Topics in the Foundations of General Relativity and Newtonian Gravitation Theory

−1
0
+1

CHICAGO LECTURES IN PHYSICS
Robert M. Wald, series editor
Henry J. Frisch, Gene R. Mazenko, and Sidney R. Nagel

Other *Chicago Lectures in Physics* titles available from the University of Chicago Press

Currents and Mesons, by J. J. Sakurai (1969)

Mathematical Physics, by Robert Geroch (1984)

Useful Optics, by Walter T. Welford (1991)

Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics, by Robert M. Wald (1994)

Geometrical Vectors, by Gabriel Weinreich (1998)

Electrodynamics, by Fulvio Melia (2001)

Perspectives in Computation, by Robert Geroch (2009)

Topics in the Foundations of General Relativity and Newtonian Gravitation Theory

David B. Malament

THE UNIVERSITY OF CHICAGO PRESS • CHICAGO AND LONDON

David B. Malament is professor in the Department of Logic and Philosophy of Science at the University of California, Irvine. He is the editor of *Reading Natural Philosophy: Essays in the History and Philosophy of Science and Mathematics*.

The University of Chicago Press, Chicago 60637
The University of Chicago Press, Ltd., London

Printed in the United States of America
21 20 19 18 17 16 15 14 13 12 1 2 3 4 5

ISBN-13: 978-0-226-50245-8 (cloth)
ISBN-10: 0-226-50245-7 (cloth)

Library of Congress Cataloging-in-Publication Data

Malament, David B.
Topics in the foundations of general relativity and Newtonian gravitation theory / David Malament.
p. cm.
Includes bibliographical references and index.
ISBN-13: 978-0-226-50245-8 (hardcover : alkaline paper)
ISBN-10: 0-226-50245-7 (hardcover : alkaline paper) 1. Relativity (Physics) 2. Gravitation. I. Title.
QC173.55.M353 2012
531'.14–dc23
2011035412

∞ This paper meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper).

To Pen

—–1
— 0
—+1

Contents

Preface

This manuscript began life as a set of lecture notes for a two-quarter (twenty-week) course on the foundations of general relativity that I taught at the University of Chicago many years ago. I have repeated the course quite a few times since then, both there and at the University of California, Irvine, and have over the years steadily revised the notes and added new material. Maybe now the notes can stand on their own.

The course was never intended to be a systematic survey of general relativity. There are many standard topics that I do not discuss—e.g., the Schwarzschild solution and the "classic tests" of general relativity. (And I have always recommended that students who have not already taken a more standard course in the subject do some additional reading on their own.) My goals instead have been to (i) present the basic logical-mathematical structure of the theory with some care, and (ii) consider additional special topics that seem to me, at least, of particular interest. The topics have varied from year to year, and not all have found their way into these notes. I will mention in advance three that did.

The first is "geometrized Newtonian gravitation theory," also known as "Newton-Cartan theory." It is now well known that one can, after the fact, reformulate Newtonian gravitation theory so that it exhibits many of the qualitative features that were once thought to be uniquely characteristic of general relativity. On reformulation, Newtonian theory too provides an account of four-dimensional spacetime structure in which (i) gravity emerges as a manifestation of spacetime curvature, and (ii) spacetime structure itself is "dynamical" in the sense that it participates in the unfolding of physics rather than being a fixed backdrop against which it unfolds. It has always seemed to me helpful to consider general relativity and this geometrized reformulation of Newtonian theory side by side. For one thing, one derives a sense of where Einstein's equation "comes from." When one reformulates the empty-space field equation of Newtonian gravitation theory (i.e., Laplace's equation

$\nabla^2\phi = 0$, where ϕ is the gravitational potential), one arrives at a constraint on the curvature of spacetime, namely $R_{ab} = \mathbf{0}$. The latter is, of course, just what we otherwise know as (the empty-space version of) Einstein's equation. And, reciprocally, this comparison of the two theories side by side provides a certain insight into Newtonian physics. For example, it yields a satisfying solution (or dissolution) to an old problem about Newtonian cosmology. Newtonian theory in a standard textbook formulation seems to provide no sensible prescription for what the gravitational field should be like in the presence of a uniform mass-distribution filling all of space. (See section 4.4.) But the problem is really just an artifact of the formulation, and it disappears when one passes to the geometrized version of the theory.

The basic idea of geometrized Newtonian gravitation theory is simple enough. But there are complications, and I deal with some of them in the present expanded form of the lecture notes. In particular, I present two different versions of the theory—what I call the "Trautman version" and the "Künzle-Ehlers version"—and consider their relation to one another. I also discuss in some detail the geometric significance of various conditions on the Riemann curvature field $R^a{}_{bcd}$ that enter into the formulation of these versions.

A second special topic that I consider is the concept of "rotation." It turns out to be a rather delicate and interesting question, at least in some cases, just what it *means* to say that a body is or is not rotating within the framework of general relativity. Moreover, the reasons for this—at least the ones I have in mind—do *not* have much to do with traditional controversy over "absolute vs. relative (or Machian)" conceptions of motion. Rather, they concern particular geometric complexities that arise when one allows for the possibility of spacetime curvature. The relevant distinction for my purposes is not that between attributions of "relative" and "absolute" rotation, but rather that between attributions of rotation that can and cannot be analyzed in terms of motion (in the limit) at a point. It is the latter—ones that make essential reference to extended regions of spacetime—that can be problematic.

The problem has two parts. First, one can easily think of different criteria for when an extended body is rotating. (I discuss two examples in section 3.2.) These criteria agree if the background spacetime structure is sufficiently simple—e.g., if one is working in Minkowski spacetime. But they do not agree in general. So, at the very least, attributions of rotation in general relativity can be ambiguous. A body can be rotating in one perfectly natural sense but not rotating in another, equally natural, sense. Second, circumstances can arise in which the different criteria—all of them—lead to determinations of rotation and non-rotation that seem wildly counterintuitive. (See section 3.3.) The upshot of this discussion is not that we cannot continue to talk about rotation in the context of general relativity. Not at all. Rather, we simply have to appreciate that it is a subtle and ambiguous notion that does not, in all cases, fully answer to our classical intuitions.

A third special topic that I consider is Gödel spacetime. It is not a live candidate for describing our universe, but it is of interest because of what it tells us about the possibilities allowed by general relativity. It represents a possible universe with remarkable properties. For one thing, the entire material content of the Gödel universe is in a state of uniform, rigid rotation (according to any reasonable criterion of rotation). For another, light rays and free test particles in it exhibit a kind of boomerang effect. Most striking of all, it admits closed timelike curves that cannot be "unrolled" by passing to a covering space (because the underlying manifold is simply connected). In section 3.1, I review these basic features of Gödel spacetime and, in an appendix to that section, I discuss how one can go back and forth between an intrinsic characterization of the Gödel metric and two different coordinate expressions for it.

These three special topics are treated in chapters 3 and 4. Much of this material has been added over the years. The original core of the lecture notes—the review of the basic structure of general relativity—is to be found in chapter 2.

Chapter 1 offers a preparatory review of basic differential geometry. It has never been my practice to work through all this material in class. I have limited myself there to "highlights" and general remarks. But I have always distributed the notes so that students with sufficient interest can do further reading on their own. On occasion, I have also run a separate "problem session" and used it for additional coaching on differential geometry. (A number of problems, with solutions, are included in the present version of the lecture notes.) I suggest that readers make use of chapter 1 as seems best to them—as a text to be read from the beginning, as a reference work to be consulted when particular topics arise in later chapters, as something in between, or not at all.

I would like to use this occasion to thank a number of people who have helped me over the years to learn and better understand general relativity. I could produce a long list, but the ones who come first, at least, are John Earman, David Garfinkle, Robert Geroch, Clark Glymour, Howard Stein, and Robert Wald. I am particularly grateful to Bob_1 and Bob_2 for allowing this interloper from the Philosophy Department to find a second home in the Chicago Relativity Group. Anyone familiar with their work, both research and expository writings, will recognize their influence on this set of lecture notes.

Erik Curiel, Sam Fletcher, David Garfinkle, John Manchak, and Jim Weatherall have my thanks, as well, for the comments and corrections they have given me on earlier drafts of the manuscript.

Matthias Kretschmann was good enough some years ago to take my handwritten notes on differential geometry and set them in TEX. I took over after that, but I might not have started without his push.

Finally, Pen Maddy has helped me to believe that this project was worth completing. I shall always be grateful to her for her support and encouragement.

1

DIFFERENTIAL GEOMETRY

1.1. Manifolds

We assume familiarity with the basic elements of multivariable calculus and point set topology. The following notions, in particular, should be familiar.

$\mathbb{R}^n$ (for $n \geq 1$) is the set of all n-tuples of real numbers $x = (x^1, \ldots, x^n)$. The Euclidean inner product (or "dot product") on $\mathbb{R}^n$ is given by $x \cdot y = x^1 y^1 + \ldots + x^n y^n$. It determines a norm, $\|x\| = \sqrt{x \cdot x}$. Given a point $x \in \mathbb{R}^n$ and a real number $\epsilon > 0$, $B_\epsilon(x)$ is the open ball in $\mathbb{R}^n$ centered at x with radius ϵ—i.e., $B_\epsilon(x) = \{y : \|y - x\| < \epsilon\}$. Clearly, x belongs to $B_\epsilon(x)$ for every $\epsilon > 0$. A subset S of $\mathbb{R}^n$ is *open* if, for all points x in S, there is an $\epsilon > 0$ such that $B_\epsilon(x) \subseteq S$. This determines a topology on $\mathbb{R}^n$. Given $m, n \geq 1$, and a map $f : O \to \mathbb{R}^m$ from an open set O in $\mathbb{R}^n$ to $\mathbb{R}^m$, f is *smooth* (or C^∞) if all its mixed partial derivatives (to all orders) exist and are continuous at every point in O.

A smooth n-dimensional manifold ($n \geq 1$) can be thought of as a point set to which has been added the "local smoothness structure" of $\mathbb{R}^n$. Our discussion of differential geometry begins with a more precise characterization.[1]

Let M be a non-empty set. An *n-chart* on M is a pair (U, φ) where U is a subset of M and $\varphi\colon U \to \mathbb{R}^n$ is an injective (i.e., one-to-one) map from U into $\mathbb{R}^n$ with the property that $\varphi[U]$ is an open subset of $\mathbb{R}^n$. (Here $\varphi[U]$ is the image set $\{\varphi(p) : p \in U\}$.) Charts, also called "coordinate patches," are the mechanism with which one induces local smoothness structure on the set M. To obtain a smooth n-dimensional manifold, we must lay down sufficiently many n-charts on M to cover the set and require that they be, in an appropriate sense, compatible with one another.

Let (U_1, φ_1) and (U_2, φ_2) be n-charts on M. We say the two are *compatible* if either the intersection set $U = U_1 \cap U_2$ is empty or the following conditions hold:

1. In this section and several others in chapter 1, we follow the basic lines of the presentation in Geroch [22].

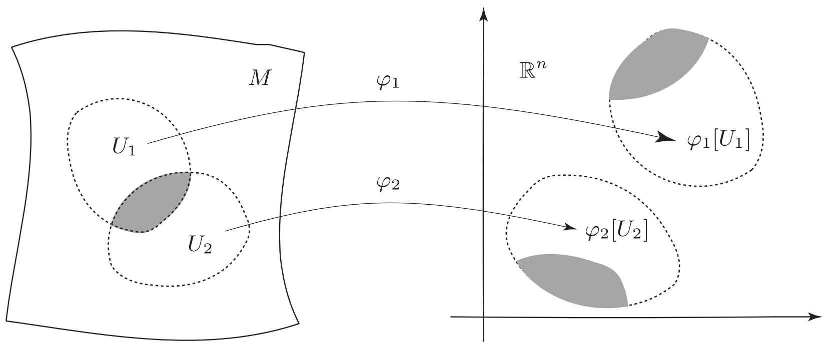

Figure 1.1.1. Two n-charts (U_1, φ_1) and (U_2, φ_2) on M with overlapping domains.

(1) $\varphi_1[U]$ and $\varphi_2[U]$ are both open subsets of $\mathbb{R}^n$.
(2) $\varphi_1 \circ \varphi_2^{-1} \colon \varphi_2[U] \to \mathbb{R}^n$ and $\varphi_2 \circ \varphi_1^{-1} \colon \varphi_1[U] \to \mathbb{R}^n$ are both smooth.

(Notice that the second makes sense since $\varphi_1[U]$ and $\varphi_2[U]$ are open subsets of $\mathbb{R}^n$ and we know what it means to say that a map from an open subset of $\mathbb{R}^n$ to $\mathbb{R}^n$ is smooth. See figure 1.1.1.)

The relation of compatibility between n-charts on a given set is reflexive and symmetric. But it need not be transitive and, hence, not an equivalence relation. For example, consider the following three 1—charts on $\mathbb{R}$:

$C_1 = (U_1, \varphi_1)$, with $U_1 = (-1, 1)$ and $\varphi_1(x) = x$
$C_2 = (U_2, \varphi_2)$, with $U_2 = (0, 1)$ and $\varphi_2(x) = x$
$C_3 = (U_3, \varphi_3)$, with $U_3 = (-1, 1)$ and $\varphi_3(x) = x^3$

Pairs C_1 and C_2 are compatible, and so are pairs C_2 and C_3. But C_1 and C_3 are not compatible, because the map $\varphi_1 \circ \varphi_3^{-1} \colon (-1, +1) \to \mathbb{R}$ is not smooth (or even just differentiable) at $x = 0$.

