^\beta + (-1)^p \omega^\alpha \wedge d\mu^\beta) \otimes [v_\alpha, v_\beta] + A^\gamma \wedge \omega^\alpha \wedge \mu^\beta \otimes ([[v_\gamma, v_\alpha], v_\beta] + [v_\alpha, [v_\gamma, v_\beta]]) \\
&= (d\omega^\alpha \wedge \mu^\beta + (-1)^p \omega^\alpha \wedge d\mu^\beta) \otimes [v_\alpha, v_\beta] \\
&\quad\quad + A^\gamma \wedge \omega^\alpha \wedge \mu^\beta \otimes [[v_\gamma, v_\alpha], v_\beta] + (-1)^p \omega^\alpha \wedge A^\gamma \wedge \mu^\beta \otimes [v_\alpha, [v_\gamma, v_\beta]] \\
&= [d\omega, \mu] + [[A, \omega], \mu] + (-1)^p ([\omega, d\mu] + [\omega, [A, \mu]] \\
&= [d_A \omega, \mu] + (-1)^p [\omega, d_A \mu]
\end{aligned}$$

When $\mathfrak{g}$ is a *matrix* Lie algebra, we can use the matrix product $v_\alpha v_\beta$ to define a wedge product of $\mathfrak{g}$-valued differential forms. If ω is a p-form, μ a q-form, we let

$$\omega \wedge \mu := (\omega^\alpha \wedge \mu^\beta) \otimes v_\alpha v_\beta$$

In this case, we can write

$$\begin{aligned}
[\omega, \mu] &= (\omega^\alpha \wedge \mu^\beta) \otimes (v_\alpha v_\beta - v_\beta v_\alpha) \\
&= (\omega^\alpha \wedge \mu^\beta - (-1)^{pq} \mu^\alpha \wedge \omega^\beta) \otimes v_\alpha v_\beta \\
&= \omega \wedge \mu - (-1)^{pq} \mu \wedge \omega
\end{aligned}$$

and view the bracket as the graded commutator for the wedge product of $\mathfrak{g}$-valued forms. In this special case where we can multiply the elements of $\mathfrak{g}$ we also have d_A acting as a graded derivation over the wedge product:

$$d_A(\omega \wedge \mu) = d_A \omega \wedge \mu + (-1)^p \omega \wedge d_A \mu$$

Bibliography

[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] C. Aneziris, A. P. Balachandran, L. Kauffman and A. M. Srivastava, Novel statistics for strings and string 'Chern–Simons' terms, *Int. J. Mod. Phys.* **A6** (1990), 2519–2558.

[4] V. I. Arnold, *Mathematical Methods of Classical Mechanics*, New York, Springer–Verlag, 1989.

[5] A. Ashtekar, *Lectures on Nonperturbative Canonical Gravity*, Singapore; River Edge, N.J., World Scientific, 1991.

[6] J. C. Baez, Spin foam models, *Class. Quantum Grav.* **15** (1998), 1827–1858. arXiv:gr-qc/9709052.

[7] 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`.

[8] J. C. Baez, Higher gauge theory, presentation at Louisiana State University, Nov. 16, 2006, http://math.ucr.edu/home/baez/highergauge/

[9] J. C. Baez and J. P. Muniain, *Gauge Fields, Knots and Gravity*, Series on Knots and Everything – Vol. 4, World Scientific, Singapore, 1994.

[10] J. Baez, A. Baratin, L. Freidel, J. Morton, D. Wise, On the representation theory of the Poincaré 2-group, to appear.

[11] J. C. Baez and A. S. Crans, Higher-dimensional algebra VI: Lie 2-algebras, *Theory and Applications of Categories* **12** (2004), 492–528.

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

[13] J. Baez and J. Dolan, Categorification, in *Higher Category Theory*, eds. Ezra Getzler and Mikhail Kapranov, Contemp. Math. **230**, American Mathematical Society, Providence, Rhode Island, 1998, 1–36.

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

[15] J. Baez and U. Schreiber, Higher gauge theory, to appear in *Categories in Algebra, Geometry and Physics*, eds. Alexei Davydov and Michael Batanin, Contemp. Math., AMS, Providence, Rhode Island.

[16] J. Baez and U. Schreiber, Higher gauge theory II: 2-connections, draft. Available as http://math.ucr.edu/home/baez/2conn.pdf.

[17] J. C. Baez, M. Ody and W. Richter, Topological aspects of spin and statistics, *J. Math. Phys.* **36** (1995), 247–257. Also available as cond-mat/9208005.

[18] A. P. Balachandran, Statistics, strings and gravity, in *Mathematical Physics: Proceedings, 3rd Regional Conference, Islamabad, Pakistan, February 17–24, 1989*, eds. F. Hussain and A. Qadir, World Scientific, Singapore, 1990, pp. 31-45.

