Numerical Relativity from a Gauge Theory Perspective

by

Will M. Farr

B.S. Physics
California Institute of Technology, 2003

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2010

© Will M. Farr, MMX. All rights reserved.

The author hereby grants to MIT permission to reproduce and distribute publicly paper and electronic copies of this thesis document in whole or in part.

Author .................................................................
Department of Physics
30 September 2009

Certified by ..............................................................
Edmund Bertschinger
Professor of Physics
Head, Department of Physics
Thesis Supervisor

Accepted by ............................................................
Krishna Rajagopal
Professor of Physics
Associate Department Head for Education
Numerical Relativity from a Gauge Theory Perspective

by

Will M. Farr

Submitted to the Department of Physics
on 30 September 2009, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

I present a new method for numerical simulations of general relativistic systems that eliminates constraint violating modes without the need for constraint damping or the introduction of extra dynamical fields. The method is a type of variational integrator. It is based on a discretization of an action for gravity (the Plebański action) on an unstructured mesh that preserves the local Lorentz transformation and diffeomorphism symmetries of the continuous action. Applying Hamilton’s principle of stationary action gives discrete field equations on the mesh. For each gauge degree of freedom there is a corresponding discrete constraint; the remaining discrete evolution equations exactly preserve these constraints under time-evolution. I validate the method using simulations of several analytically solvable spacetimes: a weak gravitational wave spacetime, the Schwarzschild spacetime, and the Kerr spacetime.

Thesis Supervisor: Edmund Bertschinger
Title: Professor of Physics
Head, Department of Physics
Contents

1 Introduction 11
   1.1 “Standard” Numerical Relativity .......................... 15
   1.2 A 1+1 Wave Equation Example ............................ 19

2 Continuous Gravity 29
   2.1 Spacetime .............................................. 30
      2.1.1 Symmetries of Spacetime ............................ 31
   2.2 Dynamical Fields ........................................ 35
      2.2.1 The Spin Connection ................................ 36
      2.2.2 The Tetrad ......................................... 45
   2.3 The Action ............................................... 47

3 Discretization of Gravity 53
   3.1 Discrete Manifolds ..................................... 53
      3.1.1 Simplexes ......................................... 54
      3.1.2 Simplicial Complexes and Dual Simplexes .......... 56
      3.1.3 Chains of Duals ................................... 60
      3.1.4 Discrete Forms .................................... 60
      3.1.5 Discrete Exterior Derivative ....................... 61
      3.1.6 Discrete Wedge Product ............................ 61
   3.2 Discretized Spacetime .................................. 62
      3.2.1 Symmetries of Discrete Spacetime ................. 62
   3.3 The Dynamical Fields ................................... 63

5
List of Figures

1-1 The submanifold of phase space satisfying the constraints is unstable in most evolution schemes. ........................................ 17

1-2 A two-dimensional discrete manifold. ................................ 21

1-3 The dual of a strut in two dimensions ................................. 23

1-4 1+1 time advancement structure. ................................... 25

1-5 The struts involved in the discrete 1+1 wave equation. ............. 26

1-6 A right-moving Gaussian wave. ...................................... 27

1-7 A right-moving Gaussian wave with hard-reflection boundary conditions. 28

2-1 Computing the path-ordered exponential of the connection around a loop. ....................................................... 43

3-1 A triangle boundary, with orientation ................................. 56

3-2 A two-dimensional complex and its dual ............................. 57

3-3 An example of dual cells. .............................................. 58

3-4 An example of the discrete connection ................................. 64

3-5 The structure of a “pizza slice.” ..................................... 66

3-6 A curvature loop involved in the discrete torsion equation. ......... 72

3-7 The two curvature loops involved in a particular continuity equation. 73

4-1 The time-advancement structure illustrated in 2+1 dimensions. .... 82

4-2 The curvature loops which center an a particular $s_2$. ............. 87

4-3 A simplex used to illustrate diffeomorphism gauge-fixing. .......... 90
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-1</td>
<td>The tile-able decomposition of the cube used to mesh domains for gravitational wave evolution.</td>
</tr>
<tr>
<td>5-2</td>
<td>Meshing of the domain $[0,1]^3$ with $3^3$ points using the meshing scheme we use for decomposing the domains for gravitational wave evolution.</td>
</tr>
<tr>
<td>5-3</td>
<td>Gravitational wave simulation</td>
</tr>
<tr>
<td>5-4</td>
<td>Gravitational wave simulation error versus time for different grid spacing</td>
</tr>
<tr>
<td>5-5</td>
<td>The space- and time-dependence of the constraint violation in the $6 \times 6 \times 20$ gravitational wave simulation.</td>
</tr>
<tr>
<td>5-6</td>
<td>The space- and time-dependence of the constraint violation in the $3 \times 3 \times 10$ gravitational wave simulation.</td>
</tr>
<tr>
<td>5-7</td>
<td>The error relative to the analytic solution for the $6 \times 6 \times 20$ gravitational wave simulation as a function of time and space.</td>
</tr>
<tr>
<td>5-8</td>
<td>The error relative to the analytic solution for the $3 \times 3 \times 10$ gravitational wave simulation as a function of time and space.</td>
</tr>
<tr>
<td>5-9</td>
<td>An example of the spherical mesh used for the Schwarzschild simulations.</td>
</tr>
<tr>
<td>5-10</td>
<td>Time versus radius for different timesteps in the Schwarzschild evolution.</td>
</tr>
<tr>
<td>5-11</td>
<td>Constraint violation in two Schwarzschild simulations.</td>
</tr>
<tr>
<td>5-12</td>
<td>Relative error in the field variables for a simulation of the Schwarzschild spacetime.</td>
</tr>
<tr>
<td>5-13</td>
<td>Relative error in the field variables for a simulation of the Schwarzschild spacetime.</td>
</tr>
<tr>
<td>5-14</td>
<td>Constraint violation in a simulation of the Kerr spacetime.</td>
</tr>
<tr>
<td>5-15</td>
<td>Relative error in the field variables for a simulation of the Kerr spacetime.</td>
</tr>
<tr>
<td>5-16</td>
<td>Relative error in the field variables for a simulation of the Kerr spacetime.</td>
</tr>
</tbody>
</table>
List of Tables

5.1 Numerically determined scalings of the residuals of the discrete equations with mesh spacing .......................... 96
Chapter 1

Introduction

General relativity is a minimalist theory. Nearly every quantity in the theory is dynamic; GR only requires that the set of events in spacetime is smooth enough to form a manifold. The theory does not depend on any notion of absolute location for these events; the relative locations of events are given by the dynamical fields for gravity—the tetrad and connection. The dynamics of matter fields respond to the state of the tetrad and connection, completing the famous cycle: “Space acts on matter, telling it how to move. In turn, matter reacts back on space, telling it how to curve.” [Misner et al., 1973, p. 5] There is no background of space and time on which the fields live—it is fields (almost) all the way down.

Such a theory poses an extreme challenge for numerical simulators. To run a simulation, we need to impose some additional arbitrary structure on the equations of general relativity. In the most prominent numerical simulation codes in use today (e.g. Pretorius (2006), Lousto & Zlochower (2008), Imbiriba et al. (2004), or Scheel et al. (2006)), that structure is a coordinate system that labels each event with a tuple of numerical coordinates. One of the tasks of the simulator using these techniques, then, is to make sure that the arbitrary labels for events on which the computer operates do not get too far “out of synch” with the actual physics of the system being simulated. The freedom to arbitrarily choose coordinates for spacetime events

\[\text{GR does not restrict us from adding structure by, say, assigning a numerical coordinate label to each event in spacetime; it only says that no physical quantity measured or calculated can depend on the way these labels are assigned.}\]
is called a \textit{gauge freedom}, and the particular choice of coordinates in a particular simulation is called \textit{fixing a gauge}.

Gauge freedoms greatly restrict the form of the field equations of GR. There is a field in the field equations corresponding to every gauge freedom; the field equations do not constrain the value of such fields. Fixing a gauge is choosing a value for these fields. There is also a constraint equation corresponding to each gauge freedom. These constraint equations require the vanishing of some components of the momentum on a three-dimensional surface. The gauge invariance of the action implies that a constraint-satisfying field configuration on a Cauchy surface continues to satisfy the constraint equations under evolution by the remaining equations. The evolution is \textit{constraint preserving}. The relationship between gauge freedoms, gauge fields, and constraints is the substance of Nöther’s theorem (see, for example, \textit{Gotay et al. (1998)}).

The techniques currently in widespread use to solve the field equations of GR numerically do not respect the gauge freedoms of GR. The discrete system which the computer is actually solving has only approximate gauge freedoms. Correspondingly, the evolution is only approximately constraint preserving\textsuperscript{2}.

Another way to think of this situation is via “backward error analysis”: consider a continuous system whose continuous evolution matches the discrete results of the numerical integration at the appropriate output times. For integrations which are accurate, this system will differ from GR by small amounts. For the algorithms currently in use, however, the arbitrary error terms mean that the corresponding continuous system is not \textit{physical}, in the sense that its description is not independent of coordinate system. For a constraint-preserving algorithm, the corresponding continuous system will differ from GR by terms which are coordinate invariant—a constraint-preserving integrator simulates a system which is \textit{physical}, just not quite

\footnote{\textit{In fact, many evolution algorithms introduce “constraint damping” terms into their discretizations. \textit{[Pretorius, 2006]} These terms tend to “dissipate” any constraint violating parts of the discrete solution. The algorithms are designed so that the dissipation wins the competition between the algorithm’s constraint violating tendencies and the damping. Though this procedure results in a (nearly) constraint-satisfying numerical evolution, the dissipative nature of the algorithm means that information is being lost.}}
The situation is analogous to discretizing Newton’s second law for a planetary system using a Runge-Kutta integrator (see, e.g., [Springel (2005) [Figure 4]]). Newton’s second law has a freedom of choice of orientation of the coordinate axes used to describe the system. The constraint corresponding to this symmetry is that angular momentum must be conserved. A standard Runge-Kutta integrator breaks this symmetry. The failure of the integrator to respect the freedom of orientation results in failure of angular-momentum conservation in the discrete system. A planet in a simulated orbit about a star will eventually spiral into the star or out of the system as it gains or loses momentum.

A solution to this problem in the simulation of planetary systems is to construct a discretization of the system which respects the symmetries of the continuous system. The symmetries of a mechanical system are most apparent in the Hamiltonian or Lagrangian formulation of the system. Both formulations have been used to construct symmetry-respecting integrators. The Hamiltonian-based integrators [Wisdom & Holman 1991; Yoshida 1993] are called mapping or symplectic\(^4\) integrators; Lagrangian-based integrators ([Marsden & West, 2001; Lew et al., 2004; Farr & Bertschinger, 2007]) are called variational. The variational approach has also been used in continuous systems to construct integrators for discretized fields in space and time in, for example, elastic systems ([Lew et al., 2004]) or electromagnetic systems ([Stern et al., 2007]).

Gauge freedoms also play an important role in quantum field theories. Here the symmetry is not primarily related to spacetime, but rather an “internal” symmetry involving transformations on field values at different points in spacetime. These gauge symmetries greatly restrict the allowed interactions between the quantum fields, and are essential to the process of renormalization, where they ensure that infinities appearing in various pieces of calculations cancel in the final computation of physical

\(^3\)In lattice QCD, minimizing the difference between the physical action and the action which corresponds to a given lattice simulation by understanding and trying to remove the backward error terms is known as the Symanzik improvement program ([Gupta, 1997]).

\(^4\)This term is misleading. Both the Hamiltonian and Lagrangian (variational) approaches yield integrators which have a symplectic flow on phase space.
Gauge freedoms are especially important in the numerical simulation of quantum field theories, in particular the simulation of quantum chromodynamics (QCD) on a numerical lattice. (For an introduction to these techniques, see Gupta (1997).) The techniques in lattice QCD have much in common with the techniques in this thesis. One of the challenges of lattice QCD is finding a way to represent the continuous theory of QCD on a discrete computational lattice in a way that preserves exactly the internal symmetries. This is important because if the symmetry is only recovered in the continuous limit, at finite lattice spacing it would not restrict the interactions among the fields; the extra un-physical interactions at finite spacing would complicate the simulation and attempts to “take the limit” of zero lattice spacing to extract physical results.

This thesis develops numerical integrators for classical general relativity which respect the gauge freedoms of that theory. I focus on variational (action-based) approaches; Hamiltonian perspectives are a possibility for future work. The variational approach requires discretizing an action integral (see Peldan (1994) for an excellent review of the various actions that have been formulated for GR) in a way that preserves its gauge freedoms. This discretization is done on a mesh borrowing techniques from lattice QCD and discrete differential geometry.

Applying Hamilton’s principle of stationary action to the discrete action generates both discrete constraints associated with the gauge freedoms and constraint-preserving evolution equations for the discrete fields in the action. The discrete field equations which result from this procedure (equations (3.47), (3.50), (3.55), (3.59), and (3.60)) in fact correspond to integrals of the continuous field equations over appropriate areas, volumes, and hypervolumes in the mesh. This ensures that the solutions to the discrete field equations correspond to integrals over mesh elements of solutions to the continuous field equations. (This property was first noticed for a similar discretization in Reisenberger (1997).) Other action-based approaches to numerical relativity, such as Regge calculus (Barrett et al., 1997; Gentle, 2002), do not have this property (Brewin, 1995; Miller, 1995).
1.1 "Standard" Numerical Relativity

As a point of comparison, this section presents an (extremely abbreviated) discussion of the techniques typically employed for numerical simulation of GR spacetimes. We will see that the gauge symmetries of GR, and the associated constraints, have presented significant difficulties for the field, particularly in the form of constraint violating modes in the evolution equations. We will describe the (ad-hoc, but successful) approaches currently used to deal with these constraint violating modes. The techniques described in this thesis are an alternative, first-principles way to deal with constraint violating modes. This section largely follows the excellent review of the present state of numerical relativity in Pretorius (2009); for more information, see that reference, and references therein.

The starting-point for standard vacuum numerical relativity is the Einstein equation,

\[
G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 0,
\]

where \(G_{\mu\nu}\) are the components of the Einstein tensor, \(R_{\mu\nu}\) are the components of the Ricci tensor, \(g_{\mu\nu}\) are the components of the metric tensor, and \(R = g^{\mu\nu} R_{\mu\nu}\) is the trace of the Ricci tensor with respect to the inverse metric \(g^{\mu\nu}\). Equation (1.1) is a system of 10 coupled, quasilinear, second-order partial differential equations for the 10 components of the metric in terms of the coordinates \(x^\mu\). However, as is well-known, and can be seen from the contracted Bianchi identity, \(\nabla^\mu G_{\mu\nu} = 0\), the four \(G_{0\nu}\) equations contain one fewer time-derivative of the metric components than the remaining 6 equations. The equations (1.1) are therefore ill-posed as a time-evolution system without additional information.

The standard way to deal with this problem is to begin by fixing a timelike \(t\) coordinate, and spacelike \(x, y,\) and \(z\) coordinates\(^5\). The metric decomposes in this coordinate system as

\[
g_{\mu\nu} dx^\mu dx^\nu = -\alpha^2 dt^2 + h_{ij} \left(dx^i + \beta^i dt\right) \left(dx^j + \beta^j dt\right),
\]

\(^5\)Some codes have used a null decomposition, choosing one or two coordinates to be light-like, instead of the 3+1 decomposition above. See Pretorius (2009) and references therein.
where the spatial component indices $i$ and $j$ range from 1 to 3, the function $\alpha$ is called the “lapse”, the three functions $\beta^i$ are called the “shift”, and $h_{ij}$ is a metric on constant $t$ surfaces. The lapse and shift specify the flow of time (the $t$ coordinate) by

$$\left( \frac{\partial}{\partial t} \right)^\mu = \alpha n^\mu + \beta^\mu, \quad (1.3)$$

where $\beta^0 = 0$, and $n^\mu$ is a unit normal to the $t = \text{const}$ surfaces. The lapse and shift encompass the coordinate degrees of freedom of the theory—they specify “where” the evolution takes each spacetime point. In the Hamiltonian analysis of this 3+1 split of the theory (see, e.g., [Arnowitt et al. (1962)]), the spatial metric, $h_{ij}$, plays the role of the coordinate, and the extrinsic curvature, defined by

$$K_{ij} = -\frac{1}{2\alpha} \left( \frac{\partial h_{ij}}{\partial t} - \mathcal{L}_\beta h_{ij} \right), \quad (1.4)$$

where $\mathcal{L}_\beta$ is the Lie derivative with respect to the vector $\beta$, plays the role of conjugate momentum. In terms of $\alpha$, $\beta^i$, $h_{ij}$, and $K_{ij}$ the equations (1.1) can be written as 12 hyperbolic evolution equations for the coordinate and conjugate momentum and four constraint equations that do not involve time-derivatives of the momentum. Evolution equations for the gauge degrees of freedom, $\alpha$ and $\beta^i$, must be added to close the system.

The standard choice for evolution of the equations (1.1) is to select a subset of the equations for the evolution. The four constraint equations are ignored, except at the initial time, where they provide constraints on the initial values of $h_{ij}$ and $K_{ij}$; the remaining 12 equations, plus the gauge conditions specifying the evolution of $\alpha$ and $\beta^i$ specify the evolution of the system. This approach, while efficient, leads to serious complications. The continuous evolution of the entire system is constraint-preserving (this is guaranteed by the gauge invariance of the theory); however, most choices for the 12 evolution equations lead to sub-systems that have continuous solutions that violate the constraints, and grow in magnitude with time. The constraint-satisfying submanifold of the phase space of solutions to the evolution equations is \textit{unstable}. This is illustrated in Figure 1-1. These growing, constraint-violating modes are sourced by
Figure 1-1: A cartoon of the phase space evolution for most choices of the 12 evolution equations in a standard numerical scheme. The vertical curve represents the subset of the phase space that satisfies the constraint equations. In the continuous evolution, a solution that begins on this surface remains on this surface permanently. However, any deviation from the surface tends to grow with evolution; this is indicated by the arrows leading away from the constraint-satisfying sub-manifold of phase space. In a numerical method, the truncation errors cause the solution to deviate from exact constraint preservation, and the resulting growing constraint-violating mode wrecks the simulation.

The issue is that the truncation errors of the discrete calculation do not, in general, satisfy the constraints. The truncation errors, as discussed above, are not physical; they do not respect the gauge freedoms of the continuous theory.

The standard solution to the problem of constraint-violating modes involves constraint damping in the codes which use generalized harmonic coordinates (e.g. Pretorius (2006)) or a careful choice of evolution variables in the BSSN method (Baumgarte & Shapiro 1998) which probably amounts to the same thing (Pretorius 2009). Constraint damping involves adding terms to the evolution equations which are multiples of the constraint equations; so long as the constraints are satisfied, the additional terms vanish and the two systems are equivalent. The terms are chosen so that the evolution in the full phase space is parabolic, becoming hyperbolic only on the constraint-satisfying surface. (In general, choosing the combination of constraints to be added is more art than science, but see Paschalidis (2008) for formulations...
different from the harmonic and BSSN that were specifically designed to have this parabolic-hyperbolic property.) In other words: the natural evolution of the system wants to move away from the constraint-satisfying submanifold of phase space, but constraint damping adds “viscosity” or dissipation terms off the constraint manifold which compensate for this tendency.

With the addition of constraint damping terms, and an appropriate choice of gauge, both the generalized harmonic algorithms and BSSN appear to be stable (and are known to be linearly stable against high-frequency perturbations \cite{Paschalidis2007}). Both techniques have been used to compute gravitational waveforms from the collapse of binary black hole systems over many orbits, for example \cite{Preto2009}, and the results are in excellent agreement. However, the addition of the constraint damping terms is ad-hoc, and the ultimate effect of the dissipation is not well understood theoretically.

This thesis presents an alternate solution to the problem of constraint-violating modes. Just as building a discrete system that respects the symmetries of a continuous mechanical system in a variational or symplectic integrator leads to conservation of momentum in the simulation, building a discrete gravity theory that respects the gauge symmetries of the continuous theory produces an integrator that respects the constraints. In such an integrator, the truncation error does not source constraint violating modes because even the truncation error of the method satisfies the constraints. In the following chapters, we describe the implementation of such an integrator.

In Chapter 2, we describe the continuous formulation of relativity we will be discretizing; it describes the geometry of spacetime in terms of local orthonormal frames which can be independently chosen at each event in spacetime. The result is a theory with both coordinate and internal symmetries. It is completely equivalent to the usual formulation of GR in terms of a metric and affine connection, but has advantages for discretization.

In Chapter 3, we describe the mathematical machinery we will need for the formulation of GR on a discrete mesh. We go on to describe the discretization of the continuous theory, and derive the discrete field equations. We demonstrate that these
correspond to the continuous field equations integrated on mesh elements, and discuss
the geometric meaning of some of the equations and the way they enforce geometric
properties of the mesh.

In Chapter 4 we present an algorithm for constructing a field configuration that
satisfies the discrete field equations given suitable initial data. We demonstrate how
the gauge freedoms introduce redundancies in the discrete field equations, and how
to deal with the resulting under-determined system by choosing a suitable gauge.
We also show how the the gauge symmetries imply constraint preservation by the
integrator. The time-advancement mesh structure we use was first described in the
context of Regge calculus in Barrett et al. (1997), and is local, easily parallelizable,
and adapts to any mesh topology. (We do not construct parallel algorithms for the
solution of the field equations here; that is a possibility for future work.)

In Chapter 5 we validate our method using small test simulations of analytically
solvable spacetimes. We demonstrate the method on the time-advancement of a weak
gravitational wave, a simulation of the Schwarzschild spacetime describing an isolated
non-spinning black hole, and a simulation of the Kerr spacetime describing a spinning
black hole.

1.2 A 1+1 Wave Equation Example

Before leaping into the discussion of general relativity, this section provides an ex-
ample of the construction of an integrator for the 1+1-dimensional massless scalar
wave equation using many of the same techniques we will apply to GR in the coming
chapters. This example serves two purposes. First, we hope to intrigue the reader
with the elegance of the discretization and the close correspondence between contin-
uous and discrete theories. Second, we hope this example can serve as a reference to
return to if the complications of the application of these techniques to GR become
overwhelming. (See Marsden et al. (1998) for an example of this technique applied
to the sine-Gordon equation, a non-linear wave equation similar to the linear wave
equation discussed here.)
The 1+1 massless scalar wave equation for a scalar field, $\phi(x)$, on a manifold $M$, can be derived from the action

$$S(\phi) = \frac{1}{2} \int_M d^2x \sqrt{-g} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi,$$

where $g^{\mu\nu}$ is the inverse metric on our 1+1 domain and $g = \det (g_{\mu\nu})$ is the determinant of the metric. This can be re-written in the language of differential forms (see Appendix A) as

$$S(\phi) = \frac{1}{2} \int_M d\phi \wedge (\star d\phi).$$

Forms are convenient to work with because they can be integrated over lengths, volumes, etc. When we discretize, we will have discrete forms that take values on pieces of the discrete mesh and correspond to the integrals of continuous forms over the volumes represented by the mesh elements.

Varying the action in equation (1.6) with respect to $\phi$, but keeping $\phi$ fixed on the boundary of the domain results in the equation of motion

$$d(\star d\phi) = 0,$$

or, in the language of the first action,

$$\Box \phi = g^{\mu\nu} \nabla_\mu \nabla_\nu \phi = \partial_\mu \left( g^{\mu\nu} \sqrt{-g} \partial_\nu \phi \right) = 0.$$

This is the 1+1 wave equation. If we specialize to the case where the manifold, $M$, is flat, and we use Cartesian coordinates so the metric $g = \eta = \text{diag}(-1, 1, 1, 1)$, then the coordinate form of the wave equation becomes

$$\Box \phi = \eta^{\mu\nu} \partial_\mu \partial_\nu \phi = \partial^\mu \partial_\mu \phi = 0.$$

To discretize this system, we begin by discretizing the domain. Figure 1-2 shows a decomposition of a two-dimensional manifold into discrete areas, enclosed by triangles. This triangulation is an example of a two-dimensional simplicial complex. The
Figure 1-2: A two-dimensional discrete manifold. The elements of the discrete manifold are zero-, one-, and two-dimensional simplexes—points, struts, and faces. Every triangle has three one-dimensional faces, and every strut has two zero-dimensional faces. The geometrical intersection of any two elements is either empty, or else a face of each.

elements of the triangulation—the points, lines, and triangles—are called simplexes. In the discrete theory, fields will take values on the simplexes. We choose the value of a discrete field on a simplex to be the integral of the corresponding continuous field over the volume represented by that simplex. The (oriented) element containing the points \( p, q, \ldots \) is denoted by \( \{ p, q, \ldots \} \). A face of an element contains a subset of the points of that element; we denote the “face-of” relation by \( \preceq \).

