GORDON BELOT

THE REPRESENTATION OF TIME AND CHANGE IN MECHANICS

ABSTRACT: This chapter is concerned with the representation of time and change in classical (i.e., non-quantum) physical theories. One of the main goals of the chapter is to attempt to clarify the nature and scope of the so-called problem of time: a knot of technical and interpretative problems that appear to stand in the way of attempts to quantize general relativity, and which have their roots in the general covariance of that theory. The most natural approach to these questions is via a consideration of more clear cases. So much of the chapter is given over to a discussion of the representation of time and change in other, better understood theories, starting with the most straightforward cases and proceeding through a consideration of cases that lead up to the features of general relativity that are responsible for the problem of time.

Keywords: Classical Mechanics; General Relativity; Symmetry; Time

Draft version 2.0. Forthcoming in J. Butterfield and J. Earman (eds.), *Handbook of the Philosophy of Physics*. North-Holland.

CONTENTS

If time is objective the physicist must have discovered that fact, if there is Becoming the physicist must know it; but if time is merely subjective and Being is timeless, the physicist must have been able to ignore time in his construction of reality and describe the world without the help of time. . . . If there is a solution to the philosophical problem of time, it is written down in the equations of mathematical physics.

Perhaps it would be better to say that the solution is to be read between the lines of the physicist's writings. Physical equations formulate specific laws . . . but philosophical analysis is concerned with statements about *the equations rather than with the content of the equations themselves.*

—Reichenbach.[1]

For many years I have been tormented by the certainty that the most extraordinary discoveries await us in the sphere of Time. We know less about time than about anything else.

—Tarkovsky.[2]

1 INTRODUCTION

This chapter is concerned with the representation of time and change in classical (i.e., non-quantum) physical theories. One of the main goals of the chapter is to attempt to clarify the nature and scope of the so-called problem of time: a knot of technical and interpretative problems that appear to stand in the way of attempts to quantize general relativity, and which have their roots in the general covariance of that theory.

The most natural approach to these questions is via consideration of more clear cases. So much of the chapter is given over to a discussion of the representation of time and change in other, better understood theories, starting with the most straightforward cases and proceeding through a consideration of cases that prepare one, in one sense or another, for the features of general relativity that are responsible for the problem of time.

Let me begin by saying a bit about what sort of thing I have in mind in speaking of the representation of time and change in physical theories, grounding the discussion in the most tractable case of all, Newtonian physics.

As a perfectly general matter, many questions and claims about the content of a physical theory admit of two construals—as questions about structural features of solutions to the equations of motion of the theory, or as questions about structural features of these equations. For instance, on the one hand time appears as an aspect of the spacetimes in which physics unfolds—that is, as an aspect of the background in which the solutions to the equations of the theory are set. On the other, time is represented via its role in the laws of physics—in particular, in its

[1] [1991, pp. 16 f.].

[2] [1991, p. 53].

role in the differential equations encoding these laws. So questions and claims about the nature of time in physical theories will admit of two sorts of reading.

Consider, for instance, the claim that time is homogeneous in Newtonian physics (or, as Newton would put it, that time flows equably). There are two sorts of fact that we might look to as grounding this claim.

1. There is a sense in which time is a separable aspect of the spacetime of Newtonian physics and there is a sense in which time, so considered, is homogenous.[3]

2. The laws of the fundamental-looking theories of classical mechanics (e.g., Newton's theory of gravity) are time translation invariant—the differential equations of these theories do not change their form when the origin of the temporal coordinate is changed—so the laws of such theories are indifferent to the identity of the instants of time.

In the Newtonian setting, these two sorts of considerations mesh nicely and provide mutual support: there is a consilience between the symmetries of the laws and the symmetries of spacetime. But in principle, the two sorts of consideration need not lead to the same sort of answer: one might consider a system in Newtonian spacetime that is subject to time-dependent forces; or one could set the Newtonian n-body problem in a spacetime which a featured a preferred instant, but otherwise had the structure of Newtonian spacetime. And as one moves away from the familiar setting of Newtonian physics, it becomes even more important to distinguish the two approaches: in general relativity, the laws have an enormous (indeed, infinite-dimensional) group of symmetries while generic solutions have no symmetries whatsoever.

In discussing the representation of time and change, this chapter will focus on structural features of the laws of physical theories rather than on features of particular solutions. To emphasize this point, I will say that I am interested in the structure of this or that theory as a *dynamical theory*.

I will approach my topics via the Lagrangian and Hamiltonian approaches to classical theories, two great over-arching—and intimately related—frameworks in which such topics are naturally addressed.[4] Roughly speaking, in each of these approaches the content of the equations of a theory is encoded in certain structures

[3](Neo)Newtonian spacetime is partitioned in a natural way by instants of absolute simultaneity, and time can be identified with the structure that the set of these instants inherits from the structure of spacetime: time then has the structure of an affine space modelled on the real numbers—so for any two instants, there is a temporal symmetry which maps one to the other.

[4]Why pursue our question within the realm of Lagrangian and Hamiltonian mechanics rather than working directly with the differential equations of theories? Because the benefits are large: these over-arching approaches provide powerful mathematical frameworks in which to compare theories. And because the costs are minimal: almost every theory of interest can be put into Lagrangian or Hamiltonian form, without any obvious change of content. And because it leads us where we want to go: current attempts to understand the content of classical physical theories are necessarily shaped by efforts to construct or understand deeper, quantum theories; and it appears that a classical theory *must* be placed in Lagrangian or Hamiltonian form in order to be quantized.

on a space of possibilities associated with the theory.[5] In the Lagrangian approach the featured space is the space of solutions to the equations of the theory, which for heuristic purposes we can identify with the space of possible worlds allowed by the theory.[6] On the Hamiltonian side, the featured space is the space of initial data for the equations of the theory, which we can in the same spirit identify with the space of possible instantaneous states allowed by the theory.

In Newtonian mechanics, the reflection within the Lagrangian framework of the time translation invariance of the laws is that the space of solutions is itself invariant under time translations: given a set of particle trajectories in spacetime obeying Newton's laws of motion, we can construct the set of particle trajectories that result if all events are translated in time by amount t; the latter set is a solution (i.e., is permitted by the laws of motion) if and only if the former set is; furthermore, the map that carries us from a solution to its time translate preserves the structure on the space of solutions that encodes the dynamics of the theory. Within the Hamiltonian framework, on the other hand, the time translation invariance of the laws is reflected by the existence of a map that sends an initial data set to the state it will evolve into in t units of time; again, this map leaves invariant the structure on the space that encodes the dynamics of the theory. So the temporal symmetry of the dynamics of the theory is reflected on the Lagrangian side by a notion of time translation and on the Hamiltonian side by a notion of time evolution.

The representation of change in Newtonian physics also takes different (but closely related) forms within the Lagrangian and Hamiltonian frameworks. Change consists in a system having different and incompatible properties at different times. We want to say, for instance, that there is a change in the observable properties of a two-body system if and only if the relative distance between the particles is different at different times.

Hamiltonian Approach. Specifying the instantaneous dynamical state of such a system suffices to specify the instantaneous relative distance between the particles. So there is a function on the space of initial data corresponding to this quantity. A history of the system is a trajectory through the space of initial data. In our simple example, observable change occurs during a given history if and only if the function corresponding to the relative distance between the particles takes on different values at different points on the trajectory in question. More generally, in any Newtonian system, any quantity of physical interest (observable or not) is represented by a function on the space of initial data, and a trajectory in this space represents such quantities as changing if the corresponding functions take on different

[5] See remark 2 the sense in which talk of possibilia is intended here.

[6] In the context of ordinary classical mechanics, one often thinks of Lagrangian mechanics as set in the velocity phase space—and thus as more closely associated with a space of initial data rather than with a space of solutions. However, this familiar approach presupposes an absolute notion of simultaneity, and for this reason it is usually dropped in favour of a spacetime covariant Lagrangian approach (in which the space of initial data plays no role) when one turns to relativistic theories. This is the point of view adopted below.

values at different points on the trajectory.

Lagrangian Approach. Clearly, no function on the space of solutions can represent a changeable quantity in the same direct way that functions on the space of initial data can. But for each t, there is a function on the space of solutions of our two-body problem that assigns to each solution the relative distance between the particles at time t according to that solution. Letting t vary, we construct a one-parameter family of functions on the space of solutions. A solution to the equations of motion represents the relative distance between the particles as changing if and only if different members of this one-parameter family of functions take on different values when evaluated on the given solution. And so on more generally: any changeable physical quantity corresponds to such a one-parameter family of functions on the space of solutions, and change is understood as in the simple two-body example.

So much for the sort of thing I have in mind in speaking of the representation of time and change in a physical theory. Before sketching the path that this chapter takes in discussing these topics, it will perhaps be helpful to say a bit about its ultimate goal—the clarification of the nature of the so-called problem of time. Discussions of the problem of time typically focus on Hamiltonian versions of general relativity, in which the focus is on the space of possible instantaneous geometries (metrics and second fundamental forms on Cauchy surfaces). This is somewhat unfortunate, since such approaches require from the start a division of spacetime into a family of spacelike hypersurfaces—which appears to be against the spirit of the usual understanding of the general covariance of the theory. In light of this fact, there is room for worry that some aspects of the problem of time as usually presented are consequences of this rather awkward way of proceeding. I take a somewhat different path, always anchoring my discussion in the Lagrangian approach, which takes as fundamental complete histories of systems rather than instantaneous states.

The view developed below is that, *roughly speaking*, the core of the problem of time is that in general relativity, when understood dynamically, there is no way to view time evolution or time translation as symmetries of the theory and, relatedly, there is no natural way to model change via functions on the spaces arising within the Lagrangian and Hamiltonian approaches.[7] This marks a respect in which gen-

[7] This formulation above only gives us a first approximation, for several reasons (each of which will be discussed more fully in following sections). (i) The problem of time only arises in those versions of general relativity most appropriate to the cosmological setting; in other applications of the theory, time is represented in a fashion very similar to that in which it is represented in special relativistic physics. (ii) In the treatment of ordinary time-dependent systems, time evolution and time translation are not symmetries of the theory—but this does not lead to any real problem in representing time and change in such theories, because one still has group actions that implement time evolution and time translation, even though they are not symmetries of the laws, and these suffice to erect an account of change very similar to that occurring in ordinary time-independent theories. (iii) In theories in which solutions are not defined globally in time, time evolution and time translation will not be implemented

eral relativity, so conceived, is very different from preceding fundamental-looking theories.

The problem of time may sound—not very pressing. To be sure, there are puzzles here. Why should general relativity differ in this way from its predecessors? In predecessors to general relativity, the representation of time and the representation of change are tied together in a very neat package—what does the general relativistic replacement for this package look like? These are interesting questions. But then of course no one should expect time to be represented in general relativity as in its predecessors—that it presents an utterly new picture of time and space is one of the glories of the theory. And one might also think: since the structure of spacetime varies from solution to solution in general relativity, it is surely more appropriate to look at the representation of time in this or that physically realistic solution, rather than in the equations of the theory, if we want to understand what the theory is telling us about the nature of time in our world.

The problem of time assumes a more pressing aspect, however, when one considers the quantization of general relativity (or of any other theory that is generally covariant in the relevant sense). The project of constructing successor theories naturally focuses our attention on structural features of the theories at hand—in constructing successors, one is in the business of laying bets as to which such features of current theories will live on (perhaps in a new form), and which ones will be left behind. And known techniques of quantization require as input not just differential equations, but theories cast in Hamiltonian or Lagrangian form. So for those interested in quantizing general relativity, questions about the structure of the theory *qua* dynamical theory naturally loom large. And lacking solutions to the puzzles mentioned above, one expects conceptual difficulties in formulating (or extracting predictions from) any quantization of general relativity. So from this perspective, the problem of time is in fact quite pressing.

This chapter takes long route to the problem of time. I begin in section 2 with the briefest of introductions to Hamiltonian and Lagrangian mechanics, by way of motivating some of what follows. In section 3, I sketch some important concepts and results of symplectic geometry, the field of mathematics that underlies classical mechanics. The concepts introduced here are crucial for what follows: for well-behaved theories, the space of solutions (on the Lagrangian side) and the space of initial data (on the Hamiltonian side) both have symplectic structures. And we will see that various symplectic (or nearly symplectic) spaces arise even when one strays away from the ideal case. In section 4, I sketch the very powerful framework of modern Lagrangian mechanics, with its apparatus of local conservation laws.

In section 5, I sketch the Lagrangian and Hamiltonian pictures for ideally well-behaved theories satisfying the following conditions: (i) the background spacetime geometry admits a group of time translations and the Lagrangian of the theory is

---

by group actions, but merely by local flows (these can be thought of as infinitesimal surrogates for group actions); these suffice for the construction of the familiar picture of change in such theories; but even these are absent in general relativity.

invariant (in a suitable sense) under the action of this group; (ii) specifying initial data for the equations of the theory suffices to determine a single maximal solution; (iii) this maximal solution is defined for all values of the time parameter. When these conditions hold, we find that there is a group of time translation symmetries operating on the space of solutions on the Lagrangian side, while on the Hamiltonian side there is a group implementing time evolution on the space of initial data. These two spaces are isomorphic, and the two group actions intertwine in a satisfying fashion. One is able to give a straightforward and appealing account of the way in which change is represented on either of the two fundamental spaces.

In section 6, I turn to the complications that must be introduced into the picture when one drops any one of the conditions (i)–(iii) of the preceding paragraph. Finally in section 7, I address the representation of time and change in general relativity. This leads directly to the problem of time.

As this outline makes clear, much of the chapter is given over to exposition of technical material. In order to keep the length reasonable, I have had to presume that the reader comes to this chapter with quite a bit of technical background. I have tried to write for an ideal reader who has previously studied general relativity or gauge theory, and hence feels comfortable with the basic concepts, results, and constructions of differential geometry (although at a few strategic points I have included discussion meant to jog the memory of such readers).

This chapter is founded upon the modern geometrical approach to Lagrangian mechanics that is presented in the barest sketch in section 4. This approach, developed relatively recently by mathematicians, provides a highly abstract framework for thinking about physical theories rather than a fully rigourous treatment of any given theory. It exists at the formal, differential-geometric level: the focus is on the geometric structure of various spaces and on the geometric content of equations and constructions; functional analytic are held in abeyance. Much of the material sketched in other sections functions at this same level.

In content, this chapter overlaps somewhat with [Malament, this volume], [Rovelli, this volume], and [Brading and Castellani, this volume]. But it is most closely related to [Butterfield, this volume]. Butterfield's chapter provides a philosophical introduction to modern geometric approaches to mechanics; the present chapter is meant as an example of the application of this approach to a philosophical problem. The present chapter is, however, meant to be self-contained. And there is in fact a considerable difference in emphasis between this chapter and Butterfield's: the latter is restricted to finite-dimensional systems, and focuses on the Hamiltonian side of things; the present chapter is primarily concerned with field theories, and focuses to a much greater extent on the Lagrangian approach.

REMARK 1 (Notation and Terminology). Elements of and structures on the space of solutions of a theory are always indicated by capital letters (Greek or Latin) while elements of and structures on the space initial data of a theory are always indicated by lower case letters (Greek or Latin). Boldface indicates three-vectors or three-vector-valued functions. In this chapter, a curve is officially a map from intervals of real numbers into a space that is a manifold or a mild generalization of a manifold—sometimes for emphasis I redundantly call a curve a *parameterized curve*. An *affinely parameterized curve* is a equivalence class of such curves, where two curves count as equivalent if they have the same image and their parameterization agrees up to a choice of origin.[8] A *unparameterized curve* is an equivalence class of curves, under the equivalence relation where curves count as equivalent if they have the same image. I sometimes conflate a curve and its image.

REMARK 2 (Possible Worlds Talk). Below, especially in section 7, I sometimes speak of points of the space of solutions (initial data) as representing possible worlds (possible instantaneous states) permitted by the theory, even though I do not pretend to be involved in fine-grained matters of interpretation here. This sort of thing is meant only in a rough and heuristic way. The idea is that in trying to understand a theory, we are in part engaged in a search for a perspicuous formulation of the theory; and it is reasonable to hope that if a formulation is perspicuous, then there will exist a *prima facie* attractive interpretation of the theory according to which there is a bijection between the space of solutions (initial data) and the space of possible worlds (possible instantaneous states) admitted by the theory under that interpretation. This is not to deny that there may be reasons for ultimately rejecting such interpretations: a Leibnizean might settle on a standard formulation of classical mechanics, even though that means viewing the representation relation between solutions and possible worlds as many-to-one in virtue of the fact that solutions related by a time translation must be seen as corresponding to the same possible world.

2 HAMILTONIAN AND LAGRANGIAN MECHANICS

This section contains a *very* brief sketch of the Hamiltonian and Lagrangian approaches to the Newtonian n-body problem.[9] The intended purpose is to motivate some of what follows in later sections.

2.1 The n-Body Problem

We consider n gravitating point-particles. Let the mass of the ith particle be m_i. Working relative to a fixed inertial frame we write: $q := (\mathbf{q}_1, \dots, \mathbf{q}_n) = (q_1, \dots, q_{3n})$ for the positions of the particles, $\dot{q} := (\dot{\mathbf{q}}_1, \dots, \dot{\mathbf{q}}_n) = (\dot{q}_1, \dots, \dot{q}_{3n})$ for their velocities, and $\ddot{q} := (\ddot{\mathbf{q}}_1, \dots, \ddot{\mathbf{q}}_n) = (\ddot{q}_1, \dots, \ddot{q}_{3n})$ for their accelerations (in this chapter, boldface always indicates a three-vector). The gravitational force

[8] That is, an affinely parameterized curve is an equivalence class of curves under the equivalence relation according to which curves $\gamma_1 : [a, b] \to M$ and $\gamma_1 : [a, b] \to M$ are equivalent if and only if there exists $s \in \mathbb{R}$ such that $\gamma_1(t) = \gamma_2(t + s)$ for all $t \in [a, b]$.

[9] For textbooks approaching classical mechanics in a variety of styles, see, e.g., [Goldstein, 1953], [Lanczos, 1986], [Singer, 2001], [Marsden and Ratiu, 1994], [Arnold, 1989], [Arnold *et al.*, 1997], and [Abraham and Marsden, 1978].

exerted on the ith particle by the jth particle is

$$\mathbf{F}_{ij} = \frac{m_i m_j}{r_{ij}^2}\mathbf{u}_{ij}, \tag{1}$$

where r_{ij} is the distance between the ith and jth particles, $\mathbf{u}_{ij}$ is the unit vector pointing from the ith to the jth particle, and units have been chosen so that Newton's constant is unity. Of course, equation 1 is not well-defined for $r_{ij} = 0$. So from now on we assume that $q \in Q := \mathbb{R}^{3n}/\Delta$, where Δ is the *collision set* $\{q \in \mathbb{R}^{3n} : \mathbf{q}_i = \mathbf{q}_j$ for some $i \neq j\}$.

The net force acting on the ith particle is

$$\mathbf{F}_i = \sum_{j \neq i} \mathbf{F}_{ij}.$$

So the *equations of motion* for our theory are: $\mathbf{F}_i = m_i \ddot{\mathbf{q}}_i$.[10] Resolving each force and acceleration vector into its components, we have $3n$ second-order differential equations. Roughly speaking, these equations have a well-posed initial value problem: specifying $3n$ values for the initial positions of our particles and $3n$ values for their initial velocities (momenta) determines a unique analytic solution to the equations of motion, which tells us what the positions and velocities (momenta) of the particles are at all other times at which these quantities are defined.[11]

2.2 The Hamiltonian Approach

The basic variables of the Hamiltonian approach are the positions of the particles and the corresponding momenta, $p := (m_1\dot{\mathbf{q}}_1, \ldots, m_n\dot{\mathbf{q}}_n) = (m_1 q_1, \ldots, m_n q_{3n})$. A state of the system, (q, p), is specified by specifying the position and momentum of each particle. To each state we can assign a *kinetic energy*

$$T(q, p) := \frac{1}{2m_i}\sum_{i=1}^{n} |\mathbf{p}_i|^2$$

and a *potential energy*,

$$V(q, p) := \sum_{i<j} \frac{m_i m_j}{r_{ij}}.$$

Note that $\mathbf{F}_i = -\nabla_i V(q)$, where ∇_i is the gradient operator $(\frac{\partial}{\partial q_{3i-2}}, \frac{\partial}{\partial q_{3i-1}}, \frac{\partial}{\partial q_{3i}})$ corresponding to the configuration variables of the ith particle. So the potential energy encodes information about gravitational forces, while the kinetic energy can be thought of as encoding information about the inertial structure of Newtonian

[10]This should be read as a differential equation constraining the allowed trajectories $q(t)$. Similarly for the other differential equations appearing in this section.

[11]Note that some solutions fail to be defined for all values of t; see example 33 below for discussion.

spacetime. So one might hope that together these quantities encode all of the physics of the n-body problem. This is indeed the case.

We introduce the space of initial data for the theory, $\mathcal{I} := \{(q,p)\colon q \in Q\}$ and the Hamiltonian $H : \mathcal{I} \to \mathbb{R}$, $H(q,p) := T(q,p) + V(q)$. The Hamiltonian is thus just the total energy.

The original equation of motion $m_i\ddot{\mathbf{q}}_i = \mathbf{F}_i$ can be rewritten as $\dot{\mathbf{p}}_i = -\nabla_i V(q)$; or, since $\nabla_i T = 0$, as $\dot{\mathbf{p}}_i = -\nabla_i H$. In another notation, this becomes $\dot{p}_i = -\frac{\partial H}{\partial q_i}$. Furthermore, since the only term in H depending on p_i is of the form $\frac{1}{2m}p_i^2$, we find that $\frac{\partial H}{\partial p_i} = \dot{q}_i$.

In this way, we move from the original Newtonian equations to *Hamilton's equations*:

$$\dot{q}_i = \frac{\partial H}{\partial p_i} \quad \text{and} \quad \dot{p}_i = -\frac{\partial H}{\partial q_i} \quad (i = 1, \ldots, 3n).$$

In fact, the ordinary Newtonian equations are equivalent to Hamilton's equations. So we see that the function $H = T + V$ encodes all of the dynamical content of the n-body problem.

Our present interest is in the geometry implicit in Hamilton's equations. Hamilton's equations gives us values for $\dot{q}_i(q,p)$ and $\dot{p}_i(q,p)$ at each point $(q,p) \in \mathcal{I}$. That is, Hamilton's equations give us a component expression for a tangent vector $X_H(q,p)$ at each point $(q,p) \in \mathcal{I}$. The vector field X_H on $\mathcal{I}$ encodes the dynamics of our theory: through each point $(q_0,p_0) \in \mathcal{I}$ there is exactly one curve $(q(t),p(t)) : \mathbb{R} \to \mathcal{I}$ such that: (i) $(q(0),p(0)) = (q_0,p_0)$; and (ii) for each s, the tangent vector to the curve $(q(t),p(t))$ at $t = s$ is given by $X_H(q(s),p(s))$. This curve tells us that if the system is in state (q_0,p_0) at time $t = 0$, then it is in state $(q(s),p(s))$ at time $t = s$.

We can rewrite Hamilton's equations as:

$$(\dot{q}_1, \ldots, \dot{q}_{3n}, \dot{p}_1, \ldots, \dot{p}_{3n}) \begin{vmatrix} 0 & I \\ -I & 0 \end{vmatrix} = \left(\frac{\partial H}{\partial q_1}, \ldots, \frac{\partial H}{\partial q_{3n}}, \frac{\partial H}{\partial p_1}, \ldots \frac{\partial H}{\partial p_{3n}} \right),$$

where I is the $3n \times 3n$ identity matrix. On the left hand side we have a vector multiplied by a matrix; on the right hand side another vector. Thinking of $\mathcal{I}$ as a manifold, we can recognize the coordinate-independent objects standing behind this equation: on the left we have the tangent vector field X_H contracted with a two-form; on the right, the differential dH (i.e., the exterior derivative of H). So we can re-write Hamilton's equations in a coordinate-independent form as:

$$\omega(X_H, \cdot) = dH,$$

where ω is the two-form on $\mathcal{I}$ that assumes the form $\sum_i dq_i \wedge dp_i$ in our coordinates.

ω is a *symplectic form* on $\mathcal{I}$: a closed, nondegenerate two-form.[12] ω can be thought of as being somewhat like an anti-symmetric metric on $\mathcal{I}$ (e.g., both sorts

[12] See section 3.2 below for further discussion and for an unpacking of this definition.

of object establish a preferred isomorphism between vector fields and one-forms). But that analogy cannot be taken too seriously in light of the following striking differences between the two sorts of objects:

1. The isometry group of a finite-dimensional Riemannian manifold is always finite-dimensional. But our symplectic form is invariant under an infinite-dimensional family of diffeomorphisms from $\mathcal{I}$ to itself. We can see this as follows. Let us think of $\mathcal{I}$ as the cotangent bundle of Q; that is, we think of a point (q, p) as consisting of a point $q \in Q$ and a covector $p \in T_q^*Q$.[13] A *cotangent coordinate system* on $\mathcal{I} = T^*Q$ arises as follows: choose arbitrary coordinates $\{q_i\}$ on Q and write $p \in T_q^*Q$ as $p = \sum p_i dq^i$, so that $\{q_i, p_j\}$ forms a set of coordinates on T^*Q. In any cotangent coordinate system,
$$\omega = \begin{vmatrix} 0 & I \\ -I & 0 \end{vmatrix}. \tag{2}$$
So ω is invariant under the transformation that carries us from one set of cotangent coordinates on $\mathcal{I}$ to another. And the set of such transformations is infinite-dimensional, since any diffeomorphism $d : Q \to Q$ generates such a transformation.
2. One does not expect any manifold or bundle to carry a natural Riemannian metric. But if M is any finite-dimensional manifold, the cotangent bundle T^*M carries a canonical symplectic form, ω, that takes the form $\omega = \sum_i dq_i \wedge dp_i$ relative to any set of local cotangent coordinates on M.[14]
3. If (M, g) and (M', g') are n-dimensional Riemannian manifolds, then for any $x \in M$ and $x' \in M'$, we know that g and g' endow the tangent spaces T_xM and $T_{x'}M'$ with the same geometry; but in general we expect that no diffeomorphism $d : M \to M'$ will give an isometry between a neighbourhood of x and a neighbourhood of x'. But the Darboux theorem tells us that if (M, ω) is a finite-dimensional manifold equipped with a symplectic form, then (M, ω) is locally isomorphic to $T^*\mathbb{R}^n$ equipped with its canonical cotangent bundle symplectic form. An immediate corollary is that every finite-dimensional symplectic manifold is even-dimensional.

Of course, for present purposes, the interest in identifying the symplectic structure lying behind the Hamiltonian version of the n-body problem lies in generalization. (1) Note that if we are interested in n particles interacting via forces that arise from a potential energy function V, as above, then we can construct a Hamiltonian treatment equivalent to the usual Newtonian one by again taking $\mathcal{I}$ as the

[13] Why regard p as a covector rather than a tangent vector? Because in general the momentum p of a system with Lagrangian L is defined as $\frac{\partial L}{\partial \dot{q}}$, which transforms as a covariant quantity under change of coordinates on Q.

[14] Where as above, a set of local coordinates on M induces a natural set of cotangent coordinates on T^*M. In example 7 below we will see a coordinate-free version of this construction that carries over to the infinite-dimensional case.

space of initial data, equipping it with the symplectic form ω as above, defining a Hamiltonian $H : \mathcal{I} \to \mathbb{R}$ as the sum of the kinetic and potential energies, and taking as our dynamical trajectories the integral curves of the vector field X_H on $\mathcal{I}$ that solves $\omega(X_H, \cdot) = dH$. (2) More generally, we can model a vast number of classical mechanical systems as follows: let the space of initial state be a symplectic manifold (M, ω) (not necessarily a cotangent bundle) and let a Hamiltonian $H : M \to \mathbb{R}$ be given; then let the dynamics be given by the vector field X_H solving $\omega(X_H, \cdot) = dH$.

2.3 The Lagrangian Approach

It is helpful to approach the Lagrangian version of the n-body problem somewhat indirectly.[15]

Critical Points in Calculus

For $f : \mathbb{R}^n \to \mathbb{R}$, the differential of f is given by $df = (\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n})$. We say that f has a *critical point* at $x_0 \in \mathbb{R}^n$ if $df(x_0) = 0$; i.e., f has a critical point at x_0 if $df(x_0) \cdot e = 0$ for each $e \in \mathbb{R}^n$ (since $\mathbb{R}^n$ is a linear space, we can identify $T_{x_0}\mathbb{R}^n$ with $\mathbb{R}^n$ itself and let $e \in \mathbb{R}^n$ here). There are a number of helpful ways of thinking of $df(x_0) \cdot e$: (i) this quantity coincides with the *directional derivative* of f at x_0 in direction e,

$$df(x_0) \cdot e = \lim_{t \to 0} \frac{f(x_0 + te) - f(x_0)}{t};$$

(ii) if we have a curve $\gamma : \mathbb{R} \to \mathbb{R}^n$ with $\gamma(0) = x_0$ and $\dot{\gamma}(0) = e$, then $df(x_0) \cdot e = \frac{d}{dt} f(\gamma(t)) \mid_{t=0}$.

The Calculus of Variations and the Euler–Lagrange Equations

We now consider an infinite-dimensional analog: we look for critical points of a function defined on a space of curves in Euclidean space. This is the foundation of the Lagrangian approach to particle mechanics.

Let $Q = \mathbb{R}^n$, let $[a, b] \subset \mathbb{R}$ be a closed interval, and let $x, y \in Q$. Let $\Gamma(a, b; x, y)$ be the set of C^2 curves $\gamma : [a, b] \to Q$ with $\gamma(a) = x$ and $\gamma(b) = y$. And let $\Gamma(a, b; 0, 0)$ be the space of C^2 curves $\gamma : [a, b] \to Q = \mathbb{R}^n$ with $\gamma(a) = (0, \dots, 0)$ and $\gamma(b) = (0, \dots, 0)$. Both $\Gamma(a, b; x, y)$ and $\Gamma(a, b; 0, 0)$ are well-behaved infinite-dimensional spaces.[16] For $\gamma \in \Gamma(a, b; x, y)$ we can think

[15] For introductions to the Lagrangian approach via the calculus of variations, see [Dubrovin *et al.*, 1992, Chapter 6], [Lanczos, 1986, Chapters II and V], and [van Brunt, 2004]. For some of the rigourous underpinnings of the calculus of variations see, e.g., [Choquet-Bruhat *et al.*, 1977, §§II.A and II.B] and [Choquet-Bruhat and DeWitt-Morrette, 1989, §II.3].

[16] Let $\Gamma(a, b)$ be the space of C^2 curves $\gamma : [a, b] \to Q$. This is a linear space under pointwise addition (i.e., $(\gamma + \gamma')(x) = \gamma(x) + \gamma'(x)$) that can be made into a Banach space in a number of ways. $\Gamma(a, b; 0, 0)$ is a linear subspace of $\Gamma(a, b)$ while $\Gamma(a, b; x, y)$ is an affine subspace modelled on $\Gamma(a, b; 0, 0)$.

of $\Gamma(a, b; 0, 0)$ as $T_\gamma\Gamma(a, b; x, y)$ (think of $h \in \Gamma(a, b; 0, 0)$ as describing a vector field along γ).[17]

The tangent bundle of Q is $TQ = \mathbb{R}^{2n}$. Let $L : TQ \to \mathbb{R}$ be a smooth function. This allows us to define a function $I_{a,b} : \Gamma(a, b; x, y) \to \mathbb{R}$ by $I_{a,b}(\gamma) := \int_a^b L(\gamma(t), \dot{\gamma}(t))\, dt$. We are interested in finding the critical points of $I_{a,b}$: these will be points in $\Gamma(a, b; x, y)$ (that is, curves $\gamma : [a, b] \to Q$) of special interest. Like any function on a well-behaved space, $I_{a,b}$ has a differential, which we denote $\delta I_{a,b}$; this can be thought of as one-form on $\Gamma(a, b; x, y)$.

DEFINITION 3 (Stationary Curves). We say that $\gamma : [a, b] \to Q$ is *stationary for* L *over* $[a, b]$ if $\delta I_{a,b}(\gamma) = 0$. We say that $\gamma : \mathbb{R} \to Q$ is *stationary for* L if its restriction to $[a, b]$ is stationary over $[a, b]$ for all closed intervals $[a, b]$.

As in the case of an ordinary function on $\mathbb{R}^n$, $\delta I_{a,b}(\gamma) = 0$ if and only if $\delta I_{a,b}(\gamma) \cdot h = 0$ for all $h \in T_\gamma\Gamma(a, b; x, y) = \Gamma(a, b; 0, 0)$. We can then calculate $\delta I_{a,b}(\gamma) \cdot h$ by finding $\frac{d}{d\varepsilon} I_{a,b}(\gamma[\varepsilon]) \mid_{\varepsilon=0}$ for $\gamma[\varepsilon]$ a curve in $\Gamma(a, b; x, y)$ with $\gamma[0] = \gamma$ and $h = \frac{d}{d\varepsilon}\gamma[\varepsilon] \mid_{\varepsilon=0}$.

Let us calculate. Fix L and $[a, b]$. Let $\gamma \in \Gamma(a, b; x, y)$ and $h \in \Gamma(a, b; 0, 0)$. For each ε in some sufficiently small neighbourhood of zero, we define a curve $\gamma[\varepsilon] : \mathbb{R} \to Q$ by $\gamma[\varepsilon](t) := \gamma(t) + \varepsilon h(t)$. So $\gamma[\varepsilon]$ is a curve in $\Gamma(a, b; x, y)$ with $\gamma[0] = \gamma$ and with tangent $h = \frac{d}{d\varepsilon}\gamma[\varepsilon] \mid_{\varepsilon=0}$. Then:

$$
\begin{aligned}
\delta I_{a,b}(\gamma) \cdot h &= \frac{d}{d\varepsilon} I_{a,b}(\gamma[\varepsilon])|_{\varepsilon=0} \\
&= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_a^b \left[L(\gamma[\varepsilon](t), \dot{\gamma}[\varepsilon](t)) - L(\gamma(t), \dot{\gamma}(t))\right] dt \\
&= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left(\int_a^b \varepsilon \left[\frac{\partial L}{\partial x}(\gamma(t), \dot{\gamma}(t) \cdot h(\gamma(t)) \right.\right. \\
&\quad \left.\left. + \ \frac{\partial L}{\partial \dot{x}}(\gamma(t), \dot{\gamma}(t) \cdot \dot{h}(\gamma(t)) \right] + O(\varepsilon^2)\, dt \right) \\
&= \int_a^b \frac{\partial L}{\partial x} \cdot h\, dt - \int_a^b \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) \cdot h\, dt + \left(h \frac{\partial L}{\partial \dot{x}} \right)\Bigg|_a^b .
\end{aligned}
$$

The first equality follows from a basic fact about the differential of a function; the second follows by definition; the third via Taylor's theorem; the fourth via an integration by parts. We now note that since h vanishes at $\gamma(a) = x$ and $\gamma(b) = y$

[17] We can think of $T_\gamma\Gamma(a, b; x, y)$ as being built as follows: one considers one-parameter family $\gamma[\varepsilon] : \varepsilon \in \mathbb{R} \mapsto \gamma[\varepsilon] \in \Gamma(a, b; x, y)$ of curves with $\gamma[0] = \gamma$, and declares such one-parameter families, $\gamma[\varepsilon]$ and $\gamma'[\varepsilon]$, to be equivalent if $\frac{d}{d\varepsilon}\gamma[\varepsilon] \mid_{\varepsilon=0} = \frac{d}{d\varepsilon}\gamma'[\varepsilon] \mid_{\varepsilon=0}$; $T_\gamma\Gamma(a, b; x, y)$ is the resulting space of equivalence classes. We construct a bijection between $\Gamma(a, b; 0, 0)$ and $T_\gamma\Gamma(a, b; x, y)$ thought of as the space of such equivalence classes by associating with $h \in \Gamma(a, b; 0, 0)$ the equivalence class containing $\gamma[\varepsilon] : \varepsilon \mapsto \gamma + \varepsilon \cdot h$.

the third term in the final line vanishes. So

$$\delta I_{a,b}(\gamma_0) \cdot h = \int_a^b \left[\frac{\partial L}{\partial x} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}} \right) \right] h \, dt.$$

To say that $\gamma : [a, b] \to Q$ is stationary over $[a, b]$ is to say that this expression vanishes for each h. So the condition that γ is stationary for L over $[a, b]$ is that the *Euler–Lagrange equation*

$$\frac{\partial L}{\partial x} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}} = 0 \tag{3}$$

holds along $\gamma(t)$ for $t \in [a, b]$. And the condition that $\gamma : [a, b] \to \mathbb{R}$ is stationary for L is just that the equation 3 is satisfied all along γ.