We now define a *smooth n-dimensional manifold* (or, in brief, an *n-manifold*) ($n \geq 1$) to be a pair $(M, \mathcal{C})$ where M is a non-empty set and $\mathcal{C}$ is a set of n-charts on M satisfying the following four conditions.

(M1) Any two n-charts in $\mathcal{C}$ are compatible.
(M2) The (domains of the) n-charts in $\mathcal{C}$ cover M; i.e., for every $p \in M$, there is an n-chart (U, φ) in $\mathcal{C}$ such that $p \in U$.
(M3) (*Hausdorff condition*) Given distinct points p_1 and p_2 in M, there exist n-charts (U_1, φ_1) and (U_2, φ_2) in $\mathcal{C}$ such that $p_i \in U_i$ for $i = 1, 2$ and $U_1 \cap U_2$ is empty.
(M4) $\mathcal{C}$ is *maximal* in the sense that any n-chart on M that is compatible with every n-chart in $\mathcal{C}$ belongs to $\mathcal{C}$.

(M1) and (M2) are certainly conditions one would expect. (M3) is included, following standard practice, simply to rule out pathological examples (though one does, sometimes, encounter discussions of "non-Hausdorff manifolds"). (M4) builds in the requirement that manifolds do not have "extra structure" in the form of distinguished n-charts. (For example, we can think of the point set $\mathbb{R}^n$ as carrying a single [global] n-chart. In the transition from the point set $\mathbb{R}^n$ to the n-manifold $\mathbb{R}^n$ discussed below, this "extra structure" is washed out.)

Because of (M4), it might seem a difficult task to specify an n-dimensional manifold. (How is one to get a grip on all the different n-charts that make up a maximal set of such?) But the following proposition shows that the specification need not be difficult. It suffices to come up with a set of n-charts on the underlying set satisfying (M1), (M2), and (M3), and then simply throw in wholesale all other compatible n-charts.

PROPOSITION 1.1.1. *Let M be a non-empty set, let $\mathcal{C}_0$ be a set of n-charts on M satisfying conditions* (*M*1), (*M*2), *and* (*M*3), *and let $\mathcal{C}$ be the set of all n-charts on M compatible with all the n-charts in $\mathcal{C}_0$. Then $(M, \mathcal{C})$ is an n-manifold; i.e., $\mathcal{C}$ satisfies all four conditions.*

Proof. Since $\mathcal{C}_0$ satisfies (M1), $\mathcal{C}_0$ is a subset of $\mathcal{C}$. It follows immediately that $\mathcal{C}$ satisfies (M2), (M3), and (M4). Only (M1) requires some argument. Let $C_1 = (U_1, \varphi_1)$ and $C_2 = (U_2, \varphi_2)$ be any two n-charts compatible with all n-charts in $\mathcal{C}_0$. We show that they are compatible with one another. We may assume that the intersection $U_1 \cap U_2$ is non-empty, since otherwise compatibility is automatic.

First we show that $\varphi_1[U_1 \cap U_2]$ is open. (A parallel argument establishes that $\varphi_2[U_1 \cap U_2]$ is open.) Consider an arbitrary point of $\varphi_1[U_1 \cap U_2]$. It is of the form $\varphi_1(p)$ for some point $p \in U_1 \cap U_2$. Since $\mathcal{C}_0$ satisfies (M2), there exists an n-chart $C = (U, \varphi)$ in $\mathcal{C}_0$ whose domain contains p. So $p \in U \cap U_1 \cap U_2$. Since C is compatible with both C_1 and C_2, $\varphi[U \cap U_1]$ and $\varphi[U \cap U_2]$ are open sets in $\mathbb{R}^n$, and the maps

$$\varphi_1 \circ \varphi^{-1} \colon \varphi[U \cap U_1] \to \mathbb{R}^n, \qquad \varphi_2 \circ \varphi^{-1} \colon \varphi[U \cap U_2] \to \mathbb{R}^n,$$
$$\varphi \circ \varphi_1^{-1} \colon \varphi_1[U \cap U_1] \to \mathbb{R}^n, \qquad \varphi \circ \varphi_2^{-1} \colon \varphi_2[U \cap U_2] \to \mathbb{R}^n,$$

are all smooth (and therefore continuous). Now $\varphi[U \cap U_1 \cap U_2]$ is open, since it is the intersection of open sets $\varphi[U \cap U_1]$ and $\varphi[U \cap U_2]$. (Here we use the fact that φ is injective.) So $\varphi_1[U \cap U_1 \cap U_2]$ is open, since it is the preimage of $\varphi[U \cap U_1 \cap U_2]$ under the continuous map $\varphi \circ \varphi_1^{-1}$. But, clearly, $\varphi_1(p) \in \varphi_1[U \cap U_1 \cap U_2]$, and $\varphi_1[U \cap U_1 \cap U_2]$ is a subset of $\varphi_1[U_1 \cap U_2]$. So we see that our arbitrary point $\varphi_1(p)$ in $\varphi_1[U_1 \cap U_2]$ is contained in an open subset of $\varphi_1[U_1 \cap U_2]$. Thus $\varphi_1[U_1 \cap U_2]$ is open.

Next we show that the map $\varphi_2 \circ \varphi_1^{-1} \colon \varphi_1[U_1 \cap U_2] \to \mathbb{R}^n$ is smooth. (A parallel argument establishes that $\varphi_1 \circ \varphi_2^{-1} \colon \varphi_2[U_1 \cap U_2] \to \mathbb{R}^n$ is smooth.) For this it suffices to show that, given our arbitrary point $\varphi_1(p)$ in $\varphi_1[U_1 \cap U_2]$, the restriction of $\varphi_2 \circ \varphi_1^{-1}$ to some open subset of $\varphi_1[U_1 \cap U_2]$ containing $\varphi_1(p)$ is smooth. But this follows easily. We know that $\varphi_1[U \cap U_1 \cap U_2]$ is an open subset of $\varphi_1[U_1 \cap U_2]$ containing $\varphi_1(p)$. And the restriction of $\varphi_2 \circ \varphi_1^{-1}$ to $\varphi_1[U \cap U_1 \cap U_2]$ is smooth, since it can be realized as the composition of $\varphi \circ \varphi_1^{-1}$ (restricted to $\varphi_1[U \cap U_1 \cap U_2]$) with $\varphi_2 \circ \varphi^{-1}$ (restricted to $\varphi[U \cap U_1 \cap U_2]$), and both these maps are smooth. $\square$

Our definition of manifolds is less restrictive than some in that we do not include the following condition.

(M5) (*Countable cover condition*) There is a countable subset $\{(U_n, \varphi_n) \colon n \in \mathbb{N}\}$ of $\mathcal{C}$ whose domains cover M; i.e., for all p in M, there is an n such that $p \in U_n$.

In fact, all the manifolds that one encounters in relativity theory satisfy (M5). But there is some advantage in not taking the condition for granted from the start. It is simply not needed for our work until we discuss derivative operators—i.e., affine connections—on manifolds in section 1.7. It turns out that (M5) is actually a necessary and sufficient condition for there to *exist* a derivative operator on a manifold (given our characterization). It is also a necessary and sufficient condition for there to *exist* a (positive definite) Riemannian metric on a manifold. (See Geroch [23]. The paper gives a nice example of a 2-manifold that violates [M5].)

Our way of defining n-manifolds is also slightly non-standard because we jump directly from the point set M to the manifold $(M, \mathcal{C})$. In contrast, one often proceeds in two stages. One first puts a topology $\mathcal{T}$ on M, forming a topological space $(M, \mathcal{T})$. Then one adds the set of n-charts $\mathcal{C}$ to form the "manifold" $\big((M, \mathcal{T}), \mathcal{C}\big)$. If one proceeds this way, one must require of every n-chart (U, φ) in $\mathcal{C}$ that U be open—i.e., that U belong to $\mathcal{T}$, so that $\varphi \colon U \to \mathbb{R}^n$ qualifies as continuous.

Given our characterization of an n-manifold $(M, \mathcal{C})$, we do not (yet) know what it means for a subset of M to be "open." But there is a natural way to use the n-charts in $\mathcal{C}$ to define a topology on M. We say that a subset S of M is *open* if, for all p in S, there is an n-chart (U, φ) in $\mathcal{C}$ such that $p \in U$ and $U \subseteq S$. (This topology can also be characterized as the coarsest topology on M with respect to which, for all n-charts (U, φ) in $\mathcal{C}$, $\varphi : U \to \mathbb{R}^n$ is continuous. See problem 1.1.3). It follows immediately that the domain of every n-chart is open.

PROBLEM 1.1.1. *Let $(M, \mathcal{C})$ be an n-manifold, let (U, φ) be an n-chart in $\mathcal{C}$, let $\widehat{O}$ be an open subset of $\varphi[U]$, and let O be its pre-image $\varphi^{-1}[\widehat{O}]$. (So, $O \subseteq U$.) Show that $(O, \varphi|_O)$, the restriction of (U, φ) to O, is also an n-chart in $\mathcal{C}$.*

PROBLEM 1.1.2. *Let $(M, \mathcal{C})$ be an n-manifold, let (U, φ) be an n-chart in $\mathcal{C}$, and let O be an open set in M such that $U \cap O \neq \varnothing$. Show that $\big(U \cap O, \varphi|_{U \cap O}\big)$, the restriction of (U, φ) to $U \cap O$, is also an n-chart in $\mathcal{C}$. (Hint: Make use of the result in problem 1.1.1. Strictly speaking, by the way, we do not need to assume that $U \cap O$ is non-empty. But that is the only case of interest.)*

PROBLEM 1.1.3. *Let $(M, \mathcal{C})$ be an n-manifold and let $\mathcal{T}$ be the set of open subsets of M. (i) Show that $\mathcal{T}$ is, in fact, a topology on M, i.e., it contains the empty set and the set M, and is closed under finite intersections and arbitrary unions. (ii) Show that $\mathcal{T}$ is the coarsest topology on M with respect to which $\varphi : U \to \mathbb{R}^n$ is continuous for all n-charts (U, φ) in $\mathcal{C}$.*

Now we consider a few examples of manifolds. Let M be $\mathbb{R}^n$, the set of all ordered n-tuples of real numbers. Let U be any subset of M that is open (in the standard topology on $\mathbb{R}^n$), and let $\varphi\colon U \to \mathbb{R}^n$ be the identity map. Then (U, φ) qualifies as an n-chart on M. Let $\mathcal{C}_0$ be the set of all n-charts on M of this very special form. It is easy to check that $\mathcal{C}_0$ satisfies conditions (M1), (M2), and (M3). If we take $\mathcal{C}$ to be the set of all n-charts on M compatible with all n-charts in $\mathcal{C}_0$, then it follows (by proposition 1.1.1) that $(M, \mathcal{C})$ is an n-manifold. We refer to it as "the manifold $\mathbb{R}^n$." (Thus, one must distinguish among the point set $\mathbb{R}^n$, the vector space $\mathbb{R}^n$, the manifold $\mathbb{R}^n$, and so forth.)

Next we introduce the manifold S^n. The underlying set M is the set of points $x = (x^1, \dots, x^{n+1}) \in \mathbb{R}^{n+1}$ such that $\|x\| = 1$. For each $i = 1, \dots, n+1$, we set

$$U_i^+ = \big\{(x^1, \dots, x^i, \dots, x^{n+1}) \in M : x^i > 0\big\},$$
$$U_i^- = \big\{(x^1, \dots, x^i, \dots, x^{n+1}) \in M : x^i < 0\big\},$$

and define maps $\varphi_i^+ \colon U_i^+ \to \mathbb{R}^n$ and $\varphi_i^- \colon U_i^- \to \mathbb{R}^n$ by setting

$$\varphi_i^+\big(x^1, \ldots, x^{n+1}\big) = \big(x^1, \ldots, x^{i-1}, x^{i+1}, \ldots, x^{n+1}\big) = \varphi_i^-\big(x^1, \ldots, x^{n+1}\big).$$

$\big(U_i^+$ and U_i^- are upper and lower hemispheres with respect to the x^i coordinate axis; φ_i^+ and φ_i^- are projections that erase the i^{th} coordinate of $\big(x^1, \ldots, x^{n+1}\big).\big)$ The $(n+1)$ pairs of the form (U_i^+, φ_i^+) and (U_i^-, φ_i^-) are n-charts on M. The set $\mathcal{C}_1$ of all such pairs satisfies conditions (M1) and (M2). For all $p \in M$ and all $\epsilon > 0$, if $B_\epsilon(p) \cap M$ is a subset of U_i^+ (respectively U_i^-), we now add to $\mathcal{C}_1$ the n-chart that results from restricting (U_i^+, φ_i^+) (respectively (U_i^-, φ_i^-)) to $B_\epsilon(p) \cap M$. The expanded set of n-charts $\mathcal{C}_2$ satisfies (M1), (M2), and (M3). If, finally, we add to $\mathcal{C}_2$ all n-charts on M compatible with all n-charts in $\mathcal{C}_2$, we obtain the n-manifold S^n.