The continuous field \( \phi \) is a scalar at each spacetime point. A scalar is a zero-form; scalar field values should be discretized on points. We will denote the value of the discrete \( \Phi \) on the point \( p \) in the discrete manifold by \( \langle \Phi, \{ p \} \rangle \).

The field \( d\phi \) is a one-form. One-forms can be integrated on one-dimensional structures. Therefore, \( d\Phi \) should live on struts in the mesh: \( \langle d\Phi, \{ p, q \} \rangle \), where \( p \) and \( q \) are two points in the mesh connected by a strut. We define

\[
\langle d\Phi, \{ p, q \} \rangle = \langle \Phi, \{ q \} \rangle - \langle \Phi, \{ p \} \rangle .
\] (1.10)

21
This is a discrete version of Stokes’ theorem. If \( \langle d\Phi, \{p, q\} \rangle \) is \( d\phi \) integrated on the strut, then its value should be \( \phi \) integrated on the boundary of the strut. This is exactly what equation (1.10) provides. Note that \( d\Phi \) changes sign when the strut changes orientation.

The pairing \( \langle d\Phi, \{p, q\} \rangle \) is the exterior derivative integrated on the strut \( \{p, q\} \). The definition in equation (1.10) is a first-order differencing scheme. With first-order finite-differencing, if \( x^\mu \) were the coordinates of \( p \), and \( y^\mu \) were the coordinates of \( q \), then we would compute an approximation to the exterior derivative of the continuous field \( \phi \) using

\[
\partial_\mu \phi \approx \frac{\phi(y) - \phi(x)}{y^\mu - x^\mu}.
\] (1.11)

Integrating the derivative along the strut from \( p \) to \( q \) would give

\[
\int_{\{p,q\}} \partial_\mu \phi dx^\mu \approx \frac{\phi(y) - \phi(x)}{y^\mu - x^\mu} (y^\mu - x^\mu) = \phi(y) - \phi(x).
\] (1.12)

Replacing the values of the continuous field at \( y \) and \( x \) by the values of the discrete field on the mesh elements \( q \) and \( p \), we obtain equation (1.10).

The boundary of a boundary is empty. Our definition of the exterior derivative respects this. Evaluating \( d^2 \Phi \) on a two-simplex, we have

\[
\langle d^2 \Phi, \{p, q, r\} \rangle = \langle d\Phi, \{p, q\} \rangle + \langle d\Phi, \{q, r\} \rangle + \langle d\Phi, \{r, p\} \rangle
\]

\[
= \langle \Phi, \{q\} \rangle - \langle \Phi, \{p\} \rangle + \langle \Phi, \{r\} \rangle - \langle \Phi, \{q\} \rangle + \langle \Phi, \{p\} \rangle - \langle \Phi, \{r\} \rangle = 0,
\] (1.13)

for any values of the field \( \Phi \), so this definition enforces the identity \( d^2 = 0 \).

The discretization of the dual form, \( \star d\phi \) involves the dual mesh of the discrete manifold. The dual mesh is composed of points which lie at the centers of elements of the original, or primal, mesh. We denote the center of an \( i \)-dimensional mesh element \( s_i \) as \( c(s_i) \). For any volume of the original mesh, \( s_2 = \{p, q, r\} \), and any face of that volume, \( s_1 = \{p, r\} \), there is a unique dual element, written \( \star s_1 \cap s_2 \) which is geometrically dual to \( s_1 \) and contained in \( s_2 \). For each strut not on the boundary of a complex, there are two incident volumes, and two corresponding duals—one for each
Figure 1-3: The triangle \( s_2 = \{p, q, r\} \) with the dual strut to \( s_1 = \{p, r\} \), \(*s_1 \cap s_2 = \{c(s_2), c(s_1)\}\), indicated. Note that the relative orientation of \( s_1 \) and \(*s_1\) is consistent with the orientation of the volume \( s_2 \): the volume is oriented in a right-handed fashion from \( p \) to \( q \) to \( r \), and the struts \( s_1 \) and \(*s_1\) are also in a right-handed orientation.

“side” of the strut. Struts on the boundary have only a single dual; the “side” of the strut outside the complex has no \( s_2 \) associated with it, and therefore no dual. The dual element of a one-dimensional simplex is itself one-dimensional, and consists of the strut connecting the center of \( s_1 \) and the center of \( s_2 \): \(*s_1 \cap s_2 = \{c(s_1), c(s_2)\}\) or \(*s_1 \cap s_2 = \{c(s_2), c(s_1)\}\) depending on the relative orientations of \( s_1 \) and \( s_2 \). The orientation of the dual element is chosen to be consistent with the orientation of \( s_2 \), so that the pair of struts \( s_1 \) and \(*s_1 \cap s_2 \) are in a relative orientation that is the same as the orientation of \( s_2 \). Figure 1-3 illustrates this. The dual of a \( n \)-dimensional simplex in a \( N \)-dimensional volume represents the \((N - n)\)-dimensional space contained in that volume but not spanned by the \( n \)-simplex.

The discretization of \(*d\phi\) lives on the duals to the discretization of \(d\phi\). We exploit the one-to-one mapping between struts and their duals in a particular volume to define

\[
\langle *d\Phi, *s_1 \cap s_2 \rangle \equiv \epsilon_{s_1} \frac{|*s_1 \cap s_2|}{|s_1|} \langle d\Phi, s_1 \rangle ,
\]

(1.14)
where $|s|$ denotes the length of the strut $s$ with respect to a discretized metric. The factor of $\epsilon_s$, accounts for the Lorentzian signature of our 1+1 manifold: if $s_1$ is timelike, then $\epsilon = -1$, otherwise it is 1. The ratio of lengths accounts for the fact that $d\Phi$ scales as the length of the struts, while $\star d\Phi$ should scale as the length of the dual struts. (Referring to Figure 1-3, note that the distance from $p$ to $r$ can be increased—thereby increasing the magnitude of $d\Phi$ on this strut—without changing the distance from $c(s_2)$ to $c(s_1)$.)

The discrete action is

$$S(\Phi) = \frac{1}{2} \sum_{s_2} \sum_{s_1 \leq s_2} \langle d\Phi, s_1 \rangle \langle \star d\Phi, \star s_1 \cap s_2 \rangle. \quad (1.15)$$

The continuous integral over volume has become a discrete sum of volume elements in the mesh.

Using equations (1.14) and (1.10), we have

$$S(\Phi) = \frac{1}{2} \sum_{s_2} \sum_{s_1 \leq s_2} \epsilon_{s_1} \frac{|\star s_1 \cap s_2|}{|s_1|} \langle d\Phi, s_1 \rangle^2$$

$$= \frac{1}{2} \sum_{s_2} \sum_{s_1 = (p,q) \leq s_2} \epsilon_{s_1} \frac{|\star s_1 \cap s_2|}{|s_1|} \left(\langle \Phi, \{q\} \rangle - \langle \Phi, \{p\} \rangle\right)^2. \quad (1.16)$$

Hamilton’s stationary action principle provides us with evolution equations for the field values $\Phi$. We wish to advance a known field configuration (initial condition) forward in time. Figure 1-4 demonstrates the structure of a mesh which results from advancing the point $\{p\}$ to $\{q\}$. Assume we know the values of $\Phi$ on the solid mesh. Hamilton’s stationary action principle, applied at the interior point $\{p\}$, states that

$$\frac{\partial S}{\partial \langle \Phi, \{p\} \rangle} = 0. \quad (1.17)$$

The action is a sum over volumes and struts. To compute the derivative, we need only consider the terms in the action corresponding to volumes and struts of which $p$
Figure 1-4: Advancing the point \{p\} forward in time to the point \{q\}. (Time runs vertically, space horizontally.) If field values are known on the solid-line mesh, then demanding that the variation of the action in equation (1.16) with respect to \(\langle \Phi, \{p\} \rangle\) vanish produces a discrete Euler-Lagrange equation which can be solved for \(\langle \Phi, \{q\} \rangle\) in terms of the known values. Once \(\langle \Phi, \{q\} \rangle\) is known, the process can be repeated for any of the points in the advanced-time boundary of the mesh.

is a face. Taking the derivative of these terms gives

\[
\sum_{s_1 \geq \{p\}} \sum_{s_2 \geq s_1} \epsilon_{s_1} \frac{|s_1 \cap s_2|}{|s_1|} \langle d\Phi, s_1 \rangle = \\
\sum_{s_2 \geq \{p,q\}} \epsilon_{\{p,q\}} \frac{|\{p,q\} \cap s_2|}{|\{p,q\}|} (\langle \Phi, \{q\} \rangle - \langle \Phi, \{p\} \rangle) + \\
\sum_{\{p,r\}, r \neq q} \sum_{s_2 \geq \{p,r\}} \epsilon_{\{p,r\}} \frac{|\{p,r\} \cap s_2|}{|\{p,r\}|} (\langle \Phi, \{r\} \rangle - \langle \Phi, \{p\} \rangle) = 0. \tag{1.18}
\]

This is the discrete version of equation (1.7). The struts involved in this sum are illustrated in Figure 1-5. Only one of the terms in this equation involves \(\langle \Phi, \{q\} \rangle\), so we can solve this equation for \(\langle \Phi, \{q\} \rangle\) in terms of the known values of \(\Phi\) on the solid mesh. This process can then be repeated for any point in the advanced-time boundary of the new mesh.

Equation (1.18) states that the volume-weighted sum of \(d\Phi\) (accounting for signs
Figure 1-5: The struts involved in equation (1.18) are indicated on the time-advancement structure with a thick line.

on timelike struts over all struts incident on the point \{p\} vanishes. If the mesh were regular, there would be one strut oriented into \{p\} for each strut oriented out of \{p\}, and the sum would reduce to the average (again, with a change of sign for the timelike struts) of field values over points in the neighborhood of \{p\}. This is the standard “averaging” algorithm for solving Laplacian or D’Lambertian equations on a grid. Our algorithm is a generalization of this to an unstructured mesh.

Figures 1-6 and 1-7 show the results of simulations performed using this technique.

Now on to general relativity!
Figure 1-6: A Gaussian wave moving in the positive $x$-direction with standard deviation of 15 spatial units beginning with a peak at the origin, simulated using the method described in the text. The simulation has a spatial discretization scale of one unit, and was advanced with a shared timestep for each point of 0.5 units. The boundary conditions are appropriate for the propagating Gaussian; the field propagates off the grid with no significant reflection.
Figure 1-7: A Gaussian wave with standard deviation of 15 spatial units beginning with a peak at $x = 50$ units moving in the positive $x$-direction, simulated using the method described in the text. The simulation has a spatial discretization scale of one unit, and was advanced with a shared timestep of 0.5 units. The boundary conditions are $\Phi = 0$; the field reflects off the boundaries with no significant loss through the boundary.
Chapter 2

Continuous Gravity

To implement an action-based discretization of GR, we must first choose an action to discretize. In this chapter I will describe two particular action-based formulations of gravity based on the Hilbert-Palatini and Plebanski actions. Much of the material is a summary of the information in the excellent review article Peldan (1994), though the approach and notation here differ significantly from that article. The Plebanski action is useful in loop quantum gravity, where its spin foam quantization is known as the Barrett-Crane model. See Perez (2004) for an introduction to these concepts.

The advantages of these formulations over the standard (and more familiar) formulation in terms of a metric and affine connection is that the dynamical entities in these formulations are forms. Forms, and the closely related exterior derivative operator, can be discretized in a coordinate-invariant way on a mesh by integrating them over curves, surfaces, volumes, etc. For more on forms and exterior derivatives see Appendix A.

By the end of this chapter, we will have in hand a formulation of General Relativity in 3+1 dimensions which will transfer nicely to a discrete manifold (discretization is the subject of Chapter 3).
2.1 Spacetime

In these formulations of gravity, spacetime consists of a 3+1 dimensional manifold of events (what we usually think of as “spacetime”) and various vector spaces attached to each event in the manifold\footnote{Formally, this structure is known as a fibre bundle.} We attach to each spacetime event the usual tangent and cotangent spaces. In addition, we attach a Minkowski space (i.e. $\mathbb{R}^4$, with the usual metric, $\eta = \text{diag}(-1,1,1,1)$) to each event. These Minkowski spaces are identified with the local frames of freely falling observers at each event. We will describe gravity by the effect it has on the frames of these observers.

The following are examples of some fields which we can define on the spacetime bundle (we denote coordinates in the event manifold by $x$):

- **Scalar fields:** $\phi(x)$.

- **Vector fields:** $v^{\mu}(x)$. These live in the tangent space of the event manifold at $x$.

- **One-form fields:** $w_\mu(x)$. These live in the co-tangent space of the event manifold at $x$.

- **General tensor fields:** $T^{\mu\nu\ldots}_{\rho\sigma\ldots}(x)$. These live in tensor products of the tangent and co-tangent spaces at $x$.

- **Minkowski vector fields:** $v^A(x)$. These live in the Minkowski vector space attached to the event at $x$.

- **Minkowski one-form fields:** $w_A(x)$. These live in the Minkowski co-vector space attached to the event at $x$.

- **General Minkowski tensor fields:** $T^{AB\ldots}_{CD\ldots}(x)$. These live in tensor products of the Minkowski tangent and co-tangent spaces at $x$.

- **General tensor fields:** $T^{AB\ldots\mu\nu\ldots}_{CD\ldots\rho\sigma\ldots}(x)$. These live in tensor products of the Minkowski and tangent spaces attached to the event at $x$.
• In the next section, we will see that anti-symmetric pairs of Minkowski indices can be decomposed into their six independent components using the generator matrices for \(SO(3,1)\). So, for example, the anti-symmetric Minkowski tensor field \(m^{AB}\) can be written as a sum of generators as

\[
m^{AB}(x) = m^{[AB]}(x) = m^a(x)t^{AB}_a, \quad a = 0, 1, \ldots, 5.
\]  

(2.1)

The adjoint field \(m^a(x)\) lives in the Minkowski space at \(x\).

We call the Greek indices \(\mu, \nu, \ldots\) “event” indices, the uppercase Latin indices \(A, B, \ldots\) “Minkowski” indices, and the lower-case Latin indices \(a, b, \ldots\) “adjoint” indices.

Because we have the usual metric on the Minkowski spaces, we can “raise” and “lower” Minkowski indices using \(\eta\).

### 2.1.1 Symmetries of Spacetime

General relativity in this formulation permits two local symmetry operations on the spacetime bundle: diffeomorphism symmetry and Lorentz symmetry. These are the gauge freedoms of GR.

**Diffeomorphism** In this work, we will take the passive view of diffeomorphism: an (infinitesimal) diffeomorphism is a change of coordinates on the event manifold, from \(x\) to \(\tilde{x} \equiv x + \xi\), where we regard \(\xi\) as small. If we have a vector field on the event manifold, \(v^\mu\), then components of \(v\) at the same spacetime point simply rotate:

\[
\tilde{v}^\mu(\tilde{x}(x)) = v^\nu(x)\frac{\partial \tilde{x}^\mu}{\partial x^\nu} = v^\nu(x)\partial_\nu \xi^\mu.
\]  

(2.2)

If we look at \(v\) and \(\tilde{v}\) as functions from coordinate tuples to component tuples, however, then we have

\[
\tilde{v}^\mu(\tilde{x}) = v^\mu(\tilde{x} - \xi) + v^\nu(\tilde{x} - \xi)\partial_\nu \xi^\mu = v^\mu(\tilde{x}) + v^\nu(\tilde{x})\partial_\nu \xi^\mu - \xi^\nu \partial_\nu v^\mu(\tilde{x}).
\]  

(2.3)
to lowest order in $\xi$. So, under a diffeomorphism parametrized by $\xi$, we have

$$\delta_\xi v^\mu = (\mathcal{L}_\xi v)^\mu = [v, \xi]^\mu = v^\nu \partial_\nu \xi^\mu - \xi^\nu \partial_\nu v^\mu.$$  \hfill (2.4)

Similar rules obtain for the transformations of forms and tensors. For example, a tensor with two lower indices transforms as

$$\delta_\xi w_{\mu\nu} = - (\xi^\alpha \partial_\alpha w_{\mu\nu} + w_{\alpha\nu} \partial_\mu \xi^\alpha + w_{\mu\alpha} \partial_\nu \xi^\alpha).$$  \hfill (2.5)

Diffeomorphisms do not touch Minkowski indices.

**Lorentz** A Lorentz transformation can be represented by a matrix, $\Lambda^A{}^B$, which leaves the Minkowski metric, $\eta = \text{diag}(-1, 1, \ldots)$ invariant:

$$\Lambda^{A'}{}_{A} \eta_{A'B'} \Lambda^{B'}{}_{B} = \eta_{AB}. \hfill (2.6)$$

If $\Lambda$ is close to the identity,

$$\Lambda^A{}^B = \delta^A{}^B + \epsilon^A{}^B + \ldots$$  \hfill (2.7)

then Equation $2.6$ implies

$$\epsilon^{AB} = \epsilon^{[AB]}, \hfill (2.8)$$

because there are $n(n+1)/2$ independent components in a $(n+1)$ by $(n+1)$ antisymmetric matrix, the $n+1$ dimensional Lorentz group is $n(n+1)/2$-dimensional. Under an infinitesimal Lorentz transformation, a Minkowski vector changes by

$$\delta_\epsilon v^A = \epsilon^{A'}{}_{B'} v^{B'}.$$  \hfill (2.9)

Similar rules obtain for Minkowski one-forms and tensors. For example, a tensor $T^A_B$ transforms as

$$\delta_\epsilon T^A_B = \epsilon^{A'}{}_{A'} T^A_B - \epsilon^{B'}{}_{B'} T^A_{B'}.$$  \hfill (2.10)

A Lorentz transformation does not touch event indices.
In 3+1 dimensions, the Lorentz group is six-dimensional, consisting of three boosts along the three spatial axes and three rotations about the spatial axes. Infinitesimal Lorentz transformations on vectors can be parametrized by introducing six generators:

\[ t^A = \{t^A, t^B\}, \quad a = 0, \ldots, 5, \]  

(2.11)

which span the space of 4x4 anti-symmetric matrices. A convenient parametrization for the generators is

\[ t^0 = -t^0 = \delta^{B}_{(a+1)}, \quad a = 0, 1, 2 \]  

(2.12)

for the three boosts and

\[ t^A = -t^A = -\epsilon_{(a-3)(A-1)(B-1)}, \quad a = 3, 4, 5, \quad A, B = 1, 2, 3 \]  

(2.13)

with \( t^A = -t^A = 0 \), for the three rotations, where \( \epsilon_{ijk} \) is the completely anti-symmetric tensor on three indices \((0 \leq i, j, k \leq 2)\). With this parametrization, the boost generators correspond to boosts along the axes \( x, y, \) and \( z \) respectively, and the three rotation generators correspond to positive rotations about the \( x, y, \) and \( z \) axes, respectively.

\( SO(3, 1) \) is a Lie group; the generators satisfy

\[ t^A t^B t^C - t^A = t^A t^C - t^B = f^c_{ab} t^a t^b, \]  

(2.14)

for some structure constants \( f^a_{bc} \). Any six matrices which satisfy the commutation relations in equation (2.14) with the same structure constants \( f \) are generators for a representation of the Lorentz group.

Due to the Jacobi identity for commutators of matrices, the structure constants
themselves satisfy the commutation relations in equation (2.14):

\[ [f_a, f_b]_d^c = f_{ae}^bf_{bd}^e - f_{be}^af_{ad}^e = f_{ab}^gf_{gd}^c. \]  

(2.15)

The representation of the Lorentz group whose generators are the structure constants is called the \textit{adjoint} representation. It is the representation of Lorentz transformations on the generators themselves. If \( \Lambda_B^A = \delta_B^A + \epsilon^a t_{aB}^A + \mathcal{O} (\epsilon^2) \) acts on \( m_B^A = v_a t_{aB}^A \), we have

\[ m_B^A \mapsto m_B^A + \epsilon^a t_{AC}^A m_C^B - \epsilon^a t_{aB}^A m_C^A + \mathcal{O} (\epsilon^2). \]  

(2.16)

Expanding in terms of the generator decomposition of \( m \), we have

\[ v_a t_{aB}^A \mapsto v_a t_{aB}^A + \epsilon^b t_{bc}^A v_a t_{aB}^C - v_a t_{aC}^A \epsilon^b t_{bB}^C + \mathcal{O} (\epsilon^2) = (v_a + \epsilon^b f_{bc}^a v^c) t_{aB}^A + \mathcal{O} (\epsilon^2). \]  

(2.17)

Thus we have

\[ v^a \mapsto v^a + \epsilon^b f_{bc}^a v^c + \mathcal{O} (\epsilon^2); \]  

(2.18)

the structure constants are the generators of transformations on adjoint indices.

In 3+1 dimensions, there are two Lorentz-invariant ways to contract the four Minkowski indices on pairs of generators. The first way is to contract both the indices using \( \eta \). This induces a metric on the adjoint space (not to be confused with the spacetime metric, \( g_{\mu\nu} \)):

\[ g_{ab} \equiv t_a^{AB} t_{bAB}. \]  

(2.19)

For the representation of the generators given in equations (2.12) and (2.13),
we have

\[ g_{ab} = \begin{pmatrix}
-2 & 0 & 0 & 0 & 0 \\
0 & -2 & 0 & 0 & 0 \\
0 & 0 & -2 & 0 & 0 \\
0 & 0 & 0 & 2 & 0 \\
0 & 0 & 0 & 0 & 2
\end{pmatrix}. \] \hfill (2.20)

(2.20)

We can also contract all the Minkowski indices in a pair of generators using \( \epsilon_{ABCD} \), the completely anti-symmetric tensor in four dimensions. This induces a second metric on the adjoint space:

\[ \tilde{g}_{ab} = \epsilon_{ABCD} t^A_a t^C_b. \] \hfill (2.21)

(2.21)

For the representation of the generators given in equations (2.12) and (2.13), we have

\[ \tilde{g}_{ab} = \begin{pmatrix}
0 & 0 & 0 & 4 & 0 & 0 \\
0 & 0 & 0 & 0 & 4 & 0 \\
0 & 0 & 0 & 0 & 0 & 4 \\
4 & 0 & 0 & 0 & 0 & 0 \\
0 & 4 & 0 & 0 & 0 & 0 \\
0 & 0 & 4 & 0 & 0 & 0
\end{pmatrix}. \] \hfill (2.22)

(2.22)

We will employ these metrics to simplify terms in the action and equations of motion in the following sections.

\section*{2.2 Dynamical Fields}

This section describes the dynamical fields we will use to describe the effects of gravity on the frames of freely-falling observers.
2.2.1 The Spin Connection

Without additional information, there is no way to relate the various Minkowski spaces attached to our event manifold—the Minkowski space at each point is independent of all the other Minkowski spaces because we are free to choose freely falling observers independently at each event. Any Minkowski space can be related to any other Minkowski space by a Lorentz transformation, so if we want to relate Minkowski spaces at points $p$ and $q$ in the spacetime manifold—because we want to compare vectors which live in them, for example—we only need to specify the Lorentz transformation which relates them.