REMARK 4 (Parsing the Euler–Lagrange Equations). Here is how to unpack equation 3.[18] Rewrite the expression for L, replacing $\dot{x}$ everywhere by ξ. Then interpret equation 3 as a differential equation for admissible trajectories $x(t)$, understanding $\frac{\partial L}{\partial \dot{x}}$ to mean $\frac{\partial L(x,\xi)}{\partial \xi_i} \,|_{\xi=\dot{x}(t)}$ and $\frac{d}{dt}\frac{\partial L}{\partial \dot{x}}$ to mean

$$\left(\frac{\partial^2 L}{\partial \xi_i \partial \xi_j} \ddot{x}_j + \frac{\partial^2 L}{\partial \xi_i \partial x_j} \dot{x}_j \right) \bigg|_{\xi=\dot{x}(t)}.$$

Hamilton's Principle

Consider a physical system with configuration space Q (i.e., consider a system whose possible spatial positions are parameterized by Q). Let the kinetic energy be a function $T : TQ \to \mathbb{R}$ that arises from a Riemannian metric g on Q via $T(x, v) := g_x(v, v)$ and let $V : Q \to \mathbb{R}$ be a potential for a force depending on the configuration degrees of freedom alone. Then the Lagrangian for the system is $L(x, v) := T(x, v) - V(x)$. Hamilton's principle states that the stationary curves for L are the physically possible trajectories. Many physically interesting systems can be cast in this form—e.g., the n-body problem considered above. For such systems that Euler–Lagrange equations derived from Hamilton's principle are equivalent to the usual Newtonian equations of motion.

Symplectic Structure of the Space of Solutions of the Euler–Lagrange Equations

Let Q be a manifold and TQ its tangent bundle. Let $L : TQ \to \mathbf{R}$ be a smooth function. The space, $\mathcal{S}$, of stationaries of L has a natural manifold structure: for those $\gamma \in \mathcal{S}$ defined at $t = 0$, we can take the values of $x = \gamma(0)$ and $v = \dot{\gamma}(0)$ relative to coordinates $\{x_i\}$ on Q as coordinates on $\mathcal{S}$; doing this for each value of t gives us a differentiable atlas for $\mathcal{S}$. It follows that dim $\mathcal{S}$ = dim TQ. We

[18]For this, see, e.g., [Dubrovin *et al.*, 1992, p. 318].

can also endow $\mathcal{S}$ with a geometric structure: consider the boundary term, $h\frac{\partial L}{\partial \dot{x}}$, discarded above in deriving the Euler–Lagrange equations; since h is to be thought of as a tangent vector to the space $\Gamma(a, b; x, y)$, we must view $\alpha = \frac{\partial L}{\partial \dot{x}}$ as a one-form on that space; taking its exterior derivative gives us a two-form, $\omega := \delta\alpha$, on $\Gamma(a, b; x, y)$; the restriction of this form to $\mathcal{S}$ is the structure we seek. In the coordinates we have introduced on $\mathcal{S}$, ω takes the form:

$$\omega = \frac{\partial^2 L}{\partial x^a \partial v^b} dx^a \wedge dx^b + \frac{\partial^2 L}{\partial v^a \partial v^b} dv^a \wedge dx^b.$$

For any L, this is a closed two-form. It is nondegenerate, and hence symplectic, so long as $\det\left[\frac{\partial^2 L}{\partial v^a \partial v^b}\right] \neq 0$.[19] For Lagrangians of the form considered above this always holds—and we then find that $(\mathcal{S}, \omega)$ is (locally) symplectically isomorphic to the corresponding space of initial data that arises from a Hamiltonian treatment of the theory.[20]

3 SYMPLECTIC MATTERS

Throughout the chapter, we are going to be investigating the representation of time and change in physical theories by asking about their representation in Lagrangian and Hamiltonian formulations of these theories. On the Lagrangian side, the focus is always on the space of solutions of the equations of our theory, while on the Hamiltonian side the focus is always on the space of initial data for those equations. It is a fact of primary importance that for well-behaved theories the space of initial data and the space of solutions share a common geometric structure—these spaces are isomorphic as symplectic manifolds. Thus the notion of a symplectic manifold and its generalizations will play a central role in our investigations.

It will be helpful to begin with a general discussion of the nature of symplectic manifolds: subsection 3.1 deals with some preliminary matters; subsection 3.2 offers a sketch of some of the basic concepts, constructions, and results of symplectic geometry as it figures in mechanics; subsection 3.3 offers the same sort of treatment of presymplectic geometry (a generalization of symplectic geometry that will play an important role in sections 6.2 and 7 below); subsection 3.4 discusses the sense in which a symplectic structure is the *sine non qua* of quantization.

3.1 Preliminaries

The spaces that we will come across below will be generalizations of ordinary n-dimensional manifolds in three respects. (i) They are allowed to be non-Haus-

[19]The symplectic structure of the space of solutions for Lagrangian theories is discussed in [Woodhouse, 1991, §§2.3 and 2.4].

[20]This follows from the fact that Lagrangians arising from kinetic and potential terms of the sort considered above are always hyperregular; see, e.g., [Abraham and Marsden, 1978, p. 226].

dorff.[21] (ii) They are allowed to be infinite-dimensional: a manifold is locally modelled on a vector space; we allow ours to be modelled on $\mathbb{R}^n$ or on an infinite-dimensional Banach space.[22] (iii) They are allowed to have mild singularities—roughly speaking, our spaces will be composed out of manifolds in the way that an ordinary cone is composed out of its apex (a zero-dimensional manifold) and mantle (a two-dimensional manifold)—but our spaces still have smooth structures and support tensors in much the same way that manifolds do.[23]

In order to avoid becoming bogged down in technicalities, I will present my sketch of the required notions and constructions of symplectic and presymplectic geometry in the context of manifolds; but when in following sections I speak of 'spaces' rather than manifolds, it should be understood that I am allowing the spaces in question to have mild singularities of the sort mentioned above.

Below we will often be interested in the actions of Lie groups on manifolds, and in vector fields as the infinitesimal generators of such actions. Let me end this discussion of preliminary matters by reviewing some pertinent definitions and constructions.

Recall that a *Lie group* is a manifold which is also a group, with the operations of group multiplication, $(g, h) \in G \times G \mapsto g \cdot h \in G$, and the taking of inverses, $g \in G \mapsto g^{-1} \in G$, as smooth maps. An *action* of a Lie group G on a manifold M is a smooth map $\Phi : G \times M \to M$ such that: (i) $\Phi(e, x) = x$ for e the identity element of G and for all $x \in M$; (ii) $\Phi(g, \Phi(h, x)) = \Phi(gh, x)$ for all $g, h \in G$ and $x \in M$. One often writes $g \cdot x$ or $\Phi_g(x)$ for $\Phi(g, x)$.[24] The *orbit* through $x \in M$ of the action is the set $[x] := \{g \cdot x : g \in G\}$. The action of a Lie group partitions a manifold into orbits.

While other Lie groups will figure below, we will most often be interested in

[21]Recall that a topological space X is *Hausdorff* if for any distinct $x, y \in X$, there exist disjoint open sets U and V with $x \in U$ and $y \in V$. While most textbooks require manifolds to be Hausdorff, all of the basic constructions and results go through without this assumption—see [Lang, 1999]. As we will see in examples 32 and 33 below, the solution spaces of even the simplest physical systems can be non-Hausdorff.

[22][Abraham *et al.*, 1988] and [Lang, 1999] provide introductions to differential geometry that cover the case of infinite-dimensional Banach manifolds. See [Milnor, 1984, §§2–4] for an introduction to a more general approach, under which manifolds are modelled on locally convex topological vector spaces. Note that the inverse function theorem and the existence and uniqueness theorem for ordinary differential equations fail under this more general approach.

[23]The spaces under consideration are Whitney stratified spaces. As suggested in the text, each such space admits a canonical decomposition into manifolds. This decomposition allows us to treat each point in such a space as lying in a manifold, which allows us to construct a space of tangent vectors and cotangent vectors at each point, and hence to construct tensors in the usual way. The dimensions of the manifold pieces (and of the tangent and cotangent spaces) will in general vary from point to point within the stratified space. See [Pflaum, 2001] or [Ortega and Ratiu, 2004, §§1.5–1.7] for a treatment of such spaces in the finite-dimensional case. The picture appears to be very similar in the infinite-dimensional examples that arise in physics: for general relativity, see [Andersson, 1989] and [Marsden, 1981, Lecture 10]; for Yang–Mills theories, see [Arms, 1981] and [Kondracki and Rogulski, 1986].

[24]Equivalently, an action of G on M is a group homomorphism $g \mapsto \phi_g$ from G to $\mathcal{D}(M)$ (the group of diffeomorphisms from M to itself) such that the map $(g, x) \in G \times M \mapsto \phi_g(x) \in M$ is smooth.

the simplest of all Lie groups: the additive group $\mathbb{R}$. A *flow* on a manifold M is a one-parameter group of diffeomorphisms from M to M. So if $\{\Phi_t\}_{t\in\mathbb{R}}$ is a flow on M, then $\Phi_0(x) = x$ and $\Phi_t \circ \Phi_s(x) = \Phi_{t+s}(x)$ for all $x \in M$ and $s, t \in \mathbb{R}$. A flow $\{\Phi_t\}$ on M and an action $\Phi : \mathbb{R} \times M \to M$ of $\mathbb{R}$ on M are more or less the same thing: given an $\mathbb{R}$-action $\Phi : \mathbb{R} \times M \to M$, one defines a flow $\{\Phi_t\}$ via $\Phi_t(x) = \Phi(t, x)$ for all $t \in \mathbb{R}$ and $x \in M$; likewise if one is given a flow and wants to define an $\mathbb{R}$-action.

Any $\mathbb{R}$-action on M induces a vector field X on M. Let $x \in M$ and consider the curve $\gamma_x(t) : \mathbb{R} \to M$ defined by $\gamma_x : t \mapsto \Phi_t(x)$. The image of γ_x in M is just the orbit $[x]$. Now suppose that $y \in [x]$—i.e., there is $t \in \mathbb{R}$ such that $y = \Phi_t(x)$. Some facts follow immediately from the group property of $\{\Phi_t\}$. We find that the image of γ_y is also $[x]$—so $[x] = [y]$. We find, in fact, that $\gamma_y(s) = \gamma_x(s + t)$ for all $s \in \mathbb{R}$; that is, each of the curves γ_y corresponding to points $y \in [x]$ agree up to choice of origin for their parameterization. So each orbit $[x]$ of our $\mathbb{R}$-action arises as the shared image of a (maximal) family of curves agreeing in their parameterization up to a choice of origin. As a convenient shorthand, we will speak of such a family of curves as an *affinely parameterized curve*, which we will think of as a curve with its parameterization fixed only up to a choice of origin. We can now construct a vector field X on M as follows: for $x \in M$ we define $X(x) = \dot{\gamma}_x(0)$ (the above discussion shows that X is a smooth vector field on M).

Now suppose that we are given a vector field X on a manifold M, and let us see whether we can think of X as generating an $\mathbb{R}$-action on M. Given $x \in M$, there is a unique curve γ_x passing through x at time $t = 0$ and such that for each value of t at which the curve is defined, its tangent vector at the point $\gamma(t) \in M$ is given by the value of X at that point.[25] Call this curve the *integral curve based at* x. We find that if y lies in the image of integral curve based at x, then the integral curves based at x and y have the same image and agree up to a choice of origin in their parameterization. So we might just as well replace these curves by the corresponding affinely parameterized curve, which we will call the *integral curve* through x (or y). So the vector field X allows us to define a family of integral curves on M, with each point in M lying on exactly one such curve. For $x \in M$ and $t \in \mathbb{R}$, let us agree that $\Phi(t, x)$ is the point that we reach by tracing t units along the integral curve through x, when this instruction is well-defined (recall that the integral curve based at x may only be defined on a subinterval of $\mathbb{R}$). This Φ will be an $\mathbb{R}$-action if and only if the domain of definition of each integral curve is all of $\mathbb{R}$. In this case, we call X a *complete* vector field, and call Φ the $\mathbf{R}$-action *generated* by X.

The picture is as follows: an $\mathbb{R}$-action Φ induces a vector field X on M, and X generates Φ. We think of the group $\{\Phi_t\}$ as consisting of the finite transformations generated by the infinitesimal transformations X (here and below, “infinitesimal”

[25] This is just a statement of the existence and uniqueness theorem for first-order ordinary differential equations.

always means "living in the tangent space"). When X is incomplete, $\Phi(t, x)$ is not defined for all pairs (t, x). In this case, Φ is known as a *local flow*. For many purposes, local flows are nearly as nice as flows, and it is still helpful to think of them as having vector fields as their infinitesimal generators.

3.2 Symplectic Manifolds

DEFINITION 5 (Symplectic Manifold). Let M be a manifold. A *symplectic form* on M is a closed nondegenerate two-form, ω. Here nondegeneracy means that at each $x \in M$ the map $\omega^\flat(x) : v \in T_xM \mapsto \omega(v, \cdot) \in T^*_xM$ is injective.[26] The pair (M, ω) is called a *symplectic manifold*.[27]

DEFINITION 6 (Symplectic Symmetry). Let (M, ω) be a symplectic manifold. A *symplectic symmetry* of (M, ω) is a diffeomorphism $\Phi : M \to M$ that preserves ω in the sense that $\Phi^*\omega = \omega$ (i.e., the pullback of ω by Φ is just ω).

EXAMPLE 7 (Cotangent Bundle Symplectic Structure). Let Q be a finite- or infinite-dimensional manifold and let T^*Q be its cotangent bundle. We define a canonical symplectic form on T^*Q as follows. Let $\pi : T^*Q \to Q$ be the canonical projection $(q, p) \mapsto q$, and let $T\pi$ be the corresponding tangent map. There is a unique one-form θ on T^*Q such that $\theta(q, p) \cdot w = p(T\pi \cdot w)$ for all $(q, p) \in T^*Q$ and all $w \in T_{(q,p)}T^*Q$. We can then define the desired symplectic form as $\omega := -d\theta$, where d is the exterior derivative on T^*Q.[28]

Let (M, ω) be a symplectic manifold, and let $C^\infty(M)$ be the set of smooth functions on M. For present purposes, the fundamental role of ω is to allow us to associate with each $f \in C^\infty(M)$ a smooth vector field X_f on M: X_f is implicitly defined by the equation $\omega(X_f, \cdot) = df$, where df is the exterior derivative of f (the nondegeneracy of ω guarantees that there is a unique solution to this equation).[29] We say that f *generates* X_f or that X_f is *generated by* f.

This basic construction has two fruits of the first importance:

1. Via the map $f \mapsto X_f$, ω allows us to define a new algebraic operation on $C^\infty(M)$: the *Poisson bracket* bracket between $f, g \in C^\infty(M)$ is $\{f, g\} := \omega(X_f, X_g)$.[30] This plays a crucial role in the theory of quantization—see

[26] Of course, for finite-dimensional M, $\omega^\flat(x)$ is surjective if and only if injective.

[27] [Abraham and Marsden, 1978] and [Arnold, 1989] are the standard treatments of mechanics from the symplectic point of view. [Schmid, 1987] covers some of the same ground for the case of infinite-dimensional manifolds. [Ortega and Ratiu, 2004] is a comprehensive reference on the geometry and symmetries of finite-dimensional symplectic spaces (including singular spaces). [Cannas da Silva, Unpublished] is a helpful survey of symplectic geometry. [Weinstein, 1981] and [Gotay and Isenberg, 1992] offer overviews of the role of symplectic geometry in mathematics and physics.

[28] In the finite-dimensional case and relative to a set of cotangent coordinates, ω is given by equation 2 above.

[29] In the infinite-dimensional case, X_f may not be defined on all of M. For well-behaved f, we can deal with this by replacing M by the subspace on which X_f is defined. Below I will suppose that this has been done. For discussion, examples, and references see [Marsden, 1981, pp. 11 ff.] and [Marsden and Ratiu, 1994, p. 106].

[30] The Poisson bracket is a Lie bracket that obeys Leibniz's rule, $\{fg, h\} = f\{g, h\} + g\{f, h\}$.

section 3.4 below.

2. Via the map $f \mapsto X_f$, ω often allows us to associate smooth functions on M with one-parameter groups of symmetries of (M, ω), and vice versa. (i) Let $f \in C^\infty(M)$ and let X_f be the vector field generated by f (via ω), and suppose that X_f is complete so that we are able to construct a corresponding flow, $\xi = \{\Phi_t\}_{t\in\mathbb{R}}$. Then each Φ_t preserves ω, in the sense that $\Phi_t^*\omega = \omega$.[31] Furthermore: f itself is invariant under each Φ_t.[32] (ii) Let $\xi = \{\Phi_t\}_{t\in\mathbb{R}}$ be a one-parameter group of symplectic symmetries of (M, ω) and let X be the vector field on M that is the infinitesimal generator of ξ. It is natural to ask whether we can find an $f \in C^\infty(M)$ that generates X. There are cases in which this is not possible.[33] But in the examples that arise in physics, this can typically be done. And by (i) above, when we can find such an f, we find that it is preserved by the flow ξ.[34]

It is perhaps easier to grasp the function of a symplectic structure if one keeps in mind the Hamiltonian application of this framework.

DEFINITION 8 (Hamiltonian System). A *Hamiltonian system*, (M, ω, h), consists of a symplectic manifold, (M, ω), called the *phase space*, and a function $h : M \to \mathbb{R}$, called the *Hamiltonian*.

We think of (M, ω) as the phase space of some physical system—such as the space of particle positions and momenta—and of h as assigning to each state of the system the total energy of that state. Together h and ω determine a flow $\{\Phi_t\}_{t\in\mathbb{R}}$ on M: each Φ_t maps each state to the state that dynamically follows from it after

[31] Indeed, we can further note that the Lie derivative of ω along X_f vanishes—and this holds even when X_f is incomplete. This provides a sense in which the local flow generated by an incomplete vector field preserves ω.

[32] Indeed, the Lie derivative of f along X_f vanishes. This also holds when X_f is incomplete—so there is a sense in which the local flow generated by such an incomplete vector field preserves f.

[33] See [Ortega and Ratiu, 2004, §4.5.16] for an example. See [Butterfield, this volume, §2.1.3] for further discussion.

[34] More ambitiously, let G be a Lie group acting on M via symplectic symmetries, with $\dim G > 1$. Such a group will contain many one-parameter subgroups—as the group of isometries of Euclidean space contains a one-parameter group of translations corresponding to each direction in Euclidean space and a one-parameter group of rotations corresponding to each axis in Euclidean space. In this case, we can hope that for each one-parameter subgroup of G it is possible to find a function on M that generates that subgroup. If all goes very well—as it does in many examples that arise in physics—we can hope that the algebra of Poisson brackets between these generators will mirror the algebra of the group (i.e., there will be a Lie algebra isomorphism here). In this case, we speak of the existence of a *momentum map* (*warning*: terminology varies—many authors call these *infinitesimally equivariant momentum maps*). If f and g are functions on M such that their Poisson bracket vanishes, then we find that f is invariant under each symplectic symmetry in the one-parameter group generated by g. In particular, if G is a group of symplectic symmetries of (M, ω) and f a function on M such that the Poisson bracket of f with any function generating a one-parameter subgroup of G vanishes, then each of these generators is invariant under the one-parameter group of symmetries generated by f. [Woodhouse, 1991, §3.4] provides a useful guide to situations under which momentum maps are or are not available. See [Butterfield, this volume, §6] for further discussion.

t units of time. h will be preserved under this group—this corresponds to the conservation of energy.[35]

3.3 *Presymplectic Manifolds*

In sections 6.2 and 7 below we will be concerned with theories whose space of solutions and space of initial data are not symplectic.

DEFINITION 9 (Presymplectic Manifold). Let M be a manifold. A *presymplectic structure* on M is a closed degenerate two-form, ω; we call (M, ω) a *presymplectic manifold*.[36] Here degeneracy means: at each point x there is a nontrivial *nullspace* $N_x \subset T_x M$ consisting of tangent vectors v such that $\omega_x(v, \cdot) = 0$.

A presymplectic structure ω on a manifold M induces a partition of M by submanifolds, $\{M_\alpha\}$ as follows. We define an equivalence relation on M by declaring $x, y \in M$ to be equivalent if they can be joined by a curve $\gamma : \mathbb{R} \to M$ each of whose tangent vectors is null—i.e., $\dot{\gamma}(t) \in N_{\gamma(t)}$ for each $t \in \mathbb{R}$. The equivalence classes, M_α, of this relation are called *gauge orbits*. For $x \in M$ we also denote the gauge orbit containing x by $[x]$. Each gauge orbit is a submanifold of M.[37] We call a function $f \in C^\infty(M)$ *gauge-invariant* if $f(x) = f(y)$ whenever x and y belong to the same gauge orbit of M (i.e., a function is gauge-invariant if and only if it is constant on gauge orbits).

We call a diffeomorphism from M to itself which preserves a presymplectic form ω a *presymplectic symmetry* of (M, ω). We say that two presymplectic symmetries, Φ and Φ', *agree up to gauge* if for each $x \in M$, $[\Phi(x)] = [\Phi'(x)]$ (i.e., for each $x \in M$, Φ and Φ' map x to the same gauge orbit); we call the set of presymplectic symmetries that agree with Φ up to gauge the *gauge equivalence class* of Φ. Similarly, we will say that two one-parameter groups, $\xi = \{\Phi_t\}$ and $\xi' = \{\Phi'_t\}$, of presymplectic symmetries agree up to gauge if Φ_t and Φ'_t agree up to gauge for each t; the *gauge equivalence class* of $\xi = \{\Phi_t\}$ comprises all ξ' that agree with it up to gauge in this sense.

If a presymplectic symmetry $\Phi : M \to M$ fixes each M_α (i.e., Φ maps points in M_α to points in M_α), then we call Φ a *gauge transformation*. Note that a gauge transformation agrees up to gauge with the identity map on M.

In the symplectic case: when all goes well, the equation $\omega(X_f, \cdot) = df$ allows one to associate each smooth function on a symplectic manifold (M, ω) with a

[35] Often it will be possible to identify a larger group G of symplectic symmetries of (M, ω) that leaves h invariant (such as the group of Euclidean symmetries acting in the obvious way in Newtonian particle mechanics). Then a momentum map (see preceding footnote) would allow one to construct $\dim G$ independent quantities, whose algebra would mirror that of G, and that would be conserved under the dynamics generated by h.

[36] Terminology varies: often (but not here) symplectic forms count as special cases of presymplectic forms; sometimes (but not here) presymplectic forms are required to have constant rank or to have well-behaved spaces of gauge orbits. On presymplectic geometry, see, e.g., [Gotay and Nester, 1980].

[37] If X and Y are null vector fields on M (i.e., $X(x), Y(x) \in N_x$ for each $x \in M$) then, $[X, Y]$ is also a null vector field. It follows (by Frobenius' theorem) that the N_x form an integrable distribution, with the M_α as the integral manifolds.

one-parameter group of symplectic symmetries of (M, ω)—and vice versa.

In the presymplectic case: when all goes well, the equation $\omega(X_f, \cdot) = df$ allows one to associate each smooth gauge-invariant function on a presymplectic manifold (M, ω) with a gauge equivalence class of one-parameter groups of presymplectic symmetries of (M, ω)—and vice versa. So in the presymplectic case: if f generates the one-parameter group $\xi = \{\Phi_t\}$ of presymplectic symmetries via the equation $\omega(X_f, \cdot) = df$, then it also generates each $\xi' = \{\Phi'_t\}$ in the gauge equivalence class of ξ.

Note an interesting special case: any solution X_f of the equation $\omega(X_f, \cdot) = df$ for f a constant function is a vector field on M consisting of null vectors; so the corresponding one-dimensional group of presymplectic symmetries of (M, ω) consists of gauge transformations. Conversely: if $\xi = \{\Phi_t\}$ is a one-parameter group of gauge transformations of (M, ω), then any function that generates ξ (via ω) is a constant function.

Given a presymplectic manifold (M, ω), we can construct M' the space of gauge orbits of M. M' inherits a topological structure from M.[38] We will call the process of passing from M to M' *reduction*, and call M' the *reduced space*. In general, M' need not be a manifold, nor anything nearly so well-behaved as the spaces we want to consider below.[39] But when all goes well (as it usually does in the sort of cases considered below) M' will inherit from M a smooth structure (so it will be a space with at most mild singularities). And so long as some further technical conditions on ω hold, M' inherits from (M, ω) a two-form ω' that is nondegenerate as well as closed.[40] So, in this case, (M', ω') is a symplectic space. Note that each gauge-invariant $f \in C^\infty(M)$ corresponds to a unique $f' \in C^\infty(M')$. While f generates an equivalence class of one-parameter groups of presymplectic transformations of (M, ω), f' generates a single one-parameter group of symplectic transformations of (M', ω').[41]

[38] We equip M' with the quotient topology, according to which a set $U' \subset M'$ is open if and only if $\pi^{-1}(U')$ is open in M (here π is the projection $x \in M \mapsto [x] \in M'$).

[39] If (M, ω, H) is a Hamiltonian system in the sense of definition 8 above, then the restriction of ω to a surface, E, of constant energy is presymplectic—with the gauge orbits of $(E, \omega \mid_E)$ being the dynamical trajectories of the Hamiltonian system. If the dynamics is ergodic, then generic trajectories come arbitrarily close to each $x \in E$. It follows that the quotient space E' has the trivial topology, according to which the only open sets are the empty set and the space itself.

[40] See [Marsden, 1981, p. 6] and [Ortega and Ratiu, 2004, §6.1.5].

[41] Each presymplectic symmetry of (M, ω) corresponds to a symplectic symmetry of (M', ω'), with two presymplectic symmetries correspond to the same symplectic symmetry if and only if they agree up to gauge. Thus each gauge equivalence class of presymplectic symmetries corresponds to a single symplectic symmetry. And each gauge equivalence class of one-parameter groups of presymplectic symmetries corresponds to a single one-parameter group of symplectic symmetries of the reduced space.

3.4 Symplectic Structures and Quantization

Quantization is the process of constructing a quantum counterpart to a given classical theory.[42] As it is presently understood, it is a process which takes as its starting point a theory in Hamiltonian or Lagrangian form (or the discrete-time analog of such a theory). One does not know how to quantize a theory *qua* differential equations directly, without passing first to a Hamiltonian or Lagrangian recasting of the theory.[43]

The following observations lend some plausibility to the idea that a symplectic structure is the *sine qua non* of quantization.

1. The core notion of quantization involves the following steps. One begins with a symplectic space (the space of classical solutions or initial data) and selects a set of functions on this space (classical physical quantities) that is closed under addition and the Poisson bracket induced by the symplectic structure. One then looks for a set of operators (quantum observables) acting on a space of quantum states, such that the algebra of these operators mirrors (or approximately mirrors, with increasingly better match as one approaches the classical limit) the algebra (under addition and the Poisson bracket) of the chosen classical quantities. One may then also need to take the further step of adding a Hamiltonian operator that implements the quantum dynamics.

2. Some classical theories have the unfortunate feature that when cast in Lagrangian or Hamiltonian form, they come to us with a space of solutions or initial data that is merely presymplectic. Typically, it is known that there is a symplectic space in the offing via reduction, as outlined above in section 3.3. But it may be difficult to construct this space, or it may happen, for one reason or another, that it seems easier to work with the presymplectic version of the theory. So a number of strategies have been developed for quantizing theories in presymplectic form: gauge fixing, Dirac constraint quantization, BRST quantization, etc. But it is very natural to think of each of these techniques as offering an indirect approach to the quantization of the underlying symplectic space.[44]

---

[42] For overviews of the literature on quantization, see [Landsman, this volume] and [Ali and Engliš, 2005].

[43] For an attempt to show that in order for a theory to be quantizable, it must be derivable from a Lagrangian, see [Hojman and Shepley, 1991].

[44] (i) Gauge fixing just amounts to finding a submanifold of the presymplectic space that is symplectically isomorphic to the reduced space; see, e.g., [Henneaux and Teitelboim, 1992, §1.4]. (ii) For Dirac's approach, see [Dirac, 2001] or [Henneaux and Teitelboim, 1992]. It is felt Dirac's algorithm should be amended when its output differs from the result of directly quantizing the reduced theory; see, e.g., [Duval *et al.*, 1990]. (iii) In the case of finite-dimensional systems, it can be shown that application of the BRST algorithm leads to a quantization of the reduced theory; see, e.g., [Loll, 1992] or [Tuynman, 1992]. (iv) On the relation between the BRST approach and the suggested amendment of the Dirac approach, see [Guillemin and Sternberg, 1990, §12].

However, there exist approaches to quantization that do not appear to employ the symplectic structure of the classical spaces at all—for example Mackey quantization (which has a somewhat limited range of application) and path integral quantization (which has very wide application, but murky foundations in its application to field theories). As emphasized in [Landsman, this volume], the relation between the classical and the quantum is far from completely understood.

4 LAGRANGIAN FIELD THEORY

Differential equations are normally given to us in the following way. We are given a set of independent variables and a set of dependent variables, and a space of functions, $\mathcal{K}$, consisting of functions, u, that map values of the independent variables to values of the dependent variables. A differential equation Δ can be thought of as a condition on a function and its derivatives that is satisfied by only some $u \in \mathcal{K}$. We call the u that satisfy Δ the *solutions* to Δ and denote the space of such solutions by $\mathcal{S}$.

In physical applications, the independent variables typically parameterize space, time, or spacetime while the dependent variables parameterize the possible values of some quantity of interest. We can think of the functions $u \in \mathcal{K}$ as describing situations that are in some sense possible and of solutions $u \in \mathcal{S}$ as describing situations that are genuinely physically possible according to the theory whose laws are encoded in Δ. Although the terminology is not wholly perspicuous, I will speak of elements of $\mathcal{K}$ as corresponding to *kinematical possibilities* and of elements of $\mathcal{S}$ as corresponding to *dynamical possibilities*.

EXAMPLE 10 (Mechanics of a Particle). Consider the theory of a particle a particle in Euclidean space subject to a position-dependent force. The independent variable parameterizes time and the dependent variables parameterize the possible positions of the particle; an arbitrary continuous functions $x(t)$ of the form $t \in \mathbb{R} \mapsto x(t) \in \mathbb{R}^3$ describes a kinematically possible pattern of behaviour of the particle; $x(t)$ describes a dynamically possible behaviour if it satisfies the Newtonian equation $\ddot{x}(t) = F(x(t))$.

EXAMPLE 11 (The Klein–Gordon Field). The usual theory of a scalar field has the following ingredients: as independent variables we take inertial coordinates $\{t, x, y, z\}$ on Minkowski spacetime, V; the theory has a single dependent variable, parameterizing the real numbers; so the kinematically possible fields are given by (suitably smooth) real-valued functions on Minkowski spacetime; the dynamically possible fields are those $\Phi : V \to \mathbb{R}$ satisfying the Klein–Gordon equation,

$$\frac{\partial^2\Phi}{\partial t^2} - \frac{\partial^2\Phi}{\partial x^2} - \frac{\partial^2\Phi}{\partial y^2} - \frac{\partial^2\Phi}{\partial z^2} - m^2\Phi = 0.$$

Our primary concern below is with field theories—those physical theories whose laws are encoded in differential equations whose independent variables parameterize spacetime.[45] We think of such a field theory as consisting of the following components: a spacetime V; a space W in which the fields take their values, a space, $\mathcal{K}$, of kinematically possible fields (i.e., of functions from V to W satisfying suitable smoothness and boundary conditions); and a set of differential equations Δ.

This section has the following structure. In the first subsection below I discuss the Lagrangian approach, in which one singles out the set of dynamical possibilities within the space of kinematical possibilities via a variational problem for a Lagrangian rather than via the direct imposition of a differential equation. In the second subsection, I discuss a very important advantage of the Lagrangian approach over the direct approach: the former but not the latter allows one to equip the space of dynamical possibilities with a (pre)symplectic form. In the third subsection, I discuss the celebrated relation between conserved quantities and symmetries in the Lagrangian approach, first discerned by Noether. The discussion of these subsections is based upon [Zuckerman, 1987] and [Deligne and Freed, 1999, Chapters 1 and 2]; see also [Woodhouse, 1991, Chapters 2 and 7].

Before beginning it will be helpful to make some more specific assumptions about the theories we will be discussing. These assumptions can will be in force throughout the remainder of the chapter.

Spacetime. Our spacetime V will always be an n-dimensional Hausdorff manifold V, with the topology $M \times \mathbb{R}$ for some $(n-1)$-manifold M. We will always think of time as having the topology of $\mathbb{R}$, so we will say that a spacetime with topology $M \times \mathbb{R}$ has *spatial topology* M. In particular, we will say that V is *spatially compact* if M is compact.

In most theories, the geometry of spacetime is fixed from solution to solution. So we typically think of V as carrying a solution-independent geometrical structure (I will be lazy, and sometimes use V to denote the manifold, sometimes the manifold and the geometry together).[46] Without worrying about precision, I will stipulate now that we will only be interested in spacetimes that are well-behaved. Examples: Newtonian spacetime, neoNewtonian spacetime, Minkowski spacetime, or other globally hyperbolic general relativistic spacetimes.

The spacetime geometries that we consider single out a distinguished class

[45] *Prima facie*, the ontology of such a classical field theory satisfies Lewis's Humean supervenience—"the doctrine that all there is to the world is a vast mosaic of local matters of particular fact, just one little thing after another. ... We have geometry: as system of external relations of spatiotemporal distance between points. ... And at those points we have local qualities: perfectly natural intrinsic properties which need nothing bigger than a point at which to be instantiated. ... And that is all" [1986, pp. ix f.]. Indeed, Lewis says that the picture was "inspired by classical physics" [1999, p. 226]. See [Butterfield, Unpublished] for doubts about the fit between Humean supervenience and classical physics.

[46] We could also allow V to carry non-geometrical solution-independent structures, corresponding to external fields etc.

of hypersurfaces in V that correspond to instants of time.[47] In prerelativistic spacetimes the instants are just the hypersurfaces of absolute simultaneity. Typically, in relativistic spacetimes the instants are just the Cauchy surfaces. Occasionally in highly symmetric relativistic spacetimes, one requires instants to be Cauchy surfaces with nice symmetries—thus one might in some contexts require instants in Minkowski spacetime to arise as hypersurfaces of simultaneity relative to an inertial observer. Furthermore, in spacetimes carrying geometrical structures, it often makes sense to speak of certain curves in V as *possible worldlines of point-particles*. In prerelativistic spacetime, a curve counts as a possible worldline of a point-particle if it is transverse to the hypersurfaces of simultaneity; in relativistic spacetime such possible worldlines are given by timelike curves.

In addition to considering field theories set in fixed general relativistic background spacetimes, we also want to consider general relativity itself as a Lagrangian field theory. In that context the spacetime metric g is itself dynamical and varies from solution to solution. With this example in mind, we will allow a bare manifold of topology $M \times \mathbb{R}$ unequipped with any geometry to count as a spacetime for present purposes, so that general relativity can be developed alongside theories set in a fixed geometrical background. Note that even in a theory like general relativity in which the spatiotemporal geometry is dynamical we can still speak of a hypersurface as being an instant *relative* to a solution g.[48]

Field Values. We will take W, the space of field values, to be a finite-dimensional vector space. However, we could afford to be more general, at the price of complicating some of the notation below. Our $\mathcal{K}$ is a space of sections of a trivial vector bundle over V; it follows that for $\Phi \in \mathcal{K}$ a tangent vector $\delta\Phi \in T_\Phi\mathcal{K}$ is also a map from V to W. We could have allowed $\mathcal{K}$ to be a space of sections of an arbitrary fibre bundle $E \to V$. The chief complication that this would introduce is that a tangent vector $\delta\Phi \in T_\Phi\mathcal{K}$ would then be a section of the bundle $\Phi^*T(E/V)$.

Kinematically Possible Fields. In setting up a rigourous classical field theory, care must be taken in selecting differentiability and boundary conditions to impose on the kinematically possible fields. We can here afford to neglect such details, and just say that for each theory considered below, $\mathcal{K}$ is taken to be a space of well-behaved functions $\Phi : V \to W$, required to satisfy appropriate conditions of differentiability and behaviour at infinity, but otherwise arbitrary. Note that while $\mathcal{K}$ will be a manifold (often even an affine

[47] The crucial point is this: one needs to choose boundary conditions and a notion of instant in such a way that for certain $(n-1)$-forms, ω, for any instant $\Sigma \subset V$, $\int_\Sigma \omega$ converges, and is independent of the instant chosen (cf. fnn. 61 and 73 below). In the standard cases, the obvious notions of instant suffice.

[48] Of course, in a theory in which the spacetime geometry is a solution-independent matter, $\Sigma \subset V$ is an instant relative to a solution Φ if and only if it is an instant according to the geometry of V.

or linear space), in general $\mathcal{S}$ will be a nonlinear subspace of $\mathcal{K}$ with mild singularities.

Differential Equations. The Lagrangian framework sketched below is very general and does not require a restriction on the order of the differential equations. However, because in later sections we will often be interested in comparing Hamiltonian and Lagrangian versions of the same theory, and because the Hamiltonian framework takes second-order equations as its point of departure, we will restrict attention to such equations beginning in section 5 below.

REMARK 12 (Finite-Dimensional Theories as Field Theories). In a classical theory of a system with finitely many degrees of freedom (finite systems of particles, rigid bodies, etc.) the configuration space Q is a manifold parameterizing the possible dispositions of the system in physical space. A history of the system is a curve $x : t \in \mathbb{R} \mapsto x(t) \in Q$. We can fit such theories into the present framework, by taking $W = Q$ and $V = \mathbb{R}$ (so the only independent variable is time). No harm comes of treating such a theory as a degenerate case of a field theory, so long as one does not forget that in this case the "spacetime" V parameterized by the independent variables of the theory is distinct from the spacetime in which the system is located.

REMARK 13 (Notation). Because a choice of V and W is implicit in a choice of $\mathcal{K}$, we can denote a field theory by $(\mathcal{K}, \Delta)$.