We thus have the manifolds $\mathbb{R}^n$ and S^n for every $n \geq 1$. From these we can generate many more manifolds by taking products and cutting holes.

Let $\mathcal{M}_1 = (M_1, \mathcal{C}_1)$ be an n_1-manifold and let $\mathcal{M}_2 = (M_2, \mathcal{C}_2)$ be an n_2–manifold. The *product manifold* $\mathcal{M}_1 \times \mathcal{M}_2$ is an $(n_1 + n_2)$–manifold defined as follows. The underlying point set is just the Cartesian product $M_1 \times M_2$—i.e., the set of all pairs (p_1, p_2) where $p_i \in M_i$ for $i = 1, 2$. Let (U_1, φ_1) be an n_1-chart in $\mathcal{C}_1$ and let (U_2, φ_2) be an n_2-chart in $\mathcal{C}_2$. We associate with them a set U and a map $\varphi \colon U \to \mathbb{R}^{(n_1+n_2)}$. We take U to be the product $U_1 \times U_2$; and given $(p_1, p_2) \in U$, we take $\varphi\big((p_1, p_2)\big)$ to be $(y^1, \ldots, y^{n_1}, z^1, \ldots, z^{n_2})$, where $\varphi_1(p_1) = (y^1, \ldots, y^{n_1})$ and $\varphi_2(p_2) = (z^1, \ldots, z^{n_2})$. So defined, (U, φ) qualifies as an $(n_1 + n_2)$–chart on $M_1 \times M_2$. The set of all $(n_1 + n_2)$–charts on $M_1 \times M_2$ obtained in this manner satisfies conditions (M1), (M2), and (M3). If we now throw in all n-charts on $M_1 \times M_2$ that are compatible with all members of this set, we obtain the manifold $\mathcal{M}_1 \times \mathcal{M}_2$. Using this product construction, we generate the 2−manifold $\mathbb{R}^1 \times S^1$ (the "cylinder"), the 2−manifold $S^1 \times S^1$ (the "torus"), and so forth.

Next, let $(M, \mathcal{C})$ be an n-manifold, and let S be a closed proper subset of M. (So M–S is a non-empty open subset of M.) Further, let $\mathcal{C}'$ be the set of all n-charts (U, φ) in $\mathcal{C}$ where $U \subseteq (M-S)$. Then the pair $(M-S, \mathcal{C}')$ is an n-manifold in its own right. (This follows as a corollary to the assertion in problem 1.1.2.)

A large fraction of the manifolds one encounters in relativity theory can be obtained from the manifolds $\mathbb{R}^n$ and S^n by taking products and excising closed sets.

We now define "smooth maps" between manifolds. We do so in two stages. First, we consider the special case in which the second manifold (i.e., the one into which the first is mapped) is $\mathbb{R}$. Then we consider the general case. Let $(M, \mathcal{C})$ be an n-manifold. We say that a map $\alpha \colon M \to \mathbb{R}$ is *smooth* (or C^∞)

if, for all n-charts (U, φ) in $\mathcal{C}$, $\alpha \circ \varphi^{-1} : \varphi[U] \to \mathbb{R}$ is smooth. (Here we use a standard technique. To define something on an n-manifold, we use the charts to pull things back to the context of $\mathbb{R}^n$ where the notion already makes sense.) Next let $(M', \mathcal{C}')$ be an m-manifold (with no requirement that $m = n$). We say that a map $\psi : M \to M'$ is *smooth* (or C^∞) if, for all smooth maps $\alpha : M' \to \mathbb{R}$ on the second manifold, the composed map $\alpha \circ \psi : M \to \mathbb{R}$ is smooth. One can check that the second definition is compatible with the first (see problem 1.1.4), and with the standard definition of smoothness that applies specifically to maps of the form $\psi : \mathbb{R}^n \to \mathbb{R}^m$.

PROBLEM 1.1.4. *Let $(M, \mathcal{C})$ be an n-manifold. Show that a map $\alpha : M \to \mathbb{R}$ is smooth according to our first definition (which applies only to real-valued maps on manifolds) iff it is smooth according to our second definition (which applies to maps between arbitrary manifolds).*

Let $(M, \mathcal{C})$ and $(M', \mathcal{C}')$ be manifolds. The definition of smoothness just given naturally extends to maps of the form $\psi : O \to M'$ where O is an open subset of M (that need not be all of M). It does so because we can always think of O as a manifold in its own right when paired with the charts it inherits from $\mathcal{C}$—i.e., the charts in $\mathcal{C}$ whose domains are subsets of O. On this understanding it follows, for example, that if a map $\psi : M \to M'$ is smooth, then its restriction to O is smooth. It also follows that given any chart (U, φ) in $\mathcal{C}$, the maps $\varphi : U \to \mathbb{R}^n$ and $\varphi^{-1} : \varphi[U] \to M$ are both smooth.

The point mentioned in the preceding paragraph will come up repeatedly. We shall often formulate definitions in terms of structures defined on manifolds and then transfer them without comment to open subsets of manifolds. It should be understood in each case that we have in mind the manifold structure induced on those open sets.

Given manifolds $(M, \mathcal{C})$ and $(M', \mathcal{C}')$, a bijection $\psi : M \to M'$ is said to be a *diffeomorphism* if both ψ and ψ^{-1} are smooth. Two manifolds are said to be *diffeomorphic*, of course, if there exists a diffeomorphism between them—i.e., between their underlying point sets. Diffeomorphic manifolds are as alike as they can be with respect to their "structure." They can differ only in the identity of their underlying elements.

1.2. Tangent Vectors

Let $(M, \mathcal{C})$ be an n-manifold and let p be a point in M. In this section, we introduce the notion of a "vector" (or "tangent vector" or "contravariant vector") at p.

We also show that the set of all vectors at p naturally forms an n-dimensional vector space.

Consider first the familiar case of $\mathbb{R}^n$. A vector ξ at a point in $\mathbb{R}^n$ can be characterized by its components $(\xi^1, \ldots, \xi^n)$ with respect to the n coordinate axes. This characterization is not available for arbitrary n-manifolds where no coordinate curves are distinguished. But an alternate, equivalent characterization does lend itself to generalization.

Let p be a point in $\mathbb{R}^n$. We take $\mathcal{S}(p)$ to be the set of all smooth maps $f: O \to \mathbb{R}$, where O is some open subset (or other) of $\mathbb{R}^n$ that contains p. If $f_1: O_1 \to \mathbb{R}$ and $f_2: O_2 \to \mathbb{R}$ are both in $\mathcal{S}(p)$, then we can define new maps $(f_1 + f_2): O_1 \cap O_2 \to \mathbb{R}$ and $(f_1\, f_2): O_1 \cap O_2 \to \mathbb{R}$ in $\mathcal{S}(p)$ by setting $(f_1 + f_2)(q) = f_1(q) + f_2(q)$ and $(f_1\, f_2)(q) = f_1(q) f_2(q)$ for all points q in $O_1 \cap O_2$.

Now suppose that ξ is a vector at p in $\mathbb{R}^n$ with components $(\xi^1, \ldots, \xi^n)$ and that f is in $\mathcal{S}(p)$. The *directional derivative* of f at p in the direction ξ is defined by

$$\xi(f) = \xi \cdot (\nabla f)_{|p} = \sum_{i=1}^{n} \xi^i \frac{\partial f}{\partial x^i}(p). \tag{1.2.1}$$

It follows immediately from the elementary properties of partial derivatives that, for all f_1 and f_2 in $\mathcal{S}(p)$,

(DD1) $\xi(f_1 + f_2) = \xi(f_1) + \xi(f_2)$.
(DD2) $\xi(f_1 f_2) = f_1(p)\, \xi(f_2) + f_2(p)\, \xi(f_1)$.
(DD3) If f_1 is constant, $\xi(f_1) = 0$.

Any map from $\mathcal{S}(p)$ to $\mathbb{R}$ satisfying these three conditions will be called a *derivation* (or *directional derivative operator*) at p. Thus, every vector at p defines, via equation (1.2.1), a derivation at p. Indeed, we shall see in a moment that equation (1.2.1) defines a bijection between vectors at p (understood as ordered n-tuples of reals) and derivations at p. This will give us our desired alternate characterization of vectors in $\mathbb{R}^n$. But first we need a lemma.

LEMMA 1.2.1. *Let $f_1: O_1 \to \mathbb{R}$ and $f_2: O_2 \to \mathbb{R}$ be elements of $\mathcal{S}(p)$ that agree on some open set $O \subseteq O_1 \cap O_2$ containing p. Then, for all derivations ξ at p, $\xi(f_1) = \xi(f_2)$.*

Proof. Let $h: O \to \mathbb{R}$ be the constant map on O that assigns 1 to all points. Certainly h is in $\mathcal{S}(p)$. The maps $h f_1$ and $h f_2$ have domain O and agree throughout

O; i.e., $hf_1 = hf_2$. So $\xi(hf_1) = \xi(hf_2)$. But by (DD2) and (DD3),

$$\xi(hf_1) = h(p)\xi(f_1) + f_1(p)\xi(h) = 1\,\xi(f_1) + f_1(p)\,0 = \xi(f_1).$$

Similarly, $\xi(hf_2) = \xi(f_2)$. So $\xi(f_1) = \xi(f_2)$, as claimed. □

PROPOSITION 1.2.2. *Equation (1.2.1) defines a bijection between vectors at p and derivations at p.*

Proof. Suppose first that $\xi = (\xi^1, \ldots, \xi^n)$ and $\eta = (\eta^1, \ldots, \eta^n)$ are vectors at p that, via equation (1.2.1), determine the same derivation at p. Then $\xi \cdot (\nabla f)_{|p} = \eta \cdot (\nabla f)_{|p}$, for all f in $\mathcal{S}(p)$. Consider the special case where f is the coordinate map $x^i : \mathbb{R}^n \to \mathbb{R}$ that assigns to a point in $\mathbb{R}^n$ its i^{th} coordinate. We have

$$(\nabla x^i)_{|p} = \left(\frac{\partial x^i}{\partial x^1}, \ldots, \frac{\partial x^i}{\partial x^i}, \ldots, \frac{\partial x^i}{\partial x^n} \right)_{|p} = (0, \ldots, 0, 1, 0, \ldots, 0),$$

where the sole 1 in the far right n-tuple is in the i^{th} position. So $\xi^i = \xi \cdot (\nabla x^i)_{|p} = \eta \cdot (\nabla x^i)_{|p} = \eta^i$. But this is true for all $i = 1, \ldots, n$. Hence $\xi = \eta$. Thus, the map from vectors at p to derivations at p determined by equation (1.2.1) is injective.

Next, suppose that ξ is a derivation at p and that the numbers $\xi^1, \ldots, \xi^n$ are defined by $\xi^i = \xi(x^i)$. We show that, for all f in $\mathcal{S}(p)$, $\xi(f) = \displaystyle\sum_{i=1}^{n} \xi^i \frac{\partial f}{\partial x^i}(p)$. That is, we show that ξ can be realized as the image of $(\xi^1, \ldots, \xi^n)$ under the map determined by equation (1.2.1). This will establish that the map is also surjective.

Let $f : O \to \mathbb{R}$ be a map in $\mathcal{S}(p)$. By the preceding lemma, we may assume that O is an open ball centered at p. (If f' is the restriction of f to an open ball centered at p, $\xi(f') = \xi(f)$. So we lose nothing by working with f' rather than f.) If x is a point in O, it follows by the fundamental theorem of calculus that

$$f(x) = f(p) + \int_0^1 \frac{d}{dt} f(p + t(x - p))\, dt.$$

(We want the domain of f to be an open ball centered at p to insure that f is defined at all points on the line segment connecting p and x.) By the "chain rule,"

$$\frac{d}{dt} f(p + t(x - p)) = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x^i}(p + t(x - p)) \right) (x^i - p^i).$$

Inserting the right side of this equation into the integrand above, we arrive at

$$f(x) = f(p) + \sum_{i=1}^{n} g_i(x)\,(x^i - p^i), \tag{1.2.2}$$

where, for all i, the map $g_i : O \to \mathbb{R}$ is given by $g_i(x) = \int_0^1 \frac{\partial f}{\partial x^i}(p + t(x-p))\,dt$. The g_i belong to $\mathcal{S}(p)$. It now follows from (DD1), (DD2), and (DD3) that

$$\xi(f) = \sum_{i=1}^{n} \left[g_i(p)\,\xi(x^i - p^i) + \Big((x^i - p^i)(p) \Big)\,\xi(g_i) \right].$$

(Here we are construing the numbers $f(p)$ and $p^1, \ldots, p^n$ as constant functions on O.) But $(x^i - p^i)(p) = p^i - p^i = 0$, and $\xi(x^i - p^i) = \xi(x^i) - \xi(p^i) = \xi^i - 0 = \xi^i$. So we have

$$\xi(f) = \sum_{i=1}^{n} \xi^i\, g_i(p).$$

But it follows from equation (1.2.2) that $\frac{\partial f}{\partial x^i}(p) = g_i(p)$. So $\xi(f) = \sum_{i=1}^{n} \xi^i\, \frac{\partial f}{\partial x^i}(p)$, as claimed. □

With proposition 1.2.2 as motivation, we now give our definition of "vectors" at points of manifolds. Given a manifold $(M, \mathcal{C})$ and a point p in M, let $\mathcal{S}(p)$ be the set of smooth maps $f : O \to \mathbb{R}$ where O is some open subset (or other) of M that contains p. (Our prior remark about adding and multiplying elements of $\mathcal{S}(p)$ carries over intact.) We take a *vector* (or *tangent vector*, or *contravariant vector*) at p to be a map from $\mathcal{S}(p)$ to $\mathbb{R}$ that satisfies (DD1), (DD2), and (DD3).