The spin connection is a field which does exactly that. It has the following index structure:

$$\omega^A_{\mu B}. \quad (2.23)$$

Given a point, $x^\mu$, and a vector, $v^\mu$, $\omega^A_{\mu B}(x)v^\mu$ gives the infinitesimal Lorentz transformation which relates the Minkowski frames at $x$ with those infinitesimally separated along the $v$ direction:

$$\Lambda(x^\mu \rightarrow x^\mu + v^\mu)^B_A = \delta^A_B + \omega^A_{\mu B} v^\mu + O(v^2). \quad (2.24)$$

Since it is an infinitesimal Lorentz transformation (i.e. a Lie-algebra element of $SO(n,1)$), we have, from equation (2.8),

$$\omega^{AB}_\mu = \omega^{[AB]}_\mu. \quad (2.25)$$

We will often decompose $\omega$ in the adjoint basis, and write $\omega^a_\mu$ for the resulting components:

$$\omega^{AB}_\mu = \omega^a_\mu t^{AB}_a. \quad (2.26)$$

For any object which transforms under some representation of the Lorentz group, contracting $\omega^a_\mu$ with the appropriate generators for that representation gives the infinitesimal transformation on that representation.
Covariant Derivative

The spin connection allows us to define a “gauge-covariant derivative” on objects which live in the Minkowski vector spaces of the spacetime manifold:

\[ D_\mu v^A = \partial_\mu v^A - \omega^A_{\mu B} v^B, \]  
(2.27)

or, acting on an object with more complicated index structure,

\[ D_\mu T^A_{BC} = \partial_\mu T^A_{BC} - \omega^A_{\mu A'} T^{A'}_{BC} + \omega^B_{\mu B'} T^A_{B' C} + \omega^C_{\mu C'} T^A_{B C'}. \]  
(2.28)

On an object with adjoint indices, we have

\[ D_\mu v^a = \partial_\mu v^a - \omega^b_{\mu f} f^a_{bc} v^c. \]  
(2.29)

The covariant derivative is the leading-order term in the comparison between an object at an advanced point, \( x^\mu + h^\mu \), and the advancement of an object at a base point, \( x^\mu \):

\[ v^A(x + h) - \Lambda(x \rightarrow x + h) T^A_B v^B(x) = h^\mu \left( \partial_\mu v^A(x) - \omega^A_{\mu B} v^B(x) \right) + O(h^2) = h^\mu \left( D_\mu v^A \right) + O(h^2). \]  
(2.30)

Typically (Wald, 1984; Misner et al., 1973), the covariant derivative is defined with the opposite sign of \( \omega \). The choice is a matter of convention; the “-” sign in the covariant derivative is more convenient for our purposes. In the standard convention, the covariant derivative of a vector, \( v^A \), is

\[ D^\text{standard}_\mu v^A = \partial_\mu v^A + \omega^A_{\mu B} v^B; \]  
(2.31)
if $v^A$ has constant coefficients, this reduces to

$$D^\text{standard}_\mu v^A = \omega^A_{\mu B} v^B.$$  \hspace{1cm} (2.32)

In our convention, the covariant derivative of a vector with constant coefficients is

$$D_\mu v^A = -\omega^A_{\mu B} v^B.$$ \hspace{1cm} (2.33)

In the standard development of GR, the “-” sign in our convention would be troublesome. However, in the standard convention, the infinitesimal Lorentz transformation between Minkowski spaces separated by the small coordinate increment $h^\mu$ is (compare to equation (2.24))

$$\Lambda_{\text{standard}}(x \to x + h)^A_B = \delta^A_B - h^\mu \omega^A_{\mu B} + O(h^2).$$ \hspace{1cm} (2.34)

In our development of GR, this transformation will play a large role, so the “-” sign in the standard convention is troublesome. Accordingly, we adopt the definition of the covariant derivative in equation (2.27).

The covariant derivative of a Minkowski object in a particular direction computes the difference between the total rate of change of the object in that direction with the rate of change expected due to the changing Minkowski frames in that direction. We need the spin connection to define the covariant derivative because the spin connection tells us how to relate neighboring frames.

The covariant derivative of a vector, $v^A$, should produce an object that transforms as a vector. For a Lorentz transformation parametrized by the field $\epsilon^A_B(x)$, we must have

$$\delta_\epsilon \left( D_\mu v^A \right) = \epsilon^A_B D_\mu v^B.$$ \hspace{1cm} (2.35)

But, $D_\mu v^A$ is composed out of pieces which can transform under Lorentz transformations, too: both the vector $v^A$ and $\omega^A_{\mu B}$ can transform. (Note that $\partial_\mu$ is unaffected by Lorentz transformations.)
It should not matter whether we transform $D_\mu v^A$ as a compound object, or whether we transform its parts first, and then form the compound object. Thus, we require that

$$\delta_\epsilon \left( D_\mu v^A \right) = - \left( \delta_\epsilon \omega^A_{\mu B} \right) v^B + D_\mu \left( \delta_\epsilon v^A \right). \tag{2.36}$$

This gives us a condition which $\delta_\epsilon \omega^A_{\mu B}$ must satisfy. Writing it out, we see that

$$\epsilon^A_B \left( \partial_\mu v^B - \omega^B_\mu C^C v^C \right) = - \left( \delta_\epsilon \omega^A_{\mu B} \right) v^B + \partial_\mu \left( \epsilon^A_B v^B \right) - \omega^A_{\mu B} \left( \epsilon^B_C v^C \right). \tag{2.37}$$

Expanding the derivative on the right hand side, we see that the $\partial_\mu v^B$ terms cancel. Since $v^A$ is arbitrary, we have

$$\delta_\epsilon \omega^A_{\mu B} = \partial_\mu \epsilon^A_B + \epsilon^A_C \omega^C_\mu B - \omega^A_\mu C \epsilon^C_B. \tag{2.38}$$

The last two terms are exactly the transformation rule we would expect with the index structure of $\omega$, but the first is unusual. The first term compensates for the derivative of $\epsilon$ which appears in $D_\mu \left( \delta_\epsilon v^A \right)$ in equation (2.36) (recall that $\epsilon$ can be an arbitrary smooth function on the event manifold). By combining two objects, $\partial_\mu$ and $\omega^A_{\mu B}$, neither of which is covariant under Lorentz transformations, we can form the covariant operator $D_\mu$.

The derivative appearing in Equation (2.38) makes the transformation rule for $\omega$ inhomogeneous. Even if $\omega$ vanishes in one frame, we can make a transformation to a new frame in which it does not vanish. Tensors do not have this behavior; a tensor which vanishes in one frame vanishes in all frames. Therefore, we conclude that $\omega$ does not live in any of the Minkowski spaces of spacetime—that is, $\omega^A_{\mu B}$ is not a tensor, even though it does carry Minkowski indices. This is to be expected: since $\omega$ relates neighboring Minkowski spaces and these spaces transform independently under local Lorentz transformations, we would expect that the transformation of $\omega$ at a point $p$ would depend on the behavior of $\epsilon$ in the neighborhood of $p$. For a proper tensor, the transformation must depend only on the value of $\epsilon$ at $p$.\(^2\)

\(^2\)Note that, for constant Lorentz transformations, such that $\partial_\mu \epsilon^A_{B} = 0$, $\omega$ transforms as a tensor.
We can use the gauge-covariant derivative to define a “gauge-covariant exterior derivative” on event-manifold forms which carry Minkowski indices:

\[
\mathcal{D} p^A = dp^A - \omega^A{}_B \wedge p^B,
\]

or with the coordinate indices written explicitly:

\[
(\mathcal{D} p^A)_{\nu\mu_1...\mu_n} = (\mathcal{D} p^A)_{[\nu\mu_1...\mu_n]} = (n + 1) \left[ \partial_{[\nu} p^A_{\mu_1...\mu_n]} - \omega^A_{[B} p^B_{\mu_1...\mu_n]} \right],
\]

where

\[
p^A_{\mu_1...\mu_n} = p^A_{[\mu_1...\mu_n]}
\]

is an \(n\)-form.

If an object is of mixed (i.e. Minkowski and event-manifold) indices, and is not a form, we do not (yet) have a covariant derivative operator for it. (In section 2.2.2 we will use the inverse tetrad to define the usual covariant derivative operator on tensors.) Fortunately, we will not need such objects to write our actions.

**Curvature and Gauge-Parallel Transport**

The connection is not covariant, but there is a covariant object we can form out of the connection, called the gauge curvature. The curvature two-form, \(R_{\mu\nu} = R_{[\mu\nu]}\), is an operator defined by

\[
R_{\mu\nu} \equiv - (D_{\mu} D_{\nu} - D_{\nu} D_{\mu}).
\]

It is common to specialize to the case where this operator acts on Minkowski vectors; in this case the curvature operator can be represented by a mixed-tensor, \(R_{\mu\nu B}^A\):

\[
(R_{\mu\nu} v)^A = R_{\mu\nu B}^A v^B,
\]

where \(R_{AB} = R_{[AB]}\) (i.e. \(R\) is an infinitesimal Lorentz transformation on vectors). If we know how a Lorentz transformation acts on vectors, we can derive its action on any other Minkowski tensors. For example, because one-forms are dual to vectors, the
corresponding Lorentz transformations must be inverses of each other; for infinitesimal transformations, this implies that
\[
(R_{\mu\nu}\tau)_A = -R^B_{\mu\nu}A^T_B, \tag{2.44}
\]
for an arbitrary Minkowski one-form \(\tau\).

Expanding out the commutator of covariant derivatives acting on a vector, we see that
\[
R^A_{\mu\nu} = (d\omega^A_B - \omega^A_C \wedge \omega_D^C)_{\mu\nu}. \tag{2.45}
\]
We can re-write this in the adjoint basis of \(SO(3,1)\) using
\[
\omega^{AB}_\mu = \omega^{aA}_{\mu b}t^B_a. \tag{2.46}
\]
We have
\[
R^a_{\mu\nu} = \left(d\omega^a_b - \frac{1}{2}f^a_{bc}\omega^b \wedge \omega^c\right)_{\mu\nu}. \tag{2.47}
\]

The physical interpretation of the gauge curvature involves the non-commutativity of parallel transport. An object with Lorentz indices \(v\), is parallel-transported along a curve, \(x^\mu = \gamma^\mu(\tau)\), with parameter \(\tau\) if the rate of change of the components \(v^A\) along the curve is exactly compensated by the connection between the Minkowski spaces at neighboring points on the curve:
\[
\frac{D}{D\tau}v^A(\tau) = 0. \tag{2.48}
\]

Fix a point, \(p\), and consider a loop, \(\gamma(\tau) : [0,1] \to M\), with \(\gamma(0) = \gamma(1) = p\). A Minkowski vector, \(v^A\), is parallel transported around this loop if
\[
\frac{D}{D\tau}v^A(\tau) = 0. \tag{2.49}
\]

\(^3\)Though it is common for such an object to be a field, defined on some region of spacetime, the following definition of \(D/D\tau\) only requires that \(v\) is defined on the curve itself.
This ODE has the formal solution
\[ v^A(1) = \mathcal{P} \exp \left[ \int_0^1 d\tau \omega^A_{\mu B}(\gamma(\tau)) \frac{d\gamma^\mu}{d\tau} \right] v^B(0), \] (2.50)
where the “path-ordering” operator, \( \mathcal{P} \) means, roughly, that \( \omega \) matrices with larger arguments (i.e. further around the curve) sit to the left of those with smaller arguments:
\[
\mathcal{P} \exp \left[ \int_0^1 d\tau \omega^A_{\mu B}(\gamma(\tau)) \frac{d\gamma^\mu}{d\tau}(\tau) \right] \equiv \delta^A_B + \int_0^1 d\tau \omega^A_{\mu B}(\gamma(\tau)) \frac{d\gamma^\mu}{d\tau}(\tau) \\
+ \int_0^1 d\tau \int_0^\tau d\tau' \omega^A_{\mu C}(\gamma(\tau))\omega^C_{\nu B}(\gamma(\tau')) \frac{d\gamma^\mu}{d\tau}(\tau) \frac{d\gamma^\nu}{d\tau'}(\tau') + \ldots \tag{2.51}
\]
Note that the \( n! \) ways to order the connection matrices so that larger \( \tau \) sit to the left of smaller \( \tau \) for integrated domain \([0,1] \times [0,1] \times \ldots \) exactly cancel the \( 1/n! \) coefficients in the exponential expansion, so we restrict the integration domain to \( \tau \in [0,1], \tau' \in [0,\tau], \tau'' \in [0,\tau'], \ldots \), and drop the \( 1/n! \) terms.

To explore the connection between the curvature two-form and the path-ordered exponential, \( \mathcal{P} \exp \left( \int_0^1 d\tau \omega^A_{\mu B} \frac{d\gamma^\mu}{d\tau} \right) \), fix a coordinate system around point \( p \). Consider the loop which first goes in coordinate direction \( \mu \) for a (coordinate) distance \( s \), and then in coordinate direction \( \nu \) for a (coordinate) distance \( t \), then along \(-\mu\) for a distance \( s \), and back to \( p \) in the \(-\nu\) direction for distance \( t \). (For this discussion only, the subscripts \( \mu \) and \( \nu \) will refer to these fixed directions, which will make the notation a bit odd-looking.) Parametrize the loop so that \( \tau \) increases linearly with the \( \mu \) and \( \nu \) coordinates around the loop; the four “straight” sides then lie in the ranges \( \tau \in [0,1/4], \tau \in [1/4,1/2], \tau \in [1/2,3/4], \text{ and } \tau \in [3/4,1]. \) (See Figure 2-1.)

We will use the quadrature formula
\[
\int_a^b dx \, f(x) = (b-a) f \left( \frac{a+b}{2} \right) + \mathcal{O} \left( (b-a)^3 \right) \tag{2.52}
\]
to evaluate the integrals in Equation (2.51) for this path to second-order accuracy in

\[^{42} \text{The path-ordered exponential plays a crucial role in quantum field theories. See, e.g., Weinberg (1995, 2005).} \]
Figure 2-1: The loop around which we are computing $P \exp \left( \int_0^1 d\tau \omega^{\mu A}_{\mu B} \frac{d\gamma^\nu}{d\tau} \right)$

$s$ and $t$. On the horizontal sides, we have $d\gamma^\mu / d\tau = \pm 4s$, and on the vertical sides, we have $d\gamma^\nu / d\tau = \pm 4t$. Because of the piecewise nature of the loop, we will use

$$\int_0^1 d\tau = \int_0^{1/4} d\tau + \int_{1/4}^{1/2} d\tau + \int_{1/2}^{3/4} d\tau + \int_{3/4}^1 d\tau, \quad (2.53)$$

and

$$\int_0^1 d\tau \int_0^\tau d\tau' = \int_0^{1/4} d\tau \int_0^\tau d\tau'$$

$$+ \int_{1/4}^{1/2} d\tau \left( \int_0^{1/4} d\tau' + \int_{1/4}^\tau d\tau' \right), \quad (2.54)$$

$$+ \int_{1/2}^{3/4} \left( \int_0^{1/4} d\tau' + \int_{1/4}^{1/2} d\tau' + \int_{1/2}^\tau d\tau' \right), \quad (2.55)$$

$$+ \int_{3/4}^1 d\tau \left( \int_0^{1/4} d\tau' + \int_{1/4}^{1/2} d\tau' + \int_{1/2}^{3/4} d\tau' + \int_{3/4}^\tau d\tau' \right). \quad (2.56)$$
Evaluating the first integral of equation (2.51), we obtain

\[
\int_0^1 d\tau \omega^A_{\mu B} \frac{d\gamma^\mu}{d\tau} = s\omega^A_{\mu B}(\gamma(1/8)) + t\omega^A_{\nu B}(\gamma(3/8)) - s\omega^A_{\mu B}(\gamma(5/8)) - t\omega^A_{\nu B}(\gamma(7/8)).
\] (2.58)

We can re-write this using a Taylor series for \(\omega\) so that, for example,

\[
\omega^A_{\mu B}(\gamma(1/8)) = \omega^A_{\mu B} + \frac{s}{2} \partial_\mu \omega^A_{\mu B},
\] (2.59)

where we are writing \(\omega^A_{\mu B} \equiv \omega^A_{\mu B}(\gamma(0))\). Then we obtain

\[
\int_0^a d\tau \omega^A_{\mu B} \frac{d\gamma^\mu}{d\tau} = st \left( \partial_\mu \omega^A_{\nu B} - \partial_\nu \omega^A_{\mu B} \right).
\] (2.60)

When evaluating the second integral of equation (2.51), all values are already second-order due to the two factors of \(d\gamma/d\tau\). Therefore, we will replace all \(\omega^A_{\mu B}(\gamma(\tau))\) with \(\omega^A_{\mu B}\). Then, most terms cancel on the opposing sides of the loop. The remaining terms evaluate to

\[
\int_0^1 d\tau \int_0^\tau d\tau' \omega^A_{\mu C}(\gamma(\tau))\omega^C_{\nu B}(\gamma(\tau')) \frac{d\gamma^\mu}{d\tau}(\tau) \frac{d\gamma^\nu}{d\tau'}(\tau') = -st \left( \omega^A_{\mu C} \omega^C_{\nu B} - \omega^A_{\nu C} \omega^C_{\mu B} \right).
\] (2.61)

So, we see that

\[
P \exp \left[ \int_0^1 d\tau \omega^A_{\mu B}(\gamma(\tau)) \frac{d\gamma^\mu}{d\tau}(\tau) \right] = \delta^A_B + st \left( d\omega^A_B - \omega^A_C \wedge \omega^C_B \right)_{\mu\nu} + \ldots = \delta^A_B + stR^A_{\mu\nu} + \ldots
\] (2.62)

Since our choice of \(\mu\) and \(\nu\) directions was arbitrary, we conclude that, to lowest order, the change in a Minkowski vector when parallel-propagated around a loop is given by
the integral of the curvature two-form over the interior of the loop:

$$\delta v^A = \left( \iint R^A_B \right) v^B.$$  \hspace{1cm} (2.63)

This is the physical interpretation of the gauge curvature. After we define the tetrad in Section 2.2.2, we will discuss how the gauge curvature is related to the usual Riemann curvature tensor.

The curvature satisfies the Bianchi identity, as can be verified by direct substitution:

$$\mathcal{D} R^A_B = dR^A_B + \omega^A_{A'} \wedge R^{A'}_B - \omega^B_{B'} \wedge R^A_{B'} = 0.$$  \hspace{1cm} (2.64)

The Bianchi identity will play a role in diffeomorphism invariance.

### 2.2.2 The Tetrad

The Minkowski frame attached to each event in the event manifold represents the possible orthonormal frames of freely-falling observers at that event. We have seen that the connection describes how these frames “twist” as we move from event to event. The other dynamical field in our theory, the tetrad, describes how to map manifold vectors (measured in an arbitrary coordinate system) into the Minkowski vector which a freely-falling observer would measure against her orthonormal axes.

The tetrad is an event-manifold one-form and a Minkowski vector:

$$e^A_\mu.$$  \hspace{1cm} (2.65)

Given an event-manifold vector, $v^\mu$, we can obtain the components of the corresponding Minkowski vector using

$$v^A = e^A_\mu v^\mu.$$  \hspace{1cm} (2.66)

Since we have a metric on the Minkowski spaces, the tetrad induces a metric on the event manifold via

$$g_{\mu\nu} \equiv \eta_{AB} e^A_\mu e^B_\nu.$$  \hspace{1cm} (2.67)
For this reason, the tetrad is sometimes called the “square-root” of the metric. For our purposes, however, the tetrad is the more fundamental field: the metric is the “square” of the tetrad.

The Inverse Tetrad, Riemann Curvature, and Affine Connection

Though we will not need it for our formulation, it is useful to consider the inverse tetrad because it allows us to make a connection with the standard formulation of GR. Define an object, $e^\mu_A$, called the “inverse tetrad”, such that

$$e^B_\mu e^\mu_A = \delta^B_A$$

and

$$e^\mu_A e^A_\nu = \delta^\mu_\nu$$

in some coordinate system ($e^\mu_A$ is called the inverse tetrad because it is the matrix inverse of $e^A_\mu$). Here we are assuming that the tetrad is not degenerate. For a discussion of degenerate tetrads and their relevance in classical and quantum gravity, see [Horowitz (1991)]. We can use the inverse tetrad to transform Minkowski vectors into vectors in the event manifold tangent space, just as we can use the tetrad to transform event manifold vectors into Minkowski vectors:

$$v^\mu = e^\mu_A v^A.$$  

(2.70)

We can use the inverse tetrad to define the usual covariant derivative in the event manifold in the following manner. We already have a covariant derivative for Minkowski vectors. Combining it with the tetrad maps, we can write

$$\nabla_\mu v^\nu = e^\nu_A D_\mu (e^A_\rho v^\rho),$$

(2.71)

where $\nabla_\mu$ is the usual covariant derivative. Using this covariant derivative to define
the affine connection\footnote{The affine connection is usually defined using
\[ \nabla_\mu v^\nu = \partial_\mu v^\nu - \Gamma^\nu_{\mu\rho} v^\rho, \] (2.72)
but we are using here the opposite sign for \( \Gamma \) to be consistent with our definition of the spin connection.} as
\[ \nabla_\mu v^\nu = \partial_\mu v^\nu - \Gamma^\nu_{\mu\rho} v^\rho, \] (2.73)
we see that
\[ \Gamma^\nu_{\mu\rho} = e^\nu_A \omega^\rho_{\mu B} e^B_\rho - e^\nu_A \partial_\mu e^A_\rho. \] (2.74)

The definition of the Riemann curvature tensor states that parallel transporting a vector \( v^\mu \) around a parallelogram whose sides have tangents \( u^\mu \) and \( w^\mu \) results in a change in \( v^\mu \) of
\[ \delta v^\mu = u^\nu w^\rho R^\mu_{\nu\rho} v^\sigma \] (2.75)
to lowest order in the area of the parallelogram, where \( R^\mu_{\nu\rho} v^\sigma \) is the Riemann curvature tensor. We can use the tetrad and inverse tetrad to map the gauge curvature (which computes the same change in the parallel-transported vector in Minkowski space as the Riemann curvature tensor does in coordinate space) into coordinate space to compute the Riemann curvature tensor in terms of the gauge curvature:
\[ R^\mu_{\nu\rho} v^\sigma = R^A_{\nu\rho B} e^\mu_A e^B_\sigma. \] (2.76)

### 2.3 The Action

We will derive the field equations governing gravitational systems from an action principle: when the fields \( e \) and \( \omega \) are in a physically acceptable configuration, small variations (which vanish on the boundary of the event manifold) result in a second-order variation in a functional of these fields, called the action. If we choose an action functional whose value is invariant under the symmetry operations described in Section 2.1.1 then physical configurations of the field variables will be gauge-covariant: any gauge transformations of a solution to the field equations will themselves be solu-
tions to the field equations because such a transformation does not change the value
of the action.

The actions we consider are often classified as “first-order” because they do not
involve second-derivatives of the dynamical fields. The alternative “second-order” ac-
tions (which include the familiar Einstein-Hilbert action) depend only on the tetrad.
The connection which determines the curvature is the fixed function of the tetrad
given in equation \((2.82)\) (sometimes called the Cartan structure equation), and therefore
the curvature depends on the second derivative of the tetrad. It is a unique
feature of the standard GR action that allowing the connection to become an inde-
pendent dynamical field reproduces the structure equation as a dynamical equation
of the first-order theory. So-called \(f(R)\) extensions to gravity do not have this prop-
erty; see, for example, [Sotiriou & Liberati (2007)] and [Sotiriou (2009)]. For a complete
discussion of both types of action and the corresponding Hamiltonian formulations,
refer to [Peldan (1994)].