4.1 The Lagrangian Approach

The role of the differential equations Δ of a theory is to cut down the space of kinematical possibilities $\mathcal{K}$ to the space of dynamical possibilities $\mathcal{S}$.[49] The key insight of the Lagrangian approach is that for the vast majority of equations that arise in classical physics, there is an alternative way of singling out the subspace of solutions.[50]

[49] The text of this section is informal. More precise statements are given in the footnotes. The following terminology and results will be helpful.

The space $V \times \mathcal{K}$ is a manifold, and so carries differential forms and an exterior derivative operator. For $0 \leq p \leq n$ and $q \geq 0$ let $\Omega^{p,q}(V \times \mathcal{K})$ be the space of q-forms on $\mathcal{K}$ that take their values in the space of p-forms on V : thus if $K \in \Omega^{p,q}(V \times \mathcal{K})$, $\Phi \in \mathcal{K}$, and $\delta\Phi_1, \ldots, \delta\Phi_q \in T_\Phi\mathcal{K}$ then $K(\Phi, \delta\Phi_1, \ldots, \delta\Phi_q)$ is a p-form on our spacetime V. Each differential form on $V \times \mathcal{K}$ belongs to some $\Omega^{p,q}(V \times \mathcal{K})$. Furthermore, we can write the exterior derivative, d, on $V \times \mathcal{K}$ as $d = D + \partial$, where D is the exterior derivative on V, mapping elements of $\Omega^{p,q}(V \times \mathcal{K})$ to elements of $\Omega^{p+1,q}(V \times \mathcal{K})$ (for $0 \leq p < n$), and ∂ is the exterior derivative on $\mathcal{K}$, mapping elements of $\Omega^{p,q}(V \times \mathcal{K})$ to elements of $\Omega^{p,q+1}(V \times \mathcal{K})$. We have $\partial D = -D\partial$.

Note that if $\Phi \in \mathcal{K}$ then a tangent vector $\delta\Phi \in T_\Phi\mathcal{K}$ is itself a map from V to W. So for each admissible p and q we can consider the subspace $\Omega^{p,q}_{loc}(V \times \mathcal{K}) \subset \Omega^{p,q}(V \times \mathcal{K})$ of *local* forms consisting of those K such that for any $\Phi \in \mathcal{K}$ and $\delta\Phi_1, \ldots, \delta\Phi_q \in T_\Phi\mathcal{K}$, the value of the p-form $K(\Phi, \delta\Phi_1, \ldots, \delta\Phi_q)$ at spacetime point $x \in V$ depends only on the values at x of $\Phi, \delta\Phi_1, \ldots, \delta\Phi_q$, and finitely many of their derivatives.

[50] For discussion of the scope of the Lagrangian approach, see [Bluman, 2005, §2.1].

DEFINITION 14 (Lagrangian). Let $\mathcal{K}$ be a space of kinematically possible fields. A *Lagrangian*, L, on $\mathcal{K}$ is a local map from $\mathcal{K}$ to the space of n-forms on V (to say that L is *local* is to say that the value of $L(\Phi)$ at a point $x \in V$ depends only on the values at x of Φ and finitely many of its derivatives).[51]

Given a Lagrangian L, one can proceed, as in the treatment of the n-body problem sketched in section 2.3 above, to look for those kinematically possible Φ with the special property that infinitesimal perturbations at Φ make no difference to the value of $\int L(\Phi)$.

DEFINITION 15 (Variational Problem). Note that for each compact $U \subset V$, $S_U : \Phi \mapsto \int_U L(\Phi)$ is a real-valued function on $\mathcal{K}$. Let us call the assignment $U \mapsto S_U$ the *variational problem* of L.

DEFINITION 16 (Stationary Fields). We call $\Phi \in \mathcal{K}$ *stationary* for L if for each compact $U \subset V$ the effect of infinitesimally perturbing Φ inside U has no effect on the value of S_U.[52]

DEFINITION 17 (Lagrangian Admitted by Δ). We call L a *Lagrangian for* $(\mathcal{K}, \Delta)$ if the set of Φ stationary for L coincides with the space $\mathcal{S}$ of solutions of Δ. In this case we also say that Δ *admits* the Lagrangian L, and speak of $\mathcal{S}$ as the space of solutions of $(\mathcal{K}, L)$.

REMARK 18 (Euler–Lagrange Equations). Given a Lagrangian, one can always find a set of equations Δ (the *Euler–Lagrange equations* for L) so that L is a Lagrangian for Δ. That is: a kinematically possible field $\Phi : V \to W$ is stationary for a Lagrangian L if and only if the Euler–Lagrange equations for L are satisfied. For Lagrangians depending only on the fields and their first-order derivatives, these equations require that

$$\frac{\partial L}{\partial \Phi^\alpha}(x_a) - \sum_{a=1}^{n} \frac{\partial}{\partial x_a}\left(\frac{\partial L}{\partial \Phi^\alpha_a}\right)(x_a) = 0 \tag{4}$$

hold at each point $x \in V$ (here a indexes coordinates on V, α indexes coordinates on W, and Φ^α_a stands for $\frac{\partial}{\partial x_a}\Phi^\alpha$).[53]

---

[51]That is, $L \in \Omega^{n,0}_{loc}(V \times \mathcal{K})$.

[52]That is, Φ is stationary for L if for each compact $U \subset V$ and for each $\delta\Phi \in T_\Phi\mathcal{K}$ whose support is contained in U we find that $\partial S_U(\delta\Phi) = \int_U \partial L(\Phi, \delta\Phi)$ vanishes. We can think of this as follows: fixing Φ, U, and $\delta\Phi$, we find a curve $\Phi[\varepsilon] : [-1, 1] \to \mathcal{K}$ such that $\Phi[0] = \Phi$ and $\frac{d}{d\varepsilon}\Phi[\varepsilon]) \mid_{\varepsilon=0} = \delta\Phi$; the requirement that $\partial S_U(\delta\Phi) = 0$ amounts to $\frac{d}{d\varepsilon}\int_U L(\Phi[\varepsilon]) \mid_{\varepsilon=0} = 0$.

[53] Of course, there is a coordinate-independent description of this. It is possible to show that $\partial L = E + DM$, where $E \in \Omega^{n,1}_{loc}(V \times \mathcal{K})$ and $M \in \Omega^{n-1,1}_{loc}(V \times \mathcal{K})$, with E determined uniquely by L and M determined up to the addition of an exact form DN, with $N \in \Omega^{n-2,1}_{loc}(V \times \mathcal{K})$. The condition that $\partial S_U(\delta\Phi) = 0$ becomes $\int_U E(\Phi, \delta\Phi) + DM(\Phi, \delta\Phi) = 0$ for all $\delta\Phi$ whose support is contained in U. Since $\delta\Phi$ vanishes along the boundary of U, Stokes's theorem tells us that the second integrand makes no contribution. So Φ is stationary if and only if $\int_U E(\Phi, \delta\Phi) = 0$ for all such U and admissible $\delta\Phi$—which is equivalent to saying that $E(\Phi, \delta\Phi) = 0$ for all such $\delta\Phi$. Relative to coordinates, this last equation is equivalent to equation 4 in the case of a Lagrangian depending only on first derivatives.

REMARK 19 (Trivially Differing Lagrangians). Let us say that Lagrangians L and L' *differ trivially* if L' is of the form $L'(\Phi) = L(\Phi) + \alpha_\Phi$ for each $\Phi \in \mathcal{K}$ with α_Φ an exact Φ-dependent n-form on V.[54] Let us say that if L and L' are Lagrangians, their variational problems $U \mapsto S_U$ and $U \mapsto S'_U$ are *equivalent* if for each compact $U \subset V$ and field $\Phi \in \mathcal{K}$, we have that any infinitesimal perturbation of Φ leaves the value of S_U unchanged if and only if it leaves the value of S'_U unchanged.[55] Lagrangians that differ trivially have equivalent variational problems.[56] It follows that trivially differing Lagrangians have the same space of solutions—indeed, they have the same Euler–Lagrange equations.[57]

REMARK 20 (Uniqueness of Lagrangians). The previous remark shows that if Δ does admit a Lagrangian, it will admit infinitely many that differ trivially. Some Δ also admit multiple Lagrangians that do not differ trivially—e.g., the Newtonian equations for a particle moving in a spherical potential in three-dimensional Euclidean space.[58]

REMARK 21 (Existence of Lagrangians). Not every set of equations Δ admits a Lagrangian.[59] A charged particle moving in the electromagnetic field of a magnetic monopole is an example of a system that does not admit of Lagrangian treatment.[60]

---

[54] I.e., $L' = L + DK$ where $K \in \Omega^{n-1,0}_{loc}(V \times \mathcal{K})$.

[55] That is: the variational problems $U \mapsto S_U$ and $U \mapsto S'_U$ for Lagrangians L and L' are equivalent if for every compact $U \subset V$, every field $\Phi \in \mathcal{K}$, and every tangent vector $\delta\Phi \in T_\Phi\mathcal{K}$ with support contained in U, we have that $\partial S_U(\Phi)(\delta\Phi) = 0$ if and only if $\partial S'_U(\Phi)(\delta\Phi) = 0$.

[56] Let $L' = L + DK$ with $K \in \Omega^{n-1,0}_{loc}(V \times \mathcal{K})$. Then for any compact $U \subset V$, $\Phi \in \mathcal{K}$, and $\delta\Phi \in T_\Phi\mathcal{K}$ with support contained in U, we have $\partial S_U(\Phi)(\delta\Phi) - \partial S'_U(\Phi)(\delta\Phi) = \int_U \partial DK(\Phi)(\delta\Phi)$. But $\partial D = -D\partial$, so the right hand side is $-\int_U D\partial K(\Phi)(\delta\Phi)$, which vanishes (by Stokes's theorem and the fact that $\delta\Phi$ vanishes on the boundary of U).

[57] That is, if Lagrangians L and L' differ by a term of the form DK, then they share the same Euler–Lagrange operator E.

[58] See [Crampin and Prince, 1988] and [Henneaux and Shepley, 1982] for this example. For field-theoretic examples, see [Nutku and Pavlov, 2002]. For a topological condition on $V \times W$ sufficient to ensure that Δ does not admit nontrivially differing Lagrangians, see [Anderson and Duchamp, 1980, Theorem 4.3.ii].

[59] The problem of determining whether a given set of differential equations admits a Lagrangian is known as *the inverse problem of the calculus of variations* among mathematicians and as *Helmholtz's problem* among physicists. [Prince, 2000] is a helpful survey of results concerning finite-dimensional systems. [Anderson and Duchamp, 1980, §5] includes examples of field theories that do not admit Lagrangian formulations.

[60] See [Anderson and Thompson, 1992, pp. 4 f.]. For other examples, see [Prince, 2000].

4.2 The Structure of the Space of Solutions

The choice of a Lagrangian L allows us to equip $\mathcal{S}$ with a closed two-form, Ω.[61] So when Ω is nondegenerate, $(\mathcal{S}, \Omega)$ is a symplectic space; otherwise, it is presymplectic.[62] Roughly speaking, one expects that Ω is nondegenerate if and only if the equations, Δ, of our theory have the property that specifying initial data determines a unique inextendible solution.[63]

The choice of a Lagrangian brings into view the sort of structure required for the construction of a quantum theory. A set of differential equations Δ alone does not appear to determine such structure, and it is not known how to quantize a differential equation directly, without the introduction of a Lagrangian or a Hamiltonian. If Δ admits a Lagrangian L, then it also admits the whole class of Lagrangians that differ trivially from L (see remark 20 above). Unsurprisingly, trivially differing Lagrangians induces the same Ω on $\mathcal{S}$.[64] But when Δ admits Lagrangians L and L' that differ nontrivially, these Lagrangians can induce distinct geometric structures on $\mathcal{S}$; and one expects that these distinct (pre)symplectic structures will lead to distinct quantizations of the given classical theory.[65] In the case mentioned above of a particle moving in a spherical potential, each of these elements is present: multiple nontrivially differing Lagrangians lead to distinct symplectic structures on the space of solutions, which lead in turn to physically distinct quantizations.[66]

---

[61] Recall from footnote 53 above that we have the decomposition $\partial L = E + DM$, with E unique and M unique up to the addition of a D-exact form. We now define $Z := \partial M$. $Z \in \Omega^{n-1,2}_{loc}(V \times \mathcal{K})$ and is uniquely determined by L up to the addition of a term of the form DY, with $Y \in \Omega^{n-2,2}_{loc}(V \times \mathcal{K})$.

Let $\Phi \in \mathcal{S}$ be a solution, let $\delta\Phi_1, \delta\Phi_2 \in T_\Phi \mathcal{S}$, and let $\Sigma \subset V$ be an instant relative to Φ. Then we define $\Omega_\Sigma(\Phi, \delta\phi_1, \delta\Phi_2) := \int_\Sigma Z(\Phi, \delta\Phi_1, \delta\Phi_2)$. We assume nice boundary conditions at infinity, so that Ω_Σ is well-defined, and so that replacing Z by $Z + DY$ makes no difference to Ω_Σ. We find that the value of Ω_Σ is independent of the instant chosen—because $Z(\Phi, \delta\Phi_1, \delta\Phi_2)$ is closed as an $(n-1)$-form on V and we have been careful in our choice of notion of instant (see fn. 47). So we drop the subscript, and think of Ω as a two-form on $\mathcal{S}$, closed because Z is ∂-exact.

[62] Lagrange appears to have been the first to equip the space of solutions to a dynamical problem with this symplectic structure; see [Weinstein, 1981, §2], [Souriau, 1986], or [Iglesias, 1998].

[63] As we will see below in 6.2, if the equations of motion admit gauge symmetries (so that uniqueness fails in a certain dramatic way), then Ω is presymplectic. I believe it is widely thought that this is the only way that Ω can fail to be symplectic—at least for the sort of examples that arise in physics.

[64] Replacing L by $L + DK$ alters Z a term of the form DY, $Y \in \Omega^{n-2,2}_{loc}(V \times \mathcal{K})$. But because it is D-exact, this new term will not contribute to the integral over space that defines Ω (by Stokes's theorem and boundary conditions).

[65] When (as in the Newtonian case) the equations of motion are second-order and the space of solutions is finite-dimensional, Lagrangians L and L' induce the same two-form on the space of solutions if and only if they differ trivially; see [Crampin and Prince, 1988, §II]. Presumably this in fact holds for a much wider range of cases.

[66] See [Henneaux and Shepley, 1982].

4.3 Symmetries and Conserved Quantities

Given a set of equations Δ and a Lagrangian L admitted by Δ, there are three distinct notions of symmetry we might consider.[67] Roughly speaking, a *symmetry of* Δ is a map from $\mathcal{K}$ to itself that fixes $\mathcal{S}$ as a set and that is generated by an object local in the fields and their derivatives.[68] We can then consider the subset of *variational symmetries*, which also leave the variational problem of L invariant, or the subset of *Lagrangian symmetries* that leave L itself invariant. The three notions are distinct: every Lagrangian symmetry is a variational symmetry, but some theories have variational symmetries that are not Lagrangian symmetries; similarly, every variational symmetry is a symmetry of the associated equations of motion, but some equations that admit Lagrangians have symmetries that are not variational symmetries of any Lagrangian for the theory.[69]

For present purposes, it is natural to focus on variational symmetries of physical theories. For, on the one hand, the class of Lagrangian symmetries excludes some physically important symmetries—and in any case, within the Lagrangian approach it is not clear that it is more natural to focus on symmetries of the Lagrangian than on symmetries of the variational problem. On the other hand, the class of symmetries of equations that are not variational symmetries does not appear to include any symmetries of absolutely central physical interest—and it is at the level of variational symmetries (rather than symmetries of equations) that the powerful results of Noether, cementing a connection between certain special types of one-parameter groups of variational symmetries and certain special types of conserved quantities in classical field theories, are naturally situated.[70]

Here is a statement of the central result. Let us call a one-parameter group, $\xi = \{g_t\}$, of diffeomorphisms from $\mathcal{K}$ to itself a *Noether group* for L if its infinitesimal generator leaves invariant the variational problem of L and is local in the appropriate sense.[71] Given a Noether group $\xi = \{g_t\}$ for $(\mathcal{K}, L)$, there is

[67] See [Olver, 1993, Chapters 2, 4, and 5] for the relevant notions. *Warning*: terminology varies—sometimes my *Lagrangian symmetries* are called *variational symmetries*, sometimes my *variational symmetries* are called *divergence symmetries*, etc.

[68] See [Olver, 1993, §5.1] for details.

[69] The wave equation in (2+1) dimensions has a dilational symmetry that is not a variational symmetry and inversion symmetries that are variational but not Lagrangian; see [Olver, 1993, Examples 2.43, 4.15, 4.36, and 5.63]. Example 4.35 of the same work shows that Galilean boosts are variational symmetries for the n-body problem but are not Lagrangian symmetries. Indeed, no Lagrangian for Newtonian particles subject to forces derived from a potential can be invariant under the full group of symmetries of neoNewtonian spacetime; see [Souriau, 1997, Remark 12.136].

[70] Note, however, that there do exist results establishing links between symmetries of equations with conserved quantities, without detouring through the Lagrangian framework; see, e.g., [Bluman, 2005].

[71] More precisely, let ξ be a one-parameter group of diffeomorphisms from $\mathcal{K}$ to itself and let X be the corresponding vector field on $\mathcal{K}$ (i.e., X is the vector field whose flow is ξ). ξ is a Noether group if the following two conditions hold. (i) X is an *infinitesimal variational symmetry of* L: there exists an $R \in \Omega^{n-1,0}_{loc}(V \times \mathcal{K})$ such that $\partial L(\Phi, (X(\Phi)) = DR(\Phi)$ for all $\Phi \in \mathcal{S}$. (ii) X is *local*: for any $\Phi \in \mathcal{K}$, $X(\Phi) \in T_\Phi\mathcal{K}$ is local on V, in the sense that at any point $x \in V$, we find that $X(\Phi)(x)$ depends only on the value at x of Φ and finitely many of its derivatives (recall that an element of $T_\Phi\mathcal{K}$ is itself a map from V to W).

a map J_ξ, called the *Noether current* associated with ξ, that maps solutions to $(n-1)$-forms on V.[72] Given an arbitrary solution, $\Phi \in \mathcal{S}$, and an instant $\Sigma \subset V$ we integrate $J_\xi(\Phi)$ over Σ to give the *Noether charge*, $Q_{\xi,\Sigma}(\Phi) := \int_\Sigma J_\xi(\Phi)$. We note that $Q_{\xi,\Sigma}(\Phi)$ is independent of the Σ chosen (so long as the integral is well-defined!).[73] That is: $Q_{\xi,\Sigma}(\Phi)$ is a quantity that is constant in time within the solution (V, Φ). Thus we might as well denote it simply $Q_\xi(\Phi)$, and think of the Noether charge, Q_ξ, associated with ξ as a function on $\mathcal{S}$.

REMARK 22 (Noether Charges Generate Symmetries). Since Ω is a closed two-form, $(\mathcal{S}, \Omega)$ is a symplectic or presymplectic space: so the results of sections 3.2 and 3.3 above apply. As one would expect, Q_ξ is in fact the symplectic/a presymplectic generator of the one-parameter group ξ (thought of now as acting on $\mathcal{S}$). The beauty of Noether's result is that it shows how to construct the generator of ξ via the integration of local objects on spacetime.

REMARK 23 (Trivial Conservation Laws). So far, nothing we have said guarantees that Q_ξ is an *interesting* function on $\mathcal{S}$—it might, for instance be a the zero function, if $J_\xi(\Phi)$ is exact as an $(n-1)$-form on V. Such trivial Noether charges do in fact occur when Ω is presymplectic and ξ is a group of gauge transformations. We will see examples of this in section 6.2 below.

5 TIME AND CHANGE IN WELL-BEHAVED FIELD THEORIES

Turn we now to the representation of time and change in physical theories. In the remaining sections Hamiltonian formulations of theories will play an important role. So we henceforth restrict attention to theories with second-order equations of motion.

In this section, we discuss ideally well-behaved theories. We impose three further assumptions, which are in effect jointly for this section only: (a) global existence of solutions; (b) uniqueness of solutions; (c) our spacetime admits a time translation symmetry under which the variational problem of our Lagrangian is invariant.

We will see that in this context, we have three $\mathbb{R}$-actions: a notion of time translation on spacetime; a notion of time translation on the space of solutions of the theory; and a notion of time evolution on the space of initial data of the theory. We also find that the space of solutions and the space of initial data are isomorphic as symplectic spaces, and that there is a natural intertwining of the notion of time

[72] The Noether current associated to L and ξ is the element of $J_\xi \in \Omega^{n-1,0}_{loc}(V \times \mathcal{K})$ given by $J_\xi(\Phi) := R(\Phi) - M(\Phi, X(\Phi))$, where X is the infinitesimal generator of ξ, R is the object introduced in the preceding footnote, and M is the object introduced in footnote 53.

[73] Because $J_\xi(\Phi)$ is closed as an $(n-1)$-form on V and because we have been careful in our choice of notion of instant (see fn. 47). Note, in fact, that so long as $\Sigma, \Sigma' \subset V$ are compact $(n-1)$-manifolds that determine the same homology class in V, we will have $\int_\Sigma J_\xi(\Phi) = \int_{\Sigma'} J_\xi(\Phi)$ (see, e.g., [Lee, 2003, p. 431] and [Lee, 2000, p. 300 f.] for relevant notions and results). Hence we get a sort of conservation law even if, e.g., Σ and Σ' are not spacelike according to the geometry of V. See [Torre, Unpublished] for an introduction to such conservation laws.

translation on the space of solutions with the notion to time evolution on the space of initial data. So in this domain one can say simply (if awkwardly) that time is represented as a symmetry of the laws—and leave it open whether one means time translation or time evolution, since in the end the two come to much the same thing.

This section has five subsections. The first is devoted to the Lagrangian picture, the second to the Hamiltonian, the third to the relation between these pictures, the fourth to a discussion of the representation of time and change. The final subsection offers an overview.

5.1 The Lagrangian Picture

Let us be more precise about the special assumptions in play in this section. We impose the following conditions on our spacetime V, equations of motion Δ, and Lagrangian L.

Global Existence of Solutions. We assume that each admissible set of initial data for Δ is consistent with a solution defined on all of V.[74]

Uniqueness of Solutions. If Φ and Φ' are solutions that agree in the initial data that they induce on an instant $\Sigma \subset V$, then they agree at any point $x \in V$ at which they are both defined.

Time Translation Invariance of the Lagrangian. We require our spacetime V to have a nontrivial geometrical structure, strong enough to single out a class of $(n-1)$-dimensional submanifolds that count as instants and a class of one-dimensional submanifolds that count as possible worldlines of point-particles. Let $\bar{\xi} = \{\bar{g}_t\}$ be a one-parameter group of spacetime symmetries of V, and consider the orbits of $\bar{\xi}$ in V (the orbit $[x]$ of $\bar{\xi}$ through $x \in V$ is the image of the curve $x(t) := \bar{g}_t \cdot x$). We call $\bar{\xi}$ a *time translation group for* V if the orbits of $\bar{\xi}$ are possible worldlines of point-particles according to the geometry of V; in this case, we call these orbits *worldlines adapted to* $\bar{\xi}$. We will typically denote time translation groups as $\bar{\tau}$.

Let $\bar{G}$ be a group of spacetime symmetries of V. Given $\bar{g} \in G$ we can define a diffeomorphism $g : \mathcal{K} \to \mathcal{K}$ via $g(\Phi(x)) = \Phi(\bar{g}^{-1} \cdot x)$. In decent Lagrangian theories, one expects that if $\bar{\xi} = \{\bar{g}_t\}$ is a one-parameter group of spacetime symmetries, then $\xi = \{g_t\}$ is a Noether group for L. In this situation, ξ will map solutions to solutions; so that each $g_t \in \xi$ restricts to a map from $\mathcal{S}$ to itself; these maps are symplectic automorphisms of $(\mathcal{S}, \Omega)$ (I won't bother introducing notation to distinguish between the action of ξ on $\mathcal{K}$ and the restriction of this action to $\mathcal{S}$). In this section we assume that each time translation group, $\bar{\tau}$, of V gives rise in this way to a Noether group, τ, of L. I will call such a τ a *dynamical time translation group.*

[74] Since we are restricting attention to theories with second-order equations of motion, specifying initial data involves specifying the field values and their time-rate of change at some initial instant.

Within the class of theories that arise in physics, it appears to be an immediate consequence of the uniqueness assumption that the form Ω induced by L on the space of solutions in nondegenerate, and hence symplectic. We denote by H the corresponding conserved quantity guaranteed by Noether's theorem (in physically realistic theories, H arises by integrating the stress-energy of the field over an arbitrary instant).[75] Of course, H generates, via Ω, the action of τ on $\mathcal{S}$.

EXAMPLE 24 (Field Theory in Newtonian Spacetime). In Newtonian spacetime, each symmetry can be written as the product of a time translation with an isometry of absolute space. In coordinates adapted to the privileged absolute frame, we can write points of spacetime as $(t, \mathbf{x})$. Then the (orientation-preserving) symmetries of V are of the form $(t, \mathbf{x}) \mapsto (t+s, R(\mathbf{x}) + \mathbf{c})$, where $s \in \mathbb{R}$ implements a time translation, R is a matrix implementing a rotation in absolute space, and $\mathbf{c} \in \mathbf{R}^3$ implements a spatial translation. Up to a choice of temporal unit, there is a unique time translation group, $\bar{\tau} : (t, \mathbf{x}) \mapsto (t+s, \mathbf{x})$; the worldlines of the points of absolute space are adapted to this group. We are supposing that the corresponding group τ acting on the space of solutions is a dynamical time translation group. The Noether charge associated with τ, $H : \mathcal{S} \to \mathbb{R}$, assigns to each solution the total energy of the system at any instant (since we are considering a theory invariant under time translations, the value of the total energy along a slice is a constant).[76]

EXAMPLE 25 (Field Theory in Minkowski Spacetime). The symmetry group of Minkowski spacetime is the Poincaré group. Each inertial frame picks out a notion of simultaneity, and a time translation group, $\bar{\tau}$; the worldlines of observers at rest in the chosen frame will be adapted to this group. (Equivalently, such group is determined by the choice of a timelike vector in spacetime.) In Poincaré-invariant field theories we can choose inertial coordinates (t, x_1, x_2, x_3) such that our chosen $\bar{\tau}$ acts via $(t, x_1, x_2, x_3) \mapsto (t+s, x_1, x_2, x_3)$. In such coordinates, the Noether current is just the component T^{00} of the stress-energy tensor of the field—the Noether charge being given, as always, by the integral of the Noether current over any instant.[77]

EXAMPLE 26 (Field Theories in a Curved Spacetime). While a generic general relativistic spacetime admits no non-trivial symmetries, a solution in which, intuitively, the geometry of space is constant in time admits a time translation group.

[75] For the stress-energy tensor and its role in the examples below, see [Choquet-Bruhat and DeWitt-Morrette, 1989, §II.7] and [Deligne and Freed, 1999, §2.9].

[76] The Noether charge generating spatial translation (rotation) in a given direction (about a given axis) assigns to a solution the corresponding component of the linear (angular) momentum of the system at an instant. In fact, we get a momentum map (see fn. 34) for the action of the group of symmetries of Newtonian spacetime—the Poisson bracket algebra of the Noether charges mirrors the Lie bracket relations between the infinitesimal generators of the corresponding one-parameter groups. It is impossible, however, to construct a momentum map for the symmetry group of neo-Newtonian spacetime; see [Woodhouse, 1991, §3.4] for this and other examples in which the construction of a momentum map is impossible.

[77] One can again construct a momentum map (see fn. 34)—with spacelike translations generated by the components of linear momentum, etc., in the familiar way.

Let V be a globally hyperbolic and time-oriented general relativistic spacetime that possess such a $\bar{\tau}$. Let X_a be the vector field tangent to the orbits of $\bar{\tau}$ (so X is a timelike Killing field). Let $T^{ab}(\Phi)$ be the stress-energy tensor of the field Φ and suppose that $\nabla_a T^{ab} = 0$ (this typically holds in cases of physical interest). Let $\Sigma \subset V$ be an instant (i.e., a Cauchy surface) and let n_a be the field of unit future-pointing normal vectors along Σ. We can define the energy-momentum vector of T^{ab} relative to X_a as $P^b := X_a T^{ab}$ and define the energy along Σ as $\int_\Sigma P^a n_a dx$. This last quantity is in fact the Noether charge, and is independent of the Σ.

5.2 *The Hamiltonian Picture*

The basic idea behind the Hamiltonian approach is to work with the space of initial data of the equations of the theory rather than with the space of solutions to the equations—roughly and heuristically speaking, this means working with the space of instantaneous states of the theory rather than with its space of possible worlds.

Deterministic equations of motion tell us what the state of the system must be at earlier and later times if it is in a given initial state. So, at least for well-behaved equations of motion, the dynamical content of the equations of motion ought to be encodable in a flow on the space of initial data, with the integral curves of this flow being the dynamically possible trajectories through the space of instantaneous states.

The special assumptions in play in this section (global existence and uniqueness of solutions and the presence of a dynamical time translation group) imply (at least for the sort of the theories that arise in physics) that the space of initial data carries a symplectic structure that generates the dynamics of the theory when supplemented by the function that assigns to an initial data set the total energy of a system in that state. The dynamics can be thought of as encoded in an $\mathbb{R}$-action on the space of initial data that implements time evolution. As we will see, these structures on the space of initial data—symplectic structure, Hamiltonian, and group action—are all closely related to the corresponding objects on the space of solutions that arise on the Lagrangian side.

Intuitively speaking, an instantaneous state of the field is a specification at each point of space of the value of the field and its time rate of change; and in giving a sequence of such instantaneous states, we describe how the values of these variables evolve through time at each point of space. So in order to construct a Hamiltonian formulation of a theory in which the total history of a system is described via a trajectory through the space of initial data, we need to effect some sort of notional decomposition of spacetime into space and time.[78]

[78] Note that we did not require any such decomposition in setting up the Lagrangian formalism in section 4 above. It is, of course, crucial to distinguish the symplectic space of solutions that arises in this formalism from the symplectic velocity phase space that arises in some Lagrangian treatments—the latter does, while the former does not, presuppose a division of spacetime into instants. Cf fn. 6 above.

Informally, we can picture what is required as being a preferred family of observers together with a notion of simultaneity. Spacetime is partitioned by the worldlines of these observers (these need not be at rest relative to one another, but we do require that the worldlines involved be possible worldlines of point-particles according to the geometry of V). Each observer carries a clock; and we suppose that the set of points at which these clocks read $t = 0$ forms an instant in V. We call such a set of observers equipped with such a notion of simultaneity a *frame*. We say that a frame is *adapted to* the time translation group $\bar{\tau}$ when the following two conditions obtain. (i) The worldline of each observer is an orbit of the group $\bar{\tau}$ acting on V. (ii) Up to a choice of origin and a choice of unit of measurement, $\bar{\tau}$ gives us a parameterization of the set of instants of V, which allows us to determine ratios of temporal intervals; we require that the clock readings respect these determinations.

Still speaking informally, we can say that relative to a choice of frame, the state of the field at time t is an assignment of field value and momentum to each observer (i.e., the values of the field and its time rate of change at the spacetime point the observer occupies at the given instant), and that a history of the field is an assignment of these quantities along the worldline of each observer. So we can take an initial data set to be a pair of functions (corresponding to the field value and its time rate of change) defined on the space of observers of our frame—this space acts as a sort of abstract instant, with the same topological and geometrical structure as the concrete instants that arise as subsets of V.

We can make this more precise as follows.

DEFINITION 27 (Slicing). Let V be a spacetime with geometry and let S be an $(n-1)$ -dimensional manifold (possibly carrying a Riemannian metric). Then an S*-slicing* of V is a diffeomorphism $\sigma : \mathbb{R} \times S \to V$ such that: (i) each $\Sigma_t := \sigma(\{t\} \times S)$, $t \in \mathbb{R}$, is an instant in V (with σ providing an isomorphism between the geometry of Σ_t and the geometry of S, if any); (ii) each $X_x := \sigma(\mathbb{R} \times \{x\})$, $x \in S$, is a possible worldline of a point-particle according to the geometry of V. We call S the *abstract instant* of σ and each Σ_t an *instant in the slicing*. When V admits a time translation group $\bar{\tau}$, we call a slicing σ of V *adapted* to $\bar{\tau}$ if the following conditions are met: (a) each X_x is an orbit of $\bar{\tau}$; (b) any two instants of the slicing are related by a time translation in $\bar{\tau}$; (c) up to a choice unit and origin, the parameterizations of each X_x given by σ and by $\bar{\tau}$ agree.

EXAMPLE 28 (Newtonian Slicings). In Newtonian spacetime there is of course a unique partition of spacetime by instants and (up to a choice of unit) a unique time translation group $\bar{\tau}$. Furthermore, in this setting it is possible to take S to be the space of worldlines of the points of absolute space.[79] So the only freedom in constructing a slicing adapted to $\bar{\tau}$ is in choosing an origin and a unit for the parameterization of the instants by the reals.

[79]This space comes equipped with a natural Euclidean structure—since the distance between points of absolute space is constant in time, we can define the distance between two worldlines of such points to be the distance between the points.

EXAMPLE 29 (Flat Minkowski Slicings). In the setting of Minkowski spacetime it is sometimes natural to restrict attention to instants which arise as surfaces of simultaneity for inertial observers. In this case, our abstract instant S will again have the structure of Euclidean space. In order to construct a slicing, we must choose an instant $\Sigma_0 \subset V$ corresponding to $t = 0$, an isometry from S to Σ_0, a unit of temporal measurement, and a notion of time translation associated to an inertial observer.

EXAMPLE 30 (Generic Minkowski Slicings). More generally, in the Minkowski spacetime setting it is possible to allow arbitrary Cauchy surfaces to count as instants—in this case, one will choose S to have some non-trivial Riemannian geometry. Now there is a truly vast—indeed, infinite-dimensional—family of instants to choose from (as we allow the geometry of S to vary). On the bright side, a generic instant admits no nontrivial isometries—so having chosen S and Σ there will be no freedom in constructing an isometry from one to the other.

Let us consider a Lagrangian theory satisfying all of the present conditions, and fix a slicing of V adapted to a notion of time translation, $\bar{\tau}$ that gives rise to a dynamical time translation group τ. We can then construct a Hamiltonian version in the following steps.

1. *Given an instant and a solution, construct the instantaneous field configuration and momentum.* Let Σ be an instant contained in the given slicing and let $\Phi : V \to W$ be a solution. We define $\phi : \Sigma \to W$, the field configuration on Σ, by $\phi := \Phi \mid_\Sigma$. And we define $\dot{\phi} : \Sigma \to W$, the field velocity on Σ, as follows: at each $x \in \Sigma$, $\dot{\phi}(x)$ is the rate of change at x of the field values along the orbit of $\bar{\tau}$ through x.[80] In order to construct the instantaneous momentum of the field, we apply the usual recipe for constructing canonical momentum variables, defining $\pi := \frac{\partial L}{\partial \dot{\phi}}$ (π is a map from Σ to W^*).

2. *Given the instantaneous field configuration and momentum, construct the corresponding initial data.* This is just a matter of using σ to pull back ϕ and π, so that we can regard them as functions on S rather than Σ. Sloppily, I will use the same names for initial data defined on S and the corresponding functions defined on $\Sigma \subset V$.

3. *Construct the space of initial data, $\mathcal{I}$.* Let $\mathcal{Q}$ be the space of all $\phi : S \to W$ that can arise via the previous two steps as we allow Φ to vary in $\mathcal{S}$.[81] The set of all pairs (ϕ, π) that can arise via these steps is just cotangent bundle, $T^*\mathcal{Q}$. This space is the space, $\mathcal{I}$, of initial data for our theory. It carries a canonical symplectic structure, ω (see example 7 above).

[80] That is, let $x_0 \in \Sigma$ and find $y_0 \in S$, $t_0 \in \mathbb{R}$ such that $\sigma(t_0, y_0) = x_0$ and define the curve $x : \mathbb{R} \to V$ by $x(t) := \sigma(t, y_0)$; then let $\dot{\phi}(x) := \lim_{h \to 0} \frac{1}{h} \left(\phi(x(t_0 + h)) - \phi(x_0)\right)$.

[81] Allowing Σ and the slicing to vary as well would make no difference in the present case, so long as S and its geometry are held fixed.

4. *Construct a Hamiltonian.* We define $h : \mathcal{I} \to \mathbb{R}$, the Hamiltonian on the space of initial data, as follows. Let $(\phi, \pi) \in \mathcal{I}$ be initial data and let $\Sigma \subset V$ be an instant (not necessarily one in our slicing). Let Φ be the solution that induces (ϕ, π) on Σ and define $h(\phi, \pi) := \int_\Sigma \pi(x)\dot{\phi}(x) - L(\Phi)(x)dx$ (in the present context, the result does not depend on the instant Σ chosen).[82]

5. *Construct the Dynamics.* Together h and ω determine a vector field χ on $\mathcal{I}$ that encodes the dynamics of our theory. The integral curves of χ are the possible dynamical trajectories—if the state is (ϕ_0, π_0) at time $t = 0$, then the state t units of time later can be found by tracing t units of time along the integral curve passing through (ϕ_0, π_0). This gives us a flow on $\mathcal{I}$, which preserves both ω and h (the flow is global rather than local because we are assuming that solutions are defined for all values of t).