The set of all vectors at p has a natural vector space structure (over $\mathbb{R}$). If ξ and η are vectors at p, and k is a real number, we can define new vectors $\xi + \eta$ and $k\xi$ by setting

$$(\xi + \eta)(f) = \xi(f) + \eta(f)$$

and

$$(k\xi)(f) = k\,\xi(f)$$

for all f in $\mathcal{S}(p)$. The vector space M_p so defined is called the *tangent space* to p. We shall soon show that M_p has dimension n; i.e., it has the same dimension as $(M, \mathcal{C})$. To do so, we give a second characterization of vectors on manifolds that is of independent interest.

Let $\gamma : I \to M$ be a smooth curve in M—i.e., a smooth map from an open interval $I \subseteq \mathbb{R}$ into M. (I is of the form (a, b), $(-\infty, b)$, $(a, +\infty)$, or $(-\infty, +\infty)$, where a and b are real numbers. We know what it means to say that $\gamma : I \to M$ is smooth since, as noted toward the end of section 1.1, we can think of I as a manifold in its own right when paired with the charts it inherits from the manifold $\mathbb{R}$.) Suppose $s_0 \in I$ and $\gamma(s_0) = p$. We associate with γ a vector $\overrightarrow{\gamma}_p$ at p by setting $\overrightarrow{\gamma}_p(f) = \frac{d}{ds}(f \circ \gamma)(s_0)$ for all f in $\mathcal{S}(p)$. (This definition makes sense since $(f \circ \gamma)$ is a smooth map from I into $\mathbb{R}$.) It is easy to check that $\overrightarrow{\gamma}_p$, so defined, satisfies (DD1) – (DD3). For example, (DD2) holds for all f_1 and f_2 in $\mathcal{S}(p)$ since

$$\begin{aligned}\overrightarrow{\gamma}_p(f_1 f_2) &= \left(\frac{d}{ds}((f_1 f_2) \circ \gamma)\right)(s_0) = \left(\frac{d}{ds}((f_1 \circ \gamma)\,(f_2 \circ \gamma))\right)(s_0)\\ &= (f_1 \circ \gamma)(s_0)\left(\frac{d}{ds}(f_2 \circ \gamma)\right)(s_0) + (f_2 \circ \gamma)(s_0)\left(\frac{d}{ds}(f_1 \circ \gamma)\right)(s_0)\\ &= f_1(p)\ \overrightarrow{\gamma}_p(f_2) + f_2(p)\ \overrightarrow{\gamma}_p(f_1).\end{aligned}$$

$\overrightarrow{\gamma}_p$ is called the *tangent vector* to γ at p.

Suppose now that (U, φ) is an n-chart in our n-manifold $(M, \mathcal{C})$. Associated with (U, φ) are coordinate maps $u^i \colon U \to \mathbb{R}$ for $i = 1, \ldots, n$ defined by $u^i(q) = (x^i \circ \varphi)(q)$. (Thus, the number that u^i assigns to a point q in M is the one that x^i assigns to the image point $\varphi(q)$ in $\mathbb{R}^n$. Equivalently, $u^i(q)$ is the i^{th} coordinate of $\varphi(q)$. So $\varphi(q) = (u^1(q), \ldots, u^n(q))$.)

Now let p be a point in U. We understand the i^{th} *coordinate curve* through $\varphi(p) = (u^1(p), \ldots, u^n(p))$ in $\mathbb{R}^n$ to be the map from $\mathbb{R}$ to $\mathbb{R}^n$ given by

$$s \mapsto \left(u^1(p), \ldots, u^{i-1}(p), u^i(p) + s, u^{i+1}(p), \ldots, u^n(p)\right). \tag{1.2.3}$$

The image of the curve is a line through $\varphi(p)$, parallel to the i^{th} coordinate axis through the origin (see figure 1.2.1). We can pull this curve back to U via φ^{-1}

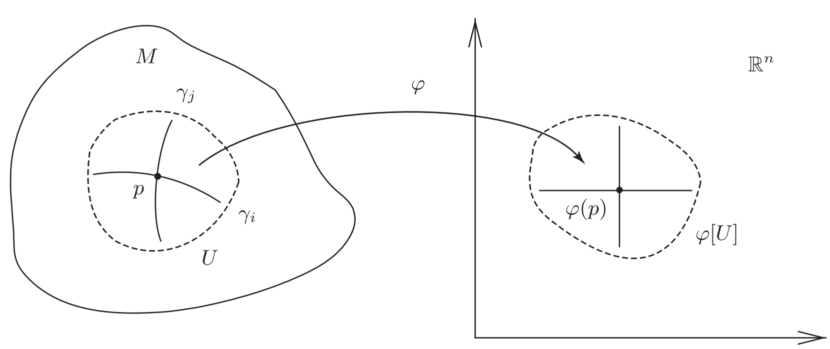

Figure 1.2.1. Coordinate curves on M with respect to (U, φ).

to obtain a smooth curve $\gamma_i : I \to U$ through p:

$$\gamma_i(s) = \varphi^{-1}\big(u^1(p), \ldots, u^{i-1}(p), u^i(p) + s, u^{i+1}(p), \ldots, u^n(p)\big). \tag{1.2.4}$$

Note that $\gamma_i(0) = p$. (We can afford to be vague about the domain I of γ_i since we are interested only in the tangent to the curve at p. All that matters is that $0 \in I$. How do we know that γ_i is smooth? This follows because φ^{-1} is smooth, and so γ_i is the composition of two smooth maps.) Extending our previous usage, we now refer to γ_i as the i^{th} *coordinate curve* through p with respect to (U, φ). (Note that coordinate curves through points in $\mathbb{R}^n$ are defined outright, but coordinate curves through points in M are necessarily relativized to n-charts.) This curve has a tangent $\overrightarrow{\gamma}_{i\,|p}$ at p. By the chain rule,

$$\overrightarrow{\gamma}_{i\,|p}(f) = \frac{d}{ds}\big(f \circ \gamma_i\big)(0) = \left(\frac{\partial\big(f \circ \varphi^{-1}\big)}{\partial x^i}\right)\big(\varphi(p)\big) \tag{1.2.5}$$

for all f in $\mathcal{S}(p)$. We note for future reference that, in particular, since $u^j = x^j \circ \varphi$,

$$\overrightarrow{\gamma}_{i\,|p}(u^j) = \left(\frac{\partial x^j}{\partial x^i}\right)\big(\varphi(p)\big) = \delta_{ij}. \tag{1.2.6}$$

(Here δ_{ij} is the Kronecker delta function that is 1 if $i = j$, and 0 otherwise.) Sometimes the tangent vector $\overrightarrow{\gamma}_i$ is written as $\dfrac{\partial}{\partial u^i}$ and $\overrightarrow{\gamma}_i(f)$ is written as $\dfrac{\partial f}{\partial u^i}$. Using this notation, and suppressing the point of evaluation p, equations (1.2.5) and (1.2.6) come out as

$$\frac{\partial f}{\partial u^i} = \frac{\partial(f \circ \varphi^{-1})}{\partial x^i} \tag{1.2.7}$$

and

$$\frac{\partial u^j}{\partial u^i} = \delta_{ij}. \tag{1.2.8}$$

Using the tangent vectors $\overrightarrow{\gamma}_{i\,|p}, i = 1, \ldots, n$, we can show that M_p is n-dimensional.

PROPOSITION 1.2.3. *Let $(M, \mathcal{C})$ be an n-manifold, let (U, φ) be an n-chart in $\mathcal{C}$, let p be a point in U, and let $\gamma_1, \ldots, \gamma_n$ be the n coordinate curves through p with respect to (U, φ). Then their tangent vectors $\overrightarrow{\gamma}_{1\,|p}, \ldots, \overrightarrow{\gamma}_{n\,|p}$ at p form a basis for M_p.*

Proof. First we show that the vectors are linearly independent. Let $a_1, \ldots, a_n$ be real numbers such that $\sum_{i=1}^{n} a_i \overrightarrow{\gamma}_{i\,|p} = \mathbf{0}$. We must show that the a_i are all 0. Now for all f in $\mathcal{S}(p)$, we have

$$0 = \left(\sum_{i=1}^{n} a_i \overrightarrow{\gamma}_{i\,|p}\right)(f) = \sum_{i=1}^{n} a_i \overrightarrow{\gamma}_{i\,|p}(f).$$

Consider the special case where f is the coordinate map $u^j = x^j \circ \varphi$ on U. Then, by (1.2.6), $\overrightarrow{\gamma}_{i\,|p}(f) = \delta_{ij}$. So the equation reduces to $0 = a_j$. And this is true for all $j = 1, \ldots, n$.

Next, suppose that ξ is a vector at p. We show that it can be expressed as a linear combination of the $\overrightarrow{\gamma}_{i\,|p}$. First we associate with ξ a vector $\hat{\xi}$ at $\varphi(p)$. (In what follows, we shall be going back and forth between the context of M and $\mathbb{R}^n$. To reduce possible confusion, we shall systematically use carets for denoting objects associated with $\mathbb{R}^n$). We take $\hat{\xi}$ to be the vector whose action on elements $\hat{f}: \hat{O} \to \mathbb{R}$ in $\mathcal{S}(\varphi(p))$ is given by $\hat{\xi}(\hat{f}) = \xi(\hat{f} \circ \varphi)$. (This makes sense since $\hat{f} \circ \varphi$ is an element of $\mathcal{S}(p)$ with domain $\varphi^{-1}[\varphi[U] \cap \hat{O}]$.) By proposition 1.2.2 (applied to $\hat{\xi}$ at $\varphi(p)$), we know that there are real numbers $\xi^1, \ldots, \xi^n$ such that

$$\hat{\xi}(\hat{f}) = \sum_{i=1}^{n} \xi^i \frac{\partial \hat{f}}{\partial x^i}(\varphi(p))$$

for all $\hat{f}$ in $\mathcal{S}(\varphi(p))$. Now let $f: O \to \mathbb{R}$ be an arbitrary element of $\mathcal{S}(p)$. Then $f \circ \varphi^{-1}: \varphi[O \cap U] \to \mathbb{R}$ belongs to $\mathcal{S}(\varphi(p))$. So, taking $\hat{f} = f \circ \varphi^{-1}$ in the preceding equation and using equation (1.2.5),

$$\hat{\xi}(f \circ \varphi^{-1}) = \sum_{i=1}^{n} \xi^i \frac{\partial (f \circ \varphi^{-1})}{\partial x^i}(\varphi(p)) = \sum_{i=1}^{n} \xi^i \overrightarrow{\gamma}_{i\,|p}(f).$$

But recalling how $\hat{\xi}$ was defined, we also have $\hat{\xi}(f \circ \varphi^{-1}) = \xi((f \circ \varphi^{-1}) \circ \varphi) = \xi(f)$. Thus, $\xi(f) = \sum_{i=1}^{n} \xi^i \overrightarrow{\gamma}_{i\,|p}(f)$ for all f in $\mathcal{S}(p)$; i.e., $\xi = \sum_{i=1}^{n} \xi^i \overrightarrow{\gamma}_{i\,|p}$. So, as claimed, ξ can be expressed as a linear combination of the $\overrightarrow{\gamma}_{i\,|p}$. □

It follows from proposition 1.2.3, of course, that every vector ξ at p has a *unique* representation in the form $\xi = \sum_{i=1}^{n} \xi^i \overrightarrow{\gamma}_{i\,|p}$. Equivalently, by equation (1.2.5),

$$\xi(f) = \sum_{i=1}^{n} \xi^i \overrightarrow{\gamma}_{i\,|p}(f) = \sum_{i=1}^{n} \xi^i \frac{\partial (f \circ \varphi^{-1})}{\partial x^i}(\varphi(p)) \tag{1.2.9}$$

for all f in $\mathcal{S}(p)$. We refer to the coefficients $\xi^1, \ldots, \xi^n$ as the *components of* ξ with respect to (U, φ).

We know that every smooth curve through p determines a vector at p, namely its tangent vector at that point. Using proposition 1.2.3, we can show, conversely, that *every* vector at p can be realized as the tangent vector of a smooth curve through p.

PROPOSITION 1.2.4. *Given an n-manifold* $(M, \mathcal{C})$*, a point p in M, and a vector* ξ *at p, there is a smooth curve* γ *through p such that* $\overrightarrow{\gamma}_p = \xi$.