We begin with the well-known Palatini\(^6\) action for gravity, which depends on the
metric \(g_{\mu \nu} = \eta_{AB} e^A_\mu e^B_\nu\) and affine connection (see equation \((2.73)\)):

\[
S(g, \Gamma) = \int d^4x \sqrt{-g} R(g, \Gamma) = \int d^4x \sqrt{-g} g^{\nu\sigma} R^\mu_{\mu\nu\sigma} (\Gamma) \tag{2.77}
\]

Re-writing this expression in terms of the tetrad and spin connection\(^7\) gives

\[
S(e, \omega) = \int d^4x \det(e) e^\mu_A e^\nu_B R^{AB}_{\mu\nu} (\omega). \tag{2.78}
\]

Applying the identity

\[
\det(e) e^\mu_A e^\nu_B = \frac{1}{2} \epsilon_{ABCD} e^C_\mu e^D_\nu, \tag{2.79}
\]

and dropping constant factors that are unimportant in vacuum gravity, we arrive at

\(^6\)The Palatini action takes the same form as the even-better-known Einstein-Hilbert action, but
the connection is an independent dynamical field in the Palatini formalism while it is a fixed function
of the tetrad field in the Einstein-Hilbert action.

\(^7\)Note equations \((2.73)\) and \((2.76)\). The curvature operator computes the infinitesimal trans-
formation resulting from parallel transport around a loop in both Minkowski space and coordinate
space.
the first-order Hilbert-Palatini action for gravity

\[ S(e, \omega) = \frac{1}{2} \int \epsilon_{ABCD} e^A \wedge e^B \wedge R^{CD}(\omega). \]  

(2.80)

The resulting Euler-Lagrange equations are

\[ \frac{\delta S}{\delta e^A} = 0 \implies \epsilon_{ABCD} e^B \wedge R^{CD} = 0 \]  

(2.81)

and

\[ \frac{\delta S}{\delta \omega^{CD}} = 0 \implies \epsilon_{ABCD} e^A \wedge \mathcal{D} e^B = 0 \implies \mathcal{D} e^B = 0. \]  

(2.82)

The first of these can be re-written:

\[ \epsilon_{ABCD} e^B \wedge R^{CD} = 0 \implies R^A_{\mu} = 0, \]  

(2.83)

where \( R^A_{\mu} \) is the Ricci tensor: \( R^A_{\mu} = R^A_{\mu\nu} e^\nu_B \). This is the Einstein equation in vacuum. The second equation expresses the vanishing of the torsion two-form, \( \mathcal{D} e^A \), and is sometimes called the first Cartan structure equation.

Consider for a moment the Hamiltonian corresponding to this action. The only term with a derivative is the curvature term, which contains \( d\omega^{AB} \). The exterior derivative is anti-symmetric in its spacetime indices, so the only time-derivatives present act on the spatial components of \( \omega \), so evidently the coordinate is \( \omega^{iAB}, i = 1, 2, 3 \). The corresponding momentum is

\[ \pi_{iAB}^i = \frac{1}{2} \epsilon_{ABCD} \epsilon^{ijk} e^D_j e^C_k. \]  

(2.84)

The coordinate has \( 6 \times 3 = 18 \) degrees of freedom per point, while the momentum has \( 4 \times 3 = 12 \) degrees of freedom per point. In the Hamiltonian formulation of this action, there are second-class constraints which restrict the extra degrees of freedom in the coordinate [Peldan, 1994]. These second-class constraints spoil the discrete theory based on a simple discretization of equation (2.80).

We can incorporate the necessary constraints into the action directly. The re-
sulting formulation is similar to the Plebański action. We allow the combination of
tetrads to become an independent field, which we will call the area field:

\[ b^{AB} \equiv e^A \wedge e^B. \]  
(2.85)

The area field can now serve as the momentum of the theory. In order to recover the
correct equations of motion, we need to add a constraint to the action to enforce the
definition in equation (2.85). See Peldan (1994) for more details.

We can re-write the Hilbert-Palatini action as

\[ S(e, \omega) = \int \epsilon_{ABCD} b^{AB}(e) \wedge R^{CD}(\omega). \]  
(2.86)

Promoting \( b \) to and independent field introduces \((6 \times 6) - (4 \times 4) = 20\) extra degrees
of freedom. Evidently, we need to impose 20 constraints which ensure that we can
always write \( b^{AB} = e^A \wedge e^B \) for some one-form field \( e^A \).

One set of constraints which works gives the Plebanski action for gravity (see Perez
(2004) for a discussion of this action in the context of spin foams in loop quantum
gravity):

\[ S(b, \omega, \phi, \mu) = \int \epsilon_{ABCD} b^{AB} \wedge R^{CD}(\omega) + \phi_{ABCD} b^{AB} \wedge b^{CD} + \mu \epsilon^{ABCD} \phi_{ABCD}, \]  
(2.87)

where the Lagrange multiplier fields \( \phi_{ABCD} \) and \( \mu \) are event-manifold scalar and four-
form, respectively. Because of the symmetries of \( b \) and \( \epsilon \), we can impose the following
symmetries on \( \phi \) without loss of generality:

\[ \phi_{ABCD} = \phi_{\{AB\}\{CD\}} = \phi_{CDAB}, \]  
(2.88)

so \( \phi \) has 21 algebraically independent components. The Euler-Lagrange equation for
\( \mu \),

\[ \phi_{ABCD} \epsilon^{ABCD} = 0, \]  
(2.89)
removes one further degree of freedom, so the Euler-Lagrange equation for $\phi$,

$$b^{AB} \wedge b^{CD} = -\mu \epsilon^{ABCD},$$  

(2.90)

imposes the necessary 20 constraints on $b_{\text{cd}}$. In terms of the tetrad, on shell, $-\mu \propto det(e)$, the volume element. The remaining two equations are

$$\epsilon_{ABCD} R^{CD} = -2\phi_{ABCD} b^{CD}$$  

(2.93)

$$\mathcal{D} b^{AB} = 0;$$  

(2.94)

the entire set is equivalent to the Hilbert-Palatini Euler-Lagrange equations (see Peldan (1994); Krasnov (2009)). We will discretize this action in Chapter 3.

Peldan (1994) shows how to eliminate the $b$ field (almost) completely to produce the CDJ action (Capovilla et al., 1991). It would be interesting to try to apply the discretization techniques in Chapter 3 to the CDJ action because it contains a Yang-Mills-like term $\text{Tr} (R_{\mu\nu} R^{\mu\nu})$ which makes the correspondence with lattice QCD (another Yang-Mills gauge theory) stronger. This, however, is beyond our scope here.

It is convenient to employ the adjoint representation of the Lorentz group and the two metrics $g$ and $\tilde{g}$ in equations (2.19) and (2.21) to re-write this action as

$$S(b, \omega, \phi, \mu) = \int \tilde{g}_{ab} b^a \wedge R^b(\omega) + g_{aa'} g_{bb'} \phi^{a'b'} b^a \wedge b^b + \mu \tilde{g}_{ab} \phi^{ab}. $$  

(2.95)

Now $\phi^{ab} = \phi^{(ab)}$ is a symmetric, traceless (with respect to $\tilde{g}$), $6 \times 6$ matrix, which has 20 algebraically independent components. This makes the number of degrees of

\footnote{The complete antisymmetry of $b^{AB} \wedge b^{CD}$ ensures that we can write it as an antisymmetric combination of tetrads. In fact, any of

$$b^{AB} = \pm e^A \wedge e^B$$

(2.91)

$$b^{AB} = \pm \epsilon^{AB}_{\ CD} e^C \wedge e^D$$

(2.92)

will solve equation (2.90). The former solution gives standard GR; the latter gives a theory with no local degrees of freedom (De Pietri & Freidel, 1999). See De Pietri & Freidel (1999) for a discussion of the interference effects of multiple solutions to these equations in quantum gravity spin-foam models. Fortunately for us, we are dealing with classical systems, and the two solutions are not continuously connected. An initial condition corresponding to the former solution will maintain that correspondence under time evolution.}
freedom in the Minkowski sector of the theory more apparent. The equations become

\[ 0 = \mathcal{D} b^a = dB^a - f^a_{bc} \omega^b \wedge b^c \quad (2.96) \]
\[ 0 = \tilde{g}_{ab} R^b + 2\phi_{ab} b^b \quad (2.97) \]
\[ 0 = \tilde{g}_{ab} \phi^{ab} \quad (2.98) \]
\[ 0 = b^a \wedge b^b + \mu \bar{g}^{ab} \quad (2.99) \]

where \( \phi_{ab} = g_{aa'} \phi^{a'b'} g_{b'b} \).
Chapter 3

Discretization of Gravity

We will discretize gravity at the level of the action. We will derive equations of motion for the discrete theory directly from the discretized action by demanding that the discrete action is stationary with respect to variations of the discrete fields. This chapter works step-by-step to show how to discretize each of the structures described in Chapter 2 which make up our theories of gravity. Reisenberger (1997) presents a similar discrete theory for left-handed Euclidean GR; this work is a generalization to full Lorentzian GR.

The formulation of gravity presented in the last chapter uses only the exterior derivative and wedge operator. These operators are independent of the metric, yet geometric, unlike in the standard formulation of GR where the partial derivative operator used to define the theory is a coordinate-dependent object. Additionally, the dynamical fields are forms, which can be integrated on regions of the appropriate dimension in a coordinate-invariant way. The improved geometric properties of the objects we use to formulate our theory will allow us to construct a diffeomorphism- and Lorentz-invariant discrete theory on a mesh in this chapter.

3.1 Discrete Manifolds

The first component of our continuous theory was spacetime (Section 2.1). Discrete spacetime is a discrete manifold; this section discusses what we mean by “discrete
Desbrun et al. (2005) contains a very thorough discussion of calculus on discrete manifolds (containing more material than we will need in this work). For a simpler discussion, see [Desbrun et al. (2008)].

3.1.1 Simplexes

We will construct a discrete manifold out of simplexes. A \( n \)-simplex is an ordered collection of \( n + 1 \) points:

\[
s_n \equiv \{p_0, p_1, \ldots, p_n\}.
\]

(3.1)

Typically, we write a simplex with a subscript indicating its dimension. This combinatorial definition of a simplex is very abstract. It can be helpful to think of an \( n \)-simplex as representing the \( n \)-dimensional region—the convex hull—enclosed by its \( n + 1 \) points. If we think of the points as vectors in \( \mathbb{R}^n \), then the region is given by

\[
\left\{ \vec{x} \mid \vec{x} = \sum_i c_i \vec{p}_i, \quad c_i > 0, \quad \sum_i c_i = 1 \right\}.
\]

(3.2)

For example, the simplex \( \{p_0\} \) is a point, the simplex \( \{p_0, p_1\} \) is the line which begins at \( p_0 \) and goes to \( p_1 \), and the simplex \( \{p_0, p_1, p_2\} \) is the interior of the triangle defined by the three points which includes the segment \( \{p_0, p_1\} \) as one of its three sides. This picture of a simplex as a convex region in \( \mathbb{R}^n \) is intuitive, but not necessary for any of the results that follow. Because the theories we are discretizing are coordinate invariant, we are free to label the points of a simplex however we like. Sticking to the combinatorial definition is a useful “check” that we are not artificially introducing a particular coordinate system.

For every simplex, \( s_n = \{p_0, p_1, \ldots, p_n\} \), there are two equivalence classes of simplexes which contain the points \( p_0, \ldots, p_n \) in orderings which differ from \( \{p_0, p_1, \ldots, p_n\} \) by, respectively, an even or odd number of interchanges. These equivalence classes consist of simplexes with, respectively, the same and the opposite orientations as \( s_n \). We denote a change in orientation with a minus sign. For example, a representative member of the opposite-orientation class is \( -s_n \equiv \{p_1, p_0, \ldots, p_n\} \).
We can form formal linear combinations of simplexes, called “chains”. The space of chains is a vector space. Each simplex is a basis vector in the space. An example of a chain, given some simplexes $s_n^{(1)}$, $s_n^{(2)}$, $s_n^{(3)}$, ..., $s_n^{(i)}$, ..., would be

$$c = \sum_i c_is_n^{(i)},$$

(3.3)

where the coefficients, $c_i \in \mathbb{Z}$. We can add and subtract chains in the obvious way:

$$c + d = \sum_i c_is_n^{(i)} + \sum_i d_is_n^{(i)} = \sum_i (c_i + d_i)s_n^{(i)}.$$  

(3.4)

The sum of a simplex and one of its opposite-orientation simplexes is “0”, the empty chain:

$$s_n + (-s_n) = 0.$$  

(3.5)

The boundary of a simplex, $s_n = \{p_0, p_1, \ldots, p_n\}$, denoted by $\partial s_n$, is the chain

$$\partial s \equiv \sum_{i=0}^n (-1)^i \{p_0, \ldots, |p_i|, \ldots, p_n\},$$

(3.6)

where the notation $\{p_0, \ldots, |p_i|, \ldots, p_n\}$ refers to the ordered collection of points $\{p_0, \ldots, p_{i-1}, p_{i+1}, \ldots, p_n\}$, i.e. the collection with point $i$ omitted. $\partial s_n$ is $n - 1$-dimensional. For example (see Figure 3-1), the boundary of the triangle $\{p_0, p_1, p_2\}$ is a chain composed of three simplexes (struts):

$$\partial \{p_0, p_1, p_2\} = \{p_1, p_2\} - \{p_0, p_2\} + \{p_0, p_1\} = \{p_1, p_2\} + \{p_2, p_0\} + \{p_0, p_1\}.$$  

(3.7)

You can see how the sign convention above ensures the correct relative orientation of the boundary.

If we extend the operation of $\partial$ linearly to chains via

$$\partial c = \partial \sum_i c_is_n^{(i)} = \sum_i c_i\partial s_n^{(i)},$$

(3.8)

we see that the topological property that the boundary of a boundary is null holds
Figure 3-1: The boundary of the triangle \{p_0, p_1, p_2\} is given by the sum of the one-simplexes in equation (3.7) for simplexes:

\[ \partial \partial \{p_0, p_1, p_2\} = \{p_1\} - \{p_2\} + \{p_2\} - \{p_0\} + \{p_0\} - \{p_1\} = 0. \quad (3.9) \]

We say that a simplex, \(s_k\), is a (proper) face of another simplex, \(S_n\), when the points of \(s_k\) are a (proper) subset of the points of \(S_n\); we denote this by \(s_k \preceq S_n\) (\(s_k \preceq S_n\)). (The operators are pronounced “face of” and “proper face of”, respectively.) For example, one of the boundary struts of a triangle is a proper face of the triangle, \(\{p_0, p_1\} \prec \{p_0, p_1, p_2\}\), as are the points which make up the triangle, \(\{p_0\}, \{p_1\}, \{p_2\} \prec \{p_0, p_1, p_2\}\).

### 3.1.2 Simplicial Complexes and Dual Simplexes

A discrete manifold will be represented by a simplicial complex. A simplicial complex is a set of simplexes, where

1. The geometrical intersection of any two simplexes in the set, \(s_n^{(1)}\) and \(s_n^{(2)}\), is either null, or is an \(n - 1\)-dimensional face of \(s_n^{(1)}\) and \(s_n^{(2)}\).

2. Every face of each simplex in the complex is also in the complex.
Figure 3-2: A 2-dimensional simplicial complex with dual simplexes. The region enclosed by the black polygon is the dual of the point in the center of the primal complex, each of the sides of the polygon are dual to the corresponding strut of the primal complex, and each point in the polygon is the dual to the corresponding 2-simplex (triangle).

The boundary of a complex is the chain that results from summing the boundaries of its constituent simplexes.

In the context of a $N$-dimensional simplicial complex, each $n$-simplex has an $N - n$-dimensional dual simplex. Roughly speaking, the dual of an $n$-simplex spans the $N - n$ orthogonal dimensions near the simplex within the complex. Figure 3-2 illustrates this idea.

More formally, in an $N$-dimensional complex, the dual of a $n$-simplex, $s_n = \{p_0, \ldots, p_n\}$, denoted $\ast s_n$, is the region spanned by simplexes constructed out of the center-points of $s_n$ and those higher-dimensional simplexes of which it is a face:

$$\ast s_n \equiv \sum_{s_N \succ s_{N-1} \succ \cdots \succ s_n} \epsilon \{c(s_n), c(s_{n+1}), \ldots, c(s_N)\}, \quad (3.10)$$
Figure 3-3: A triangle \( \{a, b, c\} \) and its dual structures. The point \( \{y\} \) is dual to the triangle \( \{a, b, c\} \). The line \( \{z, y\} \) is dual to the strut \( \{b, c\} \) (note the orientation, chosen so that \( \{b, c\} \) and \( \{z, y\} \) have a relative orientation consistent with the triangle \( \{a, b, c\} \)). The surface spanned by \( \{y, z, c\} \), and \( \{c, x, y\} \) is dual to the point \( \{c\} \) (again, the orientation of the two triangles making up this surface is consistent with the orientation of the triangle \( \{a, b, c\} \)).

where \( c(s_i) \) is the center of the simplex \( s_i \), \( \epsilon \) is an orientation factor, and the sum extends over all higher-dimensional simplexes in the complex which satisfy the “face-of” requirements. Each of the pieces of this formula will be explained below. An example of the construction of duals is given in figure 3-3. We can think of the elements of the dual sum as simplexes where the “points” are the primal simplexes (i.e. representing the center of a simplex by the simplex itself).

The dual of a simplex is a union of simplexes whose points are formed from the centers of primal simplexes (c.f. equation (3.10) and figure 3-2). Thinking combinatorially, this is a sufficient definition of an abstract entity. In typical applications, however, the discrete manifold is a meshing of some region of \( \mathbb{R}^n \) or some other background smooth manifold whose metric is known. Then the operation \( c(s_n) \) refers to a suitably defined “average” of the points of \( s_n \) —perhaps the circumcenter of \( s_n \), or the barycenter of \( s_n \). The definitions that follow depend only on the orientability of the discrete manifold, and therefore will be consistent with any definition of “center” that respects this orientation, consistent with our combinatorial approach.
The orientation of the simplexes that comprise the dual cell can be computed from the orientations of \( s_n \) and \( s_N \). Suppose that
\[
s_n = \{p_0, \ldots, p_n\}. \tag{3.11}
\]
In equation (3.10), \( s_{n+1} \) must contain all the points of \( s_n \) plus one more, which we denote by \( p_{n+1} \). Similarly, \( s_{n+2} \) must contain all the points of \( s_{n+1} \) plus one more, which we denote \( p_{n+2} \), and so on. Consider the \( N \)-simplex
\[
\{p_0, \ldots, p_N\}, \tag{3.12}
\]
with the points defined in this fashion. If
\[
\{p_0, \ldots, p_N\} \sim s_N, \tag{3.13}
\]
where \( \sim \) means “has the same orientation as”, then \( \epsilon = 1 \) in equation (3.10); otherwise, \( \epsilon = -1 \). For example, consider computing the dual of \( \{b, c\} \) in \( \{a, b, c\} \) of Figure 3-3. The triangle \( \{a, b, c\} \) has the same orientation as the triangle \( \{b, c, a\} \), so the orientation of the dual strut is \( \{c(\{b, c\}), c(\{b, c, a\})\} \).

### Boundary of Dual Cells

Given the definition of a dual of a simplex in equation (3.10),
\[
\ast s_n \equiv \sum_{s_n \succ s_{n-1} \succ \cdots \succ s_0} \epsilon \{c(s_n), c(s_{n+1}), \ldots, c(s_N)\}, \tag{3.14}
\]
we could compute the boundary of a dual cell as the sum of the boundaries of its constituent simplexes. However, there is a simpler rule for the boundary of a dual cell:
\[
\partial (\ast s_n) \equiv \sum_{s_{n+1} \succ s_n} \ast s_{n+1}. \tag{3.15}
\]
For example, refer to figure \[3-2\]. Consider the point in the center of the primal mesh, \( s_0 \). Its dual is the polygonal region surrounding it. The boundary of this region is the sum of duals of all \( s_1 \)'s (struts) of which \( s_0 \) is a face.

### 3.1.3 Chains of Duals

We can define chains of dual cells in the same way that we defined chains of primal simplexes. Each dual cell is a basis vector in a vector space over \( \mathbb{Z} \).

### 3.1.4 Discrete Forms

A discrete form is a linear map from chains to a target space. A discrete dual form is a linear map from dual chains to a target space. (Eventually, we will define some forms which take values which are not in a linear space (e.g. \( SO(n, 1) \)), but for now we restrict our attention to linear maps.) The fundamental operation for forms is “pairing” a form with a simplex or cell to get a value in the target space. We denote the pairing of the discrete \( n \)-form \( \omega \) with a simplex \( s_n \) by

\[
\langle \omega, s_n \rangle.
\]

We denote the pairing of a form with a chain in the same way:

\[
\langle \omega, c \rangle = \sum_i c_i \langle \omega, s_n^{(i)} \rangle.
\]

Pairing can be thought of as the discrete analog of integration:

\[
\langle \omega, s_n \rangle \approx \int_{V_{s_n}} \omega,
\]

where \( V_{s_n} \) is the convex hull of the simplex \( s_n \). To discretize a continuous form on a simplicial complex, we integrate it over simplexes in the complex of appropriate dimension to produce a discrete form.
3.1.5 Discrete Exterior Derivative

Following the pairing-integration analogy, we define the discrete exterior derivative, \( d \), as the dual of the boundary operator:

\[
\langle d\omega, s_{n+1} \rangle \equiv \langle \omega, \partial s_{n+1} \rangle. \tag{3.19}
\]

In words: the value of \( d\omega \) on \( s_{n+1} \) is just the sum of the values of \( \omega \) on the boundary of \( s_{n+1} \). Stokes’ theorem holds by definition:

\[
\langle d\omega, s_{n+1} \rangle \approx \int_{V_{s_{n+1}}} d\omega \equiv \int_{\partial V_{s_{n+1}}} \omega \approx \int_{\partial V_{s_{n+1}}} \omega \tag{3.20}
\]

The statement that \( dd\omega \equiv 0 \) follows directly from the topological property that the boundary of a boundary is null.

For example, fix a discrete one-form \( \omega \), and a triangle \( s_2 = \{a, b, c\} \). Then \( \langle d\omega, s_2 \rangle \) is given by

\[
\langle d\omega, s_2 \rangle = \langle \omega, \{a, b\} \rangle + \langle \omega, \{b, c\} \rangle + \langle \omega, \{c, a\} \rangle. \tag{3.21}
\]

The sum of \( \omega \) around the boundary of \( s_2 \) is a “differencing” scheme for computing the derivative of \( \omega \) on \( s_2 \). If, for example, \( \omega \) were the discretization of a constant one-form, then its value on each of the struts of \( s_2 \) would be proportional to the tangent vector along the strut. Since the triangle \( s_2 \) closes, the sum around the boundary of the tangent vectors vanishes, and we would have \( \langle d\omega, s_2 \rangle = 0 \), as expected.

3.1.6 Discrete Wedge Product

We can define a discrete wedge product: given two forms, \( v \) and \( w \), of dimension \( n \) and \( m \), their wedge product is given by

\[
\langle v \wedge w, \{p_0, \ldots, p_{n+m}\} \rangle \equiv \frac{1}{(n + m + 1)!} \times \sum_{\sigma \in \text{perm}(n+m+1)} \text{sig}(\sigma) \langle v, \{p_{\sigma_0}, \ldots, p_{\sigma_n}\} \rangle \langle w, \{p_{\sigma_n}, \ldots, p_{\sigma_{n+m}}\} \rangle. \tag{3.22}
\]
The discrete wedge product of $v$ and $w$ on a simplex, $s$, consists of an average of products of $v$ and $w$ evaluated on all non-degenerate combinations of sub-simplexes (those sharing exactly one point) of $s$ with a sign to account for the orientations of the sub-simplexes relative to $s$. Defined in this way, the discrete wedge product is not, in general, associative. However, if the forms being wedged are closed—that is, if their exterior derivatives vanish—then the product is associative [Desbrun et al., 2005]. This is sufficient to recover associativity in the continuous limit: a continuous form discretized on a sufficiently small mesh is approximately constant on the mesh elements, and therefore has zero exterior derivative.