5.3 Relation between the Lagrangian and Hamiltonian Pictures

For each instant $\Sigma_t := \sigma(\{t\} \times S)$ in our slicing σ, we define $T_{\Sigma_t} : \mathcal{S} \to \mathcal{I}$ to be the map that sends a solution Φ to the initial data set $(\phi, \pi) \in \mathcal{I}$ that results when the slicing σ is used to pullback to S the initial data induced by Φ on Σ_t. Because we are assuming global existence and uniqueness for solutions given initial data, T_{Σ_t} is in fact a bijection. Indeed, it is a diffeomorphism. Furthermore, $T^*_{\Sigma_t}\omega = \Omega$, so each T_{Σ_t} is in fact a symplectic isomorphism between $(\mathcal{S}, \Omega)$ and $(\mathcal{I}, \omega)$.

Note that in typical theories distinct instants in the slicing lead to distinct isomorphisms. If Σ_t and $\Sigma_{t'}$ are instants in our slicing and $T_{\Sigma_t} = T_{\Sigma_{t'}}$, then for each solution Φ, Φ induces the same initial data on Σ_t and $\Sigma_{t'}$—i.e., each solution is periodic with period $|t - t'|$. So if $T_{\Sigma_t} = T_{\Sigma_{t'}}$ for each Σ_t and $\Sigma_{t'}$, then every solution would have to be a constant function on V.

The maps $T_{\Sigma_t} : \mathcal{S} \to \mathcal{I}$ establish a simple relationship between our Hamiltonians $H : \mathcal{S} \to \mathbb{R}$ and $h : \mathcal{I} \to \mathbb{R} : h = H \circ T^{-1}_{\Sigma_t}$ (we could have taken this as our definition of h).

Together Ω and H determine the flow on $\mathcal{S}$ that implements time translation at the level of solutions while together ω and h determine the flow on $\mathcal{I}$ that implements time evolution of initial data. Since any T_Σ relates Ω and ω on the one hand, and H and h on the other, one would hope that it would also intertwine the group actions corresponding to these flows. This is indeed the case. Let us write $t \cdot_\mathcal{S} \Phi$ for the solution that results when we time-translate the solution Φ by t units and let us write $t \cdot_\mathcal{I} (\phi, \pi)$ for the state that initial data set (ϕ, π) evolves into after t units of time. Then we find that $t \cdot_\mathcal{I} T_\Sigma(\Phi) = T_\Sigma(t \cdot_\mathcal{S} \Phi)$.[83]

Relative to a slicing, each solution Φ on V corresponds to a curve $(\phi(t), \pi(t))$ in the space of initial data, with $(\phi(t), \pi(t)) := T_{\Sigma_t}(\Phi)$. And a curve of this form

[82]Here we use the fact that π takes values in W^* while $\dot{\phi}$ takes values in W; and we rely on the natural measure induced by the geometry of V to allow us to treat $L(\Phi)$ as a function rather than an n-form.

[83]That is, each T_Σ is equivariant for the $\mathbb{R}$-actions $\cdot_\mathcal{S}$ and $\cdot_\mathcal{I}$.

is always a dynamical trajectory in $\mathcal{I}$ (i.e., an integral curve of the flow generating time evolution on $\mathcal{I}$). Conversely, a dynamical trajectory $(\phi(t), \pi(t))$ in $\mathcal{I}$ determines a unique solution $\Phi := T_{\Sigma_0}^{-1}(\phi(0), \pi(0))$—and this solution can be viewed as the result of laying down the instantaneous field configurations $\phi(t)$ on the instants Σ_t in the slicing.

5.4 Time and Change

Change consists in a single object having a given property at a given time and a distinct and incompatible property at a different time. Within the Lagrangian approach, it is easy enough to draw a distinction between those solutions that represent change and those that do not: the changeless solutions are those which are invariant under the action of a group of time translations. Correspondingly, we will say that a dynamical trajectory in the space of initial data represents a changeless reality when the corresponding solution on V is invariant under some time translation group.[84]

This much is entirely straightforward. But it is worth pausing and thinking about how change is represented at the level at which physical quantities are represented by functions on $\mathcal{S}$ and $\mathcal{I}$. In the case of quantities defined on the space of initial data, the story is straightforward. Let $f \in C^\infty(\mathcal{I})$ correspond to some determinable physical property of instantaneous states. Then if (ϕ_0, π_0) evolves into (ϕ_1, π_1) and $f(\phi_0, \pi_0) \neq f(\phi_1, \pi_1)$ then the solution including these states manifests change with respect to the property represented by f.[85] And we can of course go on to ask, e.g., about the rate of change of f along a dynamical trajectory.

But how should we phrase this in terms of functions defined on the space of solutions?

Suppose that we are interested in the quantity that measures the volume of the spatial region on which a given field takes on non-zero values. While such a quantity is represented within the Hamiltonian framework by a function $f : \mathcal{I} \to \mathbb{R}$, there is no function on the space of solutions that can be identified with this quantity—for such functions assign values to entire physically possible histories, and thus cannot represent quantities that take on different values at different instants within a history (or rather, they cannot do so in the same direct way that

[84]Naively, one might think that a dynamical trajectory in the space of initial data should count as representing a changeless reality only if it is constant—that is, if the system is represented as being in the same instantaneous state at each instant of time. But this would be a mistake. Consider a well-behaved theory set in Minkowski spacetime, and let Φ be solution invariant under the notion of time translation associated with inertial frame A but not invariant under that corresponding to inertial frame B. Surely this counts as changeless—and ought to whether we pass to the Hamiltonian picture via a slicing adapted to frame A (which leads to a dynamical trajectory according to which the state of the system is constant) or via a slicing adapted to frame B (which leads to a picture in which the state undergoes nontrivial evolution).

[85]Even if Φ represents a state of affairs in Minkowski spacetime, changeless in virtue of being invariant under the notion of time translation associated with inertial frame A, it may still represent some physical quantities as undergoing change—such as the location of the centre of mass of a system relative to inertial frame B.

functions on the space of initial data can).

However: intuitively, for each instant $\Sigma \subset V$ there is a function $f_\Sigma : \mathcal{S} \to \mathbb{R}$ such that $f_\Sigma(\Phi)$ is the volume of the support of our field on Σ in the solution Φ. So it is tempting to say that our chosen quantity is represented as exhibiting change in a solution Φ if $f_\Sigma(\Phi) \neq f_{\Sigma'}(\Phi)$ for instants $\Sigma, \Sigma' \subset V$, and that in order to speak of the rate of change of our quantity we need to consider a parameterized family Σ_t of instants, and calculate $\frac{d}{dt} f_{\Sigma_t}(\Phi)$.[86]

Of course, in the present context, it makes sense to employ our preferred slicing in setting up this framework.[87] For each instant Σ_t in our slicing we have a symplectic isomorphism $T_{\Sigma_t} : \mathcal{S} \to \mathcal{I}$. If $f : \mathcal{I} \to \mathbb{R}$ is the function on the space of initial data that represents the quantity of interest, then $f_t := f \circ T_{\Sigma_t}$ is the desired function on the space of solutions that assigns to a solution Φ the value of f on the initial data that Φ induces on Σ_t. So each slicing σ determines a one-parameter family of functions on $\mathcal{S}$ that encodes the instantaneous values of our chosen physical quantity relative to the instants in σ. So relative to a choice of slicing, it makes sense to ask whether this quantity undergoes change, what the rate of change is, and so on.

REMARK 31 (An Alternative Approach to Constructing $\{f_t\}$). In the present setting, rather than relying on our entire one-parameter family of isomorphisms, $\{T_{\Sigma_t}\}$, to set up our one-parameter family of functions $\{f_t\}$, we could have used Σ_0 to construct f_0 then used our dynamical time translation group to define

$$f_{-t}(\Phi) := f_0(t \cdot_{\mathcal{S}} \Phi).$$

5.5 *Overview*

We have seen that if we put in place a number of very strong assumptions, we get in return a very clear picture of the representation of time and change. The assumptions are: that our equations of motion, Δ, are second-order; that these equations have ideal existence and uniqueness properties and they derive from a Lagrangian, L, that has a dynamical symmetry group, τ, that arises from time translation group, $\bar{\tau}$, on our spacetime, V; and that we have chosen a slicing σ of V that is adapted to $\bar{\tau}$.

Lagrangian Picture. The space of solutions, $(\mathcal{S}, \Omega)$, is a symplectic space. The function, $H : \mathcal{S} \to \mathbb{R}$, that assigns to each solution the total instantaneous energy relative to $\bar{\tau}$ is the symplectic generator of τ (and also the Noether conserved quantity associated with it).

[86] For this suggestion, see, e.g., [Rovelli, 1991].

[87] Otherwise we can run into trouble. Consider a Φ defined on Minkowski spacetime such that for each inertial observer the spatial volume of the region in which the field is nonzero is constant in time. Because of length contraction, relatively moving inertial observers will assign different values to this volume. So if we choose Σ and Σ' belonging to slicings corresponding to distinct inertial frames, then we find that $f_\Sigma(\Phi) \neq f_{\Sigma'}(\Phi)$ even though Φ is changeless according to each inertial observer.

Hamiltonian Picture. We are able to construct a Hamiltonian version of our theory: a symplectic space of initial data $(\mathcal{I}, \omega)$ equipped with a Hamiltonian $h : \mathcal{I} \to \mathbb{R}$ that generates the dynamics of the theory. The dynamics is encoded in an $\mathbb{R}$-action on $\mathcal{I}$ that implements time evolution.

Relation between the Pictures. To each instant Σ in our slicing corresponds the symplectic isomorphism $T_\Sigma : \mathcal{S} \to \mathcal{I}$, that maps a solution Φ to the initial data that it induces on Σ. Each such T_Σ relates H and h on the one hand and Ω and ω on the other—and intertwines the action of the group implementing time translation on $\mathcal{S}$ with the action of the group implementing time evolution on $\mathcal{I}$.

Time. Time, in one of its facets, is represented in this scheme by three $\mathbb{R}$-actions: the action via symmetries on V that implements time translation, the symplectic action implementing time translation on $\mathcal{S}$, and the symplectic action implementing time evolution on $\mathcal{I}$. Note: in some spacetimes there will only be one notion of time translation, in others there will be many.

Change. In the Lagrangian picture, changelessness is represented in a straightforward way—some solutions are invariant under a time translation group of their underlying spacetime. So change can be characterized as the absence of changelessness and the definition can then be translated into the language of the Hamiltonian approach. When it comes to representing change of given physical quantities via the behaviour of functions on the space of initial data and the space of solutions, things become a bit more interesting. Here it is the Hamiltonian picture that underwrites a straightforward approach: one finds the function on the space of initial data corresponding to the quantity of interest, and examines its behaviour as the state evolves. On the Lagrangian side, things are more complicated. No function on the space of solutions can directly represent a changeable quantity. But by employing the slicing-dependent correspondence between the two pictures one can find a one-parameter family of functions on the space of solutions, each of which describes the value of the quantity along a distinct instant from the slicing. One can use this one-parameter family to define the rate of change of the quantity; and so on.

6 COMPLICATIONS

The account of the previous section was underwritten by several very strong assumptions. I now want to consider the effect on the picture developed above if one or another of these assumptions is dropped. My strategy is to leave untouched the assumptions that we need to in order construct a Hamiltonian picture of the sort developed above—that the equations of motion be second-order and that spacetime have enough geometrical structure to support slicings—and to consider the effect of dropping the assumptions: (i) that solutions are defined globally in time; (ii) that there is a unique maximal solution consistent with any initial data set; (iii) that the Lagrangian admits a dynamical time translation group that arises from a time translation group on spacetime. I will in this section consider only the effect of dropping one of (i)–(iii) at a time—in the next section I will turn to general relativity, which is a theory in which (i)–(iii) fail, as does the assumption that spacetime has enough solution-independent geometry to support slicings.

In briefest sketch, we find that:

1. If we drop the assumption that solutions exist globally in time, then time evolution is no longer implemented by an $\mathbb{R}$-action on $\mathcal{I}$, and $\mathcal{S}$ and $\mathcal{I}$ are no longer symplectically isomorphic. But time evolution is implemented by a sort of local and infinitesimal counterpart of an $\mathbb{R}$-action and $\mathcal{S}$ and $\mathcal{I}$ are *locally* symplectically isomorphic. Overall, only small changes are required in the picture of the representation of time and change developed above.

2. If we drop the assumption that specifying initial data suffices to determine a unique solution, even locally in time, by considering the (broad and important) class of theories whose Lagrangian and Hamiltonian versions exhibit gauge freedom, then the space of solutions and the space of initial data are presymplectic spaces that are not isomorphic (even locally). Furthermore, time evolution will no longer be implemented by a one-parameter group, but by a gauge equivalence class of such groups. Difficulties also appear on the Lagrangian side. The problem appears to be that theories of this type feature nonphysical variables. The remedy is reduction—the reduced space of solutions and the reduced space of initial data are symplectic and isomorphic. Much of the picture of the representation of time and change can reappears at the reduced level.

3. If we drop the assumption that our Lagrangian is time-translation invariant, then we have to make do with time-dependent Lagrangian and Hamiltonian theories. Here the space of solutions and the space of initial data will be symplectic spaces, and will be isomorphic. But we no longer have time translation of solutions as a symmetry on the Lagrangian side, nor time evolution as a symmetry on the Hamiltonian side. Still, we are able to construct in the usual way a slicing-dependent one-parameter family of isomorphisms between the space of solutions and the space of initial data, and this allows us to reconstruct much of the familiar picture of the representation of time and change.

6.1 Singular Dynamics

Let us suppose that the condition of global existence of solutions fails for our equations of motion—there exist initial data sets that cannot be extended to solutions defined on all of V. But let us continue to suppose that our theory is otherwise well-behaved: our spacetime V has enough structure to support slicings; our equations Δ are second-order and have unique solutions; and our Lagrangian, L, admits a dynamical time translation group, τ, induced by a time translation group $\bar{\tau}$ on V. Then, at least for the sorts of cases that arise in physics, we can expect to find the following.

Lagrangian Picture. The space of solutions, $(\mathcal{S}, \omega)$, is a symplectic manifold. The dynamical time translation group, τ, acts on $\mathcal{S}$ in the usual way: each element of the group time-translates each solution by some given amount.[88] τ is generated, via Ω, by the Hamiltonian function, $H : \mathcal{S} \to \mathbb{R}$ that assigns to a solution the instantaneous energy of that solution.

Hamiltonian Picture. We can construct a Hamiltonian picture as above: given a time translation group $\bar{\tau}$, an adapted slicing σ, a solution Φ, and an instant Σ, we can construct the initial data that Φ induces on Σ relative to σ, and use σ to pull this back to our abstract instant S. We can then construct the space of initial data, $\mathcal{I}$, with its canonical symplectic form ω, use our Lagrangian to define a Hamiltonian, $h : \mathcal{I} \to \mathbb{R}$, and study the resulting dynamics. The essential novelty is that because some solutions have limited temporal domains of definition, one finds that the vector field on $(\mathcal{I}, \omega)$ generated by the h is incomplete—it has integral curves that are defined only on a subset of $\mathbb{R}$. So time evolution is not represented by an $\mathbb{R}$-action on $\mathcal{I}$: in general it does not make sense to ask of a given point in the space of initial data what state it will evolve to at arbitrarily late times. However, the vector field generated by the Hamiltonian, which as usual encodes the dynamics, can be thought of as a sort infinitesimal generator of a locally defined action of $\mathbb{R}$ on the space of initial data—in particular, if it makes sense to speak of data set x evolving into data set y after t units of time, then we find that the map that sends a state to the state t units of time later is a symplectic (and Hamiltonian-preserving) map between sufficiently small neighbourhoods of x and sufficiently small neighbourhoods of y.

Relation between the Pictures. As above, for each instant, Σ, in our slicing we can define $T_\Sigma(\Phi)$ to be the pullback to the abstract instant S of the initial data that the solution Φ induces on Σ. But now each each T_Σ is only partially defined as a function from $\mathcal{S}$ to $\mathcal{I}$ (since the value of $T_\Sigma(\Phi)$ is undefined when Φ is not defined on Σ). Nonetheless, each such T_Σ is a symplectic isomorphism between its domain of definition in $\mathcal{S}$ and $\mathcal{I}$. [89] As usual, we get a distinct such map for each instant we choose.

[88] Of course, if a solution is not defined for all time, then its domain of definition will differ from that of its time-translate in the obvious way.

[89] So, intuitively, the space of solutions is bigger than the space of initial data—we can find natural isomorphisms between the space of initial data and subspaces of the space of solutions.

Time. The representation of time becomes a bit more complicated in the present context: to each notion of time translation on spacetime corresponds a nice symmetry on the space of solutions—and a merely infinitesimal symmetry on the space of initial data.

Change. We can still represent changeable properties by functions on $\mathcal{I}$, and determine whether a given dynamical trajectory represents a change of such properties by studying the behaviour of the corresponding function along the trajectory. Despite the failure of global isomorphism between the space of solutions and the space of initial data, we find that a choice of slicing yields a one-parameter family of local isomorphisms, $\{T_{\Sigma_t}\}$, between $\mathcal{I}$ and subspaces of $\mathcal{S}$. Given a function f on the space of initial data corresponding to a quantity of interest, the family $\{T_{\Sigma_t}\}$ can be used to construct a one-parameter family of partially-defined functions $\{f_t\}$ on $\mathcal{S}$ that correspond to the given changeable physical quantity. So the representation of change in this case is much the same as in the case in which we have global existence of solutions.

The real novelty here is the lack of a global isomorphism between the space of solutions and the space of initial data. The phenomenon can be is well-illustrated by simple classical mechanical examples.

EXAMPLE 32 (The Kepler Problem). Consider a point-particle of mass m moving in the x-y plane subject to the gravitational influence of a point-particle of unit mass fixed at the origin.[90] Here our spacetime V will be $\mathbb{R}$ and the space W of field values will be the space $Q = \{(x, y)\}$ of possible positions of the moving particle. The Lagrangian is $L = \frac{1}{2}(\dot{x}^2 + \dot{y}^2) + \frac{m}{r}$, where $r^2 := x^2 + y^2$; the corresponding Hamiltonian is $H = \frac{1}{2}(\dot{x}^2 + \dot{y}^2) - \frac{1}{r}$. In order for L and H to be well-defined, we have to limit the location of the body to points in $Q := \mathbb{R}^2/(0, 0)$. We restrict attention to the case where $H < 0$. This is the case of orbits bounded in space—so, in particular, we rule out parabolic and hyperbolic motions.

We find that there are two types of solutions. (i) Regular solutions, in which the particle has non-zero angular momentum, are periodic and defined for all values of t; they represent the particle as moving along an ellipse that has the origin as a focus. (ii) Singular solutions, in which the particle has vanishing angular momentum, are defined only for $t_0 < t < t_0 + 2\varepsilon$; they represent the particle as being ejected from the origin at time t_0 (i.e., $|r(t)| \to 0$ as $t \to t_0$ from above), travelling outward along a ray from the origin with decreasing speed until reaching to a stop at time $t_0 + \varepsilon$, then falling back into the origin along the same ray, with $|r(t)| \to 0$ as $t \to t_0 + 2\varepsilon$ from below.

The space of solutions is topologically pathological. Let $\Phi(t)$ be a singular solution defined for $t \in (t_0, t_0 + 2\varepsilon)$. Let $\Lambda \subset Q$ be the line segment along which the particle moves according to Φ. It is possible to construct a sequence $\{\Phi_k\}$

[90] For the structure of the space of solutions of the Kepler problem, see [Woodhouse, 1991, §2.3] and [Marco, 1990b].

of regular solutions with the following features: each Φ_k has the same energy as Φ—it follows that each Φ_k represents the particle as moving periodically along an ellipse E_k with period 2ε; for each k, E_k is oriented so that the segment joining its foci is included in Λ; as $k \to \infty$, the eccentricity of E_k goes to infinity—so that $E_k \to \Lambda$ as $k \to \infty$. It follows that Φ is a limit of the sequence $\{\Phi_k\}$.[91] But now consider the solution $\Phi'(t) := \Phi(t + 2\varepsilon)$. This is a singular solution, defined for $t \in (t_0 - 2\varepsilon, t_0)$, that represents the particle as being emitted at time $t_0 - 2\varepsilon$, moving along Λ, and then being absorbed at time t_0. Φ' is equally a limit of $\{\Phi_k\}$. Indeed, we can generate an infinite number of limits of $\{\Phi_k\}$ by temporally translating Φ by multiples of 2ε.

Since we can find a sequence in $\mathcal{S}$ that converges to more than one limit point, $\mathcal{S}$ is not Hausdorff.[92] But, of course, the space of initial data for the Kepler problem is just T^*Q—which is Hausdorff. So the spaces are certainly not isomorphic!

Non-Hausdorff manifolds can be quite wild. But there are also relatively tame examples, such as the following. Let X be the result of excising the origin from the real line and adding in its place two new objects, a and b; a subset of X is an open ball if it coincides with an open interval in $\mathbb{R}$ that does not contain the origin, or if it arises when one takes an open interval of real numbers containing 0 and replaces 0 by one of $\{a, b\}$. We endow X with a topology by declaring that any union of open balls in X is an open set. X is a manifold according to our present standards. But it is not Hausdorff, since every neighbourhood of a overlaps with a neighbourhood of b—and, of course, a sequence like $\{\frac{1}{k}\}$ converges to both a and b.

More generally, we can construct a non-Hausdorff manifold $X_j^{n,m}$ by taking m copies of $\mathbb{R}^n$ and identifying them everywhere except on a given j-dimensional hyperplane through the origin ($1 < m \leq \infty$, $1 \leq n < \infty$, and $0 \leq j < n$).[93] The space of solutions corresponding to any fixed negative value of energy in the planar Kepler problem is assembled out of copies of $X_1^{3,\infty}$.

EXAMPLE 33 (Singularities of the n-Body Problem). For n particles in $\mathbb{R}^3$ the space of possible particle configurations is $\mathbb{R}^{3n}$. But this space includes collisions—and the potential energy for the n-body problem is singular at such points. So, as before, we let $\Delta := \{q \in \mathbb{R}^{3n} : \mathbf{q}_i = \mathbf{q}_j \text{ for some } i \neq j\}$ and let Q be $\mathbb{R}^{3n}/\Delta$ then take as our space of initial data $T^*Q = \{(q,p) \in T^*\mathbb{R}^{3n} : q \notin \Delta\}$.

We pose the initial data (q, p) at time $t = 0$. We know that this determines a maximal dynamical trajectory $t \mapsto (q(t), p(t))$, defined on an interval $[0, t^*)$, with $0 < t^* \leq \infty$ (the corresponding story for negative times is, of course, the same). Clearly it is possible to choose (q, p) so that t^* is finite—if we let $p = 0$ for $n > 1$,

[91] The topology on the space of solutions can be constructed as follows. For each $t \in \mathbb{R}$, take the position and velocity of the particle at a given time as coordinates on the space of solutions, and construct the set of open balls relative to these coordinates. Now take the union of these sets as t varies. The result is a basis for the basis for the topology we seek.

[92] Recall that a topological space X is *Hausdorff* if for any $x, y \in X$ it is possible to find disjoint open $U, V \subset X$ with $x \in U$ and $y \in V$. A sequence in a Hausdorff space has at most one limit.

[93] The example of the preceding paragraph is $X_0^{1,2}$.

for instance, the system is going to collapse and a collision will occur. Let us call our dynamical trajectory *singular* if $t^* < \infty$. It can be proved that if $t^* < \infty$, then as $t \to t^*$, $(q(t), p(t)) \to \Delta$, in the sense that $\lim_{t\to t^*} \min_{1\leq i<j\leq n} r_{ij} = 0$. Let us say that a singular trajectory ends in a *collision* if there is a point $(q_1, p_1) \in \Delta$ such that $\lim_{t\to t^*}(q(t), p(t)) = (q_1, p_1)$; otherwise, we say that it ends in a *pseudocollision*.

Consider the following cases.[94]

$\mathbf{n = 1}$. This is the case of a single free particle. The dynamics is non-singular.

$\mathbf{n = 2}$. This is the Kepler problem.[95] The only singularities are collision singularities. These occur if and only if the angular momentum of the system vanishes.

Famously, these singularities can be regularized.[96]

This is clear enough physically: one simply imposes the condition that any collisions that occur are elastic. This allows one to sew together a solution which ends with a collision at time t_0 with one, that, intuitively begins at time t_0 with the particles having interchanged their velocities. Continuing in this way, one constructs continuous and piecewise analytic solutions of eternal temporal extent. Because collision solutions are now infinite in temporal extent, the space of solutions, in this new sense, is isomorphic to the extended space of initial data that includes the collision states that lie in $T^*\Delta$ (let us interpret such states as representing the velocity that the particles will have when next emitted).

Mathematically, there are a number of underpinnings that can be given to this procedure.[97] An older one proceeds in terms of *analytic continuation*—thinking of the original collision solution as a complex function, one asks whether there is any analytic continuation of this function past the time of collision. Under a more modern approach, one looks for a way of continuing singular solutions that preserves the continuous dependence of evolution on initial data.[98] In the case of the two-body problem, either approach vin-

[94] For surveys of the singularities of the n-body problem, see [Diacu, 1992] and [Diacu, 2002]; for a popular treatment, [Diacu and Holmes, 1996, Chapter 3].

[95] Begin with the two-body problem. Restrict attention to the plane of motion of the particles; choose a frame in which the centre of mass of the two-body system is at rest at the origin and denote the positions of the bodies as $\vec{q}_1$ and $\vec{q}_2$. Obviously if we know $\vec{r} = \vec{q}_2 - \vec{q}_1$ then we know the positions of both particles (since we know their masses and the location of their centre of mass). Now note that the equation of motion for $\vec{r}$ is that for a single particle moving in an gravitational potential around the origin, if we take the origin to have unit mass, and the moving particle to have mass $m = \frac{m_1 m_2}{m_1+m_2}$.

[96] For various approaches to the regularization of two-body collisions, see [Souriau, 1982], [Marco, 1990a] and [Cushman and Bates, 1997, §II.3].

[97] See [McGehee, 1975] for these.

[98] More precisely, one excises from $T^*\mathbb{R}^{6n}$ an open set with compact closure that contains a collision; the boundary of this set falls into two pieces, corresponding to initial data of trajectories entering the set and initial data of trajectories leaving the set; evolution gives a diffeomorphism from the subset

dicates extension of singular solutions by elastic collisions as the unique tenable method of extension.

$\mathbf{n = 3.}$ Singular trajectories end in collisions. Collisions involving only two bodies can be regularized as elastic collisions. But some three body collisions are non-regularizable (according to any of several criteria).[99] Such three-body collisions are complex, and presumably make it difficult to determine the topology of the space of solutions—so in this case, unlike the $n = 2$ case, one does not have a clear picture of the relation between the global structure of the space of solutions and the structure of the space of initial data.

$\mathbf{n \geq 4.}$ For $n \geq 4$, as usual, singular trajectories can end in collisions: two-body collisions are regularizable; but at least some collisions involving larger numbers of particles are not. Furthermore, for $n > 4$ it is known that pseudocollisions can also occur—so it would appear to be more difficult then ever to determine the topology of the space of solutions.[100]

REMARK 34 (Quantization of Singular Systems). When the space of solutions and the space of initial data are isomorphic, it is, of course, a matter of indifference which space one takes as the starting point for quantization. When dynamics is singular and these spaces are no longer isomorphic one faces a real choice. And the choice is not entirely pleasant—one has to choose between the space of initial data, on which the dynamics is implemented by an incomplete vector field, and the space of solutions, which one expects to have a complicated and pathological topology. Presumably there is no guarantee that the two approaches always lead to the same quantizations in the domain of singular dynamics.[101]

6.2 Gauge Freedom

We next want to consider what happens when we drop the assumption that specifying initial data suffices to determine a unique maximal solution to our equations of motion. To this end, we are going to assume that our equations of motion *underdetermine* the behaviour of the field, in the radical sense that for given initial data

of the former corresponding to non-singular solutions to the subset of the latter corresponding to non-singular solutions; one asks whether this can be extended to a diffeomorphism of one whole piece to the other.

[99] See [McGehee, 1975].

[100] See [Saari and Xia, 1995]. The question is open for $n = 4$; but see [Gerver, 2003] for a possible example. Pseudocollisions require that the positions of at least some of the particles become unbounded as $t \to t^*$—by exploiting arbitrarily great conversions of potential energy into kinetic energy, these particles escape to infinity in a finite time. As emphasized by Earman ([1986, Chapter III] and [this volume]), this means that pseudocollisions involve a rather radical and surprising failure of determinism—which is most dramatic when one considers the time reverse of such a process, in which particles not originally present anywhere in space suddenly swoop in from infinity.

[101] For approaches to the quantization of systems with singular dynamics see, e.g., [Gotay and Demaret, 1983] and [Landsman, 1998].

the general solution consistent with that data contains at least one arbitrary function of the full set of independent variables.[102] There is a wide class of physical theories whose equations exhibit this *prima facie* pathological behaviour—including, most importantly, Maxwell's theory of electromagnetism, general relativity, and their generalizations.

In this subsection I will first sketch a little bit of the theory of Lagrangian treatments of such theories without making any special assumptions about time translation invariance, the global existence of solutions, or the structure of spacetime. These further assumptions will later be brought into play, and will underwrite a consideration of the Hamiltonian form of a theory that is well-behaved except in possessing under-determined dynamics, and of the representation of time and change in such theories. This discussion will be followed by three examples.

Let us begin by introducing the notion of a family of gauge symmetries of a Lagrangian theory. Recall that a group, G, acting on the space, $\mathcal{K}$, of kinematically possible fields is a group of variational symmetries of a Lagrangian, L, defined on $\mathcal{K}$ if the action of G is appropriately local and leaves the variational problem of L invariant.[103] We call a group, G, of variational symmetries of $(\mathcal{K}, L)$ a group of *gauge symmetries* if it can be parameterized in a natural way by a family of arbitrary functions on spacetime.[104] Roughly speaking, each function on spacetime generates a Noether group of symmetries of $(\mathcal{K}, L)$—a one-parameter group of (suitably local) symmetries of the variational problem of L.[105] Since the set of functions on spacetime is infinite-dimensional, any group of local symmetries of a theory is infinite-dimensional.

Most familiar groups of symmetries of physical theories—the group of isometries of a spacetime with non-trivial geometry, the group that acts by changing the phase of the one-particle wavefunction by the same factor at each spacetime point, etc.—are finite-dimensional, and hence not do not count as groups of gauge symmetries in the present sense.

It is easy to see that the equations of motion of a Lagrangian theory admitting such a group of gauge symmetries under-determine solutions to the theory. Let

---

[102] On the relevant notion of under-determined equations of motion, see, e.g., [Olver, 1993, pp. 170–172, 175, 342–346, and 377].

[103] For a more precise definition, see [Zuckerman, 1987, p. 274].

[104] Let us be more precise. First, let Y be a vector space, and let Γ be a space of functions from V to Y (more generally, let Γ be a space of sections of some vector bundle $E \to V$). We assume that Γ includes all smooth, compactly supported maps from V to W, but leave open the precise boundary conditions, smoothness conditions, etc., required to characterize Γ. (Special care regarding boundary conditions is required when Γ contains elements with noncompact support).

Now we define a group of *gauge symmetries* parameterized by Γ as a pair of linear and local maps, $\varepsilon \mapsto X_\varepsilon$ and $\varepsilon \mapsto R_\varepsilon$ sending elements ε of Γ to local vector fields on $\mathcal{S}$ and to elements of $\Omega^{n-1,0}_{loc}(V \times \mathcal{K})$, respectively, such that $\partial L(\Phi, X_\varepsilon(\Phi)) = DR_\varepsilon(\Phi)$ for all $\Phi \in \mathcal{S}$ and $\varepsilon \in \Gamma$. So each $\varepsilon \in \Gamma$ is associated with an infinitesimal generator of a Noether group for L (cf. fn. 71 above).

[105] A bit more precisely: the discussion of the previous footnote shows that the map $\varepsilon \mapsto (X_\varepsilon, R_\varepsilon)$ is a map from Γ to the set of generators of Noether groups of $(\mathcal{K}, L)$; in fact, the image of this map will be infinite-dimensional in nontrivial examples, although it may have a nontrivial kernel (in example 38 below, constant functions all generate the same (trivial) Noether group).

ε be a function on spacetime that vanishes everywhere but on some compact set $U \subset V$; if we allow the corresponding Noether group $\xi = \{g_t\}$ to act on a solution Φ, then for $t \neq 0$ the resulting solutions $\Phi_t = g_t \cdot \Phi$ will agree with Φ outside of U, but in general disagree with Φ inside U. Thus if we choose an instant $\Sigma \subset V$ which does not intersect U, we find that Φ and Φ_t induce the same initial data on Σ, but differ globally—so uniqueness fails for the equations of motion of the theory.

Recall from section 3.3 that a presymplectic form is a degenerate closed two-form, and that the imposition of such a form on a space serves to partition the space by submanifolds called gauge orbits. An argument very similar to that of the previous paragraph shows that if L admits a group of gauge symmetries, then the form Ω that L induces on the space of solutions is presymplectic, and that the corresponding gauge orbits are such that two solutions belong to the same gauge orbit if and only if they are related by an element of the group of gauge symmetries of L.[106] So gauge symmetries of L are gauge transformations of $(\mathcal{S}, \Omega)$, in the sense stipulated in section 3.3 above—they preserve the gauge orbits of the space of solutions.

It follows from general facts about presymplectic forms that if a function on $\mathcal{S}$ generates a one-parameter group of gauge symmetries, then that function is a constant function. In particular: the Noether conserved quantity, $Q_\xi : \mathcal{S} \to \mathbb{R}$, associated with a one-parameter group of gauge symmetries, ξ, must be a constant function.[107] Such conserved quantities are trivial, in the sense that they do not provide any means to distinguish between physically distinct solutions.

So much we can say about any Lagrangian theory admitting a group of gauge symmetries. Let us now specialize to the case where our equations of motion, Δ, are second-order, our spacetime, V, has enough geometrical structure to admit slicings, solutions exist globally in time, and our Lagrangian, L, admits a dynamical time translation group, τ. With these further assumptions in place, we can investigate the implications that giving up on local uniqueness of solutions has for the picture of time and change developed in section 5 above. We find the following.

Lagrangian Picture. We are assuming that we have a notion of time translation arising out of the structure of our background spacetime V. This notion gives rise, in the usual way, to a dynamical time translation group, τ. The corresponding conserved quantity is the usual Hamiltonian, H, which assigns to each solution its instantaneous energy. So far so good. But now recall from the discussion of section 3.3 above that in the setting of a presymplectic space, if a function generates a given one-parameter family of transformations of the space, then it also generates all one-parameter families of trans-

[106] Recall that in fn. 61 above, Ω was defined as the integral of a certain object over an arbitrary instant $\Sigma \subset V$. If we consider an infinitesimal local symmetry X_ε which has no effect on solutions along Σ, then Ω will not see X_ε—i.e., $X_\varepsilon(\Phi)$ will be a null vector at each $\Phi \in \mathcal{S}$. See [Deligne and Freed, 1999, §2.5] and [Woodhouse, 1991, p. 145].

[107] In fact, it will be the zero function, because the Noether current J_ξ will be exact as an $(n-1)$-form on V. See [Zuckerman, 1987, p. 274].

formation gauge equivalent to the given one. In the present case, this means that in addition to the dynamical time translation group, τ, H generates all one-parameter groups of transformation of $(\mathcal{S}, \Omega)$ that agree up to gauge with τ.

Hamiltonian Picture. Having fixed a notion of time translation in spacetime and an associated slicing of spacetime into instants, we can proceed as usual to construct the space of initial data that arise when the configuration and momentum variables of the field are restricted to an arbitrary instant in our slicing.[108] As in the well-behaved case, given an instant $\Sigma \subset V$ of our slicing and a solution Φ, we can construct a corresponding initial data set (ϕ, π) on our abstract instant S, by pulling back to S the initial data that Φ induces on Σ. In the well-behaved case, we found that the space of initial data had the structure $T^*\mathcal{Q}$, where $\mathcal{Q}$ was the space of all ϕ that arise as instantaneous field configurations by restricting solutions to instants. In the present case we find that the (ϕ, π) that arise as initial data sets form a subspace of $T^*\mathcal{Q}$ (where $\mathcal{Q}$ is again the space of all ϕ that arise as restrictions of solutions to instants).[109] In addition, we may also find that in order to construct consistent dynamics, we need to further restrict admissible initial data. The upshot is that we take as our space of initial data a subspace $\mathcal{I} \subset T^*\mathcal{Q}$. $\mathcal{I}$ comes equipped with a natural geometric structure: the cotangent bundle $T^*\mathcal{Q}$ comes equipped with its canonical symplectic form (see example 7 above); the restriction of this form to $\mathcal{I}$ yields a presymplectic form ω. When all goes well, the gauge orbits determined by ω have the following structure: initial data sets (ϕ, π) and (ϕ', π') arising as the initial data induced on a given instant $\Sigma \subset V$ by solutions Φ and Φ' belong to the same gauge orbit in $(\mathcal{I}, \omega)$ if and only if Φ and Φ' belong to the same gauge orbit in $(\mathcal{S}, \Omega)$.[110] One can go on to define a Hamiltonian function, h, on $(\mathcal{I}, \omega)$ in the usual way. Of course, since $(\mathcal{I}, \omega)$ is a merely presymplectic space, h generates a whole gauge equivalence class of notions of dynamics (i.e., one-parameter groups of symmetries of $(\mathcal{I}, \omega)$). Suppose that according to one such notion of dynamics, initial state x_0 evolves into state $x(t)$ at time t. Then although other notions of time evolution generated by h will in general disagree about what state x_0 evolves into at time t, they will all agree that the state at t lies in $[x(t)]$, the gauge orbit of $x(t)$.[111]

Relation between the Pictures. As usual, for each instant Σ in our slicing we

[108] On constructing the constrained Hamiltonian system corresponding to a given Lagrangian theory admitting gauge symmetries, see [Dirac, 2001], [Gotay *et al.*, 1978], and [Henneaux and Teitelboim, 1992]. For philosophical discussion, see [Earman, 2003] and [Wallace, 2003].