Proof. Let (U, φ) be an n-chart in $\mathcal{C}$ with $p \in U$, and let $u^i (i = 1, \ldots, n)$ be the corresponding coordinate maps on U. (Recall that $u^i = x^i \circ \varphi$.) By proposition 1.2.3, we know that there are real numbers $\xi^1, \ldots, \xi^n$ such that $\xi = \sum_{i=1}^{n} \xi^i \overrightarrow{\gamma}_{i\,|p}$. Now let $\gamma : I \to U$ be the smooth map defined by

$$\gamma(s) = \varphi^{-1}\left(u^1(p) + \xi^1 s, \ldots, u^n(p) + \xi^n s\right).$$

Note that $\gamma(0) = p$. (The exact specification of the domain of γ does not matter, but we may as well take it to be the largest open interval I containing 0 such that, for all s in I, $\left(u^1(p) + \xi^1 s, \ldots, u^n(p) + \xi^n s\right)$ is in $\varphi[U]$.) For all f in $\mathcal{S}(p)$,

$$\begin{aligned}
\overrightarrow{\gamma}_p(f) = \frac{d}{ds}(f \circ \gamma)(0) &= \sum_{i=1}^{n} \left(\frac{\partial\left(f \circ \varphi^{-1}\right)}{\partial x^i}(\varphi(p)) \right) \xi^i \\
&= \sum_{i=1}^{n} \xi^i \, \overrightarrow{\gamma}_{i\,|p}(f) = \xi(f).
\end{aligned}$$

(The second equality follows by the "chain rule," and the third by equation [1.2.5].) Thus, $\overrightarrow{\gamma}_p = \xi$. $\square$

So far, we have two equivalent characterizations of "vectors" at a point p of a manifold. We can take them to be derivations—i.e., mappings from $\mathcal{S}(p)$ to $\mathbb{R}$ satisfying conditions (DD1)–(DD3)—or take them to be tangents at p to smooth curves passing through p. We mention, finally, a third characterization that was the standard one before "modern" coordinate-free methods became standard in differential geometry. It requires a bit of preparation. (This third characterization will play no role in what follows, and readers may want to jump to the final paragraph of the section.)

Let (U_1, φ_1) and (U_2, φ_2) be n-charts on our background manifold $(M, \mathcal{C})$ such that $(U_1 \cap U_2) \neq \emptyset$. Let p be a point in $(U_1 \cap U_2)$. Further, for all

$i = 1, \ldots, n$, let $x'^i : \varphi_1[U_1 \cap U_2] \to \mathbb{R}$ be the map defined by

$$x'^i = x^i \circ \varphi_2 \circ \varphi_1^{-1},$$

where x^i is the i^{th} coordinate map on $\mathbb{R}^n$. We can think of the x'^i as providing a second coordinate system on $\varphi_1[U_1 \cap U_2]$ that is connected to the first by a smooth, invertible transformation,

$$(x^1, \ldots, x^n) \mapsto (x'^1(x^1, \ldots, x^n), \ldots, x'^n(x^1, \ldots, x^n)).$$

PROPOSITION 1.2.5. *Under the assumptions of the preceding paragraph, let ξ be a non-zero vector at p whose components with respect to (U_1, φ_1) and (U_2, φ_2) are $(\xi^1, \ldots, \xi^n)$ and $(\xi'^1, \ldots, \xi'^n)$. Then the components obey the transformation law*

$$\xi'^i = \sum_{j=1}^{n} \xi^j \frac{\partial x'^i}{\partial x^j}(\varphi_1(p)). \tag{1.2.10}$$

(Of course, they also obey its symmetric counterpart, with the roles of x^i and ξ^i systematically interchanged with those of x'^i and ξ'^i.)

Proof. Let f be any element of $\mathcal{S}(p)$. Then

$$\sum_{j=1}^{n} \xi^j \frac{\partial\big(f \circ \varphi_1^{-1}\big)}{\partial x^j}\big(\varphi_1(p)\big) = \xi(f) = \sum_{j=1}^{n} \xi'^j \frac{\partial\big(f \circ \varphi_2^{-1}\big)}{\partial x^j}\big(\varphi_2(p)\big). \tag{1.2.11}$$

Here we have simply expressed the action of ξ on f in terms of the two sets of components, using equation (1.2.9). Hence, in particular, if $f = x'^i \circ \varphi_1 = x^i \circ \varphi_2 \circ \varphi_1^{-1} \circ \varphi_1 = x^i \circ \varphi_2$, we get

$$\sum_{j=1}^{n} \xi^j \frac{\partial x'^i}{\partial x^j}(\varphi_1(p)) = \sum_{j=1}^{n} \xi'^j \frac{\partial x^i}{\partial x^j}(\varphi_2(p)) = \xi'^i. \qquad \square$$

In what follows, let $\mathcal{C}(p)$ be the set of charts in $\mathcal{C}$ whose domains contain p.

PROBLEM 1.2.1. *Let ξ be a non-zero vector at p, and let $(k^1, \ldots, k^n)$ be a non-zero element of $\mathbb{R}^n$. Show there exists an n-chart in $\mathcal{C}(p)$ with respect to which ξ has components $(k^1, \ldots, k^n)$.*

(Hint: Consider any n-chart (U_1, φ_1) in $\mathcal{C}(p)$, and let $(\xi^1, \ldots, \xi^n)$ be the components of ξ with respect to (U_1, φ_1). Then there is a linear map from $\mathbb{R}^n$ to itself that takes $(\xi^1, \ldots, \xi^n)$ to $(k^1, \ldots, k^n)$. Let the associated matrix have elements $\{a_{ij}\}$.

So, for all $i = 1, \ldots, n,\ \ k^i = \sum_{j=1}^{n} a_{ij}\, \xi^j$. *Now consider a new chart* (U_2, φ_2) *in* $\mathcal{C}(p)$ *where* $U_2 = U_1$ *and* φ_2 *is defined by the condition*

$$x^i \circ \varphi_2 = \sum_{j=1}^{n} a_{ij}\,(x^j \circ \varphi_1).$$

Show that the components of ξ *with respect to* (U_2, φ_2) *are* $(k^1, \ldots, k^n)$.*)*

We have just seen that each vector ξ at p (understood, say, as a derivation) determines a map from $\mathcal{C}(p)$ to $\mathbb{R}^n$ satisfying the transformation law (1.2.10). (The map assigns to each n-chart the components of the vector with respect to the n-chart.) It turns out, conversely, that every map from $\mathcal{C}(p)$ to $\mathbb{R}^n$ satisfying equation (1.2.10) determines a unique vector ξ at p. It does so as follows. Let (U_1, φ_1) be an n-chart in $\mathcal{C}(p)$. We stipulate that, for all maps f in $\mathcal{S}(p)$,

$$\xi(f) = \sum_{j=1}^{n} \xi^j \frac{\partial\big(f \circ \varphi_1^{-1}\big)}{\partial x^j}\big(\varphi_1(p)\big), \tag{1.2.12}$$

where $(\xi^1, \ldots, \xi^n)$ is the element of $\mathbb{R}^n$ associated with (U_1, φ_1). We need only verify that this definition is independent of our choice of n-chart.

Let (U_2, φ_2) be any other n-chart in $\mathcal{C}(p)$ with associated n-tuple $(\xi'^1, \ldots, \xi'^n)$. Then, by assumption, the latter are related to $(\xi^1, \ldots, \xi^n)$ by equation (1.2.10). Now consider the map $f \circ \varphi_1^{-1} : \varphi_1[U_1 \cap U_2] \to \mathbb{R}$. It can be realized as the composition of two maps, $f \circ \varphi_1^{-1} = (f \circ \varphi_2^{-1}) \circ (\varphi_2 \circ \varphi_1^{-1})$. Hence, by the chain rule,

$$\begin{aligned}
\frac{\partial\big(f \circ \varphi_1^{-1}\big)}{\partial x^j}\big(\varphi_1(p)\big) &= \sum_{k=1}^{n} \frac{\partial\big(f \circ \varphi_2^{-1}\big)}{\partial x^k}\big(\varphi_2(p)\big)\, \frac{\partial\big(x^k \circ \varphi_2 \circ \varphi_1^{-1}\big)}{\partial x^j}\big(\varphi_1(p)\big) \\
&= \sum_{k=1}^{n} \frac{\partial\big(f \circ \varphi_2^{-1}\big)}{\partial x^k}\big(\varphi_2(p)\big)\, \frac{\partial x'^k}{\partial x^j}\big(\varphi_1(p)\big),
\end{aligned}$$

for all j. Hence, by equations (1.2.12) and (1.2.10),

$$\begin{aligned}
\xi(f) &= \sum_{j=1}^{n} \xi^j \left[\sum_{k=1}^{n} \frac{\partial\big(f \circ \varphi_2^{-1}\big)}{\partial x^k}\big(\varphi_2(p)\big)\, \frac{\partial x'^k}{\partial x^j}\big(\varphi_1(p)\big)\right] \\
&= \sum_{k=1}^{n} \left[\sum_{j=1}^{n} \xi^j\, \frac{\partial x'^k}{\partial x^j}\big(\varphi_1(p)\big)\right] \frac{\partial\big(f \circ \varphi_2^{-1}\big)}{\partial x^k}\big(\varphi_2(p)\big) = \sum_{k=1}^{n} \xi'^k \frac{\partial\big(f \circ \varphi_2^{-1}\big)}{\partial x^k}\big(\varphi_2(p)\big).
\end{aligned}$$

Thus, our definition of ξ is, indeed, independent of our choice of n-chart. We could equally well have formulated equation (1.2.12) using (U_2, φ_2) and $(\xi'^1, \ldots, \xi'^n)$.

The upshot is that there is a canonical one-to-one correspondence between vectors at p and maps from $\mathcal{C}(p)$ to $\mathbb{R}^n$ satisfying equation (1.2.10). This gives us our promised third (classical) characterization of the former.

There is a helpful picture that accompanies our formal account of tangent vectors and tangent spaces. Think about the special case of a 2-manifold $(M, \mathcal{C})$ that is a smooth surface in three-dimensional Euclidean space. In this case, the tangent space to the manifold M_p at a point p is (or can be canonically identified with) the plane that is tangent to the surface at p. In traditional presentations of differential geometry, vectors at points of manifolds are sometimes called "infinitesimal displacements." The picture suggests where this term comes from. A displacement from p on the surface M is approximated by a tangent vector in M_p. The degree of approximation increases as the displacement on M shrinks. In the limit of "infinitesimal displacements," the two coincide. (Quite generally, statements about "infinitesimal objects" can be read as statements about corresponding objects in tangent spaces.)

1.3. Vector Fields, Integral Curves, and Flows

In what follows, let $(M, \mathcal{C})$ be an n-manifold. (We shall often supress explicit reference to $\mathcal{C}$.) A *vector field* on M is a map ξ that assigns to every point p in M a vector $\xi(p)$ in M_p. (Sometimes we shall write $\xi_{|p}$ for the value of the field ξ at p rather than $\xi(p)$.) We can picture it as field of arrows on M. Given any smooth map $f : M \to \mathbb{R}$, ξ induces a map $\xi(f) : M \to \mathbb{R}$ defined by $\xi(f)(p) = \xi_{|p}(f)$. If $\xi(f)$ is smooth for all such f, we say that the vector field ξ itself is *smooth*.

The proposed picture of a vector field as a field of arrows on M suggests that it should be possible to "thread" the arrows—at least when the field is smooth—to form a network of curves covering M. (See figure 1.3.1.) In fact, this *is* possible.

Let ξ be a smooth vector field on M. We say that a smooth curve $\gamma : I \to M$ is an *integral curve* of ξ if, for all $s \in I$, $\overrightarrow{\gamma}_{\gamma(s)} = \xi(\gamma(s))$—i.e., if the tangent vector to γ at $\gamma(s)$ is equal to the vector assigned by ξ to that point. Intuitively, an integral curve of ξ threads the arrows of ξ and is so parametrized that it "moves quickly" (it covers a lot of M with each unit increment of the parameter s) where ξ is large and "slowly" where ξ is small. Let us also say that a smooth curve $\gamma : I \to M$ has *initial value* p if $0 \in I$ and $\gamma(0) = p$.

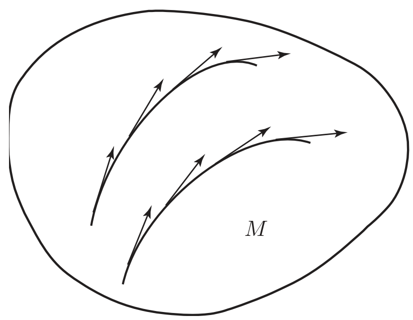

Figure 1.3.1. Integral curves "threading" the vectors of a smooth vector field.

The following is the basic existence and uniqueness theorem for integral curves.