[19] T. Bartels, Higher gauge theory: 2-bundles. Available as `math.CT/0410328`

[20] P. Bonderson, A. Kitaev, and K. Shtengel, Detecting non-abelian statistics in the $\nu = 5/2$ fractional quantum Hall state, *Phys. Rev. Lett. 96* (2006) 016803. Also available as cond-mat/0508616.

[21] E. Brieskorn, Automorphic sets and singularities, in *Braids*, *Contemp. Math.* **78**, Amer. Math. Soc., Providence, RI, 1998, pp. 45–115.

[22] N. Burgoyne and R. Cushman, Conjugacy classes in linear groups, *Jour. Alg.* **44** 339-362 (1977).

[23] F. E. Camino, W. Zhou and V. J. Goldman, Realization of a Laughlin quasiparticle interferometer: observation of fractional statistics, **Phys. Rev. B 72** (2005), 075342. Also available as cond-mat/0502406.

[24] Steven Carlip, *Quantum Gravity in 2+1 dimensions*, Cambridge; New York, Cambridge University Press, 1998.

[25] 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`.

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

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

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

[29] J. S. Carter and M. Saito, *Knotted Surfaces and their Diagrams*, Amer. Math. Soc., Providence, RI, 1998.

J. S. Carter, S. Kamada, and M. Saito, *Surfaces in 4-Space*, Springer, Berlin, 2004.

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

[31] D. M. Dahm, *A Generalisation of Braid Theory*, Ph.D. Thesis, Princeton University, 1962.

[32] S. Deser, R. Jackiw, and G. 't Hooft, Three-dimensional Einstein gravity: dynamics of flat space, *Ann. Phys.* **152**, 220-235.

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

[34] R. Fenn, R. Rimányi and C. Rourke, The braid-permutation group, *Topology* **36** (1997), 123–135.

[35] R. Fenn, C. Rourke, Racks and links in codimension two, *Journal of Knot Theory and its Ramifications* **1** (1992), 343–406.

[36] M. Forrester-Barker, Group objects and internal categories, available as `math.CT/0212065`

[37] M. H. Freedman, A. Kitaev, M. J. Larsen and Z. Wang, Topological quantum computation. Also available as quant-ph/0101025.

[38] L. Freidel and D. Louapre, Ponzano-Regge model revisited I: gauge fixing, observables and interacting spinning particles, *Class. Quant. Grav.* **21** (2004), 5685-5726. Also available as hep-th/0401076.

L. Freidel and D. Louapre, Ponzano-Regge model revisited II: equivalence with Chern-Simons, available as gr-qc/0410141

L. Freidel and E. R. Livine, Ponzano-Regge model revisited III: Feynman diagrams and effective field theory, available as hep-th/0502106.

[39] L. Freidel, J. Kowalski-Glikman and L. Smolin, 2+1 gravity and doubly special relativity, *Phys. Rev.* **D69** (2004), 044001. Also available as hep-th/0307085.

[40] 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`.

[41] L. Freidel and K. Krasnov, Spin foam models and the classical action principle, *Adv. Theor. Math. Phys.* **2** (1998), 1221–1285

[42] L. Freidel and K. Krasnov, R. Puzio, BF description of higher–dimensional gravity theories, `hep-th/9901069`

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

[44] D. I. Fuks–Rabinovich, On the automorphism groups of free products, I (Russian), *Mat. Sbornik* **8** (1940), 265–276.

D. I. Fuks–Rabinovich, On the automorphism groups of free products, II (Russian), *Mat. Sbornik* **9 9** (1941), 183–220.

[45] 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`

[46] 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`

[47] K. Gödel, An example of a new type of cosmological solutions of Einstein's field equations of gravitation, *Rev. Mod. Phys.* **21** No. 3 (1949) 447–450.

[48] W. M. Goldman, *Discontinuous Groups and the Euler Class*, Ph.D. thesis, University of California, Berkeley, 1980.

[49] W. M. Goldman, Mapping class group dynamics on surface group representations, `math.GT/0509114`

[50] W. M. Goldman, Topological components of spaces of representations

[51] D. L. Goldsmith, The theory of motion groups. *Michigan Math. J.* **28** (1981), 3–17.

[52] D. L. Goldsmith, Motion of links in the 3-sphere, *Math. Scand.* **50** (1982), 167–205.

[53] S. W. Hawking and G. F. R. Ellis, *The large scale structure of space–time*, Cambridge University Press, Cambridge, 1973.

[54] R. Hermann, *Lie groups for physicists*, New York, W. A. Benjamin, 1966.

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

[56] D. Joyce, A classifying invariant of knots, the knot quandle, *J. Pure Appl. Alg.* **23** (1982), 37–65.

[57] L. H. Kauffman, *Knots and Physics*, World Scientific, Singapore, 1991.

[58] A. Kitaev, Fault-tolerant quantum computation by anyons, *Ann. Phys.* **303** (2003), 2–30. Also available as quant-ph/9707021.