[109] This is because so-called *first-class constraints* arise: it follows from the definition of the momenta, $p_i := \frac{\partial L}{\partial \dot{q}_i}$, that some components of the momenta are required to vanish identically.

[110] *Warning:* it is not difficult to construct (unphysical) examples in which this nice picture fails—see example 36 below.

[111] In fact, for each point $y \in [x(t)]$, there is a notion of time evolution generated by h according to which x_0 evolves into y at time t.

can define $T_\Sigma : \mathcal{S} \to \mathcal{I}$, the map that sends a solution to the (pullback to S of the) initial data that it induces on Σ. In the setting of section 5, these maps gave us isomorphisms between the space of solutions and the space of initial data. But in the presence of gauge symmetries, these maps are *not* isomorphisms—since the existence of gauge symmetries implies that many solutions induce the same initial data on any given Σ. The situation is most dramatic when we consider a theory with only finitely many degrees of freedom which admits gauge symmetries: for then the space of solutions will be infinite-dimensional while the space of initial data will be finite-dimensional (see example 37 below).[112] When all goes well, we get the following picture of the relation between solutions and dynamical trajectories in the space of initial data (holding fixed a notion of time translation and a slicing adapted to it).

1. Let Φ be a solution and let $x(t) = (\phi(t), \pi(t))$ be the curve in $\mathcal{I}$ that arises by letting $x(t)$ be the initial data set that Φ induces on the instant $\Sigma_t \subset V$. Then $x(t)$ is a dynamical trajectory of the Hamiltonian version of the theory.
2. Given a dynamical trajectory, $x(t)$, of the Hamiltonian version of the theory we find that there is a unique solution $\Phi \in \mathcal{S}$ such that the curve in $\mathcal{I}$ that corresponds to Φ in the sense of the preceding clause is just $x(t)$.
3. If $\Phi, \Phi' \in \mathcal{S}$ belong to the same gauge orbit in the space of solutions, then the corresponding dynamical trajectories, $x(t)$ and $x'(t)$ in $\mathcal{I}$ agree up to gauge (in the sense that for each t, $x(t)$ and $x'(t)$ belong to the same gauge orbit in $\mathcal{I}$).
4. If dynamical trajectories $x(t)$ and $x'(t)$ in $\mathcal{I}$ agree up to gauge, then the solutions $\Phi, \Phi' \in \mathcal{S}$ to which they correspond belong to the same gauge orbit in $\mathcal{S}$.

Time. Our notion of time translation lifts in a nice way from our spacetime V to the space of solutions, $\mathcal{S}$, where we get the usual representation of time via an $\mathbb{R}$-action. Even here there is an oddity: the Hamiltonian that generates this action also generates each $\mathbb{R}$-action gauge-equivalent to it. The situation is messier still in the space of initial data: given a notion of time translation on spacetime and a slicing adapted to that notion, we can construct a Hamiltonian picture; but in the presence of gauge symmetries, we find that there are many dynamical trajectories through each point in the space of initial data. In effect, our single notion of time translation in spacetime splits into a multitude of $\mathbb{R}$-actions on the space of initial data, each with equal claim to be implementing the dynamics of the theory.

[112] So in this case we see that no map from $(\mathcal{S}, \Omega)$ to $(\mathcal{I}, \omega)$ is an isomorphism; intuitively this is true for any theory admitting local symmetries.

Change. The evolution of arbitrary quantities under the dynamics defined on the space of initial data is indeterministic: if $x_0 = (\phi_0, \pi_0)$ is an initial data set, there will be distinct dynamical trajectories $x(t)$ and $x'(t)$ passing through x_0 at time $t = 0$; for an arbitrary function $f : \mathcal{I} \to \mathbf{R}$, we have no reason to expect that $f(x(t)) = f(x'(t))$ for $t \neq 0$; so fixing the state at time $t = 0$ does not suffice to determine the past and future values of the quantity represented by f. But since in this situation $x(t)$ and $x'(t)$ will agree about which gauge orbit the state lies in at each time, we find that the evolution of gauge-invariant quantities (those represented by functions on the space of initial data that are constant along gauge orbits) is fully deterministic—given the initial state, one can predict the value of such a quantity at all times. Furthermore, our slicing allows us to associate with each gauge-invariant function, f, on the space of initial data a one-parameter family $\{f_t\}$ of gauge-invariant functions on the space of solutions: let $f_t(\Phi)$ be the value that f takes on the initial data set that Φ induces on Σ_t. In this way we can represent change of gauge-invariant quantities via functions on the space of initial data or the space of solutions in the usual way.

This last point, especially, ought to arouse the suspicion that our theory, in the form currently under consideration, contains surplus structure. For while the theory has some quite disappointing features—ill-posed initial value problem, trivial conservation laws, a merely presymplectic geometric structure, failure of even local isomorphism between the space of solutions and the space of initial data—one finds that there is a large subset of physical quantities that behave just as the quantities of a well-behaved theory do. One naturally wonders whether there might be a well-behaved theory governing the behaviour of these quantities lurking somewhere in the background.

This sort of suspicion motivates the application to $(\mathcal{S}, \Omega)$ and $(\mathcal{I}, \omega)$ of the reduction procedure discussed in section 3.3 above. When all goes well, the following picture emerges: the reduced space of solutions (i.e., the space of gauge orbits of the space of solutions) and the reduced space of initial data (i.e., the space of gauge orbits of the space of initial data) are both symplectic spaces—and these reduced spaces are isomorphic.[113] The Hamiltonian functions corresponding to time translation on the original space of solutions and time evolution on the original space of initial data project down to the reduced spaces. The resulting reduced Hamiltonians generates time translation and time evolution on their respective spaces.

In typical examples that arise in physics, one sees that the original theory's invariance under a group of gauge symmetries was in fact a sign that physically otiose variables had been included in the theory. Indeed: the fact that the original space of initial data is presymplectic with a symplectic reduced space indicates that within the original Hamiltonian formulation of the theory one can partition the set of variables parameterizing the original space of initial data into two classes, that

[113] example 36 below is an (unphysical) case where this isomorphism fails.

we will call the class of *physically relevant variables* and the class of *physically otiose variables*; specifying the initial values of all variables suffices to determine the values for all times of the physically relevant variables while leaving wholly arbitrary the evolution of the physically otiose variables.[114] At least locally the physically relevant variables can be taken to parameterize the reduced space of initial data. This provides a strong reason to think that the Hamiltonian theory defined on the reduced space of initial data gives a perspicuous representation of the physics under investigation, involving as it does exactly those quantities whose evolution is determined by the original theory. And this in turn provides good reason to think of the reduced space of solutions as representing possible histories of the system whose possible instantaneous states are represented by points in the space of initial data.[115]

REMARK 35 (Reduction and Determinism). Suppose that one is presented with a *prima facie* indeterministic theory, in which many future sequences of states are consistent with a given initial state. Then one could always construct a deterministic theory by simply identifying all of the futures consistent with a given state. As noted by Maudlin, it would be foolish to apply this strategy whenever one encountered an indeterministic theory: (i) general application of this strategy would render determinism true by fiat; and (ii) one would often end up embracing trivial or silly theories.[116] For example: in Newtonian physics, the initial state in which space is empty of particles is consistent with a future in which space remains empty, and also with a future in which particles swoop in from infinity, then interact gravitationally for all future time (see fn. 100 above); to identify these futures—to view them as mere re-descriptions of a single physical possibility—would be absurd.

Now, reduction is a special case of the general strategy that Maudlin objects to. But since just about any wise course of action is a special case of a strategy that is in general foolish, this is not in itself an objection to reduction. We ought to check whether the complaints that Maudlin quite rightly registers against the general strategy redound to the discredit of the special case. I claim that they do not. (i) It is true that reducing theories with gauge symmetries converts *prima facie* indeterministic theories into deterministic ones. But this is unobjectionable: the sort of indeterminism that is a *prima facie* feature of a theory with gauge symmetries (namely, the existence of quantities whose evolution is wholly unconstrained by the initial state of the system) appears to be unphysical. (ii) For the sort of theories

[114] In a presymplectic manifold satisfying suitable technical conditions, every point has a neighbourhood admitting a chart whose coordinates fall into two classes—those that parameterize gauge orbits and those that parameterize the directions transverse to the gauge orbits; see [Abraham and Marsden, 1978, Theorem 5.1.3]. In the space of initial data, it is natural to take the variables of the first type to be physically otiose and those of the second type to be physically relevant.

[115] Since this space arises by identifying solutions related by elements of the group of gauge symmetries of the theory, while the reduced space of initial data arises by identifying initial data that are induced on a given instant by solutions related by elements of the group of gauge symmetries of the theory.

[116] See [Maudlin, 2002, pp. 6–8].

that arise in physics, one does not have to fear that reduction will lead to a trivial or absurd result—in known cases, reduction carries one to a well-behaved symplectic space that is a suitable setting for a physical theory. Indeed, in such cases, it is (almost unanimously) agreed that the resulting symplectic space parameterizes the true degrees of freedom and provides the correct setting for the dynamics of the original theory.[117]

EXAMPLE 36 (A Pathological example). Before proceeding, it is important to emphasize that it is not hard to cook up simple (but unphysical) examples that do not follow the pattern sketched above for theories with gauge symmetries.[118] Consider a particle moving in the x-y plane with Lagrangian $L = \frac{1}{2}e^y\dot{x}^2$. The corresponding Euler–Lagrange equations tell us that x is constant in time while the evolution of y is wholly arbitrary. So the space of solutions consists of pairs $(x_0, y(t))$ where $x_0 \in \mathbb{R}$ and $y(t) : \mathbb{R} \to \mathbb{R}$ an arbitrary smooth function; two solutions $(x_0, y(t))$ and $(x_0', y'(t))$ belong to the same gauge orbit if and only if $x_0 = x_0'$. So the reduced space of solutions is just $\mathbb{R}$—which, having an odd number of dimensions, cannot carry a symplectic structure. On the Hamiltonian side one finds that the momentum conjugate to x and the momentum conjugate to y both have to vanish—which means that the space of initial data is $\mathbb{R}^2 = \{(x, y)\}$, with every point being gauge equivalent to every other.[119] So the reduced space of initial data is a single point—which is not isomorphic to the reduced space of solutions.

EXAMPLE 37 (Particles on a Line). We consider two gravitating point-particles moving on a line. For simplicity, we choose units so that Newton's constant is unity, assume that the particles have unit mass, and set aside worries about collisions and their regularization. We consider three theories of this system.

The Newtonian Theory. We denote the positions of the particles as q_1 and q_2 with $q_2 > q_1$. We interpret these as giving the positions of the particles relative to a frame at rest in absolute space. The Lagrangian for this system is $L = T - V$ where the kinetic energy is $T := \frac{1}{2}(\dot{q}_1^2 + \dot{q}_2^2)$ and the potential energy is $V = -\frac{1}{q_2 - q_1}$. The usual Newtonian equations of motion follow. It is helpful to consider a variant formulation of this theory. We define new configuration variables, $r_0 := \frac{1}{2}(q_1 + q_2)$ and $r_1 := q_2 - r_0 = \frac{1}{2}(q_2 - q_1)$ (so r_0 is the position of the centre of mass of the system, and r_1 is half the relative distance between the particles). In terms of these variables, our Lagrangian is $L(r_0, r_1, \dot{r}_0, \dot{r}_1) = \frac{1}{2}(\dot{r}_0^2 + \dot{r}_1^2) + \frac{1}{2r_1}$. The equations of motion tell

[117] General relativity provides the sole instance in which there is any dissent from the consensus view; see [Kuchař, 1986] and [Kuchař, 1993]. This is also the case that Maudlin is concerned with—he, like Kuchař, worries that unreflective application of reduction to general relativity leads to absurd conclusions about time and has hampered conceptual progress in quantum gravity. Part of the burden of section 7 below is to show that no absurdities follow from the application of reduction in that case.

[118] For the following example, see [Henneaux and Teitelboim, 1992, §1.2.2]. For further discussion of such examples, see [Gotay, 1983].

[119] This is an example where one constraint arises directly from the definition of the momenta while the other is required in order to formulate consistent dynamics.

us that r_0 is a linear function of time (since the centre of mass of an isolated system moves inertially) while $r_1(t)$ solves $\ddot{r}_1 = -\frac{1}{2r_1^2}$, and so describes the relative motion between the particles as they interact gravitationally.

The Leibnizean Theory. In this theory, space and motion are relative, and so the relative distance, r_1, between the particles is the only configuration variable (or rather, r_1 is half the relative distance). The Lagrangian for the Leibnizean theory is $L'(r_1, \dot{r}_1) := \frac{1}{2}\dot{r}_1^2 + \frac{1}{2r_1}$. The equation of motion is $\ddot{r}_1 = -\frac{1}{2r_1^2}$. So the Leibnizean theory gives the same dynamics for the relative distances between the particles as the Newtonian theory.

The Semi-Leibnizean Theory. We take both r_0 and r_1 as configuration variables, and take as our Lagrangian $L''(r_0, r_1, \dot{r}_0, \dot{r}_1) := \frac{1}{2}\dot{r}_1^2 + \frac{1}{2r_1}$ (so L'' is a function of r_0, r_1, $\dot{r}_0$, and $\dot{r}_1$ which happens to depend only on r_1 and $\dot{r}_1$). We apply the variational algorithm: as always, it leads to the conclusion that a curve $x(t) := (r_0(t), r_1(t))$ is a solution to the equations of motion if and only if $\frac{\partial L''}{\partial r_i} - \frac{d}{dt}\frac{\partial L''}{\partial \dot{r}_i} = 0$ is satisfied at each point on the curve for $i = 0, 1$. For $i = 1$, we again find that $\ddot{r}_1 = -\frac{1}{2r_1^2}$, so we get the same dynamics for the evolution of the relative distances as in the Newtonian and Leibnizean cases. But for $i = 0$, our condition on curves is empty, since L'' does not depend on either r_0 or $\dot{r}_0$. It follows that a curve $x(t) := (r_0(t), r_1(t))$ counts as a solution to our equations of motion if $r_1(t)$ describes a motion permitted by the Newtonian or Leibnizean theory and r_0 is any (continuous and appropriately differentiable) function whatsoever.

Let us contrast the structure of these three theories.

Symmetries. The group of variational symmetries of the Newtonian theory is three-dimensional, consisting of Galilean boosts and spatial and temporal translations. The group of variational symmetries of the Leibnizean theory is one-dimensional, consisting of time translations. But the variational symmetry group of the semi-Leibnizean theory is infinite-dimensional: in addition to temporal translations, it includes time-dependent spatial translations of the centre of mass as a group of gauge symmetries. If $r_0(t)$ and $r_1(t)$ are continuous functions, then $x(t) = (r_0(t), r_1(t))$ is a kinematical possibility. Let $\Lambda(t)$ be any other continuous function from $\mathbb{R}$ to itself. Then $x'(t) := (r_0(t) + \Lambda(t), r_1(t))$ is also a kinematical possibility and $L''(x(t)) = L''(x'(t))$ for all t (since L'' doesn't care at all about r_0). That is, the map $\Phi_\Lambda : (r_0(t), r_1(t)) \mapsto (r_0(t) + \Lambda(t), r_1(t))$ from the space of kinematical possibilities to itself preserves the Lagrangian, and hence is a variational symmetry. Indeed, for each such Λ we get a distinct variational symmetry of L''. So the space of continuous $\Lambda : \mathbb{R} \to \mathbb{R}$ parameterizes a group of gauge symmetries of L''.

Gauge Symmetries and the Initial Value Problem. We can exploit these symmetries to show how drastically ill-posed the initial value problem for the semi-Leibnizean theory is. Suppose that at $t = 0$ we fix values for r_0, r_1, $\dot{r}_0$, and $\dot{r}_1$. Let $x(t) = (r_0(t), r_1(t))$ be a solution satisfying those initial data. Now select $\Lambda : \mathbb{R} \to \mathbb{R}$ such that $\Lambda(0) = 0$ and $\dot{\Lambda}(0) = 0$. Since Φ_Λ is a Lagrangian symmetry, $\Phi_\Lambda(x(t)) = (r_0(t) + \Lambda(t), r_1(t))$ is also a solution—which, of course, satisfies the specified initial data at time $t = 0$. In this way, we can construct an infinite-dimensional family of solutions for each specified set of initial data.

Structure of the Space of Solutions. The spaces of solutions for the Newtonian and Leibnizean theories are symplectic spaces, of dimension four and two, respectively. As we have seen the space of solutions of the semi-Leibnizean theory is infinite-dimensional. And the form that L'' induces on this space is degenerate—the space is not symplectic. The associated gauge orbits have the following structure: $x(t) = (r_0(t), r_1(t))$ and $x'(t) = (r'_0(t), r'_1(t))$ lie in the same gauge orbit if and only if $r_1(t) = r'_1(t)$ for all t (i.e., solutions lie in the same gauge orbit if and only if they agree about the relative distances between the particles—what they say about the motion of the centre of mass is irrelevant).

Hamiltonian Picture. Writing $p_i = \dot{r}_i$, we find that the spaces of initial data for our theories are as follows.

1. For the Newtonian theory, the space of initial data is $T^*\mathbb{R}^2 = \{(r_0, r_1, p_0, p_1) : r_i, p_i \in \mathbb{R}\}$ carrying its canonical symplectic structure $\omega = \sum_{i=0,1} dr_i \wedge dp_i$. The Hamiltonian is $H(r_0, r_1, p_0, p_1) = \frac{1}{2}(p_0^2 + p_1^2) - \frac{1}{2r_1}$. The equations of motion are the usual deterministic Newtonian equations.
2. For the Leibnizean theory, the space of initial data is $T^*\mathbb{R} = \{(r_1, p_1) : r_1, p_1 \in \mathbb{R}\}$ carrying its canonical symplectic structure $\omega = dr_1 \wedge dp_1$. The Hamiltonian is $H'(r_1, p_1) := \frac{1}{2}p_1^2 - \frac{1}{2r_1}$. The equations of motion are the usual deterministic Leibnizean equations.
3. Recall that in constructing the Hamiltonian system corresponding to a given Lagrangian system, we must first construct the momentum variables corresponding to the position variables of the Lagrangian system. The semi-Leibnizean theory has two position variables, r_0 and r_1. Our recipe tells us that the corresponding momentum variables are $p_i := \frac{\partial L''}{\partial \dot{r}_i}$, for L'' the semi-Leibnizean Lagrangian. As usual, $p_1 := \dot{r}_1$. But because L'' is independent of $\dot{r}_0$, we find that $p_0 \equiv 0$. It follows that the space of initial data for this theory is the space $\Gamma = \{(r_0, r_1, p_1) : r_0, r_1, p_1 \in \mathbb{R}\}$ that arises when we restrict attention to those states in the space of initial data for the Newtonian theory in which $p_0 = 0$; restricting the symplectic structure of the

Newtonian theory to Γ yields a presymplectic structure (the vectors pointing in the r_0 direction are the null vectors). The gauge orbits have the following structure: $x = (r_0, r_1, p_1)$ and $x' = (r'_0, r'_1, p'_1)$ lie in the same gauge orbit if and only if $r_1 = r'_1$ and $p_1 = p'_1$. The Hamiltonian for this theory is $H''(r_0, r_1, p_1) := \frac{1}{2}p_1^2 - \frac{1}{2r_1}$, which determines the usual Newtonian/Leibnizean behaviour for r_1 and p_1 while leaving the evolution of r_0 wholly unconstrained. That is, if $x(t)$ and $x'(t)$ are curves in the space of initial data corresponding to solutions of this Hamiltonian problem, then one finds that in general $x(t) \neq x'(t)$ for $t \neq 0$, but $[x(t)] = [x'(t)]$ for all t. Note that each such curve $x(t)$ corresponds to a point in the space of solutions, and that the condition $[x(t)] = [x'(t)]$ for all t just says that for the points in the space of solutions that correspond to the curves x and x', themselves lie in the same gauge orbit.

Reduction. As one would expect, the reduced space of initial data of the semi-Leibnizean theory is isomorphic to the space of initial data of the Leibnizean theory, and the reduced space of solutions of semi-Leibnizean theory is isomorphic to the space of solutions of the Leibnizean theory—in both cases, this is because identifying points in the relevant gauge orbits amounts to dropping r_0 as a dynamical variable. So reduction implements our physical intuition that r_0 is an extraneous variable that ought to be excised and eliminates the pathologies of the semi-Leibnizean theory. Furthermore, the reduced space of initial data (reduced space of solutions) inherits from the original theory a Hamiltonian (Lagrangian) that is really that of the Leibnizean theory—so these reduced spaces carry dynamical theories with the correct dynamics and symmetry groups.

Of course, this is a toy example—one of the simplest possible. And it has been set up here so that it is clear from the beginning that the variables of the semi-Leibnizean theory can be segregated into the physically relevant r_1, which plays a role in the Lagrangian and whose dynamics is deterministic, and the physically otiose r_0, which plays no role in the Lagrangian, and whose evolution is completely unconstrained by the dynamics. So it has been clear from the beginning that r_0 ought to be excised from the theory—there has been no temptation to keep it on board and to conclude that we have an indeterministic theory on our hands.

But note that if we had stuck with our original Newtonian variables, q_1 and q_2 (with $q_1 < q_2$), and had written $L'' := \frac{1}{2}(\dot{q}_2 - \dot{q}_1)^2 - \frac{1}{2(q_2 - q_1)}$ then things would not have been quite so clear: the equations of evolution for q_1 and q_2 would have mixed together physically relevant information and physically otiose information and it would have taken a little bit of work to see what was going on.

When we are faced with Lagrangian theories admitting groups of local symmetries, we know (unless they exhibit the sort of pathological behaviour we saw in example 36 above) that there is some way of separating the variables into the physically relevant and the physically otiose (it is easiest to see this on the Hamiltonian side). But it is not always easy to find such a separation. This is one of several reasons why we end up working with such theories rather than with the more attractive reduced theories that stand behind them.

EXAMPLE 38 (Maxwell's Theory). We consider the electromagnetic field. Let V be Minkowski spacetime, and fix an inertial frame and an associated set of coordinates (x_0, x_1, x_2, x_3). We choose as the target space for our field $W = \mathbb{R}^4$. So the kinematically possible fields are of the form $A : V \to \mathbb{R}^4$ (subject to some unspecified differentiability and boundary conditions). A is the usual four-potential.

We define $F_{\upsilon\nu} := \frac{\partial A_\upsilon}{\partial x_\mu} - \frac{\partial A_\mu}{\partial x_\upsilon}$ $(\upsilon, \nu = 0, \ldots, 3)$. So a kinematically possible field $A(x)$ determines a matrix-valued function, F. We label the component functions making up F according to the following scheme, thus identifying components of the F with components of the electric and magnetic fields, $\mathbf{E}(x) = (E^1(x), E^2(x), E^3(x))$ and $\mathbf{B}(x) = (B^1(x), B^2(x), B^3(x))$:

$$F_{\mu\nu}(x) = \begin{vmatrix} 0 & -E^1(x) & -E^2(x) & -E^3(x) \\ E^1(x) & 0 & B^3(x) & -B^2(x) \\ E^2(x) & -B^3(x) & 0 & B^1(x) \\ E^3(x) & B^2(x) & -B^1(x) & 0 \end{vmatrix}$$

We take as the Lagrangian for our theory $L := -\frac{1}{2}\left(|\mathbf{B}(x)|^2 + |\mathbf{E}(x)|^2\right)$. Writing $A(x) = (A_0(x), A_1(x), A_2(x), A_3(x))$ and $\mathbf{A}(x) := (A_1(x), A_2(x), A_3(x))$, we find that the equations of motion for our Lagrangian are:

$$\nabla^2 A_0 + \frac{\partial}{\partial x_0}(\nabla \cdot \mathbf{A}) = 0$$
$$\nabla^2 \mathbf{A} - \frac{\partial^2 \mathbf{A}}{\partial x_0^2} = 0$$

(here $\nabla := (\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3})$ is the ordinary three-dimensional gradient operator). These equations are equivalent to the usual vacuum Maxwell equations for the electric and magnetic fields: $\dot{\mathbf{B}} = -\nabla \times \mathbf{E}$, $\nabla \cdot \mathbf{B} = 0$, $\dot{\mathbf{E}} = \nabla \times \mathbf{B}$, and $\nabla \cdot \mathbf{E} = 0$.

Let $\Lambda : V \to \mathbb{R}$ be a continuous function (appropriately differentiable and satisfying appropriate boundary conditions). Then the map $\Phi_\Lambda : A \mapsto A' := A + d\Lambda$ is a map from the space of kinematically possible fields to itself. If one calculates the matrices F' and F corresponding to A and A', one finds $F' = F$. So $\mathbf{E}$ and $\mathbf{B}$ are invariant under our gauge transformation $A \mapsto A'$. It follows that $L(\Phi_\Lambda(A)) - L(A) = 0$, so Φ_Λ is a Lagrangian symmetry—in particular, A' is a solution if and only if A is. Since Λ was arbitrary, and since Λ and Λ' lead to distinct symmetries so long as $d\Lambda \neq d\Lambda'$, we have in fact found a huge family of symmetries of our theory. Indeed, the Φ_Λ form a group of gauge symmetries of our theory in the official sense introduced above.

Of course, it follows that the initial value problem for A is ill-posed: let $A(x)$ be a solution for initial data posed on the instant $x_0 = 0$ and let Λ be a nonconstant function that vanishes on a neighbourhood of the hypersurface $x_0 = 0$; then A and $A' = A + d\Lambda$ are solutions that agree on $x_0 = 0$ but do not agree globally.

And, of course, the form that our Lagrangian induces on the space of solutions is degenerate. The corresponding gauge orbits have the following form: solutions A and A' belong to the same gauge orbit if and only if there is a $\Lambda : V \to \mathbb{R}$ such that $A' = A + d\Lambda$. An equivalent condition is that A and A' lie in the same gauge orbit if and only if they lead to the same $\mathbf{E}$ and $\mathbf{B}$ —which is just to say that the reduced space of solutions is the space of solutions to the field equations for $\mathbf{E}$ and $\mathbf{B}$ (remember, we are working in a fixed coordinate system, so these are well-defined). This reduced space is a symplectic manifold.

We can construct the Hamiltonian theory corresponding to our Lagrangian theory (our chosen inertial coordinates give us a slicing). For convenience, we take the configuration variables for our Lagrangian theory to be $A_0(x)$ and $\mathbf{A}(x)$. Let $\mathcal{Q}$ be the space of possible $(A_0, \mathbf{A})$ and $T^*\mathcal{Q}$ be the corresponding cotangent bundle, carrying its canonical symplectic structure. A point in $T^*\mathcal{Q}$ consists of a quadruple $(A_0(x), \mathbf{A}(x), \pi_0(x), \boldsymbol{\pi}(x))$ of fields on spacetime, with A_0 and π_0 taking values in $\mathbb{R}$ and $\mathbf{A}$ and $\boldsymbol{\pi}$ taking values in $\mathbb{R}^3$. Our usual procedure tells us that the momentum π_0 corresponding to A_0 is identically zero (our Lagrangian does not depend on $\dot{A}_0$); the momentum $\boldsymbol{\pi}$ corresponding to $\mathbf{A}$ is $\boldsymbol{\pi}(x) = -\mathbf{E}(x)$. So the space Γ of initial data for our theory is the subspace $T^*\mathcal{Q}$ of points of the form $(A_0(x), \mathbf{A}(x), 0, \boldsymbol{\pi}(x))$—so we can take points in Γ to be triples of the form $(A_0, \mathbf{A}, \boldsymbol{\pi})$. The presymplectic form that Γ inherits from its embedding in $T^*\mathcal{Q}$ yields gauge orbits of the following form: $(A_0, \mathbf{A}, \boldsymbol{\pi})$ and $(A_0', \mathbf{A}', \boldsymbol{\pi}')$ belong to the same gauge orbit if and only if $\boldsymbol{\pi} = \boldsymbol{\pi}'$ and $\nabla \times \mathbf{A} = \nabla \times \mathbf{A}'$. Since $\boldsymbol{\pi} = -\mathbf{E}$ and $\mathbf{B} = \nabla \times \mathbf{A}$, this tells us that two points in the space of initial data lie in the same gauge orbit if and only if they correspond to the same electric and magnetic fields. If follows that the reduced space of initial data is just the space of instantaneous states of the electric and magnetic fields. We again find that the space of reduced space of initial data is symplectically isomorphic to the reduced space of solutions.

In present case, as in the semi-Leibnizean example above, we can view the given Lagrangian theory as containing surplus unphysical variables, whose evolution is undetermined by the dynamics, alongside physically sensible variables whose evolution is fully determined by the dynamics. In the present case, however, it is a bit harder to make this division explicit: clearly the good variables are the electric and magnetic fields and the bad ones are those that encode additional information in A—all we care about is which gauge orbit A lies in, so a specification of A gives us surplus information. Reduction allows us to avoid ever mentioning this sort of surplus information.

We can formulate a Hamiltonian version of Maxwell's theory in the setting of the reduced space of initial data: points in this space specify the values of the electric and magnetic fields at points of space at a given time; this space is symplectic; and it is possible to find a Hamiltonian on this space that drives the dynamical evolution encoded in Maxwell's equations for $\mathbf{E}$ and $\mathbf{B}$.[120]

It is natural to wonder whether the reduced space of solutions also supports a Lagrangian version of Maxwell's theory. That is, is there a Lagrangian in terms of $\mathbf{E}$, $\mathbf{B}$, and their derivatives whose variational problem has as its Euler–Lagrange equations the Maxwell's equations for $\mathbf{E}$ and $\mathbf{B}$?

At first sight, it might seem that we could just employ our original Lagrangian,

$$L := -\frac{1}{2}\left(|\mathbf{B}|^2 + |\mathbf{E}|^2\right),$$

for this purpose. But this leads to the wrong equations of motion. And there is reason to worry that $\mathbf{E}$ and $\mathbf{B}$ are ill-suited to the variational approach, since their six components are not independent—they can be derived from the three-component vector potential—and so cannot be varied independently.[121] So it is not obvious that the reduced space of solutions does support a Lagrangian version of Maxwell's theory.

Whether or not this problem is insuperable in the case of Maxwell's theory in Minkowski spacetime, other problems lie ahead. Suppose that we construct our spacetime, V,by rolling up one of the spatial dimensions of Minkowski spacetime: V is locally Minkowskian but has the global structure of $\mathbb{R}^3 \times S^1$. This makes a surprising difference to our theory. It is still true that the gauge orbits in the space of solutions are of the following form: $[A] := \{A + d\Lambda\}$ for all appropriate Λ. And it is still true that specifying a gauge orbit $[A]$ determines the behaviour of the electric and magnetic fields on spacetime. But it is no longer quite true that we can go in the other direction: in order to specify a gauge orbit $[A]$, one has to specify in addition to $\mathbf{E}$ and $\mathbf{B}$ also a single complex number, which we will call the *holonomy*. Intuitively, the holonomy measures the phase change that results when an electron is transported along a given loop that wraps once around the closed dimension of space. Thus a point in the reduced space of solutions can be viewed as consisting of a specification of $\mathbf{E}$ and $\mathbf{B}$ plus the holonomy. This extra number ruins everything: for while $\mathbf{E}$ and $\mathbf{B}$ are appropriately local objects, assigning a property to each point of spacetime, the holonomy is a nonlocal item. This becomes even more clear if we look for a way of describing the reduced space of solutions that does not have the strange feature of including two very different sorts of variable: the best way to proceed appears to be to describe a point in the reduced space of solutions as an certain sort of (highly constrained) assignment of a complex number to each closed curve in spacetime. So in such a topologically nontrivial spacetime, in order to specify a gauge orbit $[A]$ we need to specify nonlocal information. The present framework requires that a Lagrangian field theory involve an assignment of a property of each point of spacetime, and so

[120] See, e.g., [Marsden and Weinstein, 1982].

[121] See [Goldstein, 1953, p. 366] for this point. See [Sudbery, 1986] for a way around this worry—which, however, requires a slight generalization of the present notion of a Lagrangian theory.

cannot accommodate this example.[122]

REMARK 39 (Lagrangians and the Reduced Space of Solutions). In the very simple particle theory considered in example 37 we saw a case in which the reduced space of solutions of a theory admitting gauge symmetries inherited from the original theory a Lagrangian that encoded the gauge-invariant aspects of the original dynamics. But in the more interesting case of Maxwell's theory, considered in example 38, it seems less likely that there is any sense in which the reduced space of solutions arises directly from a local Lagrangian, without passing through a formulation admitting gauge symmetries. And this seems very unlikely indeed if we choose our spacetime to be topologically nontrivial, because in this case the Maxwell field appears to involve a non-local degrees of freedom.

Note that things become even worse in non-Abelian Yang–Mills theories. In these theories, the space of fields is the space of connection one-forms on a suitable principal bundle $P \to V$ over spacetime, the Lagrangian is a direct generalization of the Lagrangian of Maxwell's theory, and the group of gauge symmetries is the group of vertical automorphisms of P. The reduced space of solutions is the space of connections modulo vertical automorphisms of P. Even when V is Minkowski spacetime, the best parameterization of the reduced space of solutions would appear to be one that deals with holonomies around closed curves in spacetime.[123] So it would again appear difficult (perhaps impossible) to capture this reduced space of solutions via the variational problem of a local Lagrangian.[124] Indeed, it seems plausible the prevalence of gauge freedom in physical theories is grounded in the fact that by including nonphysical variables one is sometimes able to cast an intrinsically nonlocal theory in to a local form.[125]

6.3 *Time-Dependent Systems*

Let us assume that our spacetime, V, admits a slicing, and that our equations of motion, Δ, are second-order and exhibit good existence and uniqueness properties.[126] But we now assume that our Lagrangian L is *time-dependent*, in the sense

[122] That is, we seem to be talking about properties that require something bigger than a point to be instantiated, in violation of Humean supervenience (see fn. 45 above).

[123] There is, however, considerable controversy among philosophers regarding the best interpretation of classical non-Abelian Yang–Mills theories. See [Healey, Unpublished], [Maudlin, Unpublished], and [Belot, 2003, §12].

[124] Under a usage distinct from the present one, any Hamiltonian theory on a velocity phase space (i.e., a tangent bundle) counts as a Lagrangian theory; see, e.g., [Abraham *et al.*, 1988, Chapter 8]. Under this alternative use, Lagrangians are not required to be local and a variational principle plays no necessary role. It may well be that there are treatments of theories that are Lagrangian in this sense, but not in the sense that I am concerned with here.

[125] On this point, see, e.g., [Belot, 2003, §13]. For further speculation about the importance of gauge freedom, see [Redhead, 2003].

[126] Recall from section 5.2 above that a slicing of a spacetime is a decomposition into space and time; not every slicing satisfies the stronger condition that this decomposition meshes with a time translation group on V. Only spacetimes with geometries strong enough to determine a family of instants and a family of possible point-particle worldlines admit slicings.

that it does not admit a dynamical time translation group, τ, arising from a time translation group, $\bar{\tau}$, on V.

The time-dependent Lagrangian theories that arise in physics fall under the two following cases.

Case (A): V admits a time translation group $\bar{\tau}$, but this group does not correspond to a symmetry of the equations of motion. Example: A system of particles in Newtonian spacetime, subject to forces arising from a time-dependent potential.

Case (B): V does not admit a time translation group. For example, let (V, g) be a curved general relativistic spacetime without temporal symmetries and take the Klein–Gordon equation for a scalar field on (V, g), $\nabla^a \nabla_a \Phi - m^2 \Phi = 0$, as the equation of motion (note that the metric on V plays a role in defining the derivative operators); the corresponding Lagrangian is $L = \frac{1}{2}\sqrt{-g}\nabla_a \nabla^a \Phi + m^2 \Phi^2$.

We will also need that notion of a time-dependent Hamiltonian system.

DEFINITION 40 (Time-Dependent Hamiltonian Systems). A *time-dependent Hamiltonian system* (M, ω, h) consists of a symplectic manifold (M, ω), called the *phase space*, together with a smooth function $h : \mathbb{R} \times M \to \mathbb{R}$, called the *Hamiltonian*. We often write $h(t)$ for $h(t, \cdot) : M \to \mathbb{R}$.