The discrete exterior derivative distributes through discrete wedge products exactly as in the continuous case [Desbrun et al., 2005]:

\[
\langle d(v \wedge w), s_n \rangle = \langle dv \wedge w, s_n \rangle + (-1)^{\dim(v)} \langle v \wedge dw, s_n \rangle.
\] (3.23)

### 3.2 Discretized Spacetime

We will formulate our discrete GR on a simplicial complex which represents the event manifold. At the center of each simplex in the complex we will place a Minkowski frame.

#### 3.2.1 Symmetries of Discrete Spacetime

We have the following symmetry operations in our theory:

**Diffeomorphism** A discrete diffeomorphism shifts the location of one point in the complex infinitesimally. This does not change the connectivity of the complex, but does alter the values of forms defined on the simplexes and dual cells incident to that point.

**Lorentz Transformation** A Lorentz transformation acts on each of the Minkowski spaces attached to centers of the simplexes of the discrete event manifold. It can be represented by a matrix-valued dual form, $\Lambda^A_B$, such that the transformation
of vectors, $v^A$, in the Minkowski space at the point at the center of the $k$-dimensional simplex, $c(s_k)$, is given by

$$v^A \rightarrow v'^A \equiv \langle \Lambda^A_B, c(s_k) \rangle v^B. \quad (3.24)$$

(A similar transformation equation holds for co-vectors and tensors in the Minkowski space at $p$.) Unlike the continuous case, where neighboring points are infinitesimally separated, points in the dual complex have finite separation, so we do not require that $\Lambda$-matrices transforming Minkowski spaces attached to neighboring points are in any way “close”.

### 3.3 The Dynamical Fields

#### 3.3.1 The Connection

Since the Minkowski frames at points in the dual complex are allowed to transform under arbitrary, finite transformations, the connection needs to be a full Lorentz matrix, rather than an infinitesimal Lorentz transformation. The discrete connection is a dual one-form field, living on dual struts, giving the Lorentz transformation between the frames at the centers of geometrically adjacent simplexes connected by the strut. The discrete connection, $\langle \Omega^A_B, \{ c(s_k), c(s_l) \} \rangle$, takes objects in the Minkowski frame at one $c(s_k)$ to its neighbor $c(s_l)$. We require that either $l = k - 1$ and $s_l \preceq s_k$ or $k = l - 1$ and $s_k \preceq s_l$ to ensure that the frames are geometrically adjacent. See figure 3.4 for an illustration of this in two dimensions.

It is convenient to define a vector-valued dual one-form, $\Omega^a$, such that

$$\langle \Omega^A_B, \{ c(s_k), c(s_l) \} \rangle = \exp \left( \langle \Omega^a, \{ c(s_k), c(s_l) \} \rangle t^A_a \right), \quad (3.25)$$

where the $t^A_a$, $a = 1, \ldots, n(n+1)/2$, are the generators of $SO(3,1)$. (We could choose any parametrization of the Lorentz group to take $\Omega^a \rightarrow \Omega^A_B$; we have arbitrarily
chosen the exponential map$^1$)

We can think of the discrete connection as the integral of the continuous connection over the dual strut $\{c(s_k), c(s_l)\}$:

$$\langle \Omega^A_B, \{c(s_k), c(s_l)\} \rangle = \mathcal{P} \exp \left( \int_{\{c(s_k), c(s_l)\}} \omega^A_B \right).$$  \hfill (3.26)

The connection is not a linear map on dual cells; instead, we have

$$\langle \Omega^A_B, -\{c(s_k), c(s_l)\} \rangle = \langle \Omega^{-1}_B, \{c(s_k), c(s_l)\} \rangle \langle \Omega^{-1}_B, -\{c(s_k), c(s_l)\} \rangle ;$$  \hfill (3.27)

since, if $\{c(s_k), c(s_l)\}$ runs from $c(s_k)$ to $c(s_l)$, $-\{c(s_k), c(s_l)\}$ runs from $c(s_l)$ to $c(s_k)$, and the corresponding transformations between spaces are inverses.

Because it maps objects in one Minkowski frame into objects in another frame, the connection transforms inhomogeneously under local Lorentz transformations:

$$\langle \Omega'^A_{B'}, \{c(s_k), c(s_l)\} \rangle \equiv \langle \Lambda^{A'}_{A}, \{c(s_l)\} \rangle \langle \Omega'^A_{B'}, \{c(s_k), c(s_l)\} \rangle \langle \Lambda^{-1}_{B'}B, \{c(s_k)\} \rangle ;$$  \hfill (3.28)

$^1$Here and almost always we ignore the failure of the exponential map to be one-to-one on all of $\text{SO}(n,1)$, and restrict ourselves to some (small) region about the origin.
that is, the $A$ index “lives” in the $c(s_l)$ space, while the $B$ index “lives” in the $c(s_k)$ space.

In the case that the strut $\{c(s_k), c(s_l)\}$ is short and the fields involved are smooth, the transformation rule goes over to the continuous rule, equation (2.38). Define

\begin{equation}
\langle \Omega^A_{\cdot B}, \{c(s_k), c(s_l)\} \rangle \equiv \delta^A_B + \omega^A_{\cdot B}
\end{equation}

(3.29)

\begin{equation}
\langle \Lambda^{A'}_{\cdot A}, \{c(s_l)\} \rangle \equiv \delta^{A'}_{A'} + \epsilon^{A'}_{\cdot A'}
\end{equation}

(3.30)

\begin{equation}
\langle \Omega^{A'}_{\cdot B'}, \{c(s_k), c(s_l)\} \rangle \equiv \delta^{A'}_{B'} + \omega^{A'}_{\cdot B'}
\end{equation}

(3.31)

\begin{equation}
\langle \Lambda^{-1B'}_{\cdot B}, \{c(s_k)\} \rangle \equiv \delta^{B'}_{B} - \epsilon^{B'}_{\cdot B}
\end{equation}

(3.32)

accurate to first order in the length of the strut. Then equation (3.28) implies

\begin{equation}
\omega^A_{\cdot B} = \omega^A_{\cdot B} + \epsilon^A_{\cdot A'} \omega^{A'}_{\cdot B'} - \epsilon^B_{\cdot B'} \omega^{B'}_{\cdot B} + \epsilon^{A'}_{\cdot B} - \epsilon^{A'}_{\cdot B'}
\end{equation}

(3.33)

We can simplify this by noting that $\epsilon^A_{\cdot B} - \epsilon^0_{\cdot B}$ is just $\langle d\epsilon, \{c(s_k), c(s_l)\} \rangle$. We can drop the subscript on the $\epsilon$s multiplying the $\omega$s because these terms are already first-order in the length of the strut, yielding

\begin{equation}
\delta \omega^A_{\cdot B} = d\epsilon^A_{\cdot B} + \epsilon^A_{\cdot A'} \omega^{A'}_{\cdot B'} - \epsilon^{B'}_{\cdot B} \omega^B_{\cdot B'}
\end{equation}

(3.34)

which is precisely the transformation rule for the continuous connection in equation (2.38).

**The Discrete Curvature**

The discrete curvature is a matrix-valued two-form, which we will write as $R^A_{\cdot B}$. We will always evaluate the discrete curvature in an four-dimensional discrete manifold on “pizza slices” resulting from the geometric intersection of the dual of a two-simplex
Figure 3-5: The structure of a “pizza slice,” $\star s_2 \cap s_4$, in 4 dimensions. It is the region enclosed by the struts $\{c(s_2), c(s_3)\}$, $\{c(s_3), c(s_4)\}$, $\{c(s_4), c(s'_3)\}$, and $\{c(s_3), c(s_2)\}$. The dotted lines show other pieces of $\star s_2$ which do not intersect $s_4$.

and a four-simplex, which we denote by $\star s_2 \cap s_4$ (refer to Figure 3-5):

$$\left< R^A_B, \star s_2 \cap s_4 \right> \equiv \left< \Omega^A_C, \{c(s_3), c(s_4)\} \right> \left< \Omega^C_D, \{c(s_2), c(s_3)\} \right>$$

$$\left< \Omega^D_E, \{c(s'_3), c(s_2)\} \right> \left< \Omega^E_B, \{c(s_4), c(s'_3)\} \right>$$

(3.35)

where the product proceeds in an order consistent with the relative orientations of $\star s_2$ and $s_4$. The curvature is the accumulated Lorentz transformation which results from parallel transport around the boundary of the pizza slice. Thus, when either $s_2$ or $s_4$ changes orientation, the orientation of the boundary struts reverses, and the curvature matrix inverts:

$$\left< R^A_B, -\star s_2 \cap s_4 \right> = \left< R^A_B, \star s_2 \cap -s_4 \right> = \left< R^{-1A}_B, \star s_2 \cap s_4 \right> .$$

(3.36)

The choice of initial point in the loop is arbitrary; we define the curvature loop to begin at $c(s_4)$. If we had chosen a different point the change in the curvature would correspond to a Lorentz transformation which, as we shall show later, is a gauge transformation in our theory. The curvature transforms under Lorentz transforma-
tions as
\[
\langle R^A_B, \star s_2 \cap s_4 \rangle = \langle \Lambda^A_{A'}, c(s_4) \rangle \left\langle R^{A'}_{B'}, \star s_2 \cap s_4 \langle \Lambda^{-1}_{B'} B, c(s_4) \right\rangle. \tag{3.37}
\]

The indices on the discrete curvature both live in the frame at \(c(s_4)\), so the discrete curvature transforms homogeneously under Lorentz transformations even though the connection on which it is based does not.

Based on the argument in Section 2.2.1 as the area, \(A\), of the pizza slice, \(\star s_2 \cap s_4\), goes to zero, we have
\[
\langle R^A_B, \star s_2 \cap s_4 \rangle = \delta^A_B + A R^A_B + \ldots \tag{3.38}
\]
where \(R^A_B\) on the right hand side of this equation denotes the continuous curvature evaluated at an arbitrary point on the slice.

We define the adjoint discrete curvature by
\[
\langle R^a, \star s_2 \cap s_4 \rangle t^A_B = \left\langle R^{[AB]}, \star s_2 \cap s_4 \right\rangle. \tag{3.39}
\]

The adjoint discrete curvature captures the anti-symmetric part of the discrete curvature. The discrete curvature is a full Lorentz transform, unlike the continuous curvature, which is an infinitesimal Lorentz transformation, and thus is not an anti-symmetric matrix. \(R^a\) changes sign when the orientation of the pizza slice changes\(^2\)

\[
\langle R^a, - \star s_2 \cap s_4 \rangle = - \langle R^a, \star s_2 \cap s_4 \rangle. \tag{3.41}
\]

When the curvature is small, \(R^A_B = \delta^A_B + R^{a_A}_{a_B} + \ldots \)

\(^2\)For any Lorentz transformation, \(\Lambda^A_B\), we have
\[
\Lambda^{[AB]} = -(\Lambda^{-1})^{[AB]} \tag{3.40}
\]
3.3.2 The Tetrad and Area Field

The continuous tetrad \((e^A)\) is a one-form, and the continuous area field \((b^{AB})\) is a two-form; the discrete versions of these fields should live on struts and areas, respectively. These fields also carry Minkowski indices, so we must specify which Lorentz frame the discretized fields live in. Finally, the discrete tetrad and area field should be dual to the discrete curvature in the same way as the continuous fields are dual to the continuous curvature.

The area field must contract with the curvature in the action (see equation (2.95)). The curvature lives in the Lorentz frame at the center of a four-simplex, \(s_4\), and the pizza slice on which it is evaluated is contained in \(s_4\). The corresponding area field is associated with a two-simplex face of \(s_4\), but its components are measured at the center of the four-simplex. To denote this, we write

\[
\langle B^{AB}, (s_2, s_4) \rangle = \langle B^{[AB]}, (s_2, s_4) \rangle ;
\]

for the area field (which is antisymmetric in its two Minkowski indices) corresponding to the two-simplex \(s_2\) in the frame at the center of \(s_4\). The sign of the \(B^{AB}\) matrix is sensitive to only the orientation of \(s_2\); \(s_4\) simply provides the frame for the Minkowski indices \(A\) and \(B\). For each face, \(s_2\), in the discrete manifold, there are many area fields—one for each of the volumes, \(s_4 \succ s_2\), incident on the face. Each field measures the same \(s_2\), but in a different frame. We will see that one of the equations of motion of the discrete theory is a “smoothness” condition on \(B\) relating the \(\langle B^{AB}, (s_2, s_4) \rangle\) for all \(s_4 \succ s_2\).

We can think of the discrete area field as the integral of the continuous area field over \(s_2\):

\[
\langle B^{AB}, (s_2, s_4) \rangle = \int_{s_2} \Lambda_{s_4A}^A \Lambda_{s_4B}^B b^{AB'} ,
\]

where

\[
\Lambda_{s_4B}^A(x) = \mathcal{P} \exp \left( \int_{[x,c(s_4)]} \omega_B^A \right) .
\]

\(\Lambda_{s_4B}^A(x)\) translates objects in the frame at \(x\) to the frame at \(c(s_4)\), ensuring the correct
Lorentz transformation properties of the discrete area field.

Because $B^{AB}$ is anti-symmetric, we can define an adjoint area field, $B^a$, by

$$\langle B^a, (s_2, s_4) \rangle t^{AB}_a = \langle B^{AB}, (s_2, s_4) \rangle.$$  \hspace{0.5cm} (3.45)

### 3.4 The Action

In this subsection we will discretize the 3+1 Plebanski action\textsuperscript{3}

By analogy with the continuous action, the discrete action is

$$S = \sum_{s_2} \sum_{s_4} \tilde{g}_{ab} \langle B^a, (s_2, s_4) \rangle \langle R^b, *s_2 \cap s_4 \rangle$$

$$+ \sum_{s_4} \langle B^a \wedge B^b, s_4 \rangle g_{aa'}g_{bb'} \langle \Phi^{a'b'}, *s_4 \rangle$$

$$+ \sum_{s_4} \tilde{g}_{ab} \langle \mu, s_4 \rangle \langle \Phi^{ab}, *s_4 \rangle.$$  \hspace{0.5cm} (3.46)

The action is invariant under Lorentz transformations at $c(s_2), c(s_3),$ and $c(s_4)$. The action is also invariant under shifts of manifold points—discrete diffeomorphisms—because it is expressed in terms of discrete fields that live on the simplexes. As we will show later, the action is, to lowest order, equal to the integral of the continuous action over the volumes, $s_4$. The continuous action is invariant under transformations of the fields corresponding to an infinitesimal diffeomorphism; therefore the discrete action is invariant to lowest order, too. The arguments in [Freidel & Louapre, 2003] demonstrate that the action is, in fact, invariant to all orders under a discrete diffeomorphism. These gauge symmetries will have consequences for the time-evolution equations which will be discussed in Chapter 4.

The discrete Einstein equation comes from varying the action with respect to

\textsuperscript{3}One can attempt to discretize the Einstein-Hilbert action in 3+1. The resulting discrete equations have extra unconstrained modes compared to the continuous equations. These modes appear to be related to the second-class constraints which appear in the Hamiltonian formulation of this theory [Peldan, 1994], discussed in Section 2.3. These second-class constraints correspond to the Lagrange multiplier terms in the Plebanski action. Making these terms explicit in the discrete action solves the problem of the extra modes.
\[ \langle B^a, (s_2, s_4) \rangle : \]
\[ \hat{g}_{ab} \langle R^b, \ast s_2 \cap s_4 \rangle + \frac{1}{15} \langle \Phi^{ab}, \ast s_4 \rangle g_{aa'} g_{bb'} \sum_{s'_2 \prec s_4} \text{sig} (s_2, s'_2, s_4) \langle B^{b'}, (s'_2, s_4) \rangle = 0, \quad (3.47) \]

where \( \text{sig} (s_2, s'_2, s_4) \) is 0 if \( s_2 \) and \( s'_2 \) share more than a single point (i.e. are degenerate), is 1 if \( s_2 \wedge s'_2 \) has the same orientation as \( s_4 \), and \(-1\) otherwise. This equation is analogous to the continuous Einstein equation. The sum of area fields over the \( s'_2 \) corresponds to an average of the area fields in the directions dual to the two-surface \( s_2 \). The curvature is evaluated over a loop running in the two directions dual to \( s_2 \) as well.

Comparing to the continuous Einstein equation arising from the Plebański action,
\[ \hat{g}_{ab} R^b (\omega) + 2 \phi_{ab} b^b = 0, \quad (3.48) \]
we see that the discrete Einstein equation for \( s_2 \) is, to lowest order, the continuous Einstein equation integrated over the two dimensions in \( s_4 \) dual to \( s_2 \).

The discrete torsion equation (the discrete version of the first Cartan structure equation, equation (2.96)) comes from varying \( \Omega \) on \( \{c(s_3), c(s_4)\} \). It is most convenient to consider variations of the form
\[ \langle \Omega^A_B, \{c(s_3), c(s_4)\} \rangle \rightarrow \epsilon^a t^A_{aC} \langle \Omega^C_B, \{c(s_3), c(s_4)\} \rangle ; \quad (3.49) \]
these parametrize all possible variations in \( \Omega^4 \). The corresponding Euler-Lagrange equation is
\[ \frac{\delta S}{\delta \epsilon^a} = \hat{g}_{bc} \sum_{s_2 \in \partial s_3} \langle B^b, (s_2, s_4) \rangle \langle (t_a R)^c, \ast s_2 \cap s_4 \rangle = 0, \quad (3.50) \]
where the notation \( (t_a R)^c \) is defined by
\[ t^A_{aC} R^{[C|B]} = (t_a R)^c t^{[AB]}_c, \quad (3.51) \]
i.e. project the matrix product \( t_a R \) onto the generators of \( SO(3,1) \). (The symmetric

\[ SO(3,1) \] is a Lie group, and the generators form a complete basis for the tangent vectors of the group manifold.)
part—if any—of $t_a R$ does not contribute to the action or the equation of motion.) The discrete torsion equation is a sum over all $s_2$s in the boundary of $s_3$ of the discrete area field. This corresponds to the continuous torsion equation $D b = 0$ (equation (2.96)), as we will now show.

Consider one of the four curvature loops involved in the discrete torsion equation, illustrated in Figure 3.6. Label the four connection terms appearing in the curvature product $\Omega^a_i$, $i = 1, \ldots, 4$ according to the indicated struts. The product of $t_a$ and $R$ is given by

$$ (t_a R)^c = \delta_c^a - \frac{1}{2} f_{ba}^c \left( \Omega_1^b + \Omega_2^b + \Omega_3^b + \Omega_4^b \right) + O (\Omega^2). \quad (3.52) $$

The terms with $\Omega_1$ and $\Omega_3$ both involve the direction in $s_3$ dual to $s_2$ because both involve the addition of $p_3$; the terms with $\Omega_2$ and $\Omega_4$ involve the addition of $p_4$, which is a direction common to all the loops involved in the equation (since the strut $\{c (\{p_0, p_1, p_2, p_3\}), c (\{p_0, p_1, p_2, p_3, p_4\})\}$ is common to all the loops). The common terms can be ignored—they cancel when averaged over the boundary of $s_3$. Summing over the boundary, the Kronecker delta term corresponds to $dB$. The terms with $\Omega$ in the directions dual to $s_2$ within $s_3$ correspond to $\Omega \wedge B$ on $s_3$. To lowest order it is sufficient to assume a constant connection over the volume; then the two terms $\Omega_1$ and $\Omega_3$ add, eliminating the factor of $1/2$, and the sum over the boundary becomes exactly what one would expect from discretizing

$$ D b^a = dB^a - f_{ba}^c \omega^b \wedge b^c \quad (3.53) $$

over the volume $s_3$.

A third equation results from varying $\Omega$ on $\{c (s_2), c (s_3)\}$. This equation has no continuous analog; it is an identity in the continuous limit. In a well-formed simplicial complex, with $s_3$ in the interior, there are two four-simplexes, $s_4^{(1)}$ and $s_4^{(2)}$, that have $s_3$ as a face. So, there are two curvature loops that have the dual strut $\{c (s_2), c (s_3)\}$ in them. This situation is illustrated in Figure 3.7.

---

5There are four two-simplex faces for each three-simplex.
Figure 3-6: The curvature loop for a term in the torsion equation (equation (3.50)). Here \( s_2 = \{p_0, p_1, p_2\} \), \( s_3 = \{p_0, p_1, p_2, p_3\} \), \( s_4 = \{p_0, p_1, p_2, p_3, p_4\} \), and \( s'_3 = \{p_0, p_1, p_2, p_4\} \). The loop is traversed in the indicated order: \( s_4 \rightarrow s'_3 \rightarrow s_2 \rightarrow s_3 \). The struts 1 and 3 lie primarily along the same dimension of \( s_4 \) because they involve the addition or removal of \( p_3 \); similarly, the struts 2 and 4 lie along the same dimension because they involve the addition or removal of \( p_4 \). Together these two dimensions span the dual directions to \( s_2 \) within \( s_4 \). Additionally, because \( p_3 \) belongs to \( s_3 \), the first of these dimensions spans the dual space to \( s_2 \) within \( s_3 \); this is the dimension on which \( \Omega \) should be evaluated in the discretized \( \Omega \wedge B \) on \( s_3 \).
Figure 3-7: The two curvature loops which involve the strut \( \{c(s_2), c(s_3)\} \).
If we again consider variations of the form

$$\langle \Omega^A_{\dot{B}}, \{c(s_2), c(s_3)\} \rangle \rightarrow \langle \Omega^A_{\dot{C}}, \{c(s_2), c(s_3)\} \rangle e^a t^C_{a\dot{B}}; \quad (3.54)$$

we obtain the equation

$$\frac{\delta S}{\delta \epsilon^a} = \tilde{g}_{ab} \left\langle B^a, \left( s_2, s_4^{(1)} \right) \right\rangle \left\langle \left( \tilde{R}_c \right)^b, *s_2 \cap s_4^{(1)} \right\rangle = \tilde{g}_{ab} \left\langle B^b, \left( s_2, s_4^{(2)} \right) \right\rangle \left\langle \left( \tilde{R}_c \right)^{b\prime}, *s_2 \cap s_4^{(2)} \right\rangle, \quad (3.55)$$

where

$$\tilde{R}_c = \Omega_{34} \Omega_{23} t_c \Omega_{32} \Omega_{43}, \quad (3.56)$$

(suppressing Minkowski indices for clarity) when $R$ is given by the matrix product around the curvature loop $s_4 \rightarrow s_3 \rightarrow s_2 \rightarrow s_3 \rightarrow s_4$:

$$R = \Omega_{34} \Omega_{23} \Omega_{32} \Omega_{43}, \quad (3.57)$$

and $\Omega_{ij} = \langle \Omega, \{c(s_i), c(s_j)\} \rangle$ (again suppressing the Minkowski indices). In the continuous limit, $R \rightarrow I$, and $\left( \tilde{R}_c \right)^b \rightarrow \delta_c^b$. Thus, in the continuous limit, (3.55) becomes

$$\tilde{g}_{ab} \left\langle B^a, \left( s_2, s_4^{(1)} \right) \right\rangle = \tilde{g}_{ab} \left\langle B^a, \left( s_2, s_4^{(2)} \right) \right\rangle, \quad (3.58)$$

which is to be expected. As the mesh size gets smaller, the frames at $s_4^{(1)}$ and $s_4^{(2)}$ coincide, and the area fields integrated on $s_2$ measured in each frame coincide.