PROPOSITION 1.3.1. *Let ξ be a smooth vector field on M and let p be a point in M. Then there is an integral curve $\gamma : I \to M$ of ξ with initial value p that has the following maximality property: if $\gamma' : I' \to M$ is also an integral curve of ξ with initial value p, then $I' \subseteq I$ and $\gamma'(s) = \gamma(s)$ for all $s \in I'$.*

It is clear that the curve whose existence is guaranteed by the proposition is unique. (For if $\gamma' : I' \to M$ is another, we have $I' \subseteq I$ and $I \subseteq I'$, so $I' = I$, and also $\gamma'(s) = \gamma(s)$ for all $s \in I'$.) It is called the *maximal* integral curve of ξ with initial value p. It also clearly follows from the proposition that if γ is an integral curve of ξ with initial value p, and if its domain is $\mathbb{R}$, then γ is maximal. (The converse is false. Maximal integral curves need not have domain $\mathbb{R}$. We shall soon have an example.) The proof of the proposition, which we skip, makes use of the basic existence and uniqueness theorem for solutions to ordinary differential equations. Indeed, the proposition can be understood as nothing but a geometric formulation of that theorem. (See, for example, Spivak [57, volume 1, chaper 5].)

Here are some examples. In the following, let x^1 and x^2 be the standard coordinate maps on $\mathbb{R}^2$. (So if $p = (p^1, p^2) \in \mathbb{R}^2$, then $x^1(p) = p^1$ and $x^2(p) = p^2$.)

(1) Let ξ be the "horizontal" vector field $\dfrac{\partial}{\partial x^1}$ on $\mathbb{R}^2$. $\Big($Given any point p and any function f in $\mathcal{S}(p)$, the vector $\dfrac{\partial}{\partial x^1}\big|_p$ at p assigns to f the number $\dfrac{\partial f}{\partial x^1}(p)$.$\Big)$ The maximal integral curve of ξ with initial value $p = (p^1, p^2)$ is the map $\gamma : \mathbb{R} \to \mathbb{R}^2$ with

$$\gamma(s) = \big(p^1 + s,\, p^2\big).$$

(The "vertical" vector field $\dfrac{\partial}{\partial x^2}$ is defined similarly.)

(2) Let ξ be the "rotational" vector field $-x^2 \dfrac{\partial}{\partial x^1} + x^1 \dfrac{\partial}{\partial x^2}$ on $\mathbb{R}^2$. The maximal integral curve of ξ with initial value $p = (p^1, p^2)$ is the map $\gamma : \mathbb{R} \to \mathbb{R}^2$ with

$$\gamma(s) = \big(p^1 \cos s - p^2 \sin s, p^1 \sin s + p^2 \cos s\big).$$

The image of γ is a circle, centered at $(0,0)$, that passes through p. (In the degenerate case where p is $(0,0)$, γ is the constant curve that sits at $(0,0)$.)

(3) Let ξ be the "radial expansion" vector field $x^1 \dfrac{\partial}{\partial x^1} + x^2 \dfrac{\partial}{\partial x^2}$ on $\mathbb{R}^2$. The maximal integral curve of ξ with initial value $p = (p^1, p^2)$ is the map $\gamma : \mathbb{R} \to \mathbb{R}^2$ with

$$\gamma(s) = (p^1\, e^s,\, p^2\, e^s).$$

If $(p^1, p^2) \neq (0,0)$, the image of γ is a radial line starting from, but not containing, $(0,0)$. If p is $(0,0)$, γ is the constant curve that sits at $(0,0)$.

Let us check one of these—say (2). The indicated curve is, in fact, an integral curve of the given vector field since, for all $s \in \mathbb{R}$, and all $f \in \mathcal{S}(\gamma(s))$, by the chain rule,

$$\begin{aligned}
\overrightarrow{\gamma}_{\gamma(s)}(f) &= \frac{d}{ds}\big(f \circ \gamma\big)(s) = \frac{d}{ds} f\big(p^1 \cos s - p^2 \sin s, p^1 \sin s + p^2 \cos s\big) \\
&= \frac{\partial f}{\partial x^1}(\gamma(s))\,\big(-p^1 \sin s - p^2 \cos s\big) + \frac{\partial f}{\partial x^2}(\gamma(s))\,\big(p^1 \cos s - p^2 \sin s\big) \\
&= \frac{\partial f}{\partial x^1}(\gamma(s))\,\big(-x^2(\gamma(s))\big) + \frac{\partial f}{\partial x^2}(\gamma(s))\,\big(x^1(\gamma(s))\big) \\
&= \left(-x^2 \frac{\partial}{\partial x^1} + x^1 \frac{\partial}{\partial x^2}\right)_{|\gamma(s)}(f).
\end{aligned}$$

PROBLEM 1.3.1. *Let ξ be the vector field $x^1 \dfrac{\partial}{\partial x^1} - x^2 \dfrac{\partial}{\partial x^2}$ on $\mathbb{R}^2$. Show that the maximal integral curve of ξ with initial value $p = (p^1, p^2)$ is the map $\gamma : \mathbb{R} \to \mathbb{R}^2$ with $\gamma(s) = (p^1\, e^s, p^2\, e^{-s})$. (The image of γ is a (possibly degenerate) hyperbola satisfying the coordinate condition $x^1 x^2 = p^1 p^2$.)*

Next we consider reparametrizations of integral curves.

PROPOSITION 1.3.2. *Let ξ be a smooth vector field on M, let $\gamma: I \to M$ be an integral curve of ξ, and let $\alpha: I' \to I$ be a diffeomorphism taking the interval I' to the interval I. Consider the reparametrized curve $\gamma' = \gamma \circ \alpha: I' \to M$.*

(1) *If there is a number c such that $\alpha(s) = s + c$ for all $s \in I'$, then γ' is an integral curve of ξ.*
(2) *Conversely, if γ' is an integral curve of ξ and if ξ is everywhere non-zero on $\gamma[I]$, then there is a number c such that $\alpha(s) = s + c$ for all $s \in I'$.*

Proof. For all $s \in I'$ and all functions $f \in \mathcal{S}(\gamma'(s))$, it follows by the chain rule (and the definition of tangents to curves) that

$$\begin{aligned}\overrightarrow{\gamma'}_{\gamma'(s)}(f) &= \frac{d}{ds}(f \circ \gamma')(s) = \frac{d}{ds}(f \circ \gamma \circ \alpha)(s) = \left(\frac{d}{dt}(f \circ \gamma)\right)(\alpha(s))\,\frac{d\alpha}{ds}(s) \\ &= \overrightarrow{\gamma}_{\gamma(\alpha(s))}(f)\,\frac{d\alpha}{ds}(s).\end{aligned}$$

That is, for all $s \in I'$,

$$\overrightarrow{\gamma'}_{\gamma'(s)} = \overrightarrow{\gamma}_{\gamma(\alpha(s))}\,\frac{d\alpha}{ds}(s). \tag{1.3.1}$$

Since γ is an integral curve of ξ, we also have

$$\overrightarrow{\gamma}_{\gamma(\alpha(s))} = \xi\big(\gamma(\alpha(s))\big) \tag{1.3.2}$$

for all $s \in I'$. Now γ' is an integral curve of ξ iff $\overrightarrow{\gamma'}_{\gamma'(s)} = \xi\big(\gamma'(s)\big) = \xi\big(\gamma(\alpha(s))\big)$ for all $s \in I'$. So, by equations (1.3.1) and (1.3.2), γ' is an integral curve of ξ iff

$$\xi\big(\gamma(\alpha(s))\big)\,\frac{d\alpha}{ds}(s) = \xi\big(\gamma(\alpha(s))\big) \tag{1.3.3}$$

for all $s \in I'$. This equation is the heart of the matter. If there is a c such that $\alpha(s) = s + c$ for all $s \in I'$, then $\dfrac{d\alpha}{ds} = 1$ everywhere, and so it follows immediately that equation (1.3.3) holds for all $s \in I'$. This gives us clause (1). Conversely, if equation (1.3.3) *does* hold for all $s \in I'$, it must be the case that $\dfrac{d\alpha}{ds} = 1$ everywhere. (Here we use our assumption that $\xi\big(\gamma(\alpha(s))\big)$ is non-zero for all $s \in I'$.) So, clearly, α must be of the form $\alpha(s) = s + c$ for some number c. This gives us (2). □

The qualification in the the second clause of the proposition—that ξ be non-zero on the image of γ—is necessary. (See problem 1.3.3.) The first clause of the proposition tells us that if $\gamma : I \to M$ is an integral curve of ξ, then so is the curve defined by setting $\gamma'(s) = \gamma(s+c)$. We say that γ' is derived from γ by "shifting its initial value." Several useful facts about integral curves follow from proposition 1.3.2 (together with proposition 1.3.1). We list three as problems. The first is a slightly more general formulation of the existence and uniqueness theorem.

PROBLEM 1.3.2. *(Generalization of proposition 1.3.1) Again, let ξ be a smooth vector field on M, and let p be a point in M. But now let s_0 be any real number (not necessarily 0). Show that there is an integral curve $\gamma : I \to M$ of ξ with $\gamma(s_0) = p$ that is maximal in this sense: given any integral curve $\gamma' : I' \to M$ of ξ, if $\gamma'(s_0) = p$, then $I' \subseteq I$ and $\gamma'(s) = \gamma(s)$ for all s in I'.*

(Hint: Invoke proposition 1.3.1 and shift initial values.)

PROBLEM 1.3.3. *(Integral curves that go nowhere) Let ξ be a smooth vector field on M, and let $\gamma : I \to M$ be an integral curve of ξ. Suppose that ξ vanishes (i.e., assigns the zero vector) at some point $p \in \gamma[I]$. Then the following both hold.*

(1) *$\gamma(s) = p$ for all s in I; that is, γ is a constant curve.*
(2) *The reparametrized curve $\gamma' = \gamma \circ \alpha : I' \to M$ is an integral curve of ξ for all diffeomorphisms $\alpha : I' \to I$.*

(Hint: Think about the constant curve, with domain $\mathbb{R}$, that assigns p to all s.)

PROBLEM 1.3.4. *(Integral curves cannot cross) Let $\gamma : I \to M$ and $\gamma' : I' \to M$ be integral curves of ξ that are maximal (in the sense of problem 1.3.2) and satisfy $\gamma(s_0) = \gamma'(s'_0)$. Then the two curves agree up to a parameter shift: $\gamma(s) = \gamma'\big(s + (s'_0 - s_0)\big)$ for all $s \in I$.*

Again, let ξ be a smooth vector field on M. We say that ξ is *complete* if, for every point p in M, the maximal integral curve of ξ with initial point p has domain $\mathbb{R}$—i.e., is a curve of the form $\gamma : \mathbb{R} \to M$. For example, let M be the restriction of $\mathbb{R}^2$ to the vertical strip $\big\{(p^1, p^2) : -1 < p^1 < 1\big\}$, let ξ be the restriction of the "horizontal" vector field $\dfrac{\partial}{\partial x^1}$ (discussed above) to M, and let $p = (0, 0)$. The maximal integral curve of ξ with initial value p is the map $\gamma : (-1, 1) \to M$ with $\gamma(s) = (s, 0)$. So ξ is not complete. (Intuitively, moving along any maximal integral curve of ξ in either direction, one "runs out of space” in finite parameter time.) In contrast, the “vertical field” $\dfrac{\partial}{\partial x^2}$ is complete on M. And $\dfrac{\partial}{\partial x^1}$ itself is complete when construed as a field on (all of) $\mathbb{R}^2$.

Next, let M be the punctured manifold $\mathbb{R}^2 - \{(0,0)\}$, and let ξ be the restriction of the radial vector field (the third in our list of examples) to M. Then ξ is complete. This follows directly from our determination of the maximal integral curves of ξ. It also follows from the assertion in the next problem. (Intuitively, the vectors of ξ rapidly get small as one approaches the puncture point, and so—moving “backward” along a maximal integral curve of ξ—one cannot reach that point in finite parameter time.)

PROBLEM 1.3.5. *Let ξ be a smooth vector field on M that is complete. Let p be a point in M. Show that the restriction of ξ to the punctured manifold $M - \{p\}$ is complete (as a field on $M - \{p\}$) iff ξ vanishes at p.*

The maximal integral curves of a smooth vector field suggest the flow lines of a fluid. It turns out to be extremely useful to think of them this way. As above, let ξ be a smooth vector field on the manifold M. We associate with ξ a set $D_\xi \subseteq \mathbb{R} \times M$ and a “flow map” $\Gamma \colon D_\xi \to M$ as follows. We take D_ξ to be the set of all points (s, p) with the property that if $\gamma \colon I \to M$ is the maximal integral curve of ξ with initial value p, then $s \in I$; and in this case we set $\Gamma(s, p) = \gamma(s)$. (That is, if we start at p, and move s units of parameter distance along the maximal integral curve with initial value p, we arrive at $\Gamma(s, p)$.) So, in particular, $(0, p)$ is in D_ξ for all p in M, and $\Gamma(0, p) = p$ for all such. If the vector field ξ is complete, $D_\xi = \mathbb{R} \times M$. But, in general, D_ξ is a proper subset of the latter. (Starting at a point p, it may not be *possible* to move s units of parameter distance along the maximal integral curve with initial value p.) We have the following basic result.

PROPOSITION 1.3.3. *Let ξ be a smooth vector field on M, and let $\Gamma \colon D_\xi \to M$ be as in the preceding paragraph. Then D_ξ is an open subset of $\mathbb{R} \times M$, and Γ is smooth.*

The proposition asserts, in effect, that solutions to ordinary differential equations depend smoothly on initial conditions. (See Spivak [57, volume 1, chapter 5].)