Ordinary Hamiltonian systems (see definition 8 above) are special cases of time-dependent Hamiltonian systems in which $h(t)$ is the same function on M for each value of t; we will also call such systems *time-independent Hamiltonian systems.* In a time-independent system, the dynamical trajectories could be thought of as curves in in the phase space, parameterized up to a choice of origin, with exactly one such curve passing through each point of the space. In the time-dependent case, the situation is more complicated. For each value of t, we can solve $\omega(X_{h(t)}, \cdot) = dh(t)$ for the vector field $X_{h(t)}$ generated by $h(t)$. We can then declare that a curve $\gamma : \mathbb{R} \to \mathcal{I}$ is a dynamical trajectory of $(\mathcal{I}, \omega, h)$ if for each $t \in \mathbb{R}$, $\dot{\gamma}(t) = X_{h(t)}(\gamma(t))$ —that is, for each t, the tangent vector to γ at $x = \gamma(t)$ is given by the value of the vector field $X_{h(t)}$ at x. Notice that while in the case of a time-independent Hamiltonian system, there is a single dynamical trajectory through each point of the phase space, in the present case there will in general be many such trajectories through each point (since which states come immediately after $x \in \mathcal{I}$ depends on the tangent to the dynamical trajectory through x; and in the time-dependent case, this tangent will vary as we consider posing initial data x at different possible instants).

Given the set of assumptions that we have in play, we expect to to find the following when we investigate a time-dependent Lagrangian theory.

Lagrangian Picture. One can apply the usual variational procedure to pass from a Lagrangian to a set of equations of motion. We can also follow the usual procedure in order to equip the corresponding space of solutions, $\mathcal{S}$, with a two-form, Ω—and, as usual, one presumes that for the sort of examples that arise in physics, uniqueness of solutions to the equations of motion implies that Ω is symplectic.[127] Note, however, that in time-dependent theories of the types under consideration, $\mathcal{S}$ does *not* carry a one-parameter group implementing time translations: in theories falling under Case (A) above, such a group acts on the space of kinematically possible fields, but (in general) maps solutions to non-solutions; in theories falling under Case (B), there is no available notion of time translation. We can as usual use the stress-energy tensor of the field to define the energy of the field along any given instant—but the result is no longer independent of the instant chosen.

Hamiltonian Picture. A choice of slicing for our spacetime V leads to a Hamiltonian picture which is in many ways similar to that which emerges in the time-independent case. Let S be a manifold homeomorphic to an arbitrary instant $\Sigma \subset V$ (and with the geometry, if any, shared by such instants) and let σ be a slicing of V employing S as an abstract instant: it is helpful to think of the choice of σ as the choice of a preferred family of observers equipped with a notion of simultaneity. Then we can set about constructing a Hamiltonian version of our theory, following in so far as possible the recipe from the time-independent case.

1. Given an instant Σ in our slicing and a solution Φ, we define: ϕ, the restriction of the field to Σ; $\dot{\phi}$ the time rate of change of Φ along Σ relative to the observers and clocks that define σ; and $\pi := \frac{\partial L}{\partial \dot{\phi}}$, the field momentum along Σ relative to the slicing σ.
2. Given a solution Φ and instant Σ in our slicing, we use σ to pull back to S the initial data (ϕ, π) induced by Φ on Σ, and henceforth think of ϕ and π as functions defined on S, when convenient.
3. Let $\mathcal{Q}$ be the space of all $\phi : S \to W$ that arise in this way; then $T^*\mathcal{Q}$ is the space of all pairs (ϕ, π) that arise in this way. This is our space of initial data, $\mathcal{I}$. It carries a canonical symplectic form, ω (see example 7 above).
4. The construction of the Hamiltonian is the first stage at which we run into any novelty.[128] Let Σ_t be an instant in our slicing, and define $h(t) : \mathcal{I} \to \mathbb{R}$ (t fixed, for now) as $h(t)(\phi, \pi) := \int_{\Sigma_t} \pi\dot{\phi} - L(\Phi) dx$, where Φ is the solution that induces (ϕ, π) on Σ_t, and $\dot{\phi}$ is the field velocity that Φ induces on Σ_t. In general, this construction yields a different real-valued function on $\mathcal{I}$ for each value of t. One expects that $h(t)(\phi, \pi)$ gives the total instantaneous energy when initial data

[127] For a discussion of the construction of $(\mathcal{S}, \Omega)$ in the time-dependent case, see [Woodhouse, 1991, §2.4].

[128] For this construction see, e.g., [Kay, 1980, §1].

(ϕ, π) are posed on instant Σ_t. But imposing the same initial data at distinct times in general leads to states with different total energies (since, roughly speaking, we are dealing with systems subject to time-dependent potentials).

5. Now considering t as a variable, we see that we have defined a smooth $h : \mathbb{R} \times \mathcal{I} \to \mathbb{R}$. So $(\mathcal{I}, \omega, h)$ is a time-dependent Hamiltonian system in the sense of definition 40 above. The resulting dynamics can be thought of as follows. Suppose that we are interested in the dynamics that results when we pose our initial data on a fixed instant Σ_{t_0} in our slicing. Then, for each $s \in \mathbb{R}$ we can ask what state $x \in \mathcal{I}$, posed on Σ_{t_0}, evolves into after s units of time; we call the result $g_s^{t_0}(x)$. This gives us a map $g_s^{t_0} : \mathcal{I} \to \mathcal{I}$ for each s; and the set $\{g_s^{t_0}\}_{s \in \mathbb{R}}$ forms a one-parameter group; each of $g_s^{t_0}$ is a symplectic automorphism of $\mathcal{I}$ but does not leave h invariant. So here we have the dynamics implemented by symmetries of $(\mathcal{I}, \omega)$ that are not symmetries of $(\mathcal{I}, \omega, h)$. Letting t_0 vary gives us a one-parameter family of such one-parameter dynamics-implementing groups.

Relation between the Pictures. As in the time-independent setting, for each Σ in our slicing, we can define the map $T_\Sigma : \mathcal{S} \to \mathcal{I}$ that sends a solution to the initial data it induces on Σ. Because we are assuming global existence and uniqueness for solutions to our equations of motion, each such map is a bijection. Furthermore, as in the time-independent case, each such T_Σ is in fact a symplectic isomorphism between $\mathcal{S}$ and $\mathcal{I}$. We can use these maps to show that the time-dependent Hamiltonian system constructed above encodes the correct dynamics for our equations of motion: let Φ be a solution and let x_0 be the initial data induced by Φ on the instant Σ_0, and let $x_0(t)$ be the corresponding dynamical trajectory in the space of initial data; then for each $t \in \mathbb{R}$, $x_0(t)$ is the initial data that Φ induces on Σ_t.

In the time-independent case, we also found that the maps T_Σ intertwined the actions of the group implementing time translation on the space of solutions and time evolution on the space of initial data. In the present case, we have, so far, nothing corresponding to time translation on the space of solutions, while on the space of initial data, we have a whole family of notions of time evolution (indexed by a choice of instant upon which initial data are to be posed). Now note that for each instant Σ_t in our slicing and each $s \in \mathbb{R}$ we can define $\hat{g}_s^t := T_{\Sigma_t}^{-1} \circ g_s^t \circ T_{\Sigma_t}$; the family $\{\hat{g}_s^t\}_{s \in \mathbb{R}}$ is a one-parameter group of symplectic automorphisms of $(\mathcal{S}, \Omega)$ which is *not* a group of variational symmetries of our Lagrangian. The result of applying $\hat{g}_s^t$ to a solution Φ is the solution that would result if the initial data induced by Φ on Σ_t had been posed instead on the instant Σ_{t-s}.[129]

[129] Of course, in the time-independent case, this reduces to time translation of solutions—so we can regard the transformations $\hat{g}_s^t$ as generalizing the ordinary notion of time translation of solutions.

Time. In the present context, time translation may or may not be a symmetry of our spacetime. But even if it is, there is no corresponding symmetry of the dynamics. And so our picture is hobbled—we do not get nice actions of the real numbers on the space of solutions and on the space of initial data that implement time translation and time evolution. On the space of initial data, for each instant at which we might choose to pose initial data, we get a one-parameter group implementing time evolution—but this is not a symmetry of the Hamiltonian. On the space of solutions, we have no natural group action corresponding to time-translation. If we choose a slicing and an instant, then we can get an $\mathbb{R}$-action that gives us information not about time translation of solutions, but about what solution results if we take the initial data that a given solution induces on that given instant and re-pose it on another instant.

Change. Some physical quantities will be represented by functions on the space of initial data: for example, in a theory of two Newtonian particles subject to time-dependent external forces, relative distance between the particles will be encoded in a function on the space of initial data. But some quantities will be represented by one-parameter families of functions on the space of initial data: energy will be an example of such a quantity in any time-dependent system.[130] As we have done above with the Hamiltonian, let us use the symbol $f(t)$ to denote such a one-parameterone-parameter family—we can think of an ordinary function as being a degenerate case, where $f(t)$ is the same function on $\mathcal{I}$ for each $t \in \mathbb{R}$. Let $x(t)$ be a dynamical trajectory in $\mathcal{I}$. Then $x(t)$ represents the quantity modelled by $f(t)$ as changing if and only if $\exists t_1, t_2 \in \mathbb{R}$ such that $f(t_1)(x(t_1)) \neq f(t_2)(x(t_2))$.

On the space of solutions, we expect that, once we have chosen a slicing, each quantity of interest will be represented as usual by a one-parameter family of functions—as usual, we denote such a family of functions on $\mathcal{S}$ by $\{f_t\}$. Suppose that a quantity of interest is represented by $f(t)$ on the space of initial data, and let Σ_{t_0} be an instant in our slicing. Then we define $f_{t_0} := f(t_0) \circ T_{\Sigma_{t_0}} : \mathcal{S} \to \mathbb{R}$. Carrying this out for each $t \in \mathbb{R}$ gives us our desired $\{f_t\}$. As usual, we view a solution $\Phi \in \mathcal{S}$ as representing our quantity as changing if $\exists t_1, t_2 \in \mathbb{R}$ such that $f_{t_1}(\Phi) \neq f_{t_2}(\Phi)$.

REMARK 41 (Artificially Time-Dependent Theories). If we have a time-independent Lagrangian theory but perversely choose a slicing that is not adapted to our notion of time translation, then we the result of following the above procedure would be a time-dependent Hamiltonian system.

[130] It is not hard to find other examples. In a field theory set in a nonstationary spacetime, the abstract instant will not carry a Riemannian metric (since the instants $\Sigma \subset V$ do not share a Riemannian geometry). In this case, we find that an initial data set that represents the field as having two sharp peaks will correspond to instantaneous states in which the peaks are different distances apart, depending on the instant Σ_t in the slicing upon which the initial data are posed. So in this sort of example, even relative distance is represented by a family of functions on the space of initial data.

REMARK 42 (Quantization of Time-Dependent Systems). There is no special difficulty in quantizing a time-dependent Hamiltonian treatment of a system with finitely many degrees of freedom. But it is not in general possible to construct a well-behaved quantum Hamiltonian for a time-dependent field theory.[131] For this reason, the standard construction of free quantum field theories on curved spacetimes take as their starting point the space of solutions rather than the space of initial data.[132]

7 THE PROBLEM OF TIME IN GENERAL RELATIVITY

General relativity differs from the theories considered above in being generally covariant. It is widely accepted that this leads to certain characteristic technical and conceptual problems, grouped together under the rubric *the problem of time.* This section forms an extended commentary on the problem of time in general relativity. The first subsection below is devoted to a discussion of the general covariance of general relativity and some of its direct consequences. The following subsection contains a discussion of the problem of time itself—essentially that change cannot be represented in the theory in the way familiar from the discussion of sections 5 and 6 above. The final subsection discusses a strategy for finding time and change in general relativity (this discussion is intended by way of further clarification of the problem of time, rather than as a suggested resolution).

It is important to emphasize that while the present discussion focuses on general relativity, the problems under discussion arise whenever one has a theory that is generally covariant in an appropriate sense.

7.1 The General Covariance of General Relativity

Let V be a spacetime manifold, with or without geometrical structure. Recall that a C^k ($0 < k \leq \infty$) diffeomorphism $d : V \to V$ is a C^k bijection with C^k inverse.[133] Leaving the degree of differentiability unspecified, we will denote by $\mathcal{D}(V)$ the group of diffeomorphisms from V to itself.[134]

---

[131] See, e.g., [Kay, 1980, §2.1].

[132] See [Wald, 1994, Chapter 4].

[133] A diffeomorphism $d : V \to V$ is called *small* if it is homotopic to the identity, otherwise it is *large*. For ease of exposition, I implicitly restrict attention to small diffeomorphisms below. I will often speak of the pullback of a tensor by a diffeomorphism. The most important case will be the pullback d^*g of a spacetime metric g by a diffeomorphism d. Intuitively (V, d^*g) is the spacetime geometry that results if we lift g off V, then use d to permute the identities of points of V, then lay g back down. (V, g) and (V, d^*g) share a set of spacetime points and have isomorphic geometries; they differ only as to which points in V play which geometric roles—unless d is a symmetry of the metric, in which case they do not differ even about this.

[134] Special care is required in dealing with groups of diffeomorphisms: on the one hand, the group of C^k diffeomorphisms from a compact manifold to itself has a nice differentiable structure—it is a Banach manifold—but is not a Banach Lie group because the operation of group multiplication is not smooth; on the other hand, the the group of smooth diffeomorphisms from a compact manifold to

Roughly speaking, we want to say that a theory is generally covariant when it has $\mathcal{D}(V)$ as a symmetry group.[135] So for each of the several notions of symmetry of a theory, we have a corresponding notion of general covariance. Following [Earman, Unpublished], I will single out the following two as the most important for our purposes:

Weak General Covariance: $\mathcal{D}(V)$ is a group of symmetries of the equations Δ of the theory.

Strong General Covariance: $\mathcal{D}(V)$ is a group of gauge symmetries of the Lagrangian L of the theory.

Of course, Strong General Covariance implies Weak General Covariance (since every gauge symmetry is a variational symmetry, and therefore a symmetry of the equations of motion). But the converse is not true (a theory may be weakly generally covariant even if it does not admit a Lagrangian, and hence is not eligible to be strongly generally covariant).

We have been allowing V to carry a fixed geometrical background, encoded in some tensors that do not vary from solution to solution. We could have allowed V to carry further nongeometric solution-independent structure.[136] On the other hand, in any nontrivial theory the fields governed by the equations of motion will of course vary from solution to solution. So we have have a distinction between theories in which is equipped with nontrivial solution-independent structure and theories in which it is not.[137]

Intuitively, a theory is weakly generally covariant if and only if its solutions carry no solution-independent tensors (or spinors, or . . .)—for it is precisely when we have some fixed background tensors painted on V that the equations of motion can "care" about the distinction between a solution Φ and its pullback $d^*\Phi$ by a diffeomorphism $d : V \to V$.

Of course, general relativity is weakly generally covariant—indeed, in the important vacuum sector of general relativity the spacetime metric is the only basic

itself has a less desirable differentiable structure—it is a mere Fréchet manifold—but it is a Fréchet Lie group; see [Adams *et al.*, 1985] and [Milnor, 1984] for details. The situation is even worse for groups of diffeomorphisms from a noncompact manifold V to itself: it appears that one needs to presuppose some geometrical structure on V in order to give the group a differentiable structure; see [Cantor, 1979] and [Eichhorn, 1993]. See [Isenberg and Marsden, 1982] for tactics for circumventing these difficulties.

[135] See [Norton, 1995] for the tangled history of the notion of general covariance.

[136] For example, in studying in the motion of charged matter in a strong external electromagnetic field, we might employ a theory in which the Maxwell field as well as the spacetime geometry was solution-independent and only the motion of the matter varied from solution to solution.

[137] Note that the distinction made here between solution-independent and solution-dependent structures does not coincide with the Anderson–Friedman distinction between absolute and dynamical objects (see [Friedman, 1983, §II.2]): solution-independent objects are required to be the same from solution to solution while absolute objects are only required to be the same from solution to solution *up to diffeomorphism*.

quantity of the theory, and it is solution-dependent.[138]

The question whether general relativity satisfies Strong General Covariance is a bit more subtle. Intuitively, it ought to: for at the formal level diffeomorphisms of V are variational symmetries of the Lagrangian of the theory, and the group of such diffeomorphisms is parameterized in a suitable sense by the set of vector fields on V. But, as we will soon see, this is a point at which technicalities about boundary conditions cannot be ignored.

But we can block out such technicalities by restricting attention to the subset $\mathcal{D}_c(V) \subset \mathcal{D}(V)$, consisting of compactly supported diffeomorphism from V to itself.[139] $\mathcal{D}_c(V)$ turns out to be a group of gauge symmetries of the Lagrangian of general relativity ($\mathcal{D}_c(V)$ is parameterized by the family of compactly supported vector fields on V). So the counterpart of condition (2) above goes through when $\mathcal{D}(V)$ is replaced by $\mathcal{D}_c(V)$.

In order to say more, and to approach the question of the significance of general covariance for questions about time, we turn below to two special cases: (i) general relativity in the spatially compact domain; (ii) general relativity in the domain in which asymptotic flatness is imposed at spacelike infinity. The first case is central to cosmology: by requiring space to be compact, one eliminates worries about boundary conditions at spatial infinity; this permits one to investigate universes packed with matter while maintaining control over technical issues. The second case is of more strictly mathematical and conceptual interest (the asymptotic boundary conditions of greatest physical interest impose asymptotic flatness at null infinity rather than spatial infinity; these allow one to investigate gravitational radiation). After discussing these cases, I briefly turn to the question whether every theory can be given a generally covariant formulation.

General Relativity as a Cosmological Theory

We restrict attention to vacuum general relativity in which the spacetime metric, g, is the only field. So we take as our space of kinematically possible fields the space of Lorentz signature metrics on some fixed n-dimensional spacetime manifold V.[140] The equation of motion for this theory is $R_{ab} - \frac{1}{2}Rg_{ab} = 0$, where R_{ab} is the Ricci curvature of g and R is the scalar curvature of g; here and throughout we require the cosmological constant to vanish.

Recall that a subset $\Sigma \subset V$ is called a *Cauchy surface* of (V, g) if every inextendible timelike curve in (V, g) intersects Σ exactly once; it follows that a Cauchy surface is an $(n-1)$-dimensional spacelike submanifold on V. We call (V, g) ***globally hyperbolic*** if it possesses a Cauchy surface. If (V, g) is globally hyperbolic,

[138] In this regime, the Einstein Field equations just tell us that if metric g on V counts as a solution if and only if the Ricci curvature tensor of g vanishes. And clearly g is Ricci-flat if and only if d^*g is. So $\mathcal{D}(V)$ maps solutions to solutions.

[139] That is, a diffeomorphism $d : V \to V$ is in $\mathcal{D}_c(V)$ if and only if there exists a compact set $U \subset V$ such that d acts as the identity on V/U.

[140] So a kinematically possible field is a section of the bundle of symmetric bilinear forms of Lorentz signature over V.

then it can be foliated by Cauchy surfaces, and all of its Cauchy surfaces are homeomorphic to one another. Indeed, if (V, g) is globally hyperbolic, then V is homeomorphic to a manifold of the form $S \times \mathbb{R}$ for some $(n-1)$-dimensional manifold S, with all the Cauchy surfaces of (V, g) homeomorphic to S. For the purposes of this discussion of general relativity as a cosmological theory we restrict attention to solutions with compact and orientable Cauchy surfaces.[141]

We can proceed to construct Lagrangian and Hamiltonian versions of our theory.

Lagrangian Picture. The Lagrangian for general relativity is the *Einstein-Hilbert Lagrangian*, $L = \sqrt{-g}R$. The space of solutions is, of course, infinite-dimensional. Let us call a solution *well-behaved* if it admits a foliation by Cauchy surfaces with constant mean curvature.[142] It is believed that the set of well-behaved solutions forms a large open subset of the full space of solutions; and it is known that within the space of well-behaved solutions the only singularities that occur are mild ones at metrics that admit Killing fields (these are vector fields that can be thought of as the infinitesimal generators of spacetime symmetries).[143] The group $\mathcal{D}(V)$ is a group of gauge symmetries of the Einstein-Hilbert Lagrangian: each one-parameter group of diffeomorphisms from V to itself is a group of variational symmetries of this Lagrangian, and the group $\mathcal{D}(V)$ can be parameterized by arbitrary vector fields on V.[144] So, in accord with the theory of gauge theories developed in section 6.2 above, we find that the space, $\mathcal{S}$, of well-behaved solutions carries a presymplectic form, Ω (henceforth I drop the qualifier and speak of $\mathcal{S}$ as the space of solutions).[145] As usual, this presymplectic form induces a partition of the space of solutions by gauge orbits. Two metrics, g and g', belong to the same gauge orbit if and only if there exists a diffeomorphism $d : V \to V$ such that $g' = d^*g$. Of course, the conserved quantities associated with one-parameter groups of diffeomorphism are trivial—each is the zero function on $\mathcal{S}$. Indeed, in this context, general relativity has no non-trivial Noether quantities—beyond diffeomorphisms, the only continuous, local symmetries of the laws are metric rescalings, which are not variational symmetries.[146]

[141] The restriction to globally hyperbolic solutions is not required for construction of a Lagrangian version of general relativity, but is required for the Hamiltonian treatment and plays a role in some of the results cited below concerning the structure of the space of solutions. The requirement that the spatial topology be orientable is required for the Hamiltonian treatment.

[142] Mean curvature will be defined below on p. 69, in the course of the discussion of the space of initial data.

[143] For the structure of the space of well-behaved solutions, see [Isenberg and Marsden, 1982].

[144] See [Crnković and Witten, 1987] and [Woodhouse, 1991, pp. 143–146]; the latter provides an argument that non-compactly supported diffeomorphisms belong in the group of gauge symmetries of the Lagrangian.

[145] See also [Frauendiener and Sparling, 1992] for a construction of the presymplectic form on the space of solutions which does not proceed via the Lagrangian formalism.

[146] See [Torre and Anderson, 1996, esp. p. 489].

Reduced Space of Solutions. The space, $\mathcal{S}'$, of gauge orbits of the space of solutions of general relativity is a symplectic space with mild singularities at points corresponding to solutions with Killing fields.[147] Let us call a point $[g]$ in the reduced space of solutions a *geometry*—since distinct representatives of $[g]$ represent V as having the same spacetime geometry, but differ as to the distribution of geometrical roles to points of V. So far as I know, it makes no sense to speak of this reduced space as the space of solutions as arising from the variational problem for a local Lagrangian. Indeed, a geometry $[g]$ would not appear to assign any particular local property to any point $x \in V$.

Hamiltonian Picture. The construction of the corresponding Hamiltonian picture requires a bit of care.[148] We want to mimic as much of the procedure of section 5 above as we can, given that we do not have available a slicing (which requires that spacetime have a nontrivial solution-independent geometry). We proceed as follows.[149]

1. *Construct the space of initial data*. Up until now, we have been able to proceed as follows: (i) choose a slicing σ of V and an instant $\Sigma \subset V$ in σ, then construct the space of possible instantaneous field configurations, $\mathcal{Q}$, by looking at all the $\phi : \Sigma \to W$ that arise by restricting solutions Φ to Σ; (ii) construct the space of initial data $\mathcal{I} \subseteq T^*\mathcal{Q}$ by finding all pairs (ϕ, π) that are induced as initial data on Σ (where π is the instantaneous field momentum, defined via $\pi := \frac{\partial L}{\partial \dot{\phi}}$, with $\dot{\phi}$ is the time rate of change of the field according to the observers associated with the slicing $\sigma)$. We found that $\mathcal{I}$ was a proper subset of $T^*\mathcal{Q}$ whenever the Lagrangian L of the theory admitted a group of gauge symmetries.

   Without relying on a notion of slicing, we can construct a space of initial data via a procedure surprisingly close to the usual one.

   If $\Sigma \subset V$ is an instant and g is a solution to the Einstein field equations, then $q := g\mid_\Sigma$ is a symmetric covariant tensor of rank two. But in the present setting, the restriction of a solution to an arbitrary instant is not a good candidate for an instantaneous configuration of the field: intuitively, since the gravitational field of general relativity is a spacetime geometry, an instantaneous configuration of this field should be a spatial geometry. But, of course, $q := g\mid_\Sigma$ is a Riemannian metric on

[147] See [Isenberg and Marsden, 1982].

[148] For the constructions that follow see [Wald, 1984, Appendix E.1] or [Beig, 1994].

[149] The construction sketched here does not rely on the lapse and shift fields. Fixing the behaviour of these nonphysical fields allows one to pass from initial data on an abstract instant S to a solution on $S \times I$ for some (possibly small) interval I of real numbers. As such they allow one to establish a bijection between the space of initial data and a set of solutions of limited temporal extent. I avoid the lapse and shift here because I want to concentrate on global results and on physical fields. For a very helpful introduction to the lapse and shift formalism, see [Marsden *et al.*, 1972, §III].

Σ if and only if Σ is spacelike according to g. So $q = g \mid_\Sigma$ represents an instantaneous state of the field if and only if Σ is spacelike.[150] So it seems reasonable to take as the space of possible instantaneous field configurations, $\mathcal{Q}$, the space of Riemannian metrics q that arise by restricting each solution to the hypersurfaces that it renders spacelike.[151]

The definition of instantaneous field momenta is more complicated. In the familiar case, the slicing σ plays an important role. But nothing like that is available in the present case: it is awkward to introduce a solution-independent notion of slicing in the context of general relativity, considered as a dynamical theory.[152] There is, however, a way around this difficulty. Consider a solution g and an instant $\Sigma \subset V$ that g represents as being spacelike. Relative to g we can choose a slicing of V in the usual sense (since relative to g we can single out the instants and possible worldlines of point-particles as submanifolds of V). We call such a slicing *Gaussian* for Σ if it corresponds to a set of freely falling observers whose clocks all read zero as they pass through Σ, and whose worldlines are all orthogonal to Σ. For sufficiently small t the hypersurfaces of constant t according to the Gaussian observers will be Cauchy surfaces carrying Riemannian metrics $q(t) := g \mid_{\Sigma_t}$. So given a Gaussian slicing for Σ, we can define $\dot{q}_{ab} := \frac{\partial q_{ab}(t)}{\partial t} \mid_{t=0}$, which is a symmetric covariant tensor of rank two on Σ. In fact, $\dot{q}_{ab}$ is independent of the Gaussian slicing chosen, and can be viewed as telling us about the geometry of the embedding of Σ in (V, g). We can take a similar view of the *extrinsic curvature* of Σ in (V, g), $k_{ab} := \frac{1}{2}$ $\dot{q}_{ab}$, and the *mean curvature* along Σ in (V, g), $k := q^{ab} k_{ab}$. Now: relative to our Gaussian slicing, the tensor $\dot{q}_{ab}(0)$ represents the velocity of the gravitational field, in the sense that it encodes information about the time rate of change of the field; as usual, we can define the corresponding momentum as $\pi^{ab} := \frac{\partial L}{\partial \dot{q}}$. For the Einstein-Hilbert Lagrangian we have $\pi^{ab} = \left(\sqrt{q} k^{ab} - k q^{ab}\right)$ (so the momentum is a symmetric contravariant tensor of rank two).[153] We take as our space of initial data

[150] In the spatially compact globally hyperbolic regime, a submanifold of (V, g) with the topology of a Cauchy surface for (V, g) is a Cauchy surface if and only if it is spacelike according to g; see [Budic *et al.*, 1978, Theorem 1].

[151] As usual, the field q is taken to be defined on an abstract instant, S, diffeomorphic to the concrete instants $\Sigma \subset V$. In order to construct $\mathcal{Q}$ we choose an instant $\Sigma \subset V$ and a diffeomorphism $d : S \to \Sigma$, and use d to pull back to S all of the q that arise as restrictions of Σ of solutions that render Σ spacelike. $\mathcal{Q}$ is of course independent of the choice of Σ and d.

[152] Suppose that σ is a slicing relative to a metric g on V. Then the restriction of g to the instants Σ_t in σ will be reasonable instantaneous field configurations, and so relative to σ the solution g ought to correspond to a curve in the space of initial data of the theory. But what happens if we look at another solution g' relative to σ? In general, the result of restricting this new solution to an instant Σ_t in σ will not be an instantaneous state of the field—so σ will not give us the means to associate with each spacetime solution g a trajectory in the space of initial data.

[153] If we calculate $\dot{q}$ relative to a non-Gaussian slicing of (V, g), then we will in general get an answer

the space $\mathcal{I} \subset T^*\mathcal{Q}$ of pairs (q, π) that arise as the field configuration and momentum induced by solutions on instants they render spacelike (with q and π functions living on an abstract manifold S). As is to be expected in a theory with gauge symmetries, $\mathcal{I}$ is a proper subspace of $T^*\mathcal{Q}$ and the restriction of the canonical symplectic form on $T^*\mathcal{Q}$ equips $\mathcal{I}$ with a presymplectic form, ω. The gauge orbits of ω have the following structure: initial data sets (q, π) and (q', π') belong to the same gauge orbit if and only if they arise as initial data for the same solution g.[154]

2. *Construct a Hamiltonian*. Application of the usual rule for constructing a Hamiltonian given a Lagrangian leads to the Hamiltonian $h \equiv 0$.
3. *Construct dynamics*. Imposing the usual dynamical equation, according to which the dynamical trajectories are generated by the vector field(s) X_h solving $\omega(X_h, \cdot) = dh$, leads to the conclusion that dynamical trajectories are those curves generated by null vector fields. So a curve in $\mathcal{I}$ is a dynamical trajectory if and only if it stays always in the same gauge orbit. This is, of course, physically useless—since normally we expect dynamical trajectories for a theory with gauge symmetries to encode physical information by passing from gauge orbit to gauge orbit. But in the present case, nothing else could have been hoped for. A non-zero Hamiltonian would have led to dynamical trajectories which passed from gauge orbit to gauge orbit—but this would have been physical nonsense (and worse than useless). For such dynamics would have carried us from an initial state that could be thought of as an instantaneous state for solution g to a later instantaneous state that could not be thought of as an instantaneous state for solution g. In doing so, it would have turned out to encode dynamical information very different from that encoded in Einstein's field equations.

Reduced Space of Initial Data. We can pass to, $\mathcal{I}'$, the space of reduced initial data: a point in this space consists of a gauge equivalence class of points in the space of initial data. Like the reduced space of solutions, the reduced space of initial data is a symplectic space with mild singularities.[155] Indeed, it is presumed that the two reduced spaces are canonically isomorphic as symplectic spaces, under the map that takes a gauge orbit of initial data

quite different from that generated by a Gaussian slicing. But if we use this new notion of the field velocity in our definition of the field momentum, we find that our new observers agree with our original Gaussian observers about the value of the field momentum at each point of Σ. So, rather surprisingly, in general relativity the field momentum depends on the instant chosen, but not on a slicing.

[154] More precisely: (q_1, π_1) and (q_2, π_2) belong to the same gauge orbit if and only if there is a solution g, instants $\Sigma_1, \Sigma_2 \subset V$, and diffeomorphisms $d_1 : S \to \Sigma_1$ and $d_2 : S \to \Sigma_2$ such that for $i = 1, 2$ (q_i, π_i) is the pull back to S by d_i of the initial data that g induces on Σ_i. Note that if $\Sigma_1 = \Sigma_2$ but $d_1 \neq d_2$, then (q_1, π_1) and (q_2, π_2) will be distinct but gauge-equivalent descriptions of the geometry of a single Cauchy surface in (V, g).

[155] See [Fischer and Moncrief, 1996].

to corresponding gauge orbit of solutions.[156] $\mathcal{I}'$ inherits from $\mathcal{I}$ the trivial Hamiltonian $h \equiv 0$; this induces the trivial dynamics on $\mathcal{I}'$ according to which the dynamical trajectories are constant curves of the form $x(t) = x_0$ for all values of t.

Relation between the Pictures. The space of solutions and the space of initial data are not isomorphic—this is a general feature of theories with gauge symmetries. On the other hand, as we just noted, the reduced space of solutions and the reduced space of initial data are believed to be isomorphic. In the case of a theory on a fixed background spacetime, a slicing yields a one-parameter family of symplectic isomorphism between the space of solutions and the space of initial data that serves the dual purposes of intertwining the temporal symmetries of their respective spaces and allowing us to construct a representation of change on the space of solutions. In the present case we have only a single canonical isomorphism between the two spaces.

Time. On neither the space of solutions nor reduced space of solutions do we find an action of the real numbers implementing time translation. Nor do we find a non-trivial action implementing time evolution on the reduced space of initial data, since the Hamiltonian trajectories are all trivial there. On the space of initial data, we do have non-trivial Hamiltonian trajectories. But a dynamical trajectory on the space of initial data cannot in general be viewed as encoding time evolution: there is nothing, for instance, to prevent such a trajectory from being periodic, even when the solution corresponding to the gauge orbit the trajectory lives in is not periodic in any sense.

There is, however, a class of dynamical trajectories on the space of initial data that can be viewed as encoding dynamics—those trajectories that correspond to sequences of initial data that could be stacked to form sensible spacetime geometries (when this is possible, the result is always a solution of the field equations). Through each point of the space of initial data there are in fact many such trajectories. But, as is usual in theories with gauge symmetry, there is no privileged way of cutting down this multitude to a distinguished subset that encode time evolution via an $\mathbb{R}$-action.

Change. Let us take some changeable physical quantity like the instantaneous spatial volume of the universe. How would we represent such a quantity on the various spaces in play? On both the space of solutions and the reduced space of solutions, we face our usual problem: points in these spaces represent history timelessly, so no function on such a space can represent in a direct way a changeable physical quantity. In the past, we were able to get around this problem using one of the following strategies. (i) We could find a function on f on a space arising on the Hamiltonian side, then use a slicing-dependent one-parameter family of isomorphisms between this space and

[156] If (q, π) is the geometry of a Cauchy surface in g, then canonical isomorphism between $\mathcal{I}'$ and $\mathcal{S}'$ sends $[q, \pi]$ to $[g]$.

the (reduced) space of solutions to find a one-parameter family of functions on the latter space encoding the behaviour of the given quantity. (ii) Or we could find a function on the (reduced) space of solutions encoding the value of the quantity of interest at a given instant, then use a dynamical time translation group on the (reduced) space of solutions to generate a one-parameter family of such functions. Neither of these strategies will work this time: we do not have a one-parameter family of isomorphisms indexed by instants, nor a notion of time translation on the (reduced) space of solutions.

We are in fact no better off on the reduced space of initial data: there too points correspond to entire histories of the system, and individual functions are ill-suited to represent changeable quantities. And on the space of initial data we face an unattractive dilemma: if we seek to represent changeable quantities via non-gauge invariant functions, then we face indeterminism; if we employ gauge-invariant functions, then we are faced with essentially the same situation we met in the reduced space of initial data.

General Relativity in the Asymptotically Flat Regime

It is illuminating to consider a second sector of general relativity, in which one requires solutions to be asymptotically flat at spatial infinity. This case is of marginal physical interest, but it helps us to clarify the source of the problems we ran into in the spatially compact case.

In this regime our spacetime is $\mathbb{R}^4$ and kinematically possible fields are assignments of Lorentz signature metrics to V that are required to be, in an appropriate sense, asymptotically flat at spatial infinity.[157] Instants are also required to satisfy asymptotic conditions.

In this setting it is natural to consider $\mathcal{D}^\infty(V)$, the group of diffeomorphisms that leave the boundary conditions invariant, rather than the full group of diffeomorphisms. We find that the subgroup, $\mathcal{D}_0^\infty(V)$, of $\mathcal{D}^\infty(V)$ consisting of diffeomorphisms asymptotic to the identity at infinity is the largest group of gauge symmetries of the Lagrangian formulation of the theory and that $\mathcal{D}^\infty(V)$ is the semi-direct product of $\mathcal{D}_0^\infty(V)$ with the Poincaré group (every element of $\mathcal{D}^\infty(V)$ can be thought of as a product of an element of $\mathcal{D}_0^\infty(V)$ and a Poincaré symmetry acting at infinity).[158] The space of solutions of this theory carries a presymplectic form and breaks into gauge orbits, with two solutions in the same gauge orbit if

[157]There are several notions of asymptotic flatness at spatial infinity. In this section, results are cited that are derived using three distinct but closely related approaches: (i) that of [Andersson, 1987]; (ii) that of [Ashtekar *et al.*, 1991]; and that of [Beig and Ó Murchadha, 1987]. For ease of exposition, I gloss over the differences in these approaches in the text—I do not believe that the result is misleading. For the relations between approaches (i) and (iii), see [Andersson, 1987, Definitions 2.3 and 2.4] and [Andersson, 1989, p. 78]. Both of approaches (ii) and (iii) are situated by their protagonists with respect to that of [Beig and Schmidt, 1982]; see [Ashtekar and Romano, 1992, §7] and [Beig and Ó Murchadha, 1987, §§4 and 5].

[158]See [Andersson, 1987, Theorem 2.2] and [Ashtekar *et al.*, 1991, §3.3].

and only if they differ by a diffeomorphism in $\mathcal{D}_0^\infty(V)$.[159] Diffeomorphisms in $\mathcal{D}_0^\infty(V)$ fix the gauge orbits; those in $\mathcal{D}^\infty(V)$ but not $\mathcal{D}_0^\infty(V)$ permute them. The significance of this is most clear at the level of the reduced space of solutions: this is a symplectic space carrying a representation of the Poincaré group—and in particular, for each notion of time translation at spatial infinity, this space carries a non-zero Hamiltonian generating this notion.[160]

One can also give a Hamiltonian treatment of this sector of general relativity.[161] One constructs the space of initial data as in the spatially compact case, except that conditions must be imposed on the asymptotic behaviour of the instantaneous field configuration and momentum. The resulting space carries a presymplectic form. Initial data sets (q, π) and (q', π') belong to the same gauge orbit if and only if there is a solution g and instants $\Sigma, \Sigma' \subset V$ such that Σ and Σ' are related by an element of $\mathcal{D}_0^\infty(V)$ and g induces (q, π) on Σ and (q', π') on Σ'.[162] Just as on the space of solutions, we have a set of functions that can be viewed as the infinitesimal generators of the Poincaré group at infinity. Corresponding to a generator of time translations at infinity is a Hamiltonian on the space of initial data that generates a gauge equivalence class of notions of dynamics, each of which carries one from gauge orbit to gauge orbit (compare with the notion of dynamics on the space of initial data of an ordinary theory with gauge symmetries). So a generic dynamical trajectory, $x(t)$, generated by such a Hamiltonian will represent a nontrivial trajectory through the space of initial data; the same Hamiltonian will generate many trajectories through each point in the space of initial data; but each of these trajectories will agree for each value of t about the gauge orbit in which the state of the system dwells at that time.