Variation of equation (3.46) with respect to $\langle \mu, s_4 \rangle$ enforces the traceless condition on $\Phi^{ab}$:

$$\tilde{g}_{ab} \langle \Phi^{ab}, *s_4 \rangle = 0. \quad (3.59)$$

Varying with respect to $\Phi$ gives the constraint equations on the area field:

$$g_{aa'} g_{bb'} \left\langle B^{a'}, B^{b'}, s_4 \right\rangle + \tilde{g}_{ab} \langle \mu, s_4 \rangle = 0. \quad (3.60)$$

74
In the continuous case, this equation implied that we could write \( b^{AB} = e^A \wedge e^B \) because the product \( b^{AB} \wedge b^{CD} \) was completely antisymmetric. The discrete constraint is not quite so simple. It still implies that \( B^{AB} \wedge B^{CD} \) is completely anti-symmetric, but because the discrete wedge product is not in general associative (Desbrun et al., 2005),

\[
(E^A \wedge E^B) \wedge (E^C \wedge E^D) \neq E^A \wedge E^B \wedge E^C \wedge E^D,
\]

for a one-form \( E \). The discrete constraint, however, serves the same geometric purpose. The simplexes on which \( B \) is evaluated in each term of the sum (see equation (3.22)) computing \( B \wedge B \) share exactly one point. For each of the five points of a given \( s_4 \) which the two-simplexes on which \( B \) lives can share, there are \( \binom{4}{2} = 6 \) distinct pairings of two-simplexes that share this point. These correspond to the six independent two-dimensional sub-spaces anchored at that point. Equation (3.60) says that, averaged over these points, and over all sub-spaces at each point, the discrete area fields are antisymmetric.

In the special case that the discrete manifold is flat, then we have \( R^a = I \) everywhere, and the torsion equation implies that \( dB^a = 0 \). Writing \( B^{AB} = E^A \wedge E^B \) for one-forms \( E \) which live on struts in a particular \( s_4 \), we have \( dE = 0 \). Therefore the wedge product of \( E \) is associative, and

\[
B^{AB} \wedge B^{CD} = E^A \wedge E^B \wedge E^C \wedge E^D \propto \epsilon^{ABCD}
\]

provides a solution to the constraints. For more discussion of the discrete constraint in the context of quantum gravity models, see De Pietri & Freidel (1999); Perez (2003).

The complete set of discrete field equations is: the Einstein equation (equation (3.47)),

\[
\bar{g}_{ab} \left< R^b, *s_2 \cap s_4 \right> + \frac{1}{15} \left< \Phi^{ab}, *s_4 \right> g_{ad} g_{bd} \sum_{s_2' < s_4} \text{sig} \left( s_2, s_2', s_4 \right) \left< B^{b'}, (s_2', s_4) \right> = 0,
\]

75
the torsion equation (equation (3.50)),

$$\tilde{g}_{bc} \sum_{s_2 \in \partial s_3} \langle B^b, (s_2, s_4) \rangle \langle (t_a R)^c, \star s_2 \cap s_4 \rangle = 0, \quad (3.64)$$

the continuity equation (equation (3.55)),

$$\tilde{g}_{ab} \left\langle B^a, \left( s_2, s_4^{(1)} \right) \right\rangle \left\langle \left( \tilde{R}_c \right)^b, \star s_2 \cap s_4^{(1)} \right\rangle = $$

$$\tilde{g}_{ab} \left\langle B^b, \left( s_2, s_4^{(2)} \right) \right\rangle \left\langle \left( \tilde{R}_c \right)^b, \star s_2 \cap s_4^{(2)} \right\rangle, \quad (3.65)$$

the constraint equation (equation (3.60)),

$$g_{aa'} g_{bb'} \left\langle B^{a'}, B^{b'}, s_4 \right\rangle + \tilde{g}_{ab} \langle \mu, s_4 \rangle = 0, \quad (3.66)$$

and the tracelessness equation (equation (3.59)),

$$\tilde{g}_{ab} \langle \Phi^{ab}, \star s_4 \rangle = 0. \quad (3.67)$$

3.4.1 The “Geometric” Equations

The discrete constraint, torsion, and continuity equations (equations (3.60), (3.50), and (3.55)) have a geometric interpretation. Consider a single $s_4$. There are $6 \times 10 = 60$ degrees of freedom in its area fields. The discrete constraints eliminate 20 of those degrees of freedom. Of the remaining 40, 24 are eliminated by the four independent torsion equations on $s_4$. The remaining 16 degrees of freedom are appropriate for a simplex in 4 dimensions, which can be defined by an origin and four independent four-vectors emanating from that origin.

To be a bit more precise, consider $s_4 = \{p_0, p_1, p_2, p_3, p_4\}$. We can use the four torsion equations (equation (3.50)) to eliminate the four $B$ fields on two-simplexes which do not contain $p_0$. For such a simplex, $s_2 = \{p_i, p_j, p_k\}$, with $i, j, k > 0$, consider

---

6We will see in the next section that the gauge freedom to make Lorentz transformations at the center of $s_4$ implies that, of the five torsion equations for the five struts $\{c(s_3), c(s_4)\}$, only four are independent.
the torsion equation on the three-simplex \( s_3 = \{ p_0, p_i, p_j, p_k \} \). The torsion equation involves a sum over boundary two-simplexes of \( s_3 \). Only one of these boundary simplexes does not contain \( p_0 \) : \( s_2 \). If \( s_4 \) were flat, the torsion equation would state that \( s_3 \) “closes”: \( dB = 0 \). In the presence of curvature, this interpretation is spoiled (slightly) by the presence of the \( R \) terms in equation (3.50), but nevertheless it is possible to solve for \( \langle B^a, \{ p_i, p_j, p_k \} \rangle \) in terms of the area fields on two-simplexes containing \( p_0 \) and the connection.

Having eliminated all area forms on two-simplexes that do not contain the point \( p_0 \), we have \( 6 \times 6 = 36 \) remaining degrees of freedom in the area forms. The discrete constraint, equation (3.60), now takes the form

\[
\sum \left\langle \tilde{B}_a, \{ p_a, p_b, p_0 \} \right\rangle \left\langle \tilde{B}_b, \{ p_0, p_c, p_d \} \right\rangle + \tilde{g}_{ab} \left\langle \mu, s_4 \right\rangle = 0, \tag{3.68}
\]

where we have lowered the indices on the area forms using \( g_{ab} \), and \( \tilde{B} \) involves both the area form and curvature terms arising from using the torsion equation to eliminate area fields on simplexes without \( p_0 \). The solution to this equation takes the form

\[
\left\langle \tilde{B}^{AB}, \{ p_a, p_b, p_0 \} \right\rangle = 2 \left\langle \tilde{E}^A, \{ p_0, p_a \} \right\rangle \left\langle \tilde{E}^B, \{ p_0, p_b \} \right\rangle, \tag{3.69}
\]

for some four-vectors, \( \tilde{E} \), on the struts emanating from \( p_0 \). These four four-vectors describe the geometry of \( s_4 \) in terms of the four struts anchored at \( p_0 \).

The continuity equation (equation (3.55)) relates the area fields on all four-simplexes incident on a particular two-simplex. Any two adjacent four-simplexes share a three-simplex face. The continuity equation ensures that the area fields in each of the four-simplexes associated with two-simplexes on the shared face “match up” modulo curvature corrections, and therefore that the four-simplexes “fit together” properly to form a simplicial complex.
Chapter 4

Time Evolution

In Chapter 3, we assumed that a complete discrete spacetime was available to us for computing variations of the action. In this chapter, we will specialize to the case of 3+1 evolution, and will describe the methods we use to incrementally construct a complete spacetime given an initial 3-dimensional surface and fields defined on that surface.

4.1 Time-Advancement

Consider a three-dimensional simplicial complex. (Such a complex could, for example, be the boundary of a four-dimensional discrete spacetime.) The only fields we defined in Chapter 3 that live in a three-dimensional space are the connection fields on \( \{c(s_2), c(s_3)\} \) dual struts. These are the configuration variables of our formulation. (The area field is measured at the center of a four-simplex, the connection on \( \{c(s_3), c(s_4)\} \) dual struts similarly requires a four-simplex, the \( \Phi \) field lives at \( c(s_4) \), and the \( \mu \) field also lives on an \( s_4 \).)

What fields will play the role of momenta? Consider now a four-dimensional discrete spacetime. Separate it into two volumes using a spacelike three-dimensional surface. The action sum now splits into two terms:

\[
S = S_1 + S_2, \tag{4.1}
\]
where $S_1$ involves fields on the surface and one half of the spacetime, and $S_2$ involves fields on the surface and the other half of the spacetime. Action extremization requires that we have for fields, $\phi^\alpha$, in the surface

$$\frac{\partial S}{\partial \phi^\alpha} = 0 = \frac{\partial S_1}{\partial \phi^\alpha} + \frac{\partial S_2}{\partial \phi^\alpha}. \quad (4.2)$$

The surface separating the two volumes is part of the boundary of both volumes; it is well-known that variations of the action with respect to $\phi^\alpha$ on the boundary give the conjugate momenta, $\pi_\alpha$. If we choose an orientation for the surface consistent with being the boundary of region 1, it will be oppositely oriented relative to the boundary of region 2. Therefore, we have

$$\frac{\partial S}{\partial \phi^\alpha} = 0 = \pi_1^\alpha - \pi_2^\alpha. \quad (4.3)$$

That is, the stationary action condition is equivalent to the equality of the momenta computed with respect to the two volumes.

Consider a region of four-dimensional spacetime bounded by an initial Cauchy surface and a final Cauchy surface. Let $\phi^\alpha_i$ be the field variables that live on the initial Cauchy surface, $\psi^\alpha$ be the fields living in the interior, and $\phi^\alpha_f$ be the fields living on the final Cauchy surface. Similarly, write $\pi^\alpha_i$ for the momentum on the initial surface, and $\pi^\alpha_f$ for the momentum on the final surface. We can re-write the stationary action condition on this region as

$$\frac{\partial S}{\partial \phi^\alpha_i}(\phi_i, \psi, \phi_f) = -\pi^\alpha_i \quad (4.4)$$

$$\frac{\partial S}{\partial \psi^\alpha}(\phi_i, \psi, \phi_f) = 0 \quad (4.5)$$

$$\frac{\partial S}{\partial \phi^\alpha_f}(\phi_i, \psi, \phi_f) = \pi^\alpha_f. \quad (4.6)$$

Joining several such regions together, the fact that $\pi^\alpha_f$ from an earlier region is the $\pi^\alpha_i$ for the next, implies the stationary action condition, as in equation (4.3). To evolve the initial field configuration $(\phi^\alpha_i, \pi^\alpha_i)$ to the final configuration $(\phi^\alpha_f, \pi^\alpha_f)$, we first solve
equations (4.4) and (4.5) for $\psi$ and $\phi_f$, and then use equation (4.6) to find $\pi^f$ on the final boundary.

In our case, the fields $B$, $\Phi$, $\mu$, and $\Omega$ on $\{c(s_3), c(s_4)\}$ correspond to $\psi$ because they exist only in the four-dimensional complex (since they all involve four-simplexes). The field $\Omega$ on $\{c(s_2), c(s_3)\}$ corresponds to $\phi_\alpha^i$. The momentum on a strut $\{c(s_2), c(s_3)\}$ is given by (see equation (3.55))

$$\pi_c = \tilde{g}_{ab} \langle B^a, (s_2, s_4) \rangle \left\langle \left( \tilde{R}_c \right)^b, \ast s_2 \cap s_4 \right\rangle$$

(4.7)

where $s_4$ is one of the volumes in the mesh incident on $s_2$ and $s_3$. (The continuity equation, equation (3.55), ensures that all such momenta are equal for any $s_4$ incident on $s_2$, so it does not matter which volume we choose.) In the continuous limit, $\left( \tilde{R}_c \right)^b \to \delta_c^b$ at lowest order, so the momentum becomes

$$\pi_c \to \tilde{g}_{ab} b^a.$$  

(4.8)

Compare to the continuous situation. The configuration field is $\omega_i^a$, the spatial part of the connection. $\Omega$ on $\{c(s_2), c(s_3)\}$ corresponds exactly to the integral of the continuous connection along the direction dual to the two-dimensional surface $s_2$ in the three-dimensional volume $s_3$. The momentum, $\pi_c$, corresponds to the integral of the continuous momentum, $\tilde{g}_{ab} b^a$, integrated over a two-simplex, $s_2$.

4.2 Incrementally Advancing the Mesh

In this section, we describe how to take a three-dimensional mesh describing an initial Cauchy surface and advance it, point by point, forward in time. This procedure is exactly the one described in Barrett et al. (1997).

Choose a point, $\{p_0\}$, in the Cauchy surface. Consider all three-simplexes incident on $\{p_0\}$. Let these simplexes be of the form $\{p_0, q_i, q_j, q_k\}$. The struts $\{p_0, q_i\}$, $\{p_0, q_j\}$, and $\{p_0, q_k\}$ form a positively oriented basis for the three-dimensional space spanned by the three-simplex.
Choose a point, \(\{p_1\}\), which is intended to represent the advancement of \(\{p_0\}\). Construct a four-dimensional simplicial complex out of the simplexes \(\{p_0, p_1, q_i, q_j, q_k\}\) (note the orientation—chosen so that the timelike strut \(\{p_0, p_1\}\) and the three spacelike struts form a positively oriented basis for the four-dimensional volume spanned by the simplex). These four-simplexes are the three-simplexes incident on \(\{p_0\}\) from the Cauchy surface “dragged” up the tent-pole strut \(\{p_0, p_1\}\). This configuration is illustrated in 2 + 1 dimensions in Figure 4-1.

Once the mesh structure is specified, we can apply the discrete field equations to solve for the field variables in the new volume, using the procedure described in the last section. Obtaining the discrete connection and corresponding momentum on the new Cauchy surface, we can discard the old field values, and repeat the advancement procedure for a new point.
The mesh advancement procedure is carefully designed to preserve the primal and dual topology of the mesh. Since the connection and its momentum live on the struts of the dual mesh, it is important to preserve the topology of the Cauchy surface under advancement, or else the number of Euler-Lagrange equations will not match the number of variables, and the system will become singular.

The end result of a series of these advancements is a series of spacelike surfaces whose neighbors share all but a few volumes with each other. Barrett et al. (1997) describes it as a “puff pastry”: each point advanced corresponds to a bubble in the pastry.

It is possible to advance different points in the mesh in parallel, as long as the points do not share any adjacent volumes. This is a possibility for future work.

4.3 Gauge Symmetries and Evolution Equations

The action in equation (3.46) has a Lorentz symmetry at each \( c(s_2), c(s_3), \) and \( c(s_4), \) and a diffeomorphism symmetry at each point in the mesh. These symmetries have consequences for the Euler-Lagrange equations of the theory.

We begin abstractly. Suppose the discrete action has a symmetry operation, with parameters \( \epsilon^a \), which changes the values of the fields \( \phi_i, \psi, \) and \( \phi_f \). By definition, the action is unchanged as the parameter \( \epsilon^a \) changes, so

\[
\left. \frac{\partial S}{\partial \epsilon^a} \right|_{\epsilon=0} = 0 = \left[ \frac{\partial S}{\partial \phi_i^\alpha} \frac{\partial \phi_i^\alpha}{\partial \epsilon^a} + \frac{\partial S}{\partial \psi^\alpha} \frac{\partial \psi^\alpha}{\partial \epsilon^a} + \frac{\partial S}{\partial \phi_f^\alpha} \frac{\partial \phi_f^\alpha}{\partial \epsilon^a} \right]_{\epsilon=0}.
\] (4.9)

Evaluating this on solutions to the Euler-Lagrange equations, we have

\[
\pi_f^i \frac{\partial \phi_f^j}{\partial \epsilon^a} - \pi_i^\alpha \frac{\partial \phi_i^\alpha}{\partial \epsilon^a} = 0.
\] (4.10)

This is the discrete Nöther’s theorem: combinations of momenta along the symmetry directions of the action are conserved under evolution. If the symmetry operation does not affect the final Cauchy surface, then equation (4.10) states that the correspond-
ing initial momentum vanishes. We say that such symmetries generate “first-class constraints.” There will be a similar constraint in the final Cauchy surface for the corresponding transformation there.

For each symmetry operation, equation (4.9) also implies that a particular linear combination of the Euler-Lagrange equations vanishes (whether evaluated on a solution to the Euler-Lagrange equations or not). This means that the Euler-Lagrange equations are redundant\(^1\) and there is one degree of freedom for each symmetry that is unconstrained by the evolution equation. We must come up with another way to fix these degrees of freedom; such a procedure is called “gauge fixing.” The next subsections discuss gauge fixing the Lorentz transformation and diffeomorphism freedoms of our discrete GR.

### 4.3.1 Lorentz Gauge Symmetries

The Lorentz transformation freedoms apply at each \(c(s_2), c(s_3), \text{ and } c(s_4)\) in our mesh (these are the frames that enter action through the curvature term in equation (3.46)). A Lorentz transformation, \(\Lambda\), acting at a center of a \(k\) simplex, \(c(s_k)\), modifies all the connections incident on that simplex:

\[
\langle \Omega^A_{\ B}, \{c(s_i), c(s_k)\} \rangle \mapsto \langle \Lambda^A_{\ C}, \{c(s_k)\} \rangle \langle \Omega^C_{\ B}, \{c(s_i), c(s_k)\} \rangle.
\] (4.11)

(The Lorentz transform multiplies on the left in this case because \(c(s_k)\) is the target space for the connection—the other index of the connection lives in the frame at \(c(s_i)\), and is therefore unaffected by the transformation.) In addition, if the transformation acts at the center of four-simplexes, \(c(s_4)\), then all area fields incident on that four-simplex are transformed:

\[
\langle B, (s_2, s_4) \rangle \mapsto \langle \Lambda, \{c(s_4)\} \rangle \langle B, (s_2, s_4) \rangle \langle \Lambda^T, \{c(s_4)\} \rangle.
\] (4.12)

\(^1\)I first found this stated explicitly in Gotay et al. (1998), though it is well-known. It is a truly strange feature of gauge theories that they are simultaneously under- (gauge freedoms) and over-determined (constraints).
In addition, for transformations at the center of $s_4$, the $\Phi$ field transforms:

$$\langle \Phi^{ab}, c(s_4) \rangle \mapsto \langle \Lambda^a_{\alpha'}, c(s_4) \rangle \langle \Phi^{\alpha'b'}, c(s_4) \rangle,$$

where $\Lambda^a_{\alpha'}$ is the matrix that implements Lorentz transformations on adjoint vectors. For Lorentz transformations not at the center of four-simplexes, only the connection transforms; the other fields in the theory live in the Minkowski space at the centers of four-simplexes.

Equation (4.11) corresponds to the variations we have used to derive the torsion and continuity equations. For symmetry transformations at $c(s_4)$, equation (4.9) implies that the sum of all five torsion equations on $s_3$s which are faces of $s_4$ vanishes identically:

$$\sum_{s_3 \in \partial s_4} \sum_{s_2 \in \partial s_3} \langle B^b, (s_2, s_4) \rangle \langle (t_a R)^c, \star s_2 \cap s_4 \rangle = 0.$$

This redundancy in the torsion equations can also be understood from the topological principle that the boundary of a boundary is null: the torsion equation states that a certain two-form, summed over the boundary of an $s_3$ is zero. This automatically implies that this two-form summed over the boundaries of $s_3$s which are themselves the boundary of an $s_4$ vanishes. Of the five torsion equations on the five faces of an $s_4$, only four are independent.

Symmetry transformations at an $c(s_3)$ involve connections between both $c(s_2)$s and $c(s_3)$ and connections between $c(s_3)$ and $c(s_4)$s. The action of the symmetry is a combination of the variations we have used to derive the torsion and continuity equations. The symmetry requires that a particular combination of these equations vanishes; one of the set of the torsion and continuity equations incident that $s_3$ is redundant. Each $s_3$ not on the boundary of a simplicial complex has two $s_4$s incident on it; we can think of the redundancy implied by the symmetry at $c(s_3)$ as implying that the torsion equation on $s_3$ in one of these $s_4$s and the continuity equations for the $s_2$s on both the $s_4$s imply the torsion equation on $s_3$ and the other $s_4$.

Because Lorentz transformations at $c(s_3)$ live in a Cauchy surface, they provide
a first-class constraint on the momenta:

$$\sum_{s_2<s_3} \pi_c = 0,$$

(4.15)

which is just the torsion equation applied to $s_3$.

The Lorentz transformation freedoms at $c(s_2)$ imply that the continuity equations relating the area fields on $s_2$ and its incident $s_4$s are redundant. This is illustrated in Figure [4-2]. Because $s_2$ lives in a Cauchy surface, this symmetry also provides a first-class constraint on the momenta. This constraint is similarly the equality of the product of the area and curvature in equation (3.55) over all $s_4$s incident on $s_2$. This constraint is solved if we assign a unique momentum to each $s_2$ (equation (4.7)) from which the area field on incident $s_4$s can be derived using (3.55).

The Euler-Lagrange equations (4.4), (4.5), and (4.6) are redundant due to the Lorentz symmetries in the action and therefore do not have a unique solution for a given initial condition. Rather, there are families of solutions related by gauge transformations. To remove this ambiguity, we must specify enough information about the fields to select a unique member of this family. This procedure is called “gauge fixing.”

There are many ways to gauge fix. The simplest is to note that Lorentz transformations act in a simple way on the connection. For every site where we have a Lorentz transformation freedom, fixing the value of the connection on a dual strut incident on that site is sufficient to remove the freedom to act on that site with a Lorentz transformation. This approach is particularly useful if one wishes to compare a numerical solution with an analytical solution—fixing the one connection incident on each Lorentz symmetry site to be equal to the discretization of the continuous connection ensures that the numerical solution “tracks” the local frames of the continuous solution. In this way, the solutions obtained for the unconstrained field variables are comparable to the continuous solution. This is the procedure used in the simulations in the next chapter.

In the cases that there is no analytic solution to compare with, we propose the
Figure 4-2: The curvature loops centering on a particular $s_2$. The continuity equations relate area fields in loop 1 with 2, 2 with 3, 3 with 4, and 4 with 1. The Lorentz symmetry at $c(s_2)$ implies that one of these relations is implied by the others. This can also be understood by transitivity: if $B_1 \tilde{R}_1 = B_2 \tilde{R}_2$, $B_2 \tilde{R}_2 = B_3 \tilde{R}_3$, and $B_3 \tilde{R}_3 = B_4 \tilde{R}_4$, then $B_4 \tilde{R}_4 = B_1 \tilde{R}_1$. 
following scheme\footnote{This scheme is inspired by the Harmonic gauge for fixing the diffeomorphism symmetry in standard GR. The harmonic gauge is defined by the following condition} \footnote{Recall that $\Omega^a = 0$ implies that $\Omega^A_B = \exp (\Omega^A_{\alpha B}) = I$.} We want to ensure by our gauge fixing that no Lorentz frame gets too mis-aligned with its neighbors. Accordingly, we fix the gauge at $c(s_4)$ using the condition

$$\sum_{s_3 < s_4} \langle \Omega^a, \{ c(s_4), c(s_3) \} \rangle = 0. \quad (4.19)$$

The condition is that, on average, the transformation from the frame at $c(s_4)$ to the frames of its faces, the $c(s_3)s$, is the identity\footnote{Recall that $\Omega^a = 0$ implies that $\Omega^A_B = \exp (\Omega^A_{\alpha B}) = I$.} A similar condition fixes the gauge at the $c(s_3)$:

$$\sum_{s_2 < s_3} \langle \Omega^a, \{ c(s_3), c(s_2) \} \rangle = 0. \quad (4.20)$$

To fix the gauge at $c(s_2)$, we use

$$\sum_{c(s_3) \in \star s_2} \langle \Omega^a, \{ c(s_2), c(s_3) \} \rangle = 0. \quad (4.21)$$

This condition states that the frame at $s_2$ is “centered” with respect to the frames surrounding it in the surface dual to $s_2$.

### 4.3.2 Diffeomorphism Symmetry

When we advance a point using the procedure in Section \footnote{Recall that $\Omega^a = 0$ implies that $\Omega^A_B = \exp (\Omega^A_{\alpha B}) = I$.} we need to specify \textit{where} in the event manifold it lies. This is a manifestation of the diffeomorphism symmetry of the discrete theory. Traditionally in numerical relativity this is done
by specifying the scalar lapse, $\alpha$, and three-vector shift, $\beta^i$. The lapse specifies the elapsed proper time between corresponding points in the initial and final surfaces, and the shift measures the deviation of the vector connecting corresponding points and the normal to the initial surface.