[59] S. Kobayashi, Chap. 1 §2: Flat bundles and flat connections, *Differential Geometry of Complex Vector Bundles*, Princeton U. Press, Princeton, New Jersey, 1987, pp. 4–7.

[60] Kobayashi and Nomizu, *Foundations of Differential Geometry* vol. II, New York, Wiley, 1996.

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

[62] K. Krasnov, Quantum gravity with matter via group field theory, available as hep-th/0505174.

[63] X.-S. Lin, The motion group of the unlink and its representations, preprint, 2005.

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

[65] 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.

[66] S. Mac Lane, *Categories for the Working Mathematician*, 2 Ed., Springer, 1998.

[67] L. McCool, On basis-conjugating automorphisms of free groups, *Canad. J. Math.* **38** (1986), 1525–1529.

[68] D. McCullough and A. Miller, *Homeomorphisms of 3-Manifolds with Compressible Boundary*, Chapter 5, *Memoirs AMS* **344**, American Mathematical Society, Providence, 1986.

[69] J. Milnor, Toward the Poincaré conjecture...

[70] M. Montesinos, Alternative symplectic structures for $\mathrm{SO}(3,1)$ and $\mathrm{SO}(4)$ four-dimensional BF theories, `gr-qc/0903076`.

[71] J. W. Morgan, Recent progress on the Poincaré conjecture and the classification of 3-manifolds, Bulletin (New Series) of the A.M.S. **42**, 1, 5778
http://www.ams.org/bull/2005-42-01/S0273-0979-04-01045-6/home.html

[72] T. Needham, *Visual Complex Analysis*, Clarendon Press, Oxford, 1997.

[73] A. Niemi, The exotic statistics of leapfrogging smoke rings, *Phys. Rev. Lett.* **94** (2005), 124502. Also available as cond-mat/0410212.

[74] 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`.

[75] K. Noui and A. Perez, Three dimensional loop quantum gravity: coupling to point particles, *Class. Quant. Grav.* **22** (2005), 4489–4514

[76] D. Oriti, *Spin Foam Models of Quantum Spacetime*, Ph.D. thesis, University of Cambridge. Also available as gr-qc/0311066.

[77] P. Pedlan, Actions for gravity, with generalizations, *Class. Quantum Grav.* **11** (1994) 1087-1132.

[78] Roger Penrose and Wolfgang Rindler, *Spinors and Space-Time*, Vol. 1, Cambridge U. Press, Cambridge, 1985.

[79] A. Perez, Spin foam models for quantum gravity, *Class. Quant. Grav.* **20** (2003), R43–R104.

[80] J. Plebanski, On the separation of Einsteinian substructures, *Jour. Math. Phys.* **18** (1977) 2511-2520.

[81] Carlo Rovelli, *Quantum Gravity*, Cambridge, UK ; New York, Cambridge University Press, 2004.

[82] R. L. Rubinsztein, On the group of motions of oriented, unlinked and unknotted circles in $\mathbb{R}^3$, I. Preprint, Uppsala University, 2002.

[83] R. L. Rubinsztein, Topological quandles and invariants of links, available as math.GT/0508536.

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

[85] P. Scott, The geometries of 3-manifolds, *Bull. London Math. Soc.* **15** (1983), 401-487.

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

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

[88] P. de Sousa Gerbert, On spin and (quantum) gravity in $2 + 1$ dimensions, *Nucl. Phys.* **B346** (1990) 440-472.

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

[90] D. Sullivan and W. Thurston, Manifolds with canonical coordinate charts: some examples, *Enseign. Math.* **29** (1983) 15–25.

[91] S. Surya, Cyclic statistics in three dimensions, *J. Math. Phys.* **45** (2004), 2515–2525. Also available as hep-th/0308011.

[92] R. J. Szabo, Topological field theory and quantum holonomy representations of motion groups, *Ann. Phys.* **280** (2000), 163–208. Also available as hep-th/9908051.

[93] W. P. Thurston *Three-Dimensional Geometry and Topology*, Vol. 1, Princeton University Press, Princeton, 1997. also: http://msri.org/publications/books/gt3m

[94] 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`

[95] V. V. Vershinin, On homology of virtual braids and Burau representation, available as math.GT/9904089.

[96] F. Wattenberg, Differentiable motions of unknotted, unlinked circles in 3-space, *Math. Scand.* **30** (1972), 107–135.

[97] R. Wald, *General Relativity*, Chicago, University of Chicago Press, 1984.

[98] D. K. Wise, p-form electromagnetism on discrete spacetimes, *Class. Quantum Grav.* **23** (2006) 5129–5176.

[99] E. Witten, (2+1)-dimensional gravity as an exactly soluble system, *Nucl. Phys. B* **311** (1988) 46-78.

[100] E. Witten, Quantum field theory and the Jones polynomial, *Commun. Math. Phys.* **121** (1989) 351–399.