One expects that the reduced space of initial data should be a symplectic space isomorphic to the reduced space of solutions and carrying a representation of the Poincaré group. Choosing a notion of time translation at infinity should pick out a Hamiltonian on the reduced space of initial data whose dynamical trajectories encode the dynamics of the theory: fixing a notion of the time translation, the corresponding Hamiltonian, and an arbitrary point in the reduced space of initial data, we should find that the Hamiltonian trajectory through this point encodes a sequence of equivalence class of instantaneous data, and that any way of picking representatives of these classes that stack to form a sensible spacetime geometry encodes a solution of the theory.

So the situation in this case is very different from that we saw above in the spatially compact case. We have representations of the Poincaré group on the reduced space of solutions and on the reduced space of initial data, and we have these representations encoded in structures on the space of solutions and the space

[159] See [Ashtekar *et al.*, 1991, §3].

[160] See[Andersson, 1987].

[161] See [Beig and Ó Murchadha, 1987].

[162] As in the spatially compact case, distinct points in the space of initial data can correspond to the same field configuration and momentum induced by g on $\Sigma \subset V$, if we use distinct diffeomorphisms to pull back these tensors to the abstract instant S.

of initial data.

And we can represent changeable physical quantities in a quite familiar way, via smooth functions on the reduced space of initial data. Special cases aside, such functions change their values as the state moves along the dynamical trajectories in that space. And we can calculate the rate of change of such quantities, etc. The situation is more complicated if we seek to represent change via functions on the space of solutions—this requires some of the apparatus to be developed below in section 7.3. But at least at the intuitive level, it is clear what needs to be done: because for each point in the reduced space of solutions, there is, for each notion of time translation at infinity, a one-parameter family of points in this space that correspond to the time translates of the given point, it ought to be possible to find, for any function on the reduced space of initial data that corresponds to a changeable quantity, a one-parameter family of functions on the reduced space of solutions that encode the value of that quantity at different moments of time.

Is General Covariance Special?

Einstein believed that the general covariance of general relativity was a very special feature with momentous physical consequences. Motivated by the observation that in special relativity there is a tight connection between the fact that the laws assume the same form in every inertial frame and the fact that all inertial observers are equivalent (so that there is no notion of absolute velocity), Einstein hoped that because the laws of his theory of gravity held in arbitrary coordinates the theory would be one in which *all* observers were equivalent (so that there would no notion of absolute motion whatsoever).

But, notoriously, the means were inadequate to the end: in general relativity there is a perfectly cogent (and coordinate-independent) distinction between those observers who are accelerated and those who are unaccelerated, between those who are rotating and those who are not.[163]

Einstein's requirement that the laws of his theory should hold in arbitrary coordinate systems is just the translation into the language of coordinates of our first, weak, sense of general covariance. The preceding paragraph points out that this requirement does not have the powerful consequences that Einstein believed it to. Even worse, it was pointed out already by Kretschmann in 1917 that this weak sense of general covariance is not a very unusual feature: many pre-general relativistic theories can be given a weakly generally covariant formulation.[164] Indeed, there is a recipe that takes as input a Lagrangian field theory on a fixed background spacetime and gives as output a strongly generally covariant reformulation/relative

[163]Einstein's line of thought founders on the following observation: in special relativity Lorentz transformations are symmetries of the spacetime metric that is used to determine the state of motion of an observer, in general relativity an arbitrary diffeomorphism is certainly not a symmetry of the spacetime geometry of a given solution—but this geometry again plays a role in determining the state of motion of an observer. See [Friedman, 1983, Chapters II and V].

[164]On Kretschmann, see [Rynasiewicz, 1999].

of the given theory.[165]

EXAMPLE 43 (Artificial General Covariance). Let T_0 be the theory of a massless Klein–Gordon scalar field, Ψ propagating on a fixed background spacetime, (V_0, g_0). The Lagrangian for T_0 is $L_0(\Psi) := \frac{1}{2} g_0^{ab} \nabla_a \Psi \nabla_b \Psi$ and the corresponding equation of motion is $\Box_0 \Psi = 0$, where $\Box_0$ is the d'Alembertian corresponding to g_0.[166] Given T_0 we can construct a strongly generally covariant theory T as follows.[167] Let V be a manifold diffeomorphic to V_0. The spacetime of T is the bare manifold, V, unequipped with any geometry. T involves two fields, X and Φ : X takes values in V_0 while Φ takes values in $\mathbb{R}$. A pair (X, Φ) counts as kinematically possible only if $X : V \to V_0$ is a diffeomorphism.[168] The Lagrangian L of T is constructed as follows: for any kinematically possible (X, Φ), the n-form $L(X, \Phi)$ on V is the pullback to V by X of the n-form $L_0(\Psi)$ on V_0, where $\Psi := \Phi \circ X^{-1}$. L admits $\mathcal{D}(V)$ as a group of gauge symmetries—so T is strongly generally covariant. Note that a kinematically possible pair (X, Φ) is a solution of T if and only if $\Psi = \Phi \circ X^{-1}$ is a solution of T_0. This is equivalent to saying that a pair (X, Φ) is a solution if and only if Φ is a solution of the massless Klein–Gordon equation $\Box \Phi = 0$, with $\Box$ the d'Alembertian corresponding to the metric $g := X^* g_0$ on V.

This shows that there are relatively ordinary theories, like the theory of Klein–Gordon field, that can be given strongly generally covariant formulations. So even strong general covariance fails to distinguish general relativity from perfectly pedestrian theories.

Nonetheless, it is difficult to shake the feeling that the special nature of general relativity among physical theories has something to do with its general covariance. Indeed, it would appear that at the present time the best that can be said is that what makes general relativity special is that its most natural and perspicuous formulations are generally covariant. But that is just to say that we do not yet understand the matter, I think.

In this connection, it is natural to ask whether the difficulties that we encounter in representing time and change in general relativity arise for the artificially strongly generally covariant theory of example 43.

EXAMPLE 44 (Artificial General Covariance and the Problem of Time). Let us return to the theories T_0 and T of example 43, and let us assume for convenience that the spacetime, (V_0, g_0), of T_0 does not admit any isometries. Suppose that we

---

[165] It is not obvious how one should individuate theories in the present context. For discussion and suggestions, see [Sorkin, 2002, p. 698] and [Earman, Unpublished, §4].

[166] The d'Alembertian corresponding to a Lorentz metric g is defined just as the Laplacian of a Riemannian metric g : as $\mathrm{div}_g \circ \mathrm{grad}_g$ where div_g is the divergence operator of g and grad_g is the gradient operator of g.

[167] See [Lee and Wald, 1990, p. 734] or [Torre, 1992, §II]. The same procedure will work for any scalar field with a first-order Lagrangian that features a non-derivative coupling of the field to the spacetime metric.

[168] Strictly speaking, this takes us outside of our official framework for Lagrangian field theories, since the value that X takes at distinct points of V are not independent of one another.

were simply handed T. Would there be any way to represent changeable quantities via functions on a symplectic space associated with T?

Let $\mathcal{S}$ be the space of solutions of T, and let $\mathcal{S}'$ be the corresponding reduced space (i.e., the space of gauge orbits of $\mathcal{S}$). As one would expect, two solutions (X, Φ) and (X', Φ') lie in the same gauge orbit of $\mathcal{S}$ if and only if there exists $d \in \mathcal{D}(V)$ such that $X' = X \circ d$ and $\Phi' = \Phi \circ d$. The space $\mathcal{S}$ is, of course, presymplectic while the space $\mathcal{S}'$ is symplectic. But since solutions in the same gauge orbit will not agree about the value of Φ or X at any point of V, it is difficult to view a diffeomorphism equivalence class of solutions as assigning properties to points of V, and so it would appear to be impossible to think of $\mathcal{S}'$ as the space of solutions corresponding to some local Lagrangian. By following a procedure like that used in the discussion of general relativity above, we can construct the space of initial data, $\mathcal{I}$, of T, and the corresponding reduced space, $\mathcal{I}'$. The latter will be a symplectic space. But note that the Hamiltonians on $\mathcal{I}$ and $\mathcal{I}'$ vanish. So although we have been able to construct symplectic spaces, we do not have the nontrivial flows associated with time translation or time evolution that we require to set up our representation of change via functions on these spaces. So far, the present case looks very much like the case of spatially compact general relativity.

But now note that from knowledge of T alone we can reconstruct T_0. The field X has as its target space the manifold V_0. We take T_0 to be the theory of a scalar field Ψ on V_0 with Lagrangian L_0 given as follows: let Ψ be a kinematically possible field of T_0 and let $X : V \to V_0$ be an arbitrary diffeomorphism; then we define $L_0(\Psi)$ to be the n-form on V_0 that results when we use X^{-1} to pullback to V_0 the n-form $L(X, \Psi \circ X)$; the result is independent of the X chosen. The resulting equations of motion is $\Box_0 \Psi = 0$. Noting that $\Box_0$ arises as the d'Alembertian of a unique metric g_0 on V_0 and that field propagates causally relative to g_0, it is natural for us to view g_0 as the geometrical structure of V_0, and go on to consider slicings relative to g_0, etc.

With T_0 in hand, we can construct the space of solutions, $\mathcal{S}_0$. Relative to a slicing of (V_0, g_0), we can represent any changeable quantity—e.g., the volume of the support of the scalar field—via functions on $\mathcal{S}_0$ in the usual way.

Finally, note that $\mathcal{S}'$ is canonically symplectically isomorphic to $\mathcal{S}_0$.[169] So we can transfer our representation of change from the latter space to the former. So there is a way to avoid the problem of time in this case.[170]

There is, however, an obvious worry about this approach. Let $\tilde{g}_0$ be a metric on V distinct from g_0. Then $\Box_0$ is not the d'Alembertian of $\tilde{g}_0$; but presumably this operator is still definable in terms of $\tilde{g}_0$. So according to $\tilde{g}_0$ the Euler–Lagrange equations of L_0 on V_0 are not the Klein–Gordon equations, but some less famous

[169]Via the map that sends an equivalence class, $[X, \Phi]$, of solutions T to the solution $\Psi = \Phi \circ X^{-1}$ of T_0. It is at this point that we require the assumption the g_0 does not admit isometries: in general, $\mathcal{S}'$ is isomorphic to the quotient of $\mathcal{S}_0$ by the action of the isometry group of g_0.

[170]Note that we must choose a slicing of (V_0, g_0) in order to get a family of functions on $\mathcal{S}'$ corresponding to a changeable physical quantity. Such functions tell us things like how large the volume of the support of the field is at the instant when the geometry of space assumes a given form.

equations. Now, the above strategy amounts to thinking of T as *really* the theory of a Klein–Gordon field on a spacetime isomorphic to (V_0, g_0). But it was no part of our data that T is a Klein–Gordon theory. So what is to stop us from thinking of T as *really* a theory of a field obeying some less famous equations on a spacetime isomorphic to $(V_0, \tilde{g}_0)$? In this case we would use slicings of $(V_0, \tilde{g}_0)$ to set up our representation of change, etc.

Here are two things one might say in response to this worry. (1) We sought and found a natural way of representing change via functions on $\mathcal{S}'$. It is no problem if there are others. (2) We normally demand that of a physically reasonable theory that its field propagate along the nullcones of the spacetime metric. This will be true of T_0 only for metrics $\tilde{g}_0$ conformally equivalent to g_0.[171] Every slicing of (V_0, g_0) is also a slicing of $(V_0, \tilde{g}_0)$ for each $\tilde{g}_0$ conformally related to g_0 (since conformally related methods agree about which lines are timelike and which hypersurfaces are spacelike). So relative to such a slicing we can consider a quantity that is conformally invariant in the sense that for each $\Sigma \subset V$ in our slicing, this quantity has the same value on Σ in $(V_0, \tilde{g}_0, \Psi)$ for each $\tilde{g}_0$ conformally related to g_0. Such a quantity is represented by the same one-parameter family of functions on the reduced space of solutions of T whether we view T as secretly a theory of a Klein–Gordon field on a spacetime isomorphic to (V_0, g_0) or as secretly a theory of some other sort of field on a spacetime isomorphic to to $(V_0, \tilde{g}_0)$.

7.2 The Problem of Time

In each of the theories considered in sections 5 and 6 above, the dynamical content of the theory was encoded in a flow (possibly time-dependent, possible merely local) on a symplectic space of states within the Lagrangian or Hamiltonian formulation of the theory. That this fails in general relativity, conceived of as a theory of the universe as a whole, is what sets that theory apart. And, of course, this feature means that the standard strategies for representing change also fail for this theory: since one does not have a flow corresponding to time evolution on the reduced space of initial data, no function on that space can represent a changeable physical quantity; it follows that one does not have the apparatus required to represent changeable quantities via functions on the reduced space of solutions either.

This nexus is the problem of time: time is not represented in general relativity by a flow on a symplectic space and change is not represented by functions on a space of instantaneous or global states.[172]

Before proceeding to discuss the significance of this problem it is important to be clear about its nature and sources.

[171] Recall that metrics g_0 and g_1 on V are conformally related if there is a positive scalar $\Omega : V \to \mathbb{R}$ such that $g_1 = \Omega g_0$.

[172] The canonical presentations of the problem of time are [Kuchař, 1992] and [Isham, 1993]. For philosophical discussions, see [Belot and Earman, 2001], [Butterfield and Isham, 2000], and [Earman, 2002]. For critical reactions to this literature, see [Maudlin, 2002] and [Healey, 2002].

- If one approaches the problem of time via a focus on the transition from the space of initial data to the reduced space of initial data, the problem can appear especially urgent. For in passing from the space of initial data to the reduced space of initial data, one identifies initial data sets that correspond to distinct Cauchy surfaces within a single solution. *Prima facie*, this involves treating the current state of the universe and its state just after the Big Bang as the *same state*. Moral: according to general relativity, change is an illusion.

  But this is too hasty. For of course the reduced space of initial data is canonically isomorphic to the reduced space of solutions.[173] And in this latter space, some points represent worlds in which there is change (e.g., worlds which begin with a Big Bang) and some represent changeless worlds (e.g., world modelled by Einstein's static solution). So it is hard to see how general relativity teaches us the moral announced.

  So I would like to disavow formulations of the problem of time that rely on this way of speaking. More constructively, I would like to suggest that it is helpful to concentrate on the reduced space of solutions rather than on the reduced space of initial data in setting up the problem of time. In the well-behaved theories of section 5 the space of initial data and the space of solutions are symplectically isomorphic, but we nonetheless think of these two spaces as having distinct representational functions—roughly and heuristically speaking, one is suited to represent possible instantaneous states while the other is suited to represent possible worlds. This distinction is grounded by the fact that relative to a slicing one finds that for each $t \in \mathbb{R}$, the map T_{Σ_t} that sends a solution to the initial data that it induces on the instant $\Sigma_t \subset V$ defines a distinct isomorphism between the space of solutions and the space of initial data. This makes it natural to think of points of the latter space as representing states (universals) that can occur at distinct times and to think of points in the space of solutions as representing possible worlds composed out of such states. The elements of this story survived more or less unscathed the introduction of various complicating factors in section 6. But in the case of cosmological general relativity we have only a single canonical isomorphism between the reduced space of initial data and the reduced space of solutions. In this context, it is difficult to deny that the reduced space of solutions and the reduced space of initial data are representationally equivalent. And it seems straightforward that we should interpret points in the reduced space of solutions as representing general relativistic worlds rather than instantaneous states—so we should say that same thing about points in the reduced space of initial data. Thus, we should resist any temptation to think of the reduction procedure as telling us to think of an early state of the universe and a late state of the universe as being the same

[173] Under the map that sends $[q, \pi]$ to $[g]$ if (q, g) describes the instantaneous state on some Cauchy surface of (V, g).

instantaneous state.

- Since we have been focussing on the Lagrangian rather then the Hamiltonian picture, but have nonetheless run straight into the problem of time, we can conclude that this problem is not an artifact of the 3+1 decomposition involved in the Hamiltonian approach. Likewise, the problem of time is a feature of general relativity as a cosmological theory, but not of general relativity in the regime of asymptotic flatness at spatial infinity, nor of field theories on fixed relativistic backgrounds, nor, I think, of the artificial strongly generally covariant theory of examples 43 and 44 above. From this we can conclude that the following are *not* sufficient conditions for the problem of time: the lack of a preferred slicing; the jiggleability of admissible slicings; the invariance of the theory under a group of spacetime diffeomorphisms. It appears that the problem arises when we employ a diffeomorphism-invariant theory to model a situation in which we take geometry to be fully dynamical (i.e., we do not smuggle in any background structure, at spatial infinity or elsewhere).

For everything that I have said so far, the Problem of Time may sound like no more than a diverting puzzle. Granted, time does not appear as a symmetry in general relativity as it did in earlier theories (even in the infinitesimal sense involved in a local flow). But, of course, part of the allure of the theory is that it changes the nature of time in a fundamental way. And since successful applications of the theory involve the representation of changeable physical quantities (e.g., the perihelion of Mercury), it would seem that there *must* be some way of way of generalizing the picture of the previous sections to cover general relativity. And while it will be granted that a search for this generalization might turn out to be enlightening, it may well not seem a very pressing project.

This puzzle begins to look far more urgent when we turn our attention to quantization. The good news is that upon reduction, one ends up with a symplectic space representing the true degrees of freedom of general relativity. Without something along these lines, quantization would be impossible. But the vanishing of the Hamiltonian for cosmological general relativity means that two looming difficulties block the road to the successful quantization of general relativity.

1. What is one to do next? Normally a Hamiltonian or a Lagrangian plays a crucial role in quantization. One defines quantum dynamics via these objects. In the case of spatially compact general relativity the reduced space of initial data inherits from the original space of initial data a Hamiltonian—which vanishes, so that the corresponding dynamics is trivial. And it does not appear to make any sense to speak of a local Lagrangian field theory of the true degrees of freedom of the gravitational field. The way forward is unclear.

2. And it is not clear how one would make sense of a quantization of general relativity. While in the classical theory one can find change in solutions even

> without being able to find it at the dynamical level (in terms of quantities on the space of solutions, etc.), it is not obvious how this could be done at the quantum level. Perhaps the best that one can hope for is to be able to speak of approximate time and change in a subset of quantum states that approximate classical solutions. That seems perfectly acceptable—what one should be aiming at, even, in a theory in which the geometry of space and time are themselves quantized. But the usual techniques of semi-classical approximation require a Hamiltonian.[174]

7.3 Finding Time in General Relativity

This final section discusses what is probably the most obvious way around the problem of time. In the cases discussed in sections 5 and 6, we were able to represent change via functions on the (reduced) space of solutions of the theory because we had a slicing, $\sigma : S \times \mathbb{R} \to V$, that decomposed spacetime into space and time, and thereby allowed us to identify functions on the (reduced) space of solutions that corresponded to the values of a given quantity at different instants. But the notion of a solution-independent decomposition of spacetime into space and time makes no sense in general relativity, since solutions differ as to which curves count as timelike and which hypersurfaces count as spacelike. Somewhat surprisingly, it turns out to be possible to construct a Hamiltonian version of general relativity without employing slicings. But— unsurprisingly—without some sort of decomposition of spacetime into instants, it makes no sense to ask which states follow a given state (so there is no real dynamics on the Hamiltonian side) nor to try to construct a one-parameter family of functions on the reduced space of solutions that corresponds to the instantaneous values of a quantity of interest. So it is natural to look for a surrogate of the notion of a slicing that applies to diffeomorphism equivalence classes of solutions, rather than to individual solutions—and to hope that this will lead to familiar-looking accounts of the representation of time and change.

Throughout this final subsection, unless otherwise noted, I restrict attention to spatially compact vacuum general relativity in four spacetime dimensions with vanishing cosmological constant.

Let me begin with some definitions.

DEFINITION 45 (Geometry). A point in the reduced space of solutions of general relativity is called a *geometry*. A geometry is an orbit of the action on the space of solutions of the group $\mathcal{D}(V)$ of diffeomorphisms from V to itself. We write $[g]$ for the geometry corresponding to a solution g; we speak of a solution in $[g]$ as a solution *with geometry* $[g]$.

[174]Thus, the WKB method aims to construct approximate eigenstates for the quantum Hamiltonian. Analyses based upon decoherence, coherent states, etc., aim to show that the dynamics driven by the quantum Hamiltonian approximates that of the corresponding classical system, and so on. See [Landsman, this volume].

DEFINITION 46 (Instantaneous Geometry). The group $\mathcal{D}(S)$ of diffeomorphisms from the abstract instant S to itself acts on the space of initial data. We call an orbit of this action an *instantaneous geometry*. We denote the instantaneous geometry corresponding to an initial data point (q, π) by $\langle q, \pi \rangle$. We speak of initial data (q, π) as having the geometry $\langle q, \pi \rangle$.[175]

DEFINITION 47 (Time for a Solution). Let (V, g) be a solution. A *time for* (V, g) is a partition, $\{\Sigma\}$, of (V, g) by Cauchy surfaces, called the *instants* of the time. A *parameterized time* is a time together with a preferred parameterization of the set of instants. An *affinely parameterized time* is a time whose instants are parameterized up to the choice of origin.[176]

DEFINITION 48 (Absolute Time). Let g be a solution. A time $\{\Sigma\}$ for (V, g) is called *absolute* if every isometry of g maps instants in $\{\Sigma\}$ to instants in $\{\Sigma\}$. An affinely parameterized time for g is called *absolute* if each isometry of g satisfies the preceding condition and preserves the parameter difference between each pair of instants. A parameterized time for g is called *absolute* if each isometry of g maps each instant of the time to itself.

Every globally hyperbolic solution admits a parameterized time (since each globally hyperbolic solution can be foliated by Cauchy surfaces, which can be given an arbitrary parameterization). But it is not always possible to find absolute times for solutions with large isometry groups. Minkowski spacetime does not admit an absolute time.[177] If a spacetime admits time translation or inversion as a symmetry, then it does not admit an absolute parameterized time.

DEFINITION 49 (Time for General Relativity). A *(plain, affinely parameterized, or parameterized) time for general relativity* is a map defined on a subset of the space of solutions that assigns to each solution in its domain a (plain, affine, or parameterized) time for that solution, and does so in an appropriately smooth manner.

DEFINITION 50 (Geometric Time for General Relativity). A (plain, affinely parameterized, or parameterized) time for general relativity is called *geometric* if it satisfies the following conditions. (i) Its domain of definition is closed under the action of $\mathcal{D}(V)$ on the space of solutions. (ii) If g and g' are in the domain of the time and $g' = d^*g$ for some diffeomorphism $d : V \to V$, then the foliation assigned to g' is the image under d^{-1} of the foliation assigned to g (if the time is affinely parameterized, then we require that such a d preserve the time difference between any two instants; if the time is parameterized, then we require that such

[175] Note that a instantaneous geometry is *not* a point in the reduced space of initial data: initial data induced by a given solution on distinct Cauchy surfaces correspond to the same point in the reduced space of initial data, but (in general) to distinct points in the space of instantaneous geometries.

[176] We can think of a time for (V, g) as an unparameterized curve in the space of Cauchy surfaces of (V, g); a parameterized time is a parameterized curve of this type; an affinely parameterized time is an affinely parameterized curve of this type.

[177] A time invariant under the notion of time translation associated with a given frame will fail to be invariant under boosts relative to that frame. The same argument will work in de Sitter spacetime, or in other spacetimes admitting boost symmetries; see [Moncrief, 1992] for examples.

a d map the instant labelled by t to the instant labelled by t). I will often shorten *geometric time for general relativity* to *geometric time*.

REMARK 51 (Geometric Times are Absolute). The (plain, affinely parameterized, or parameterized) time that a geometric time for general relativity assigns to a solution g is always absolute. For if $d : V \to V$ is an isometry of g, then condition (ii) in the preceding definition tells us that d preserves the time assigned to g, together with its parameterization properties, if any. It follows that Minkowski spacetime is not in the domain of definition of any geometrical time for general relativity, and that no solution invariant under time translation or inversion is in the domain of definition of any parameterized geometric time for general relativity.

We can think of a (parameterized, affinely parameterized, or unparameterized) geometric time for general relativity as a means of associating a geometry $[g]$ in the reduced space of solutions with a (parameterized, affinely parameterized, or unparameterized) curve $\langle q(t), \pi(t)\rangle$ in the space of instantaneous geometries; we call such a curve a *dynamical trajectory*. The correspondence between geometries and dynamical trajectories is set up in the obvious way: let g be a solution in the domain of definition of a given geometric time, and let $(q(t), \pi(t))$ be the (parameterized, affinely parameterized, or unparameterized) curve in the space of initial data that results when we look at the initial data induced by g on the instants in the time assigned to g; $\langle q(t), \pi(t)\rangle$ is the dynamical trajectory we seek.[178] If g_1 and g_2 are solutions with the same geometry, then they are related by some diffeomorphism $d : V \to V$. In this case d also relates the foliations assigned to them by our geometric time, so g_1 and g_2 will correspond to the same dynamical trajectory in the space of instantaneous geometries.

A number of interesting examples of geometric times are known. Most have very small domains of definition: (i) within the class of nonrotating dust solutions, a geometric time is given by foliating each solution by the unique family of hypersurfaces everywhere orthogonal to the dust worldlines; (ii) within the class of solutions whose isometry groups are three-dimensional with spacelike orbits, a geometric time is given by foliating each solution by the orbits of its isometry group.[179] Examples of wider scope are harder to come by but do exist.

EXAMPLE 52 (CMC Time). Recall that if $\Sigma \subset V$ is a Cauchy surface for (V, g), then we can define tensors q^{ab} and k_{ab} on Σ with the following meaning: $q^{ab} := g^{ab} \mid_\Sigma$ is the Riemannian metric that g induces on Σ and $2k_{ab}$ is the rate of change of this metric according to freely falling observers whose worldlines intersect Σ orthogonally. Out of these tensors we can construct the ***mean curva-***

[178] Strictly speaking, in order to construct the curve $(q(t), \pi(t))$ in the space of initial data, we need to introduce a slicing of (V, g) whose instants coincide with those of the given time, so that we can pullback states on concrete instants to states on our abstract instant S; the arbitrariness involved in a choice of slicing washes out when we quotient the space of initial data by the action of $\mathcal{D}(S)$ to reach the space of instantaneous states.

[179] Scheme (i) generalizes Einstein's simultaneity convention to the context of dust cosmology; see [Sachs and Wu, 1977, §5.3]. Note that schemes (i) and (ii) need not coincide within their shared domain of definition; see [King and Ellis, 1973].

ture, $k : \Sigma \to \mathbb{R}$, defined by $k := q^{ab}k_{ab}$ (so $k(x)$ is just the trace of the matrix that encodes information about k_{ab} at x). A Cauchy surface $\Sigma \subset V$ for a solution g is called a *surface of constant mean curvature,* or simply a *CMC surface,* if k is a constant function on Σ. Recall that unless otherwise noted, we restrict attention to (3+1) spatially compact globally hyperbolic vacuum solutions with vanishing cosmological constant.

Applicability. It is widely believed that a large class of solutions to Einstein's field equations can be foliated by CMC surfaces.

1. It is known that the set of solutions containing a CMC slice is an open set in the space of solutions.[180]
2. It was once conjectured that all solutions contain at least one CMC surface, but it is now known that this is not so.[181]
3. It is was once conjectured that all solutions admitting a CMC slice can be foliated by such slices.[182] This is now believed to hold only for certain spatial topologies.[183]
4. It is believed that within the class of solutions foliated by CMC slices, all solutions of a given spatial topology will exhibit the same range of values of constant mean curvature, with the only exceptions being stationary solutions (recall that a solution is stationary if it admits a timelike Killing field—roughly speaking, the infinitesimal generator of a time translation group).[184]

Invariance Properties. CMC foliations behave superbly well under isometries.[185] Let (V, g) be a solution, $\{\Sigma\}$ a set of CMC surfaces that foliates V, and $d : V \to V$ an isometry of g. Then d leaves the foliation $\{\Sigma\}$ invariant.[186] If (V, g) is non-stationary, then: (a) any symmetry d of g preserves each leaf

[180] See, e.g., [Isenberg and Marsden, 1982, p. 195].

[181] See [Bartnik and Isenberg, 2004, p. 32]. The corresponding conjecture for spatially compact dust solutions is also false; see[Bartnik, 1988].

[182] For the original form of the conjecture, see, e.g., [Isenberg and Marsden, 1982, Conjecture 3.2]. This conjecture is known to be true for flat spacetimes ([Barbot, 2005, §12]) and the corresponding conjecture is known to be true in the (2+1) case ([Andersson *et al.*, 1997]). The counterpart of this conjecture is known to be false for spatially compact dust solutions ([Isenberg and Rendall, 1998]) and in the asymptotically flat vacuum case, where the Schwarzschild solution provides a counterexample ([Eardley and Smarr, 1979, §III]).

[183] For the current conjecture, see [Rendall, 1996, Conjectures 1 and 2]. It is now believed that for some spatial topologies, behaviour analogous to that of the Schwarzschild solution can occur; see [Rendall, 1996] and [Andersson, 2004, p. 81]. In the (3+1) case, the revised conjecture is known to be true for some types of highly symmetric solutions, even when some forms of matter are allowed; see [Rendall, 1996, Theorems 1 and 2], [Andersson, 2004, pp. 81 f. and 95], and the references therein.

[184] For this, see [Rendall, 1996, Conjectures 1 and 2]. For the situation in highly symmetric cases and in (2+1) dimensions, see the references of the previous two footnotes.

[185] See [Isenberg and Marsden, 1982, §3].

[186] This would fail for spacetimes admitting boost symmetries, such as Minkowski spacetime and de Sitter spacetime. (Note that since we require vanishing cosmological constant, de Sitter spacetime does

in $\{\Sigma\}$; and (b) for any real number κ, there is at most one Cauchy surface with constant mean curvature κ. If (V, g) is stationary then: g is flat and any CMC surface in (V, g) has vanishing mean curvature.[187]

CMC Time. Foliating each solution by its CMC slices, when possible, determines a geometric time within the class of solutions we are considering. We can render this an affinely parameterized geometric time as follows: for non-stationary solutions, the parameter difference between slices of mean curvature κ_1 and κ_2 is $|\kappa_2 - \kappa_1|$; for stationary solutions, the parameter difference between two slices is the proper time elapsed between those slices. If we restrict attention to non-stationary solutions, and assign to each slice the parameter value given by its mean curvature, then we arrive at a parameterized geometric time.

EXAMPLE 53 (Cosmological Time). Given a solution (V, g), the *cosmological time function* for g is the map $\tau : V \to \mathbf{R} \cup \{\infty\}$ that assigns to each $x \in V$ the supremum over the length of all past-directed causal curves starting at x. Obviously there are many well-behaved spacetimes in which $\tau(x)$ is badly behaved—e.g., in Minkowski spacetime, $\tau(x) = \infty$ for all events. We say that the cosmological time function of a solution is *regular* if: (a) $\tau(x) < \infty$ for all x and (b) $\tau \to 0$ along each past inextendible causal curve. If τ is a regular cosmological time function on (V, g) then: (i) (V, g) is globally hyperbolic; (ii) τ is a time function for the solution in the usual sense (i.e., it is continuous and strictly increasing along future-directed causal curves); and (iii) the level surfaces of τ are future Cauchy surfaces in (V, g) (i.e., these surfaces have empty *future* Cauchy horizons).[188] In spatially compact vacuum (2+1)-dimensional general relativity, it is known: (a) that the cosmological time is regular for almost all spacetime topologies; and (b) that in one important class of solutions the cosmological time coincides with the CMC time.[189] On the class of spacetimes with regular cosmological time functions whose level surfaces are Cauchy surfaces, we construct a geometric time for general relativity by foliating each solution by the surfaces of constant cosmological time; parameterizing these foliations by the value the cosmological time function takes on each leaf yields a parameterized geometric time for general relativity, so long as we exclude solutions with a time reflection symmetry.

A geometric time for general relativity is, in effect, a means of separating out from the infinite number of variables of the theory one relative to which the oth-

---

not count as a spatially compact vacuum solution for present purposes.) Note that in the asymptotically flat case, the question of the invariance of CMC slices is much more involved; see [Bartnik *et al.*, 1990, §5].

[187] Of course, in general a timelike Killing vector does not guarantee flatness. But it does so within the class solutions presently under consideration.

[188] See [Andersson *et al.*, 1998, Propositions 2.2 and 2.5 and Corollary 2.6].

[189] See [Benedetti and Guadagnini, 2001]. In general, however, surfaces of constant cosmological time are less smooth than CMC surfaces, so the two notions of time do not coincide; see [Benedetti and Guadagnini, 2001, p. 331] or [Barbot and Zeghib, 2004, §5.4.1].

ers are to be seen as evolving, by allowing us to pass from a point in the reduced space of solutions to a (possibly unparameterized) curve in the space instantaneous geometries. Instantaneous physical quantities such as the volume of the universe, or the number of stars, or the size of the solar system can be represented by real-valued functions on the space of instantaneous geometries. And so the choice of a geometric time allows us to talk about change in the familiar way: we can check to see whether a function on the space of instantaneous geometries that represents a quantity of interest takes on different values at points corresponding to the different instantaneous geometries that occur in a given spacetime geometry. If our geometric time for general relativity is affinely parameterized, we can calculate the rate of change of quantities of interest (since we then have an affinely parameterized curve through the space of instantaneous geometries corresponding to a given spacetime geometry). If we have a parameterized geometric time for general relativity, we can even mimic the construction we used in earlier sections to represent changeable quantities by one-parameter families of functions on the reduced space of solutions of the theory: given the function f on the space of instantaneous geometries that represents our quantity of interest, and a real number $t,$ we define a partially defined function f_t on the reduced space of solutions by setting $f_t[g]$ equal to the value that f takes on the instantaneous geometry corresponding to t in $[g].$

As delineated above, the problem of time in general relativity had two major aspects.

1. Time is not represented in spatially compact general relativity, as it was in earlier theories, via a flow on a symplectic space of states;

2. Change is not represented, as it was in earlier theories, via functions on symplectic spaces corresponding to the spaces of possible instantaneous states and worlds.

We now see that if we go as far as introducing a parameterized geometric time, we can address the second of these worries by representing a changeable quantity by a one-parameter family of functions on the reduced space of solutions, in the usual way.

Does the introduction of a geometric time suffice to address the first worry? Any geometric time singles out a subspace of the space of instantaneous geometries, consisting of those $\langle q, \pi \rangle$ that arise as instantaneous geometries of the Cauchy surfaces picked out by that geometric time—e.g., only instantaneous geometries portraying space as having constant mean curvature can arise according to the CMC slicing scheme. If we introduce an affinely parameterized geometric time for general relativity, then we do get a flow on the space of instantaneous geometries that arise according to this geometric time (since this space is partitioned by the affinely parameterized dynamical trajectories corresponding to geometries in the domain of definition of the given geometric time). But one does not expect this space to be symplectic nor to be isomorphic to the reduced space of solutions.[190] So a flow on the space of instantaneous geometries associated with our geometric time is not a flow on a symplectic space. And since each dynamical trajectory on the space of instantaneous geometries corresponds to a singe point in the reduced space of solutions, we have no means of carrying our flow on the former space over to a nontrivial flow on the latter.

A natural strategy to set up a representation of time via a flow on a symplectic manifold is to attempt to parlay a choice of geometric time for general relativity into a reformulation of of the theory as a nontrivial (but possibly time-dependent) Hamiltonian system. In one important case, it is known that this can achieved.

EXAMPLE 54 (CMC dynamics). We consider the CMC time introduced in example 52 above.[191] We impose restrictions on the topology of our abstract instant S.[192] Let $\mathcal{M}$ be the space of Riemannian metrics on S with constant scalar curvature -1.[193] The cotangent bundle $T^*\mathcal{M}$ is a symplectic space; an element of $T^*\mathcal{M}$ is of the form (γ, p) where $\gamma \in \mathcal{M}$ and p is a symmetric contravariant tensor density of rank two on S that is divergenceless and traceless according to γ. We consider $(\gamma, p), (\gamma', p') \in T^*\mathcal{M}$ to be equivalent if there is a diffeomorphism $d : S \to S$ such that $(\gamma', p') = (d^*\gamma, d^*p)$. The space $\mathcal{I}^* := T^*\mathcal{M}/\mathcal{D}(S)$ that results when we quotient out by this equivalence relation inherits a symplectic structure from $T^*\mathcal{M}$. We will call points in $\mathcal{I}^*$ *conformal initial data*. For each $t < 0$ there is a geometrically natural symplectic isomorphism between $\mathcal{I}^*$ and the space of instantaneous geometries with constant mean curvature t.[194] And there is a natural symplectic isomorphism between the latter set and the reduced space

[190]Intuitively, the space of instantaneous geometries that arise according to a given geometric time can be thought of as the product of the reduced space of solutions with the real line (since each geometry corresponds to a one-parameter family of instantaneous geometries relative to the geometric time). So the space of instantaneous geometries of the given geometric time is not isomorphic to the space of solutions—nor can it be symplectic, since it is the product of a symplectic space with an odd-dimensional space.