The diffeomorphism symmetry also generates constraints on the Cauchy surface. These constraints are not as easy to express as the constraints from Lorentz symmetry, however, because we do not have a dynamical field whose variations correspond directly to the symmetry action. Perhaps the geometry reduction approach of Section 3.4.1 could be combined with the approach in Freidel & Louapre (2003) (see also discussion in Horowitz (1991)); this is a subject for future work.

If we had a tetrad-based formalism, the lapse and shift would be directly encoded in the tetrad evaluated on the “vertical” strut connecting the initial and the advanced point. Referring to Figure 4-3, the value of $E^A$ on $\{p_0, p_1\}$ measures the direction of the vertical strut. Unfortunately, our theory uses the area field, not the tetrad. In the continuous theory, equation (2.90) implies that we can derive the tetrad directly from the area field. In the discrete case, however, its analog, equation (3.60), does not imply that we can write $B^{AB} = E^A \wedge E^B$ for some tetrad $E^A$ that lives on the struts of a four-simplex. So, we will have to find a way to derive an approximate tetrad basis to fix the diffeomorphism gauge. In the following, we describe how to do this. Applying gauge conditions to the approximate tetrad basis gives us conditions on the area field that fix the diffeomorphism gauge.

Consider a single four-simplex, $s_4 = \{p_0, p_1, p_i, p_j, p_k\}$, which results from advancing the point $p_0$ to the point $p_1$. Figure 4-3 illustrates a corresponding configuration in 2+1 dimensions (where the point $p_k$ is omitted). The spacelike three-simplex $s_3 = \{p_0, p_i, p_j, p_k\}$ forms the “base” of $s_4$ in the initial Cauchy surface (in Figure 4-3, the base is $s_2 = \{p_0, p_i, p_j\}$). We will first construct a three-dimensional space-like basis for this surface. Three of the four two-dimensional faces in the surface contain the point $p_0$: $s_2^{(ij)} = \{p_0, p_i, p_j\}$, $s_2^{(jk)} = \{p_0, p_j, p_k\}$, and $s_2^{(ki)} = \{p_0, p_k, p_i\}$ (referring to Figure 4-3, these three two-dimensional faces are analogous to the two one-dimensional struts $\{p_0, p_i\}$ and $\{p_0, p_j\}$). The spatial parts of the tetrad field
Figure 4-3: A simplex that is part of a time-advancement structure in 2+1 dimensions. Time runs vertically; the point $p_0$ has been advanced to the point $p_1$. The face that is the base of the simplex, $\{p_0, p_i, p_j\}$, lies in the initial surface.

evaluated on the struts $\{p_0, p_i\}$, $\{p_0, p_j\}$, and $\{p_0, p_k\}$, denoted $\vec{e}_{(i)}$, $\vec{e}_{(j)}$, and $\vec{e}_{(k)}$, respectively, provide a spatial basis for the three-volume spanned by $s_3$ (in Figure 4-3, there are two corresponding tetrads on the struts $\{p_0, p_i\}$ and $\{p_0, p_j\}$).

Consider the area field evaluated on the three two-dimensional faces of $s_3$:

\[
B^a_{(ij)} \equiv \left\langle B^a, \left( s^{(ij)}_2, s_4 \right) \right\rangle \quad (4.22)
\]

\[
B^a_{(jk)} \equiv \left\langle B^a, \left( s^{(jk)}_2, s_4 \right) \right\rangle \quad (4.23)
\]

\[
B^a_{(ki)} \equiv \left\langle B^a, \left( s^{(kl)}_2, s_4 \right) \right\rangle . \quad (4.24)
\]

Using the generator basis given in equations (2.12) and (2.13), the first three components of the area field transform as a three-vector under Lorentz transformations, as do the last three. Write these vectors as

\[
\tilde{B}_{(ij)} = (B^0_{(ij)}, B^1_{(ij)}, B^2_{(ij)}) \quad (4.25)
\]

\[
\tilde{B}_{(ij)} = (B^3_{(ij)}, B^4_{(ij)}, B^5_{(ij)}) . \quad (4.26)
\]

for $B^a_{(ij)}$, and similarly for the other two area fields. If $B^{AB} = E^A \wedge E^B$, the second

---

4Neither orthogonal, nor normalized, but a basis.
triple of components would be the cross product of the spatial parts of the tetrad:

\[
\vec{B}_{(ij)} = \vec{e}_{(i)} \times \vec{e}_{(j)} \quad (4.27)
\]

\[
\vec{B}_{(jk)} = \vec{e}_{(j)} \times \vec{e}_{(k)} \quad (4.28)
\]

\[
\vec{B}_{(ki)} = \vec{e}_{(k)} \times \vec{e}_{(i)} \quad (4.29)
\]

In non-flat field configurations, where the discrete constraint equation (equation (3.60)) does not imply that \( B^{AB} = E^A \wedge E^B \), these formulas are only approximate, but a set of three three-vectors satisfying equations (4.27), (4.28), and (4.29) can always be found. We will use the vectors \( \vec{e}_{(i)}, \vec{e}_{(j)}, \) and \( \vec{e}_{(k)} \) satisfying equations (4.27), (4.28), and (4.29) for the spatial basis on \( s_3 \). When the curvatures of the simplex are small, they will be close to the tetrad basis.

Now consider the three “vertical” two-simplexes, containing both \( p_0 \) and \( p_1 \) and living in \( s_4 \): \( s^{(i)}_2 = \{p_0, p_1, p_i\} \), \( s^{(j)}_2 = \{p_0, p_1, p_j\} \), and \( s^{(k)}_2 = \{p_0, p_1, p_k\} \). (There is no correspondence here with Figure 4-3 because there is only a single vertical one-simplex in 2+1 dimensions, \( \{p_0, p_1\} \).) We will use the area field on these three two-simplexes to fix the tetrad on the vertical strut \( \{p_0, p_1\} \). Write \( B^a_{(i)}, B^a_{(j)}, \) and \( B^a_{(k)} \) for the area fields on these two-simplexes.

Each of the vertical two-simplexes contains one strut from \( p_0 \) to another point in the base: \( s^{(i)}_2 \) contains \( \{p_0, p_i\} \), \( s^{(j)}_2 \) contains \( \{p_0, p_j\} \), and \( s^{(k)}_2 \) contains \( \{p_0, p_k\} \). Let the tetrad on the vertical strut be

\[
\langle E^A, \{p_0, p_1\} \rangle = (\delta, \vec{\gamma}) \quad (4.30)
\]

(\( \delta \) represents the lapse degree of freedom, and \( \vec{\gamma} \) the three shift degrees of freedom.)

If we could write \( B^{AB} = E^A \wedge E^B \), then we would have

\[
\vec{B}_{(i)} = \delta \vec{e}_{(i)} - e^0_{(i)} \vec{\gamma}, \quad (4.31)
\]

for the triplet involving the 0, 1, 2 components of \( B^a_{(i)} \), since \( s^{(i)}_2 \) contains the vertical strut and the strut \( \{p_0, p_i\} \). In general, this relation will only be approximate. If we
further assume\footnote{This assumption is reasonable if the “vertical” strut of Figure 4-1 is to be timelike.} that $\gamma^i \ll \delta$, then we have

$$\bar{B}_i \approx \delta \bar{e}_i. \quad (4.32)$$

This equation relates the area field on $s_2^{(i)}$, the spatial parts of the tetrad on the strut $\{p_0, p_i\}$ as determined above, and the unknown lapse. To fix the lapse degree of freedom, we need only impose a scalar relation on one of the $\bar{B}_i$, say $\bar{B}_i \cdot \bar{e}_i = \delta_0 \bar{e}_i \cdot \bar{e}_i$ for a fixed choice of $\delta_0$.

The triplets involving the 3, 4, 5 components of the vertical area fields, $\bar{B}_i$, $\bar{B}_j$, and $\bar{B}_k$ depend on the shift. If $B^{AB} = E^A \wedge E^B$, then

$$\bar{B}_i = \bar{\gamma} \times \bar{e}_i, \quad (4.33)$$

and similarly for the other vertical faces. These equations relate the area field on the vertical faces, the spatial parts of the tetrad as determined above, and the unknown shift. To fix the shift, we need only impose three conditions on the vertical area fields, say $\bar{e}_j \cdot \bar{B}_i = \bar{e}_j \cdot (\bar{\gamma}_0 \times \bar{e}_i)$ for the three pairings $(i, j)$, $(j, k)$, and $(k, i)$, where $\bar{\gamma}_0$ is a fixed vector giving the desired spatial displacement of the vertical strut in the local frame.

Because the continuity and torsion equations (equations (3.55) and (3.50)) relate the area fields on one simplex incident on the vertical strut to the the area fields on the other simplexes incident on the strut, we need only apply the lapse and shift conditions above to the area fields on one four-simplex incident on the vertical strut; the conditions are “carried” to the other four-simplexes incident on the vertical strut by the continuity and torsion equations.

### 4.4 Initial Conditions

Due to the discrete Nöther theorem, equation (4.10), if a field configuration satisfies the constraints of the last section on its initial surface, evolution via equations (4.4),
[4.5], and [4.6] will result in another configuration that satisfies the constraints. By designing a discrete action with the symmetries of the continuous action, we have automatically built a constraint-preserving integrator.

There is a corollary to the discrete Nöther theorem, equation [4.10]: it is not possible to exactly satisfy the discrete Euler-Lagrange equations when advancing a non-constraint-preserving initial condition. Consider again equation [4.9], which expresses the change in the action under a symmetry transformation parametrized by $\epsilon^a$:

$$\left.\frac{\partial S}{\partial \epsilon^a}\right|_{\epsilon=0} = 0 = \left[\frac{\partial S}{\partial \phi_i^\alpha} \frac{\partial \phi_i^\alpha}{\partial \epsilon^a} + \frac{\partial S}{\partial \psi^\alpha} \frac{\partial \phi_i^\alpha}{\partial \epsilon^a} + \frac{\partial S}{\partial \phi_j^\alpha} \frac{\partial \phi_j^\alpha}{\partial \epsilon^a}\right]_{\epsilon=0}. \tag{4.34}$$

Even in the presence of constraint violation, we still solve the evolution Euler-Lagrange equations exactly, so the logic leading to equation [4.10] still holds, and we have

$$\pi_i^f \frac{\partial \phi_j^\alpha}{\partial \epsilon^a} - \pi_i^\alpha \frac{\partial \phi_i^\alpha}{\partial \epsilon^a} = 0. \tag{4.35}$$

However, constraint violation implies that

$$\pi_i^\alpha \frac{\partial \phi_i^\alpha}{\partial \epsilon^a} \neq 0, \tag{4.36}$$

so equation [4.35] now implies that the constraint violation present in the initial surface propagates unchanged to the final surface. We will see this explicitly in Section 5.3 when the constraint violation due to a gravitational wave initial condition follows the wave through the evolution.

Ideally, one would like to build a constraint-satisfying initial condition by solving the constraint equations (which are elliptic) on an initial surface numerically. However, that is beyond the scope of the current work. (See, however, Aksoyulu et al. [2008] for recent work generating solutions to the initial-value constraints of GR using finite element techniques.) Instead, we exploit that the discrete equations are integrals of the continuous equations over corresponding surfaces in the mesh plus higher-order correction terms. Integrating an analytic solution to the continuous equations over mesh elements will satisfy the discrete constraints to lowest order. The higher-order
constraint violation is of the same magnitude as the inherent discretization error in the method, and the constraint (violation) preservation of the algorithm guarantees that it cannot grow during a simulation. Accordingly, the numerical simulations described in the next chapter are of systems for which we have an analytic solution.

Given an analytic solution, we produce a discrete connection by integrating the continuous connection along the struts connecting the centers of elements of our mesh. We produce a discrete momentum (see equation (4.7)) by integrating the area field over the faces of our mesh, and lower its index using $\tilde{g}$.

### 4.5 Boundary Conditions

For the simulations described in the next chapter, we use Dirichlet boundary conditions. We fix the value of the connection on $\{c(s_2), c(s_3)\}$ struts when $s_3$ is a timelike volume consisting of a two-dimensional spatial boundary “dragged forward” in time. The momentum is not fixed on the boundary. In particular, we must still supply a lapse and shift to fix the diffeomorphism freedom when advancing points in the spatial boundary, since the diffeomorphism freedom primarily affects the momentum. Proper out-going wave boundary conditions are a matter for future work.
Chapter 5

Numerical Simulations

This chapter describes numerical simulations we have conducted to validate our evolution scheme. The simulations evolve an initial condition computed from an exact, analytical solution to the continuous GR equations, using the gauge choices described in the previous chapter. Comparison with the analytical solutions shows agreement, up to the discretization error in the initial condition.

The three analytical solutions discussed in this section are weak gravitational waves, the Schwarzschild solution for an isolated non-spinning black hole, and the Kerr solution for an isolated, spinning black hole. We use the analytic solution to produce an initial condition and fix boundary conditions on the connection as described in the last chapter. We do not attempt to solve the discrete constraint equations; the constraint violation is of order the truncation error of the method, and guaranteed not to grow, so it does not spoil the correspondence at leading order between the discrete and continuous solutions.

5.1 Convergence of the Discrete Equations

For both the analytical solutions discussed here, we have verified numerically that the leading order term in the discrete equations corresponds to the continuous equations, as discussed in Section 3.4. That is, the residuals of the discrete equations vanish faster on mesh refinement when working with a discretized solution to the continuous
Table 5.1: Scaling of the residual of the discrete equations with mesh spacing, \( h \), on generic discrete fields and discretizations of solutions to the continuous equations. The traceless condition on \( \Phi \), equation \((3.59)\), is identically satisfied on all discretizations.

equations than when working with a discretization of generic continuous fields.

For example, the discrete Einstein equation (equation \((3.47)\)),

\[
\bar{g}_{ab} \left( \mathcal{R}^b (\mathcal{H} \cap s_4) + \frac{1}{15} \left( \Phi_{ab} (\mathcal{H} \cap s_3) \right) g_{aa'} g_{bb'} \sum_{s_2 < s_4} \left| \langle \pi_a, s_2 \rangle - \bar{g}_{ab} \int_{s_2} b^b \rangle \right| \right) = 0, \tag{5.1}
\]

relates discrete fields \((B \text{ and } R)\) that scale as the area of the mesh elements on which they live. Thus, when the typical mesh spacing, \( h \), decreases by a factor of two, one would expect that, generically, the discrete Einstein residual would decrease by a factor of 4. But, when \( B, R, \text{ and } \Phi \) are discretizations of solutions to the continuous Einstein equation, we find that the residual decreases by a factor of 8, that is, scaling like \( h^3 \). Table 5.1 gives the scaling of the equation residuals on generic fields and on fields which are discretizations of solutions to the continuous equations.

In Sections 5.3, 5.4, and 5.5, we will be plotting the relative error in the field variables of the numerical solution compared to the analytical solutions on each timeslice. Let the relative error in the fields on a particular three-simplex, \( s_3 \), in the timeslice be \( \langle \delta, s_3 \rangle \). We have

\[
\langle \delta, s_3 \rangle = \langle \delta_\pi, s_3 \rangle + \langle \delta_\Omega, s_3 \rangle, \tag{5.2}
\]

where

\[
\langle \delta_\pi, s_3 \rangle \equiv \sum_{s_2 \leq s_3} \left| \langle \pi_a, s_2 \rangle - \bar{g}_{ab} \int_{s_2} b^b \right| \tag{5.3}
\]

and

\[
\langle \delta_\Omega, s_3 \rangle \equiv \sum_{s_2 \leq s_3} \left| \langle \Omega^a, \{ c (s_2), c (s_3) \} \rangle - \int_{\{ c (s_2), c (s_3) \}} \omega^a \right| \tag{5.4}
\]
where the expression \( |v^a| \) refers to the two-norm of the vector \( v^a \):
\[
|v^a| \equiv \sqrt{\sum_{a=0}^{5} (v^a)^2},
\]
(5.5)
and the fields \( b^a \) and \( \omega^a \) are the continuous area two-form and connection one-form, respectively. Note that the two-norm is not a Lorentz invariant quantity, so the quantity \( \delta \) is not gauge-independent; nevertheless, it scales as \( \mathcal{O}(h) \) in any gauge.

We will also plot the relative constraint violation for the solutions in Sections 5.3, 5.4, and 5.5. This quantity is defined on a single three-simplex, \( s_3 \), as
\[
\langle \delta^{(c)}, s_3 \rangle \equiv \left| \sum_{s_2 \in \partial s_3} \langle \pi^a, s_2 \rangle \right| \sum_{s_2 \in \partial s_3} |\langle \pi^a, s_2 \rangle|.
\]
(5.6)
(Compare to equation (4.15).) The integrated relative constraint violation is defined on a hypersurface as
\[
\Delta^{(c)} \equiv \sum_{s_3} \langle \delta^{(c)}, s_3 \rangle,
\]
(5.7)
where the sum runs over all three-simplexes in the surface.

### 5.2 Solution Method

Equations (3.47), (3.50), and (3.60), applied on the structure that results from the advancement procedure in Section 4.2 and coupled with a choice of gauge fixing as described in Section 4.3 form a non-linear system of equations that can be solved for the values of \( \Omega \) and \( \pi \) on the advanced-time boundary of the advancement structure. The dimensionality of the system depends on the connectivity of the mesh, but for typical meshings there are a few hundred variables that must be solved for to evolve each point. The equations are relatively sparse: the Jacobian matrix of the equations with respect to the unknown field variables has a typical filling fraction of 2% to 5%.

As discussed in Section 4.4, the constraint-preserving nature of the evolution implies that any initial constraint violation is “frozen in” to the solution. There is a
limit to how accurately we can satisfy the time-evolution equations set by the error in the initial discretization. We have found that an initial guess for the field variables based on the known analytic solution followed by one round of Newton-Raphson refinement is sufficient to approach this limit; therefore this is the solution algorithm we adopt for the simulations described in the following sections. Technically, fixing the number of iterations of the Newton-Raphson algorithm makes the method explicit, but because the accuracy limit imposed by the initial discretization is always reached after a single iteration, the method is identical to an implicit method that iterates to convergence. In particular, the method does not become unstable in the limit of long timesteps.

5.3 Weak Gravitational Waves

In this section we discuss the evolution of a weak gravitational wave. We choose a gravitational wave of the “+” polarization propagating in the positive $\hat{z}$ direction. The metric components in the $(t, x, y, z)$ coordinate system in the transverse-traceless gauge are

$$g_{\mu\nu} = \eta_{\mu\nu} + \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & f(t - z) & 0 & 0 \\ 0 & 0 & -f(t - z) & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad (5.8)$$

where $f$ is the function which gives the shape of the waveform, and we assume $f \ll 1$, so terms proportional to $f^2$ can be ignored. The corresponding tetrad is

$$e^0 = dt \quad (5.9)$$
$$e^1 = dx \left( 1 + \frac{1}{2} f(t - z) \right) \quad (5.10)$$
$$e^2 = dy \left( 1 - \frac{1}{2} f(t - z) \right) \quad (5.11)$$
$$e^3 = dz, \quad (5.12)$$
where \((dt, dx, dy, dz)\) are the coordinate basis one-forms. The area field is given by

\[ b^\alpha t^A = e^A \wedge e^B. \]  

(5.13)

The torsion-free connection corresponding to this tetrad is given by

\[ \omega^1_0 = -\frac{1}{2} f'(t - z) dx \]  

(5.14)

\[ \omega^3_1 = \frac{1}{2} f'(t - z) dx \]  

(5.15)

\[ \omega^2_0 = \frac{1}{2} f'(t - z) dy \]  

(5.16)

\[ \omega^2_3 = -\frac{1}{2} f'(t - z) dy, \]  

(5.17)

with all other components either trivially related by symmetries or zero. Upon discretization, \(\Phi\) can be computed from the requirement that the \(O(h^2)\) terms in the discrete Einstein equation vanish.

We work with cubical spatial domains of the form \([0, 1] \times [0, 1] \times [0, 1]\). To mesh this domain, we use a grid of points with a uniform spacing along each dimension. The grid decomposes the domain into cubical cells. We further decompose each cell into tetrahedrons using the tile-able decomposition of the cube shown in Figure 5-1, an example of the resulting mesh for \(N_{\text{points}} = 3^3\) appears in Figure 5-2.

We fix a Lorentz gauge on the mesh such that the frames at each \(c(s_2), c(s_3),\) and \(c(s_4)\) agree with the continuous solution. We then evolve each point in the local timelike direction with an increment that is \(1/2\) the length of the smallest spatial spacing between points. That is, we set \(\delta = h_{\text{min}}/2\) and \(\vec{\gamma} = 0\) in equation (4.30), where \(h_{\text{min}}\) is the smallest spatial separation of points on the domain. This choice of gauge implies that the coordinate evolution of each point is \(\Delta t = \delta, \Delta \vec{x} = 0\).

We provide Dirichlet boundary conditions by discretizing the continuous connection on the boundary of the domain. Because this does not produce a constraint-satisfying boundary condition, the boundary becomes a potential source or sink for constraint violation, as we will see in Figure 5-4.

Figure 5-3 shows the gravitational wave content, determined as the magnitude of
Figure 5-1: The tile-able decomposition of the cube used to mesh domains for gravitational wave evolution.

Figure 5-2: Meshing of the domain $[0, 1]^3$ with $3^3$ points using the meshing scheme we use for decomposing the domains for gravitational wave evolution.
Figure 5-3: Gravitational wave content in a simulation of a weak gravitational wave initially peaking at the origin with Gaussian profile ($\sigma = 0.25$) over two crossing times on a $6 \times 6 \times 20$ grid.

the deviation of the area field from a flat area field, in a simulation with $f$ a Gaussian with amplitude $10^{-3}$ and standard deviation 0.25 on a $6 \times 6 \times 20$ grid over two crossing times.

To demonstrate the accuracy, stability, and convergence properties of the algorithm on this problem, we use the relative constraint violation defined in equation (5.6). Figure 5-4 shows the value of the constraint equation residuals integrated over spacelike hypersurfaces (equation (5.7)) versus time for two different discretizations of the domain $[0, 1]^3$. The fine grid is $6 \times 6 \times 20$ points, as in the above simulation; the coarse grid is $3 \times 3 \times 10$ points. As the point spacing decreases by a factor of two, the total constraint violation also decreases by a factor of two—the individual residuals go down by a factor of $2^4$, but there are a factor of $2^3$ more individual volumes, for a net gain of a factor of 2. This is consistent with the equation residual scaling in Table 5.1.

Figure 5-5 shows the evolution of the constraint violation in space and time for the $6 \times 6 \times 20$ simulation, while Figure 5-6 shows the same quantity in the $3 \times 3 \times 10$ simulation. The constraint violation is carried with the gravitational wave; as the
Figure 5-4: Constraint violation integrated over spacelike hypersurfaces (see equation (5.7)) versus time for two different resolutions in a simulation of a weak gravitational wave. The dashed line refers to a simulation with $3 \times 3 \times 10$ points discretizing the domain, while the solid line refers to a simulation discretizing the domain with $6 \times 6 \times 20$; when the resolution increases by a factor of two in each dimension, the constraint violation decreases by a factor of two, consistent with the residual scaling in Table 5.1. As the gravitational wave propagates across the domain, it carries the constraint violation with it; as the wave propagates off the grid, the constraint violation tends to zero as the solution approaches flat space.
wave propagates off the computational grid, the constraint violation is carried out the boundary, and tends to zero as the spacetime on the grid tends to flat.

Figure 5-7 shows the relative error in all field variables (see equation (5.2)) with respect to the analytic solution for the gravitational wave as a function of time and space in the $6 \times 6 \times 20$ simulation; Figure 5-8 shows the same quantity for the $3 \times 3 \times 10$ simulation. The errors in the lower-resolution simulation are approximately twice those of the higher-resolution simulation, as would be expected from scaling. The solution error tracks the constraint error because the source of both is essentially the truncation error in the discretization of the continuous solution.