Assume for the moment that our smooth vector field ξ on M is complete. (So $D_\xi = \mathbb{R} \times M$.) In this case, given any $s \in \mathbb{R}$, we can define a map

$\Gamma_s\colon M \to M$ by setting $\Gamma_s(p) = \Gamma(s, p)$. It follows from proposition 1.3.3 that Γ_s is smooth. (Γ_s can be realized as a composite map $M \to \mathbb{R} \times M \to M$ with action $p \mapsto (s, p) \mapsto \Gamma(s, p)$, and each of the component maps is smooth.) Furthermore, the indexed set $\{\Gamma_s\}_{s\in\mathbb{R}}$ has a natural group structure under the operation of composition $(\Gamma_s \circ \Gamma_t = \Gamma_{s+t})$, with the identity map Γ_0 playing the role of the unit element. (See the next paragraph.) It follows that Γ_s is injective and that its inverse $(\Gamma_s)^{-1} = \Gamma_{-s}$ is smooth. So each Γ_s is a diffeomorphism that maps M onto itself. We say that $\{\Gamma_s\}_{s\in\mathbb{R}}$ is a *one-parameter group of diffeomorphisms* of M generated by ξ. Note that, for all p in M, the map from $\mathbb{R}$ to M defined by $s \mapsto \Gamma_s(p)$ is just the maximal integral curve of ξ with initial value p.

That $\Gamma_s \circ \Gamma_t = \Gamma_{s+t}$ for all s and t follows as a consequence of the assertion in problem 1.3.4. Given any point p in M, and any $t \in \mathbb{R}$, let $\gamma : I \to M$ be the maximal integral curve of ξ with initial value $\Gamma_t(p)$. Then $\gamma(s) = \Gamma_s(\Gamma_t(p))$ for all s. Let $\gamma'\colon I' \to M$ be the maximal integral curve of ξ with initial value p. Then $\gamma'(t) = \Gamma_t(p) = \gamma(0)$ and $\gamma'(s+t) = \Gamma_{s+t}(p)$ for all s. Since $\gamma(0) = \gamma'(t)$, it follows from the assertion in the problem that $\gamma(s) = \gamma'(s+t)$ for all s. So we have $\Gamma_s(\Gamma_t(p)) = \gamma(s) = \gamma'(s+t) = \Gamma_{s+t}(p)$ for all p, t, and s.

Now recall the three complete vector fields on $\mathbb{R}^2$ considered above. Each defines a one-parameter group of diffeomorphisms $\{\Gamma\}_{s\in\mathbb{R}}$ on $\mathbb{R}^2$. The pattern of association is as follows.

Field	Associated Diffeomorphisms
$\dfrac{\partial}{\partial x^1}$	$\Gamma_s(p^1, p^2) = (p^1 + s,\ p^2)$
$-x^2\dfrac{\partial}{\partial x^1} + x^1\dfrac{\partial}{\partial x^2}$	$\Gamma_s(p^1, p^2) = \big(p^1 \cos s - p^2 \sin s, p^1 \sin s + p^2 \cos s\big)$
$x^1\dfrac{\partial}{\partial x^1} + x^2\dfrac{\partial}{\partial x^2}$	$\Gamma_s(p^1, p^2) = (p^1\,\mathrm{e}^s,\ p^2\,\mathrm{e}^s)$

In the three cases, respectively, Γ_s is a displacement by the amount s in the x^1 direction, a (counterclockwise) rotation through s radians with center point $(0, 0)$, and a radial expansion by the factor e^s with center point $(0, 0)$.

Let us now drop the assumption that ξ is complete. Then the "flow maps" $\Gamma_s\colon M \to M$ will not, in general, be defined for all s. But by paying attention to domains of definition, we can still associate with ξ a set of "local flow maps." It follows from proposition 1.3.3 that, given any point p in M, there are both an open interval $I \subseteq \mathbb{R}$ containing 0 and an open subset $U \subseteq M$ containing p such that $I \times U \subseteq D_\xi$. If we set $\Gamma_s(q) = \Gamma(s, q)$ for all $(s, q) \in I \times U$, then the following all hold.

(1) $\Gamma_s \colon U \to \Gamma_s[U]$ is a diffeomorphism for all $s \in I$.
(2) $(\Gamma_s \circ \Gamma_t)(q) = \Gamma_{s+t}(q)$ for all $s, t,$ and q such that $\{s, t, s+t\} \subseteq I$ and $\{q, \Gamma_t(q)\} \subseteq U$.
(3) For all q in U, the map $\gamma : I \to M$ defined by $\gamma(s) = \Gamma_s(q)$ is a smooth integral curve of ξ with initial value q.

In this case, we say that the collection $\{\Gamma_s \colon U \to \Gamma_s[U]\}_{s \in I}$ is a *local one-parameter group of diffeomorphisms* generated by ξ.

1.4. Tensors and Tensor Fields on Manifolds

We start with some linear algebra. We shall return to manifolds shortly.

Let V be an n-dimensional vector space. (Throughout this book, "vector spaces" should be understood to be vector spaces over $\mathbb{R}$.) *Linear functionals* (or *covariant vectors* or *co-vectors*) over V are linear maps from V to $\mathbb{R}$. The set of all linear functionals on V has a natural vector space structure. Given two linear functionals α and β, and a real number k, we take $\alpha + \beta$ and $k\alpha$ to be the linear functionals defined by setting

$$(\alpha + \beta)(\xi) = \alpha(\xi) + \beta(\xi),$$

$$(k\alpha)(\xi) = k\alpha(\xi),$$

for all ξ in V. The vector space V^* of linear functionals on V is called the *dual space* of V. It is easy to check that V^* has dimension n. (If $\overset{1}{\xi}, \overset{2}{\xi}, \ldots, \overset{n}{\xi}$ form a basis for V, then the elements $\overset{1}{\alpha}, \overset{2}{\alpha}, \ldots, \overset{n}{\alpha}$ in V^* defined by

$$\overset{i}{\alpha}(\overset{j}{\xi}) = \delta_{ij}$$

form a basis for V^* called the *dual basis* of $\overset{1}{\xi}, \overset{2}{\xi}, \ldots, \overset{n}{\xi}$.)

The vector space V^* has its own dual space V^{**}, consisting of linear maps from V^* to $\mathbb{R}$. V^{**} is naturally isomorphic to V under the mapping $\varphi \colon V \to V^{**}$, defined by setting $\varphi(\xi)(\alpha) = \alpha(\xi)$ for all ξ in V and all α in V^*; i.e., we require that $\varphi(\xi)$ make the same assignment to α that α itself makes to ξ.

In our development of tensor algebra we shall use the "abstract index notation" introduced by Roger Penrose. (See Penrose and Rindler [51] for a more complete and systematic treatment.) We start by considering an infinite sequence of vector spaces $V^a, V^b, \ldots, V^{a_1}, V^{b_1}, \ldots,$ all isomorphic to our original n-dimensional vector space V. Here a, b, ..., a_1, b_1, ... are elements of some (unspecified) infinite labeling set and are called "abstract indices." They must be distinguished from more familiar "counting indices."

We think of isomorphisms being fixed once and for all, and regard $\xi^a, \xi^b, \ldots$ as the respective images in $V^a, V^b, \ldots$ of ξ in V. The spaces $V^a, V^b, \ldots$ have their respective dual spaces $(V^a)^*, (V^b)^*, \ldots$. We designate these with lowered indices: $V_a, V_b, \ldots$. Our fixed isomorphisms between V and $V^a, V^b, \ldots$ naturally extend to isomophisms between V^* and $V_a, V_b, \ldots$. Given α in V^* we take its image in V_a to be the unique element α_a satisfying the condition $\alpha_a(\xi^a) = \alpha(\xi)$ for all ξ in V. It is convenient to drop parentheses and write $\alpha_a(\xi^a)$ as $\alpha_a\,\xi^a$ or $\xi^a\,\alpha_a$. Thus we have $\alpha_a\,\xi^a = \xi^a\,\alpha_a = \alpha_b\,\xi^b = \xi^b\,\alpha_b$, and so forth. (In what follows, our notation will be uniformly commutative. In a sense, the notation incorporates the canonical isomorphism of V with V^{**}. Rather than thinking of $\xi^a\,\alpha_a$ as $\alpha_a(\xi^a)$, we *can* think of it as the "action of ξ^a on α_a" and understand *that* as the action on α_a of the vector in $(V^a)^{**}$ canonically isomorphic to ξ^a.)

Indices tell us where vectors and linear functionals reside. So rather than writing, for example, "for all vectors ξ^a in $V^a \ldots$," it will suffice to write "for all vectors $\xi^a \ldots$."

We have introduced vector spaces $V^a, V^b, \ldots, V_a, V_b, \ldots$. Now we jump to a larger collection of indexed spaces $V^{a_1\ldots a_r}_{b_1\ldots b_s}(r, s \geq 1)$ where the indices $a_1, \ldots, a_r, b_1, \ldots, b_s$ are all distinct. (The order of superscript indices here will make no difference; nor will that of subscript indices. So, for example, $V^{ad}_{bc} = V^{da}_{bc} = V^{ad}_{cb} = V^{da}_{cb}$. But it will make a difference whether particular indices appear in superscript or subscript position—e.g., $V^a_b \neq V^b_a$.) To keep the notation under control, we shall work first with a representative special case: V^{ab}_c.

The elements of this space are multilinear maps that assign real numbers to unordered triples of the form $\{\mu_a, \nu_b, \gamma^c\}$—i.e., triples containing one element each from V_a, V_b, and V^c. (We shall write these triples, indifferently, as $\mu_a\,\nu_b\,\gamma^c$ or $\nu_b\,\mu_a\,\gamma^c$ or $\gamma^c\,\nu_b\,\mu_a$ or $\nu_b\,\gamma^c\,\mu_a$, and so forth.) By "multilinearity" we mean that if λ is in V^{ab}_c, then

$$\lambda\big((\mu_a + k\,\rho_a)\nu_b\,\gamma^c\big) = \lambda(\mu_a\,\nu_b\,\gamma^c) + k\,\lambda(\rho_a\,\nu_b\,\gamma^c),$$

$$\lambda\big(\mu_a(\nu_b + k\,\tau_b)\gamma^c\big) = \lambda(\mu_a\,\nu_b\,\gamma^c) + k\,\lambda(\mu_a\,\tau_b\,\gamma^c),$$

$$\lambda\big(\mu_a\,\nu_b(\gamma^c + k\,\delta^c)\big) = \lambda(\mu_a\,\nu_b\,\gamma^c) + k\,\lambda(\mu_a\,\nu_b\,\delta^c),$$

for all $\mu_a, \rho_a, \nu_b, \tau_b, \gamma^c, \delta^c$ and all real numbers k. The set V^{ab}_c has a natural vector space structure. If λ and λ' are two elements of V^{ab}_c and k is a real number, we can define new elements $(\lambda + \lambda')$ and $(k\,\lambda)$ in V^{ab}_c by setting

$$(\lambda + \lambda')(\mu_a\,\nu_b\,\gamma^c) = \lambda(\mu_a\,\nu_b\,\gamma^c) + \lambda'(\mu_a\,\nu_b\,\gamma^c),$$

$$(k\,\lambda)(\mu_a\,\nu_b\,\gamma^c) = k\,\lambda(\mu_a\,\nu_b\,\gamma^c),$$

for all μ_a, ν_b, γ^c. The vector space V^{ab}_c has dimension n^3. To see this, first note that any triple of vectors $\{\varphi^a, \psi^b, \chi_c\}$ determines an element in V^{ab}_c under the rule of association

$$\{\varphi^a, \psi^b, \chi_c\}\colon \mu_a\, \nu_b\, \gamma^c \longmapsto (\varphi^a\, \mu_a)(\psi^b\, \nu_b)(\chi_c\, \gamma^c).$$

We write this element as $\varphi^a\, \psi^b\, \chi_c$ or $\chi_c\, \varphi^a\, \psi^b$ or $\psi^b\, \chi_c\, \varphi^a$, and so on. The order of the terms makes no difference. Next, let $\overset{1}{\xi}{}^a, \overset{2}{\xi}{}^a, \ldots, \overset{n}{\xi}{}^a$ be a basis for V^a with dual basis $\overset{1}{\alpha}_a, \overset{2}{\alpha}_a, \ldots, \overset{n}{\alpha}_a$. (Here we have abstract and counting indices side by side.) One can easily verify that the set of all triples of the form $\overset{i}{\xi}{}^a\, \overset{j}{\xi}{}^b\, \overset{k}{\alpha}_c$, with i, j, k ranging from 1 to n, forms a basis for V^{ab}_c . Thus, every element of V^{ab}_c can be uniquely expressed in the form

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} \overset{ijk}{c}\,\overset{i}{\xi}{}^a\overset{j}{\xi}{}^b\,\overset{k}{\alpha}_c.$$

Sometimes it will be convenient to recast sums such as this in terms of a single summation index and absorb coefficients—i.e., in the form

$$\sum_{i=1}^{n^3} \overset{i}{\mu}{}^a\, \overset{i}{\nu}{}^b\, \overset{i}{\tau}_c.$$

(Rather than three indices that range from 1 to n, we have one index that ranges from 1 to n^3.)