[191]For an overview of the (3+1) case, see [Fischer and Moncrief, Unpublished, §§2 and 3]; for details see [Fischer and Moncrief, 1996], [Fischer and Moncrief, 1997], and the references therein. For the (2+1) case see [Moncrief, 1989] and [Andersson *et al.*, 1997]. The construction described below is an example of deparameterization. For this notion and for finite-dimensional applications, see [Beig, 1994, §2].

[192]We impose two conditions. (i) S must be of Yamabe type -1, i.e., the only constant scalar curvature Riemannian metrics that S admits have negative scalar curvature. This is essential for the constructions employed in the papers cited. (ii) S must not admit any Riemannian metrics with isometry groups of positive dimension. This saves us from having to worry about singular quotient spaces.

[193]Because S is of Yamabe type -1, every Riemannian metric on S is conformally equivalent to a metric in $\mathcal{M}$.

[194]Let us ignore the $\mathcal{D}(S)$ symmetry for a moment. Given a pair (γ, p) and a time $t < 0$ there is a unique positive scalar ϕ on S solving the the Lichernowicz equation for (γ, p, t),

$$\Delta_\gamma \phi - \frac{1}{8}\phi + \frac{1}{12}t^2\phi^5 - \frac{1}{8}(p \cdot p)\mu^{-2}\phi^{-7} = 0$$

(here Δ_γ is the Laplacian for γ and μ is the volume form for γ). Our desired (q, π) is given by $q := \phi^4\gamma$ and $\pi := \phi^{-4}p + \frac{2}{3}t\phi^2\gamma^{-1}$.

of solutions of general relativity (under which an instantaneous state is sent to the unique geometry that it occurs in). So for each $t < 0$ we have a symplectic isomorphism between the space of conformal initial data and the reduced space of solutions.

Conversely, given a geometry $[g]$ and a time $t < 0$ we can look for the point in $\mathcal{I}^*$ that corresponds to $[g]$ according to the isomorphism labelled by t. Doing this for each $t < 0$ gives us a curve in $\mathcal{I}^*$ corresponding to $[g]$. A generic point in $\mathcal{I}^*$ will lie on many such trajectories: in general if $x \in \mathcal{I}^*$ and $t_1 \neq t_2$ then the instantaneous geometry of constant mean curvature t_1 corresponding to (x, t_1) and the instantaneous geometry of constant mean curvature t_2 corresponding to (x, t_2) will reside in different spacetime geometries. If we look at the complete family of trajectories in $\mathcal{I}^*$ corresponding to all of the geometries in the reduced space of solutions, then we find that these are generated by the symplectic structure of $\mathcal{I}^*$ together with a time-dependent Hamiltonian $h(t)$ that is a simple function of t and of spatial volume.[195]

Taking this example as our model, we can introduce the notion of a *Hamiltonianization* of general relativity associated with a given parameterized geometric time for the theory. Suppose that we are given such a parameterized geometric time. Suppose further that we are able to construct a symplectic space $\mathcal{I}^*$ whose points are ($\mathcal{D}(S)$-equivalence classes of) tensors on the abstract instant S, and that for each value of t we are able to construct a geometrically natural isomorphism between $\mathcal{I}^*$ and the set of instantaneous geometries corresponding to t according to our parameterized geometric time. Composing these isomorphisms with the canonical map from the space of instantaneous geometries to the reduced space of solutions gives us a one-parameter family of symplectic isomorphisms between $\mathcal{I}^*$ and $\mathcal{S}'$.[196] This allows us to associate each geometry $[g]$ with a curve $x(t)$ in $\mathcal{I}^*$: for each t, $x(t)$ is the point in $\mathcal{I}^*$ that gets mapped to $[g]$ by the isomorphism labelled by t. We call $x(t)$ the dynamical trajectory associated with $[g]$. We now consider the class of dynamical trajectories on $\mathcal{I}^*$ that arise in this way, and ask whether there is a (possibly time-dependent) Hamiltonian on $\mathcal{I}^*$ that generates them in concert with the symplectic structure of $\mathcal{I}^*$. If there is, then the resulting (possibly time-dependent) Hamiltonian system is a Hamiltonianization of general relativity based upon the given parameterized geometric time.

As we have seen, given a parameterized geometric time for general relativity we can represent changeable quantities in the familiar way via one-parameter families of functions on the reduced space of solutions. And if we go further and introduce an associated Hamiltonianization of the theory, then we can represent time in the familiar way via a (possibly time-dependent) Hamiltonian flow on the symplectic space $\mathcal{I}^*$, whose points we can think of as initial data posable at different times.

[195] The spatial volume is itself a t-dependent function on $\mathcal{I}^*$, since the same conformal data will lead to instantaneous geometries with different volumes when supplemented by different values of t.

[196] Strictly speaking, these isomorphisms will be merely local (as in section 6.1) if the range of values taken on by the time parameter varies from geometry to geometry.

So these notions allow us to circumvent the problem of time by playing the same roles that a slicing played in sections 5 and 6 when we considered theories set in fixed background spacetimes.

Does the introduction of a geometric time or of an associated Hamiltonianization violate general covariance? In one sense there is no violation—for these notions are situated at the level of the reduced space of solutions, and so cannot, e.g., treat diffeomorphic solutions differently.

But it remains true that the introduction of a geometric time violates the spirit of general relativity, as the theory is generally understood today—most would like to think of special relativity as dissolving any privileged distinction between time and space and of general relativity as generalizing special relativity in a way that does nothing to reinstate such a distinction.[197]

Note, however, that this is really an objection to the privileging of one geometric time over others. It seems entirely in the spirit of general relativity to think of the content of the theory as being elucidated by each of its Hamiltonianizations and as being exhausted by the set of all Hamiltonianizations (that is, if we ignore spacetimes with time translation or reflection symmetries).

Still, it is natural to ask what sort of considerations could lead us to recognize a geometric time or associated Hamiltonianization as being the *correct* one.[198]

Classical Considerations. In the CMC Hamiltonianization sketched in example 54 above general relativity is recast as a time-dependent system. This is a bit unsettling: we are used to thinking that time-dependent Hamiltonians only arise when an open system is subject to external forces. So it is surprising to encounter a time-dependent Hamiltonian system in a fundamental context. Perhaps this is something we have to learn to live with: we are here in effect singling out one of general relativity's infinitely many variables and treating it as time—and we expect there to be all sorts of complicated nonlinear interactions between the variables of general relativity. However, some interesting special cases are known of geometric times that lead to time-independent Hamiltonianizations of general relativity.[199]

So we cannot rule out the possibility that there may be a geometric time of wide scope that that allows us to reformulate general relativity as a time-independent Hamiltonian theory with non-trivial dynamics.[200] Clearly the

[197]On the other hand, many early relativistic cosmologists were happy to take the natural foliation of nonrotating dust cosmologies by surfaces orthogonal to the dust worldlines as a sign that the distinction between space and time, banished in Einstein's account of electromagnetism, was reinstated at the astronomical level. See [Belot, 2005, §3.2] for discussion and references.

[198]Note that some approaches in the philosophy of time and some approaches to the interpretation of quantum mechanics would appear to require something like a preferred foliation of spacetime by instants of time.

[199]This happens with the CMC time in the case where space has the topology of a two-torus; see [Moncrief, 1989, p. 2913]. It can also be achieved for general relativity coupled to a perfect fluid—in this case the conserved quantity that drives the dynamics is total baryon number; see [Moncrief, 1977] and [Moncrief and Demaret, 1980].

[200]Note that given a non-trivial time-independent Hamiltonian on $\mathcal{I}^*$, we can use our t-dependent

construction of such a geometric time would be of the first interest: it might well seem that we had happened on the *correct* time, previously concealed from us by the unperspicuous formulations of the theory that we had been working with—much as it would have if classical mechanics had first been given a time-reparameterization invariant formulation, and it had then been discovered that a certain family of parameterizations allowed the equations to be rewritten in a much simpler form.

Quantum Considerations. The question whether to privilege one geometric time or to treat them all equally can be expected to have repercussions for quantization (which project provides the main motivation for looking for a Hamiltonian formulation of general relativity with nontrivial dynamics in the first place). For one certainly does not expect that distinct Hamiltonian formulations of general relativity corresponding to distinct choices of geometric time should have equivalent quantizations—at least not if equivalent quantizations are required to be unitarily equivalent.[201]

So what can we hope for? For long shots like the following. (1) Perhaps only one geometric time will lead to an empirically adequate quantum theory of gravity. (2) Perhaps there will be a natural class of geometric times (e.g., the ones that lead to time-independent Hamiltonians) that can be seen as underwriting the equivalent quantum theories (perhaps in a liberalized sense of "equivalent").

Far more plausibly, the solution to the difficulties in quantizing general relativity will come from some other direction entirely. But hopefully it will in any case be worthwhile to be clear about the nature of the problem of time.

ACKNOWLEDGEMENTS

I would like to thank: the editors of this volume for their invitation, their sage advice, and their great patience; audiences in Montreal, New York, and Pittsburgh who sat through chaotic presentations of this material; Vincent Moncrief for helpful correspondence; Richard Thomas for unhelpful correspondence; and Frank Arntzenius, Cian Dorr, Alexandre Guay, and Laura Ruetsche for helpful conversations.

isomorphisms between $\mathcal{I}^*$ and $\mathcal{S}'$ to construct a corresponding (possibly time-dependent) Hamiltonian on $\mathcal{S}'$. Unless the latter function were constant, it would generate a nontrivial flow on the reduced space of solutions. Of course, this could not be interpreted as time translation, although generated by the counterpart of the time-independent Hamiltonian generating time evolution.

[201] The Stone-von Neumann theorem guarantees that unless one of them does something odd, two people setting out to quantize a Hamiltonian theory with a linear, finite-dimensional phase space will end up with unitarily equivalent quantizations. But as soon as one considers infinite-dimensional or nonlinear phase spaces the situations changes radically—for example, what look like equivalent formalisms at the classical level lead to distinct quantum theories. See [Ruetsche, Unpublished] and [Gotay, 2000] for discussion, examples, and references. See [Gotay and Demaret, 1983] for a minisuperspace cosmological model that admits competing deparameterizations that lead to physically distinct quantizations.

BIBLIOGRAPHY

[Abraham and Marsden, 1978] Ralph Abraham and Jerrold Marsden. *Foundations of Mechanics*. Perseus Books, Cambridge, MA, 2nd edition, 1978.

[Abraham *et al.*, 1988] Ralph Abraham, Jerrold Marsden, and Tudor Ratiu. *Manifolds, Tensor Analysis, and Applications*. Springer-Verlag, Berlin, 2nd edition, 1988.

[Adams *et al.*, 1985] Malcolm Adams, Tudor Ratiu, and Rudolf Schmid. The Lie group structure of diffeomorphism groups and invertible Fourier operators, with applications. In Victor Kac, editor, *Infinite-Dimensional Groups with Applications*, pages 1–69. Springer-Verlag, Berlin, 1985.

[Ali and Engliš, 2005] Twareque Ali and Miroslav Engliš. Quantization methods: A guide for physicists and analysts. *Reviews in Mathematical Physics*, 17:391–490, 2005.

[Anderson and Duchamp, 1980] Ian Anderson and Thomas Duchamp. On the existence of global variational principles. *American Journal of Mathematics*, 102:781–868, 1980.

[Anderson and Thompson, 1992] Ian Anderson and Gerard Thompson. *The Inverse Problem of the Calculus of Variations for Ordinary Differential Equations*. American Mathematical Society, Providence, RI, 1992.

[Andersson *et al.*, 1997] Lars Andersson, Vincent Moncrief, and Anthony Tromba. On the global evolution problem in 2+1 gravity. *Journal of Geometry and Physics*, 23:191–205, 1997.

[Andersson *et al.*, 1998] Lars Andersson, Gregory Galloway, and Ralph Howard. The cosmological time function. *Classical and Quantum Gravity*, 15:309–322, 1998.

[Andersson, 1987] Lars Andersson. Momenta and reduction in general relativity. *Journal of Geometry and Physics*, 4:289–314, 1987.

[Andersson, 1989] Lars Andersson. Momenta and reduction in general relativity. II. The level sets. In Robert Bartnik, editor, *Conference of Mathematical Relativity*, pages 73–88. Australian National University, Canberra, 1989.

[Andersson, 2004] Lars Andersson. The global existence problem in general relativity. In Piotr Chruściel and Helmut Friedrich, editors, *The Einstein Equations and the Large Scale Behavior of Gravitational Fields*, pages 71–120. Birkhaüser, Basel, 2004.

[Arms, 1981] Judith Arms. The structure of the solution set for the Yang–Mills equations. *Mathematical Proceedings of the Cambridge Philosophical Society*, 90:361–372, 1981.

[Arnold *et al.*, 1997] V.I. Arnold, V.V. Kozlov, and A.I. Neistadt. *Mathematical Aspects of Classical and Celestial Mechanics*. Springer-Verlag, Berlin, 2nd edition, 1997.

[Arnold, 1989] V.I. Arnold. *Mathematical Methods of Classical Mechanics*. Springer-Verlag, Berlin, 2nd edition, 1989.

[Ashtekar and Romano, 1992] Abhay Ashtekar and Joseph Romano. Spatial infinity as a boundary of spacetime. *Classical and Quantum Gravity*, 9:1069–1100, 1992.

[Ashtekar *et al.*, 1991] Abhay Ashtekar, Luca Bombelli, and Oscar Reula. The covariant phase space of asymptotically flat gravitational fields. In Mauro Francaviglia, editor, *Mechanics, Analysis, and Geometry: 200 Years After Lagrange*, pages 417–450. Elsevier, Amsterdam, 1991.

[Barbot and Zeghib, 2004] Thierry Barbot and Abdelghani Zeghib. Group actions on Lorentz spaces, mathematical aspects: A survey. In Piotr Chruściel and Helmut Friedrich, editors, *The Einstein Equation and the Large Scale Behavior of Gravitational Fields*, pages 401–439. Birkhaüser, Basel, 2004.

[Barbot, 2005] Thierry Barbot. Flat globally hyperbolic spacetimes. *Journal of Geometry and Physics*, 53:123–165, 2005.

[Bartnik and Isenberg, 2004] Robert Bartnik and James Isenberg. The constraint equations. In Piotr Chruściel and Helmut Friedrich, editors, *The Einstein Equations and the Large Scale Behavior of Gravitational Fields*, pages 1–38. Birkhaüser, Basel, 2004.

[Bartnik *et al.*, 1990] Robert Bartnik, Piotr Chruściel, and Niall Ó Murchadha. On maximal surfaces in asymptotically flat spacetimes. *Communications in Mathematical Physics*, 130:95–109, 1990.

[Bartnik, 1988] Robert Bartnik. Remarks on cosmological spacetimes and constant mean curvature surfaces. *Communications in Mathematical Physics*, 117:615–624, 1988.

[Beig and Ó Murchadha, 1987] Robert Beig and Niall Ó Murchadha. The Poincaré group as the symmetry group of canonical general relativity. *Annals of Physics*, 174:463–498, 1987.

[Beig and Schmidt, 1982] Robert Beig and Bernd Schmidt. Einstein's equations near spatial infinity. *Communications in Mathematical Physics*, 87:65–80, 1982.

[Beig, 1994] Robert Beig. The classical theory of canonical general relativity. In Jürgen Ehlers and Helmut Friedrich, editors, *Canonical Gravity: From Classical to Quantum*, pages 59–80. Springer-Verlag, Berlin, 1994.

[Belot and Earman, 2001] Gordon Belot and John Earman. Pre-Socratic quantum gravity. In Craig Callender and Nick Huggett, editors, *Physics Meets Philosophy and the Planck Scale*, pages 213–255. Cambridge University Press, Cambridge, 2001.

[Belot, 2003] Gordon Belot. Symmetry and gauge freedom. *Studies in History and Philosophy of Modern Physics*, 34:189–225, 2003.

[Belot, 2005] Gordon Belot. Dust, time, and symmetry. *British Journal for the Philosophy of Science*, 55:255–291, 2005.

[Benedetti and Guadagnini, 2001] Riccardo Benedetti and Enore Guadagnini. Cosmological time in (2+1)-gravity. *Nuclear Physics B*, 613:330–352, 2001.

[Bluman, 2005] George Bluman. Connections between symmetries and conservation laws. *Symmetry, Integrability and Geometry: Methods and Applications*, 1:011, 2005.

[Brading and Castellani, this volume] Katherine Brading and Elena Castellani. Symmetries and invariances in classical physics. Chapter 13, this volume.

[Budic *et al.*, 1978] Robert Budic, James Isenberg, Lee Lindblom, and Philip Yasskin. On the determination of cauchy surfaces from intrinsic properties. *Communications in Mathematical Physics*, 61:87–95, 1978.

[Butterfield and Isham, 2000] Jeremy Butterfield and Chris Isham. On the emergence of time in quantum gravity. In Jeremy Butterfield, editor, *The Arguments of Time*, pages 111–168. Oxford University Press, Oxford, 2000.

[Butterfield, this volume] Jeremy Butterfield. On symplectic reduction in mechanics. Chapter 1, this volume.

[Butterfield, Unpublished] Jeremy Butterfield. Against *Pointillisme* about mechanics. Unpublished.

[Cannas da Silva, Unpublished] Ana Cannas da Silva. Symplectic geometry. Unpublished.

[Cantor, 1979] Murray Cantor. Some problems of global analysis on asymptotically simple manifolds. *Compositio Mathematica*, 38:3–35, 1979.

[Choquet-Bruhat and DeWitt-Morrette, 1989] Yvonne Choquet-Bruhat and Cécile DeWitt-Morrette. *Analysis, Manifolds, and Physics. Part II: 92 Applications*. North-Holland, Amsterdam, 1989.

[Choquet-Bruhat *et al.*, 1977] Yvonne Choquet-Bruhat, Cécile DeWitt-Morrette, and Margaret Dillard-Bleick. *Analysis, Manifolds, and Physics*. North-Holland, Amsterdam, 1977.

[Crampin and Prince, 1988] M. Crampin and G.E. Prince. Alternative Lagrangians for spherically symmetric potentials. *Journal of Mathematical Physics*, 29:1551–1554, 1988.

[Crnković and Witten, 1987] Čedomir Crnković and Edward Witten. Covariant description of canonical formalism in geometrical theories. In Stephen Hawking and Werner Israel, editors, *Three Hundred Years of Gravitation*, pages 676–684. Cambridge University Press, Cambridge, 1987.

[Cushman and Bates, 1997] Richard Cushman and Larry Bates. *Global Aspects of Classical Integrable Systems*. Birkhaüser, Basel, 1997.

[Deligne and Freed, 1999] Pierre Deligne and Daniel Freed. Classical field theory. In Pierre Deligne, Pavel Etingof, Daniel Freed, Lisa Jeffrey, David Kazhdan, John Morgan, David Morrison, and Edward Witten, editors, *Quantum Fields and Strings: A Course for Mathematicians*, volume 1, pages 137–225. American Mathematical Society, Providence, RI, 1999.

[Diacu and Holmes, 1996] Florin Diacu and Philip Holmes. *Celestial Encounters*. Princeton University Press, Princeton, 1996.

[Diacu, 1992] Florin Diacu. *Singularities of the N-Body Problem*. Les Publications CRM, Montréal, 1992.

[Diacu, 2002] Florin Diacu. Singularities of the N-body problem. In Hildeberto Cabral and Florin Diacu, editors, *Classical and Celestial Mechanics: The Recife Lectures*, pages 35–62. Princeton University Press, Princeton, 2002.

[Dirac, 2001] P.A.M. Dirac. *Lectures on Quantum Mechanics*. Dover, New York, 2001.

[Dubrovin *et al.*, 1992] B.A. Dubrovin, A.T. Fomenko, and S.P. Novikov. *Modern Geometry—Methods and Applications. Part I. The Geometry of Surfaces, Transformation Groups, and Fields*. Springer-Verlag, Berlin, 2nd edition, 1992.

[Duval *et al.*, 1990] Christian Duval, J. Elhadad, Mark Gotay, and Gijs Tuynman. Nonunimodularity and the quantization of the pseudo-rigid body. In John Harnad and Jerrold Marsden, editors, *Hamiltonian Systems, Transformation Groups, and Spectral Transform Methods*, pages 149–160. Les Publications CRM, Montréal, 1990.

[Eardley and Smarr, 1979] Douglas Eardley and Larry Smarr. Time functions in numerical relativity: Marginally bound dust collapse. *Physical Review D*, 19:2239–2259, 1979.

[Earman, 1986] John Earman. *A Primer on Determinism*. Reidel, Dordrecht, 1986.

[Earman, 2002] John Earman. Thoroughly modern McTaggart, or what McTaggart would have said if he had read the general theory of relativity. *Philosophers' Imprint*, 2(3):1–28, 2002.

[Earman, 2003] John Earman. Tracking down gauge. In Katherine Brading and Elena Castellani, editors, *Symmetries in Physics*, pages 140–162. Cambridge University Press, Cambridge, 2003.

[Earman, this volume] John Earman. Aspects of determinism in modern physics. Chapter 15, this volume.

[Earman, Unpublished] John Earman. Two challenges to the requirement of substantive general covariance. Unpublished.

[Eichhorn, 1993] Jürgen Eichhorn. The manifold structure of maps between open manifolds. *Annals of Global Analysis and Geometry*, 11:253–300, 1993.

[Fischer and Moncrief, 1996] Arthur Fischer and Vincent Moncrief. A method of reduction of Einstein's equations of evolution and a natural symplectic structure on the space of gravitational degrees of freedom. *General Relativity and Gravitation*, 28:207–219, 1996.

[Fischer and Moncrief, 1997] Arthur Fischer and Vincent Moncrief. Hamiltonian reduction of Einstein's equations of general relativity. *Nuclear Physics B (Proceedings Supplement)*, 57:142–161, 1997.

[Fischer and Moncrief, Unpublished] Arthur Fischer and Vincent Moncrief. Nonautonomous dynamical systems and the phase portrait of the reduced einstein equations. Unpublished.

[Frauendiener and Sparling, 1992] Jörg Frauendiener and George Sparling. On the symplectic formalism for general relativity. *Proceedings of the Royal Society. London. Series A*, 436:141–153, 1992.

[Friedman, 1983] Michael Friedman. *Foundations of Spacetime Theories*. Princeton University Press, Princeton, 1983.

[Gerver, 2003] Joseph Gerver. Noncollision singularities: Do four bodies suffice? *Experimental Mathematics*, 12:187–198, 2003.

[Goldstein, 1953] Herbert Goldstein. *Classical Mechanics*. Addison–Wesley, Cambridge, MA, 1953.

[Gotay and Demaret, 1983] Mark Gotay and Jacques Demaret. Quantum cosmological singularities. *Physical Review D*, 28:2404–2413, 1983.

[Gotay and Isenberg, 1992] Mark Gotay and James Isenberg. La symplectification de la science. *La Gazette des Mathématiciens*, 54:59–79, 1992.

[Gotay and Nester, 1980] Mark Gotay and James Nester. Generalized constraint algorithm and special presymplectic manifolds. In Gerald Kaiser and Jerrold Marsden, editors, *Geometric Methods in Mathematical Physics*, pages 79–104. Springer-Verlag, Berlin, 1980.

[Gotay *et al.*, 1978] Mark Gotay, James Nester, and George Hinds. Presymplectic manifolds and the Dirac–Bergmann theory of constraints. *Journal of Mathematical Physics*, 19:2388–2399, 1978.

[Gotay, 1983] Mark Gotay. On the validity of Dirac's conjecture regarding first-class secondary constraints. *Journal of Physics A*, 16:L141–L145, 1983.

[Gotay, 2000] Mark Gotay. Obstructions to quantization. In *Mechanics: From Theory to Computation*, pages 171–216. Springer-Verlag, Berlin, 2000.

[Guillemin and Sternberg, 1990] Victor Guillemin and Shlomo Sternberg. *Variations on a Theme by Kepler*. American Mathematical Society, Providence, RI, 1990.

[Healey, 2002] Richard Healey. Can physics coherently deny the reality of time? In Craig Callender, editor, *Time, Reality, & Experience*, pages 293–316. Cambridge University Press, Cambridge, 2002.

[Healey, Unpublished] Richard Healey. Gauging what's real. Unpublished.

[Henneaux and Shepley, 1982] Marc Henneaux and Lawrence Shepley. Lagrangians for spherically symmetric potentials. *Journal of Mathematical Physics*, 11:2101–2107, 1982.

[Henneaux and Teitelboim, 1992] Marc Henneaux and Claudio Teitelboim. *Quantization of Gauge Systems*. Princeton University Press, Princeton, 1992.

[Hojman and Shepley, 1991] Sergio Hojman and Lawrence Shepley. No Lagrangian? No quantization! *Journal of Mathematical Physics*, 32:142–146, 1991.

[Iglesias, 1998] Patrick Iglesias. Les origines du calcul symplectique chez Lagrange. *L'Enseignement Mathématique*, 44:257–277, 1998.

[Isenberg and Marsden, 1982] James Isenberg and Jerrold Marsden. A slice theorem for the space of solutions of Einstein's equations. *Physics Reports*, 89:179–222, 1982.

[Isenberg and Rendall, 1998] James Isenberg and Alan Rendall. Cosmological spacetimes not covered by a constant mean curvature slicing. *Classical and Quantum Gravity*, 15:3679–3688, 1998.

[Isham, 1993] Chris Isham. Canonical quantum gravity and the problem on time. In L.A. Ibort and M.A. Rodriguez, editors, *Integrable Systems, Quantum Groups, and Quantum Field Theories*, pages 157–288. Kluwer, Dordrecht, 1993.

[Kay, 1980] Bernard Kay. Linear spin-zero quantum fields in external gravitational and scalar fields. II. Generally covariant perturbation theory. *Communications in Mathematical Physics*, 71:29–46, 1980.

[King and Ellis, 1973] A.R. King and G.F.R. Ellis. Tilted homogeneous cosmological models. *Communications in Mathematical Physics*, 31:209–242, 1973.

[Kondracki and Rogulski, 1986] Witold Kondracki and Jan Rogulski. On the stratification of the orbit space for the action of automorphisms on connections. *Dissertationes Mathematicae*, CCL:1–62, 1986.

[Kuchař, 1986] Karel Kuchař. Hamiltonian dynamics of gauge theories. *Physical Review D*, 34:3031–3043, 1986.

[Kuchař, 1992] Karel Kuchař. Time and interpretations of quantum gravity. In G. Kunstatter, D. Vincent, and J. Williams, editors, *Proceedings of the 4th Canadian Conference on General Relativity and Astrophysics*, pages 211–314. World Scientific, Singapore, 1992.

[Kuchař, 1993] Karel Kuchař. Canonical quantum gravity. In R. Gleiser, C. Kozameh, and O. Moreschi, editors, *General Relativity and Gravitation 1992*, pages 119–150. Institute of Physics Publishing, Philadelphia, 1993.

[Lanczos, 1986] Cornelius Lanczos. *Variational Principles of Mechanics*. Dover, New York, 4th edition, 1986.

[Landsman, 1998] N.P. Landsman. Quantization of singular systems and incomplete motions. In M. Rainer and H.-J. Schmidt, editors, *Current Topics in Mathematical Cosmology*, pages 256–263. World Scientific, Singapore, 1998.

[Landsman, this volume] N.P. Landsman. Between classical and quantum. Chapter 5, this volume.

[Lang, 1999] Serge Lang. *Fundamentals of Differential Geometry*. Springer-Verlag, Berlin, 1999.

[Lee and Wald, 1990] Joohan Lee and Robert Wald. Local symmetries and constraints. *Journal of Mathematical Physics*, 31:725–743, 1990.

[Lee, 2000] John Lee. *Introduction to Topological Manifolds*. Springer-Verlag, Berlin, 2000.

[Lee, 2003] John Lee. *Introduction to Smooth Manifolds*. Springer-Verlag, Berlin, 2003.

[Lewis, 1986] David Lewis. *Philosophical Papers*, volume II. Oxford University Press, Oxford, 1986.

[Lewis, 1999] David Lewis. Humean Supervenience debugged. In *Papers in Metaphysics and Epistemology*, pages 224–247. Cambridge University Press, Cambridge, 1999.

[Loll, 1992] Renate Loll. Canonical and BRST-quantization of constrained systems. In Mark Gotay, Jerrold Marsden, and Vincent Moncrief, editors, *Mathematical Aspects of Classical Field Theory*, pages 503–530. American Mathematical Society, Providence, RI, 1992.

[Malament, this volume] David Malament. Classical relativity theory. Chapter 3, this volume.

[Marco, 1990a] Jean-Pierre Marco. Chiurgie et régularisation des variétés de mouvements. *Comptes Rendus des Séances de l'Académie des Sciences*, 310:125–127, 1990.

[Marco, 1990b] Jean-Pierre Marco. Variétés de mouvements et régularisation des systèmes dynamiques. *Comptes Rendus des Séances de l'Académie des Sciences*, 310:65–67, 1990.

[Marsden and Ratiu, 1994] Jerrold Marsden and Tudor Ratiu. *Introduction to Mechanics and Symmetry*. Springer-Verlag, Berlin, 1994.

[Marsden and Weinstein, 1982] Jerrold Marsden and Alan Weinstein. The Hamiltonian structure of the Maxwell–Vlasov equations. *Physica D*, 4:394–406, 1982.

[Marsden *et al.*, 1972] Jerrold Marsden, David Ebin, and Arthur Fischer. Diffeomorphism groups, hydrodynamics and relativity. In Ray Vanstone, editor, *Proceedings of the Thirteenth Biennial Seminar of the Canadian Mathematical Congress on Differential Geometry and Applications*, volume 1, pages 135–279. Canadian Mathematical Congress, Montreal, 1972.

[Marsden, 1981] Jerrold Marsden. *Lectures on Geometric Methods in Mathematical Physics*. SIAM, Philadelphia, 1981.

[Maudlin, 2002] Tim Maudlin. Thoroughly muddled McTaggart, or how to abuse gauge freedom to generate metaphysical monstrosities. *Philosophers' Imprint*, 2(4):1–19, 2002.

[Maudlin, Unpublished] Tim Maudlin. Suggestions from physics for deep metaphysics. Unpublished.

[McGehee, 1975] Richard McGehee. Triple collisions in gravitational systems. In Jürgen Moser, editor, *Dynamical Systems, Theory and Application*, pages 550–572. Springer-Verlag, Berlin, 1975.

[Milnor, 1984] John Milnor. Remarks on infinite-dimensional Lie groups. In Bryce DeWitt and Raymond Stora, editors, *Relativity, Groups, and Topology II*, pages 1007–1057. North-Holland, Amsterdam, 1984.

[Moncrief and Demaret, 1980] Vincent Moncrief and Jacques Demaret. Hamiltonian formalism for perfect fluids in general relativity. *Physical Review D*, 21:2785–2793, 1980.

[Moncrief, 1977] Vincent Moncrief. Hamiltonian formalism for relativistic perfect fluids. *Physical Review D*, 16:1702–1705, 1977.

[Moncrief, 1989] Vincent Moncrief. Reduction of the Einstein equations in 2+1 dimensions to a Hamiltonian system over Teichmüller space. *Journal of Mathematical Physics*, 30:2907–2914, 1989.

[Moncrief, 1992] Vincent Moncrief. Boost symmetries in spatially compact spacetimes with a cosmological constant. *Classical and Quantum Gravity*, 9:2515–2520, 1992.

[Norton, 1995] John Norton. Did Einstein stumble: The debate over general covariance. *Erkenntnis*, 42:223–245, 1995.

[Nutku and Pavlov, 2002] Yavuz Nutku and Maxim Pavlov. Multi-Lagrangians for integrable systems. *Journal of Mathematical Physics*, 43:1441–1459, 2002.

[Olver, 1993] Peter Olver. *Applications of Lie Groups to Differential Equations*. Springer-Verlag, Berlin, 2nd edition, 1993.

[Ortega and Ratiu, 2004] Juan-Pablo Ortega and Tudor Ratiu. *Momentum Maps and Hamiltonian Reduction*. Birkhaüser, Basel, 2004.

[Pflaum, 2001] Markus Pflaum. *Analytic and Geometric Study of Stratified Spaces*. Springer-Verlag, Berlin, 2001.

[Prince, 2000] G.E. Prince. The inverse problem in the calculus of variations and its ramifications. In Peter Vassiliou and Ian Lisle, editors, *Geometric Approaches to Differential Equations*, pages 171–200. Cambridge University Press, Cambridge, 2000.

[Redhead, 2003] Michael Redhead. The interpretation of gauge symmetry. In Katherine Brading and Elena Castellani, editors, *Symmetries in Physics*, pages 124–139. Cambridge University Press, Cambridge, 2003.

[Reichenbach, 1991] Hans Reichenbach. *The Direction of Time*. University of California Press, Berkeley, 1991.

[Rendall, 1996] Alan Rendall. Constant mean curvature foliations in cosmological spacetimes. *Helvetica Physica Acta*, 69:490–500, 1996.

[Rovelli, 1991] Carlo Rovelli. Time in quantum gravity: A hypothesis. *Physical Review D*, 43:442–456, 1991.

[Rovelli, this volume] Carlo Rovelli. Quantum gravity. Chapter 12, this volume.

[Ruetsche, Unpublished] Laura Ruetsche. Quantizing. Unpublished.

[Rynasiewicz, 1999] Robert Rynasiewicz. Kretschmann's analysis of general covariance. In Hubert Goenner, Jürgen Renn, Jim Ritter, and Tilman Sauer, editors, *The Expanding Worlds of General Relativity*, pages 431–462. Birkhaüser, Basel, 1999.

[Saari and Xia, 1995] Donald Saari and Zhihong Xia. Off to infinity in finite time. *Notices of the American Mathematical Society*, 42:538–546, 1995.

[Sachs and Wu, 1977] Rainer Sachs and Hung-Hsi Wu. *General Relativity for Mathematicians*. Springer-Verlag, Berlin, 1977.

[Schmid, 1987] Rudolf Schmid. *Infinite Dimensional Hamiltonian Systems*. Bibliopolis, Naples, 1987.

[Singer, 2001] Stephanie Singer. *Symmetry in Mechanics: A Gentle, Modern Introduction*. Birkhaüser, Basel, 2001.

[Sorkin, 2002] Rafael Sorkin. An example relevant to the Kretschmann–Einstein debate. *Modern Physics Letters A*, 17:695–700, 2002.

[Souriau, 1982] Jean-Marie Souriau. Géométrie globale du problème à deux corps. *Atti della Accedemia delle Scienze di Torino. Classe di Scienze Fisiche, Matematiche e Naturali*, 117(Supplement 1):369–418, 1982.

[Souriau, 1986] Jean-Marie Souriau. La structure symplectique décrite par lagrange en 1811. *Mathématiques et Sciences Humaines*, 94:45–54, 1986.

[Souriau, 1997] Jean-Marie Souriau. *Structure of Dynamical Systems*. Birkhaüser, Basel, 1997.

[Sudbery, 1986] Anthony Sudbery. A vector Lagrangian for the electromagnetic field. *Journal of Physics A*, 19:L33–L36, 1986.

[Tarkovsky, 1991] Andrei Tarkovsky. *Time Within Time: The Diaries, 1970–1986*. Seagull, Calcutta, 1991.

[Torre and Anderson, 1996] Charles Torre and Ian Anderson. Classification of local generalized symmetries for the vacuum Einstein equations. *Communications in Mathematical Physics*, 176:479–539, 1996.

[Torre, 1992] Charles Torre. Covariant phase space formulation of parameterized field theories. *Journal of Mathematical Physics*, 33:3802–3811, 1992.

[Torre, Unpublished] Charles Torre. Local cohomology in field theory (with applications to the Einstein equations). Unpublished.

[Tuynman, 1992] Gijs Tuynman. What are the rules of the game called BRST? In Mark Gotay, Jerrold Marsden, and Vincent Moncrief, editors, *Mathematical Aspects of Classical Field Theory*, pages 625–633. American Mathematical Society, Providence, RI, 1992.

[van Brunt, 2004] Bruce van Brunt. *The Calculus of Variations*. Springer-Verlag, Berlin, 2004.

[Wald, 1984] Robert Wald. *General Relativity*. University of Chicago Press, Chicago, 1984.

[Wald, 1994] Robert Wald. *Quantum Field Theory on Curved Spacetime*. University of Chicago Press, Chicago, 1994.

[Wallace, 2003] David Wallace. Time-dependent symmetries: The link between gauge symmetries and indeterminism. In Katherine Brading and Elena Castellani, editors, *Symmetries in Physics*, pages 163–173. Cambridge University Press, Cambridge, 2003.

[Weinstein, 1981] Alan Weinstein. Symplectic geometry. *Bulletin of the American Mathematical Society*, 5:1–13, 1981.

[Woodhouse, 1991] N.M.J. Woodhouse. *Geometric Quantization*. Oxford University Press, Oxford, 2nd edition, 1991.

[Zuckerman, 1987] Gregg Zuckerman. Action principles and global geometry. In Shing-Tung Yau, editor, *Mathematical Aspects of String Theory*, pages 259–284. World Scientific, 1987.