### 5.4 Schwarzschild

The Schwarzschild spacetime describes an isolated, static, non-rotating black hole. A convenient representation of the Schwarzschild spacetime can be found in [Doran (2000)]. The metric is

$$g = \left( -1 + \frac{1}{r} \right) dt^2 + \frac{x_i}{r^{3/2}} dt dx^i + \delta_{ij} dx^i dx^j,$$

(5.18)
Figure 5-6: Relative constraint violation (see equation (5.6)) versus space and time in the $3 \times 3 \times 10$ gravitational wave simulation. The violation of the constraint is largest where the curvatures peak on either side of the wave peak, and is carried along with the wave through the domain and out the boundary. The violation is approximately twice the magnitude of the violation in Figure 5-5, as expected from the scaling of the constraint violation in the discretization.

Figure 5-7: Relative error in the field variables (see equation (5.2)) as compared to the analytic solution for the $6 \times 6 \times 20$ gravitational wave simulation. The solution error tracks the constraint error because the source of both is essentially the truncation error in the discretization of the continuous solution.
Figure 5-8: Relative error in the field variables (see equation (5.2)) as compared to the analytic solution for the $3 \times 3 \times 10$ gravitational wave simulation. The solution error tracks the constraint error because the source of both is essentially the truncation error in the discretization of the continuous solution. The scale of the errors in this simulation is approximately twice those in Figure 5-7, as expected from scaling.

where $r = \sqrt{x^i x_i} = \sqrt{x^2 + y^2 + z^2}$, we have adopted units where the Schwarzschild radius, $r_s = 2GM/c^2$, is one, and, for fixed time coordinate $t$, the spatial coordinates $x, y, z$ are Cartesian on the spatial hypersurfaces. The coordinate basis one-forms are $dt$ and $dx^i = (dx, dy, dz)$. (In these formulas, we make no distinction between the spatial coordinates with raised index, $x^i$, and lowered index, $x_i$.) The $t$ coordinate is the proper time for observers freely falling along radial trajectories toward the singularity at the spatial origin who begin at rest at infinity. The tetrad corresponding to this metric is

\begin{align}
e^0 &= dt \\
e^i &= \frac{x^i}{r^{3/2}} dt + dx^i,
\end{align}

and the torsion-free connection corresponding to this tetrad is

\begin{align}
\omega^0_i &= \omega^i_0 = -\frac{x^i}{2r^{3/2}} dt + \frac{2 (x^j x_j) dx_i - 3x^i x_j dx^j}{2r^{7/2}},
\end{align}

with all $\omega^j_i = 0$. The area field, $b^{AB} = e^A \wedge e^B$. Upon discretization, we solve for the discrete $\Phi$ field by requiring that the $O(h^2)$ terms in the discrete Einstein equation
(equation \((3.47)\)) vanish.

We could use the Cartesian mesh from the last section to represent the discrete domain for our simulations of the Schwarzschild spacetime, but this would be inefficient. In order to resolve details of the solution at small \(r\), such a mesh allocates many more points than needed to represent this spherically-symmetric solution at large \(r\). Instead, we generate a mesh with approximately constant \textit{angular} resolution as follows.

Begin with a decomposition of the surface of a 2-sphere into triangles (e.g. by bisecting the six sides of a cube diagonally, and then refining the resulting triangles until a desired angular resolution is reached). Scale the points in this decomposition to some fixed, minimum \(r\). Now use the advancement procedure described in Section 4.2 to advance each point outward in \(r\) by a distance which is approximately the length of the struts in the spherical shell (if the step in \(r\) is too small or too large, this procedure creates tetrahedrons with large aspect ratios which leads to inefficient evolution, since the largest timestep for a point is limited by the minimum length of the struts in the initial surface incident on that point). This produces two triangulated spherical shells, with appropriate links between them meshing the contained volume. Repeat the process, advancing outward in \(r\) (with correspondingly larger increments to match the longer struts in the spherical shells) until the mesh crosses a desired maximum \(r\). The result of this procedure is a meshing of a spherical annulus between the minimum and maximum radius with approximately constant angular size of its elements. An example mesh appears in Figure 5-9.

As in the gravitational wave simulations, we provide Dirichlet boundary conditions by discretizing the continuous connection on the inner and outer boundaries; these boundaries can act as sources or sinks of constraint violation from the discretization.

We fix the Lorentz gauge so that the frames at the \(c(s_2), c(s_3), \text{ and } c(s_4)\) correspond to the continuous solution given in Doran (2000). We advance each point forward in time along the local timelike basis vector—that is, we set \(\vec{\gamma} = 0\) in equation \((4.30)\). This induces a coordinate increment \(\Delta t = \delta, \Delta x^i = -\delta x^i/r^{3/2}\). We choose the magnitude of the advancement in the local frame, \(\delta\), to satisfy the following two
Figure 5-9: The result of the procedure for generating spherical meshes described in the text. The final mesh contains 104 points and 432 tetrahedrons at approximately constant angular resolution. The 8 radial shells extend from a minimum radius of 0.9 to a maximum radius of 3.7.

conditions:

1. The length of the “vertical” strut, $\delta$, must be small enough that it is the only timelike strut in the advancement structure. To ensure this, we enforce that $\delta = l_{\text{min}}/2$, where $l_{\text{min}}$ is the smallest of the lengths of the struts incident on the point to be advanced.

2. Because the timelike motion induces a coordinate shift toward the origin, we must ensure that the points stay away from the central singularity. We choose $\delta$ small enough that the advanced point does not come any closer to the origin than 90% of the retarded point’s radius.

Because $\delta$ differs at different points, the coordinate $t$ as a function of $r$ on a timestep is not constant, but rather a curve like those given in Figure 5-10.

Simulations of Schwarzschild spacetimes in this gauge remain stable as long as the stretching of the spacelike hypersurfaces remains small. In fact, because the total volume of the simulation’s spacelike hypersurfaces is decreasing, the integrated
Figure 5-10: Time coordinate \((t)\) versus radius \((r)\) for some representative points in the Schwarzschild mesh during evolution using the gauge choices in the text. All points in the mesh “fall in” toward the singularity, regardless of timestep, but the points at large radius do so much more slowly, and can therefore take larger timesteps.

constraint violation (induced by the imperfect discretization of the initial condition, and preserved by the algorithm) decreases for a time. As time goes on, however, the spacelike hypersurfaces stretch (see Figure 5-10), and the outer boundary in particular begins to act as a source for constraint violation. Figure 5-11 shows the integrated constraint violation for two simulations of the Schwarzschild spacetime. One simulation has approximately \(10^3\) points in 32 radial shells in the simulation volume; the other has approximately \(10^4\) points in 64 radial shells. Both span from \(0.9r_s\) to \(10r_s\) initially. As would be expected with approximately \(2^3\) as many points, the error in the higher-resolution simulation is about a factor of two better than the lower-resolution simulation.

Figure 5-12 plots the relative error in the field variables compared to the known analytic solution (see equation (5.2)) versus \(r\) and \(t\) in the simulation using \(10^4\) points. The solution is quite accurate, with most of the error occurring near the singularity at late times. Figure 5-13 plots the same quantity, but for the lower-resolution simulation with \(10^3\) points; the error here is about twice the error in the higher-resolution
Figure 5-11: Constraint violation integrated over the spacelike hypersurfaces at each timestep (see equation (5.7)) in two simulations of Schwarzschild spacetime. The solid line is a simulation with approximately $10^4$ points in 64 radial shells; the dashed line is a simulation with approximately $10^3$ points in 32 radial shells. The lower-resolution simulation has been stretched along the timestep axis by a factor of 2 because its timesteps are approximately twice as long as the higher-resolution simulation. The overall error initially decreases as the outer boundary moves in, shrinking the total physical volume of the simulation, and then increases as the spatial hypersurfaces stretch out. The higher resolution simulation has approximately half the error of the lower resolution simulation as would be expected with a factor of approximately $2^3$ more points.
Figure 5-12: The sums of the relative error in the field variables compared to the known analytic solution (see equation (5.2)) for the Schwarzschild spacetime simulation with $10^4$ points and 64 radial shells. The solution is quite accurate, with most of the error occurring near the singularity at late times. The ridges in the error profile are due to the radial shells introduced by the meshing algorithm described in the text.

The gauge chosen in this simulation is sub-optimal, in that one does not need to advance in a purely timelike direction, but it is sufficient to demonstrate that the evolution of the Schwarzschild spacetime can be done stably in this scheme; there is ample time for instability to develop around the horizon or propagate outward from the interior of the black hole, yet no instabilities appear. A better choice of gauge is a matter for future work.

5.5 Kerr

The Kerr spacetime represents a black hole with non-zero angular momentum. This section reports on simulations of the Kerr spacetime using the algorithms described in this thesis. A convenient tetrad for the Kerr spacetime can be found in [Doran (2000)]:

$$e^A_\mu = \delta^A_\mu - \frac{\alpha}{\rho} v_\mu a_\beta \eta^{\beta A},$$

(5.22)
Figure 5-13: The sums of the relative error in the field variables compared to the known analytic solution (see equation (5.2)) for the Schwarzschild spacetime simulation with $10^3$ points in 32 radial shells. The ridges in the error profile are due to the radial shells introduced by the meshing algorithm described in the text. The error in this lower-resolution simulation is about twice that in the higher-resolution simulation in Figure 5-12, as expected from scaling.

where the coordinates are $t, x, y,$ and $z,$ and

$$\alpha = \frac{\sqrt{r}}{\rho},$$  \hspace{1cm} (5.23)

with

$$\rho^2 = r^2 + \frac{a^2 z^2}{r}.$$  \hspace{1cm} (5.24)

The variable $r$ is defined implicitly by

$$r^4 - r^2 \left( x^2 + y^2 + z^2 - a^2 \right) - a^2 z^2 = 0,$$  \hspace{1cm} (5.25)

and the forms $v$ and $a$ are given by

$$v = dt + \frac{ay}{a^2 + r^2} dx - \frac{ax}{a^2 + r^2} dy$$  \hspace{1cm} (5.26)

and

$$a = \sqrt{r^2 + a^2} \left( \frac{rx}{r^2 + a^2} dx + \frac{ry}{r^2 + a^2} dy + \frac{z}{r} dz \right).$$  \hspace{1cm} (5.27)
with \( dt, dx, dy, \) and \( dz \) coordinate basis one-forms. The parameter \( a \) is the dimensionless angular momentum (restricted to the range \( 0 \leq a < 1 \)); we have chosen the mass \( M = 1/2 \) for correspondence with the Schwarzschild simulations in the last section. The torsion free connection can be found from the Cartan structure equation (equation (2.82)); the exact expressions are fairly complicated, and not illuminating, so we will not give them here. Once again, we solve for \( \Phi \) using the discrete Einstein equation (equation (3.47)) after discretization.

We use the same mesh to simulate the Kerr spacetime as we did for the Schwarzschild spacetime (see Figure 5-9). For modest values of the spin parameter, \( a \), the Kerr solution is approximately spherically symmetric, and the mesh is relatively efficient. As \( a \) approaches its limiting value of 1, the solution becomes more and more oblate; for large \( a \) the mesh we use does not conform to the spacetime well.

We provide Dirichlet boundary conditions on both the inner and outer boundaries of the mesh by discretizing the continuous connection on these boundaries. Because the discretization does not yield an exact solution to the discrete constraints, the boundaries can act as sources or sinks for constraint violation. We fix the Lorentz gauge so that the frames at the \( c(s_2), c(s_3), \) and \( c(s_4) \) correspond to the continuous solution given above. We advance each point along the local timelike basis vector—that is, we set \( \vec{\gamma} = 0 \) in equation (4.30). As in the Schwarzschild case, this introduces a coordinate increment for each point toward the black hole, but in Kerr there is an additional coordinate increment in the direction of rotation of the hole due to frame dragging. This frame-dragging effect means that, in addition to stretching as different points advance with different timesteps as in Figure 5-10 there is a net shear between points at different distances between the hole. The shear could be ameliorated with a better lapse and shift choice, but cannot be eliminated entirely. In the Kerr solution, there is a region called the \textit{ergosphere}, outside the horizon, where all timelike vectors have a shear component relative to infinity; in this region one cannot help but advance with shear.

Due to the stretching of the mesh from shear and differing time coordinate increments, the outer and inner boundaries introduce progressively more constraint
Figure 5-14: Constraint violation in two simulations of the Kerr spacetime as described in the text integrated over the spacelike hypersurface at each timestep (see equation (5.7)). The dashed line is a simulation with approximately $10^4$ points in 64 radial shells, and the solid line is a simulation with approximately $10^5$ points in 128 radial shells. The errors are plotted versus timestep number (since different points in the mesh advance forward in time with different increments); the smaller simulation takes fewer steps, but has been scaled horizontally by an appropriate factor so the physical times of the two simulations match. As in the Schwarzschild simulations, we find that the error initially decreases as the total volume of the simulation space shrinks. As the simulation progresses, the errors introduced at the inner and outer boundary due to the stretching of the spacelike hypersurface come to dominate the constraint error.

violation up to the end of the simulation, when the introduced constraint violation destroys the accuracy of the simulation.

Figure 5-14 shows the constraint violation as a function of evolution time in two simulations of the Kerr spacetime. The two simulations involve approximately $10^4$ and $10^5$ points, distributed in roughly constant angular resolution in radial shells, as illustrated in Figure 5-9. The initial spatial surface spans the range $x^2 + y^2 + z^2 = 0.9^2$ to $x^2 + y^2 + z^2 = 10^2$. We choose the spin parameter, $a = 0.5$, so that the mesh can represent the non-spherical Kerr solution adequately.

Figure 5-15 plots the relative error in the field variables compared to the analytically known Kerr solution as a function of the radius-like value $\sqrt{x^2 + y^2 + z^2}$ and
Figure 5-15: The relative error in the field variables compared to the analytically known Kerr solution (see equation (5.2)) in the simulation with $10^5$ points in 128 radial shells described in the text. The error is plotted as a function of the radius-like value $\sqrt{x^2 + y^2 + z^2}$ and the $t$ coordinate; the ridges in the error profile are due to the radial shells in the mesh.

the $t$ coordinate. The simulation is quite accurate, and, as with the Schwarzschild simulations, the error is largest near the singularity. Figure 5-16 shows the same quantity, but for the smaller simulation with $10^4$ points in 64 radial shells. The error in this lower-resolution simulation is approximately a factor of two larger than the higher-resolution one, as expected from scaling.

5.6 Boundary Conditions, and Future Work

The numerical simulations in the preceding sections demonstrate that the methods described in this thesis are promising for the constraint-preserving numerical evolution of general relativistic systems. The major issue in the simulations presented in this chapter is the failure of the boundary conditions to satisfy the constraints. Particularly in the Schwarzschild and Kerr simulations, the constraint violation coming in from the boundaries eventually caused the numerical method to break down.

The issue of constraint-violating boundary conditions is related to the issue of constraint-satisfying initial conditions, discussed briefly in Section 4.4. It would be nice to be able to generate exactly constraint-satisfying initial data and boundary
Figure 5-16: The relative error in the field variables compared to the analytically known Kerr solution (see equation (5.2)) in the simulation with $10^4$ points in 64 radial shells described in the text. The error is plotted as a function of the radius-like value $\sqrt{x^2 + y^2 + z^2}$ and the $t$ coordinate; the ridges in the error profile are due to the radial shells in the mesh. The error in this simulation is about a factor of two larger than the simulation in Figure 5-15, as expected from scaling.

conditions even for cases where an analytic solution is not available, perhaps through techniques similar to those described in Aksoylu et al. (2008). Probably generating constraint-satisfying boundary conditions is much easier than generating fully constraint-satisfying initial conditions.

To move beyond the simulations described here, it would also be useful to implement some performance improvements in the code used in this chapter. Currently, the simulations of the Kerr spacetime with $10^5$ points in the previous section take several hours to run on a single CPU of a standard workstation. The performance could most easily be improved by finding an analytic or iterative solution method to solve the non-linear field equations instead of the Newton iteration used in the current code, as the matrix inversion in the Newton algorithm dominates the run-time.

It would also be interesting to try to translate some of the common gauge choices in finite-difference gravitational simulations to this framework—a poor choice of gauge is another limiting factor in the black hole simulations in this chapter.

Another interesting way to make a connection to the current simulations, which are all based on finite-differencing approaches, and this work would be to try to derive finite difference equations from the discrete field equations that the discrete fields
satisfy identically. Presumably these equations would be, at leading order, equivalent to a standard finite-differencing of the continuous field equations. However, since the discrete equations are constraint-preserving, the higher-order corrections to the finite-difference equations (i.e. the truncation error) would also be constraint-preserving. Even if it is not possible to find such equations explicitly, perhaps a calculation of the higher-order corrections to the standard finite-differencing of the field equations would give insight for how to modify current codes to better preserve the constraints.

These are just a few possibilities for future work. Variational approaches to the numerical simulation of GR spacetimes are in their infancy. Nevertheless, this thesis demonstrates that these approaches are promising; we hope to see continued work in this area in the future.
Appendix A

Forms, Exterior Derivatives and Integration

This section discusses our conventions for forms, exterior derivatives of forms, and integration of forms over (sub)manifolds. A classic reference on this topic is Spivak (1971).

In this section we will make extensive use of the symmetrization and antisymmetrization operators on the indices of tensors. The symmetrization operator, denoted with parenthesis, is defined by

\[
T_{(\mu_1...\mu_n)} \equiv \frac{1}{n!} \sum_{\sigma \in \text{perm}(n)} T_{\mu_{\sigma_1}...\mu_{\sigma_n}}.
\] (A.1)

The symmetrization can also occur over upper indices. The leading factor of $1/n!$ means that a tensor which is invariant under interchange of any two indices satisfies

\[
T_{\mu_1...\mu_n} = T_{(\mu_1...\mu_n)}. 
\] (A.2)

Similarly, we can define an antisymmetrization operator, denoted by square brackets:

\[
T_{[\mu_1...\mu_n]} \equiv \frac{1}{n!} \sum_{\sigma \in \text{perm}(n)} \text{sig}(\sigma) T_{\mu_{\sigma_1}...\mu_{\sigma_n}}.
\] (A.3)
where \( \text{sig}(\sigma) \) is the signature\(^1\) of the permutation \( \sigma \). A tensor which changes sign under exchange of any two indices satisfies

\[
T_{\mu_1...\mu_n} = T_{[\mu_1...\mu_n]},
\]  
(A.4)

An \( n \)-form is a tensor with \( n \) lower indices that is completely antisymmetric:

\[
\omega_{\mu_1...\mu_n} = \omega_{[\mu_1...\mu_n]},
\]  
(A.5)

In a manifold with \( m \) dimensions, the space of \( n \)-forms at a point has dimension \( \binom{m}{n} \).

In particular, all \( m \) forms are equivalent up to scaling.

We can integrate an \( n \)-form over an \( n \)-dimensional manifold in a coordinate-independent manner. Define

\[
\int_M \omega \equiv \int_M d^n x \omega_{01...(n-1)} = \frac{1}{n!} \int_M d^n x \epsilon^{\mu_1...\mu_n} \omega_{\mu_1...\mu_n},
\]  
(A.6)

where \( \epsilon^{\mu_1...\mu_n} \) is the completely antisymmetric tensor with \( \epsilon^{01...(n-1)} = 1 \), and the integrals on the right hand side are performed in \( \mathbb{R}^n \) using some coordinate system \( x^\mu \) which covers the manifold \( M \).\(^2\) Under a change of coordinates to \( \tilde{x}(x) \), the integral becomes

\[
\int_M d^n \tilde{x} \epsilon^{\mu_1...\mu_n} \omega_{\mu_1...\mu_n} = \int_M d^n \tilde{x} \epsilon^{\mu_1...\mu_n} (J^{-1})_{\mu_1}^{\mu'_1} \times ... \times (J^{-1})_{\mu_n}^{\mu'_n} \omega_{\mu'_1...\mu'_n},
\]  
(A.7)

where \( J^\mu_\nu = \partial \tilde{x}^\mu / \partial x^\nu \) is the Jacobian of the coordinate transformation. The contraction of the inverse Jacobian with the \( \epsilon \) tensor corresponds to an expansion of the determinant by minors, so we obtain

\[
\int_M d^n \tilde{x} \epsilon^{\mu_1...\mu_n} \omega_{\mu_1...\mu_n} = \int_M d^n \tilde{x} \frac{\epsilon^{\mu_1...\mu_n}}{\det(J)} \omega_{\mu_1...\mu_n} = \int_M d^n x \epsilon^{\mu_1...\mu_n} \omega_{\mu_1...\mu_n},
\]  
(A.8)

\(^1\)The signature of a permutation is \((-1)^s\), where \( s \) is the number of pairwise interchanges required to put the elements of the permutation into sorted order.

\(^2\)Integrals over regions which cannot be covered by a single coordinate system (i.e. the sphere) can be defined similarly using a partition of unity, provided the manifold has certain countability properties (for the details, see \cite{Wald1984}).
because the volume element transforms by multiplication by the Jacobian. A metric, \( g_{\mu\nu} \), on the manifold induces a volume \( n \)-form via

\[
v_{\mu_1...\mu_n} = \sqrt{g} \epsilon_{\mu_1...\mu_n}, \tag{A.9}\]

where \( g = \det (g_{\mu\nu}) \), and \( \epsilon_{\mu_1...\mu_n} = \epsilon_{[\mu_1...\mu_n]} \) with \( \epsilon_{01...(n-1)} = 1 \). Given such a volume \( n \)-form, we can define the integrals of scalars over the manifold by:

\[
\int_M f \equiv \int_M f v_{\mu_1...\mu_n}, \tag{A.10}\]

Using a metric, we can also define the Hodge-dual operation (written as *) on forms. In spaces of dimension \( n \), the Hodge-dual takes \( k \)-forms to \((n - k)\)-forms via

\[
\star \omega_{\mu_1...\mu_k} \equiv \sqrt{g} \epsilon_{\nu_1...\nu_{n-k} \mu_1...\mu_k} \omega^\nu_{\mu_1...\mu_k} g^{\mu_1'\mu_1} ... g^{\mu_k'\mu_k}, \tag{A.11}\]

where \( g^{\mu\nu} \) is the inverse metric (whose components are the matrix inverse of \( g_{\mu\nu} \)) so that \( g_{\mu\rho} g^{\rho\nu} = \delta^\nu_\mu \).

The exterior derivative of a form is defined by

\[
(d\omega)_{\mu_1...\nu_n} \equiv (n + 1) \partial_{[\nu} \omega_{\mu_1...\nu_n]}. \tag{A.13}\]

For example, if \( A_\mu \) is a one-form,

\[
(dA)_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{A.14}\]

Because partial derivatives commute, we have

\[
dd \omega \equiv 0 \tag{A.15}\]

\footnote{Note the consistency of \( \star \) with equation \(\text{(A.10)}\). If \( f = \star \omega_{\mu_1...\mu_n} \), then
\[
\int_M f = \int_M \omega_{\mu_1...\mu_n}. \tag{A.12}\]}
for all forms $\omega$. The exterior derivative has the property that

$$\int_M d\omega = \int_{\partial M} \omega$$

(A.16)

for all $n$-forms $\omega$ and simply connected $n$-dimensional manifolds $M$. This is a generalization of Stokes’ theorem.
Bibliography


Capovilla, R., Dell, J., Jacobson, T., & Mason, L. 1991, Classical and Quantum Gravity, 8, 41


Marsden, J. E., Patrick, G. W., & Shkoller, S. 1998, Communications in Mathematical Physics, 199, 351

Marsden, J. E., & West, M. 2001, Acta Numerica, 357

Miller, M. A. 1995, Classical and Quantum Gravity, 12, 3037


Peldan, P. 1994, Class. Quantum Grav., 11, 1087, arXiv:gr-qc/9305011


Wald, R. M. 1984, General Relativity (The University of Chicago Press)


Yoshida, H. 1993, Celestial Mechanics and Dynamical Astronomy, 56, 27