Generalizing now, the *tensor space* $V^{a_1\ldots a_r}_{b_1\ldots b_s}(r, s \geq 1)$ consists of multilinear maps assigning real numbers to unordered $(r+s)$-tuples, containing one element each from $V_{a_1}, \ldots, V_{a_r}, V^{b_1}, \ldots, V^{b_s}$. It is a vector space with dimension $n^{(r+s)}$, and its elements can be realized as linear combinations of the form

$$\sum_{i=1}^{n^{(r+s)}} \overset{i}{\mu}{}^{a_1} \cdots \overset{i}{\nu}{}^{a_r}\, \overset{i}{\gamma}_{b_1} \cdots \overset{i}{\lambda}_{b_s}.$$

We have assumed $(r, s \geq 1)$. But the definition scheme we have given makes sense, too, when $r = 0$ and $s = 1$, and when $r = 1$ and $s = 0$. In the former case, we recover indexed dual spaces as previously characterized. (The elements of V_b, recall, are just linear maps from V^b to $\mathbb{R}$). And in the latter case, we recover our initial indexed vector spaces, at least if we allow for the identification of those spaces with their "double duals." We can even allow $r = s = 0$ and construe the tensor space over V with no indices as just $\mathbb{R}$. The elements of tensor spaces are called *tensors*. Tensor indices in superscript (respectively, subscript) position are sometimes called "contravariant" (respectively, "covariant") indices.

We have noted that abstract indices give information about where vectors and co-vectors reside; e.g., μ^a belongs to the space V^a and ν_b belongs to V_b. We can extend this pattern of "residence labeling" to elements of arbitrary tensor spaces. For example, we can attach the index configuration ${}^{ab}_{c}$ to elements of V^{ab}_{c} and make statements of the form "for all $\lambda^{ab}_{c}\ldots$." But things are a bit delicate in the case where the total number of indices present is greater than one.

Though the order of superscript indices and the order of subscript indices make no difference when it comes to labeling tensor *spaces*, they do make a difference when it comes to labeling tensors themselves. For example, though $V_{ab} = V_{ba}$, for an arbitrary element α_{ab} of that space it need not be the case that $\alpha_{ab} = \alpha_{ba}$. (The latter equality captures the condition, not true in general, that the tensor α_{ab} is "symmetric.") To see why, suppose, once again, that $\overset{1}{\xi}{}^a, \overset{2}{\xi}{}^a, \ldots, \overset{n}{\xi}{}^a$ is a basis for V^a and $\overset{1}{\alpha}_a, \overset{2}{\alpha}_a, \ldots, \overset{n}{\alpha}_a$ is its dual basis. Let α_{ab} be the element $\overset{i}{\alpha}_a \overset{j}{\alpha}_b$, for some particular i and j. Then, according to the Penrose notation (as will be explained below), α_{ba} is the element $\overset{i}{\alpha}_b \overset{j}{\alpha}_a$. It follows from what has been said so far that the tensors $\overset{i}{\alpha}_a \overset{j}{\alpha}_b$ and $\overset{i}{\alpha}_b \overset{j}{\alpha}_a$ are simply not equal unless $i = j$. (Why? Assume they are equal. Then they have the same action on all pairs $\mu^a \nu^b$. So, in particular, they have the same action on $\overset{i}{\xi}{}^a \overset{j}{\xi}{}^b$. But

$$\overset{i}{\alpha}_a \overset{j}{\alpha}_b \big(\overset{i}{\xi}{}^a \overset{j}{\xi}{}^b\big) = \overset{i}{\alpha}_a \big(\overset{i}{\xi}{}^a\big) \overset{j}{\alpha}_b \big(\overset{j}{\xi}{}^b\big) = 1$$

and

$$\overset{i}{\alpha}_b \overset{j}{\alpha}_a \big(\overset{i}{\xi}{}^a \overset{j}{\xi}{}^b\big) = \overset{j}{\alpha}_a \big(\overset{i}{\xi}{}^a\big) \overset{i}{\alpha}_b \big(\overset{j}{\xi}{}^b\big) = (\delta_{ij})^2.$$

So $\delta_{ij} = 1$; i.e., $i = j$.)

A second point about the delicacy of the index notation should be mentioned, though it will not concern us until we reach section 1.9 and work with tensors in the presence of a (non-degenerate) metric g_{ab}. We will then want to follow standard practice and use the metric and its inverse g^{ab} to "lower and raise indices." (The rest of this paragraph can be skipped. It is included only for readers who already know about lowering and raising indices and who may anticipate the problem mentioned here.) For example, we shall write ${\alpha_a}^b$ as an abbreviation for $\alpha_{an}\, g^{nb}$. A problem will arise, though, when we try to lower or raise an index on a tensor that has indices in *both* subscript and superscript position. For example, do we write $\lambda^{ab}_{c}\, g^{cd}$ as λ^{dab} or as λ^{adb} or as λ^{abd} ? The latter three will not, in general, be equal (for the reasons given in the preceding paragraph). To cope with the problem, when the time comes, we shall adopt the convention that superscript indices should never be aligned with subscript indices. Instead, each index will have its own vertical "slot." So, for example, the elements of the space V^{ab}_{c} will carry the index structure ${}^{ab}{}_{c}$ or ${}^{a}{}_{c}{}^{b}$ or ${}_{c}{}^{ab}$ (or ${}^{ba}{}_{c}$ or ${}^{b}{}_{c}{}^{a}$ or ${}_{c}{}^{ba}$), and we will not assume, for example, that $\lambda^{ab}{}_{c} = \lambda^{a}{}_{c}{}^{b}$. (For the rest of this section—indeed until section 1.9—we shall not bother with index slots.)

One final preliminary remark about notation is called for. As mentioned before, we want the notation to be uniformly commutative, at least as regards the order of tensors within an expression (in contrast to the order of indices within a tensor). So, for example, the number $\lambda^{ab}_{c}(\mu_a\, \nu_b\, \gamma^c)$ that the tensor λ^{ab}_{c} assigns to a triple $\mu_a\, \nu_b\, \gamma^c$ will be written as $\lambda^{ab}_{c}\, \mu_a\, \nu_b\, \gamma^c$ or as $\mu_a\, \nu_b\, \gamma^c\, \lambda^{ab}_{c}$ or as $\nu_b\, \gamma^c\, \lambda^{ab}_{c}\, \mu_a$, and so forth. Furthermore, if λ^{ab}_{c} is the tensor $\varphi^a\, \psi^b\, \chi_c$, we shall write $\lambda^{ab}_{c}(\mu_a\, \nu_b\, \gamma^c)$ as $\varphi^a\, \psi^b\, \chi_c\, \mu_a\, \nu_b\, \gamma^c$ or as $\chi_c\, \mu_a\, \psi^b\, \nu_b\, \varphi^a\, \gamma^c$ or as any other string with the individual vectors in some order or other. The order does not matter because it is the indices here that determine the crucial groupings: φ^a with μ_a, ψ^b with ν_b, χ_c with γ^c.

We now have in hand the various tensor spaces $V^{a_1\ldots a_r}_{b_1\ldots b_s}$. Within each one (just because it is a vector space), there is an addition operation that is associative and commutative. We will be interested in three other tensor operations: outer multiplication, index substitution, and contraction. We will consider them in turn.

"Outer multiplication" (or, perhaps, "tensor multiplication"), first, is an operation of structure

$$V^{a_1\ldots a_r}_{b_1\ldots b_s} \times V^{c_1\ldots c_m}_{d_1\ldots d_n} \to V^{a_1\ldots a_r c_1\ldots c_m}_{b_1\ldots b_s d_1\ldots d_n},$$

where the indices $a_1, \ldots, a_r, b_1 \ldots b_s, c_1, \ldots, c_m, d_1, \ldots, d_n$ are all distinct. It is defined in an obvious way. Consider a representative special case:

$$V^{ab}_{c} \times V_{fd} \to V^{ab}_{cfd}.$$

The outer product of α^{ab}_{c} and ξ_{fd}, written $\alpha^{ab}_{c}\, \xi_{fd}$ or $\xi_{fd}\, \alpha^{ab}_{c}$, is defined by setting

$$\big(\alpha^{ab}_{c}\, \xi_{fd}\big)\big(\lambda_a\, \rho_b\, \delta^c\, \mu^f\, \nu^d\big) = \big(\alpha^{ab}_{c}\, \lambda_a\, \rho_b\, \delta^c\big)\big(\xi_{fd}\, \mu^f\, \nu^d\big)$$

for all $\lambda_a, \rho_b, \delta^c, \mu^f, \nu^d$. As usual, generally we shall drop parentheses and write terms in any order. So the action of $\alpha^{ab}_{c}\, \xi_{fd}$ on $\lambda_a\, \rho_b\, \delta^c\, \mu^f\, \nu^d$ will be expressed, indifferently, as $\alpha^{ab}_{c}\, \lambda_a\, \rho_b\, \delta^c\, \xi_{fd}\, \mu^f\, \nu^d$ or as $\alpha^{ab}_{c}\, \xi_{fd}\, \lambda_a\, \rho_b\, \delta^c\, \mu^f\, \nu^d$ or as $\lambda_a\, \rho_b\, \xi_{fd}\, \delta^c\, \mu^f\, \alpha^{ab}_{c}\, \nu^d$, and so forth. It should be clear that outer multiplication, as defined here, is commutative, associative, and distributive over addition. Notice, also, that our notation is consistent. Consider, for example, the expression $\tau^a\, \varepsilon^b\, \varphi_c\, \alpha_a\, \beta_b\, \gamma^c$. We can construe it as the action of $\tau^a\, \varepsilon^b\, \varphi_c$ on

$\alpha_a\,\beta_b\,\gamma^c$, or as the action of $\varepsilon^b\,\varphi_c\,\alpha_a$ on $\tau^a\,\beta_b\,\gamma^c$, or as the action of $\gamma^c\,\varphi_c\,\tau^a\,\varepsilon^b$ on $\alpha_a\,\beta_b$, and so on. (The third reading makes sense: $\gamma^c\,\varphi_c\,\tau^a\,\varepsilon^b$ is the element of V^{ab} that arises if one multiplies the element $\tau^a\,\varepsilon^b$ by the number $\gamma^c\,\varphi_c$.) Each of these functional operations yields the same number, so no consistency problem arises.

The operation of "$(x \to y)$ index substitution" has the structure

$$V^{x\ a_1\ldots a_r}_{\ \ b_1 b_2\ldots b_s} \to V^{y\ a_1\ldots a_r}_{\ \ b_1 b_2\ldots b_s} \quad \text{or} \quad V^{a_1 a_2\ldots a_r}_{x\ b_1\ldots b_s} \to V^{a_1 a_2\ldots a_r}_{y\ b_1\ldots b_2},$$

where the indices $x, y, a_1, \ldots, a_r, b_1, \ldots, b_s$ are all distinct. In defining the operation, it is, again, easiest to consider a representative special case, say $V^{ab}_c \to V^{db}_c$. Given a tensor α^{ab}_c, it can be expressed as a sum of the form

$$\alpha^{ab}_c = \sum_{i=1}^{n^3} \overset{i}{\mu}{}^a\, \overset{i}{\nu}{}^b\, \overset{i}{\tau}_c.$$

We take the result of $(a \to d)$ index substitution on α^{ab}_c, which we write as α^{db}_c, to be the sum

$$\alpha^{db}_c = \sum_{i=1}^{n^3} \overset{i}{\mu}{}^d\, \overset{i}{\nu}{}^b\, \overset{i}{\tau}_c.$$

(This makes sense because we already have a fixed isomorphism between V^a and V^d that takes each $\overset{i}{\mu}{}^a$ to $\overset{i}{\mu}{}^d$.) Of course, it must be checked that this definition is independent of the choice of expansion for α^{ab}_c. That is, one must check that if

$$\sum_{i=1}^{n^3} \overset{i}{\mu}{}^a\, \overset{i}{\nu}{}^b\, \overset{i}{\tau}_c = \sum_{i=1}^{n^3} \overset{i}{\delta}{}^a\, \overset{i}{\varepsilon}{}^b\, \overset{i}{\rho}_c,$$

then

$$\sum_{i=1}^{n^3} \overset{i}{\mu}{}^d\, \overset{i}{\nu}{}^b\, \overset{i}{\tau}_c = \sum_{i=1}^{n^3} \overset{i}{\delta}{}^d\, \overset{i}{\varepsilon}{}^b\, \overset{i}{\rho}_c.$$

But this follows from the fact that $\overset{i}{\mu}{}^a\,\lambda_a = \overset{i}{\mu}{}^d\,\lambda_d$ and $\overset{i}{\delta}{}^a\,\lambda_a = \overset{i}{\delta}{}^d\,\lambda_d$ for all i and all λ_a.

It can easily be checked that index substitution commutes with addition, outer multiplication, and other index substitutions. For example, if $\alpha^{ab}_c = \beta^{ab}_c + \gamma^{ab}_c$, then $\alpha^{db}_c = \beta^{db}_c + \gamma^{db}_c$. If $\lambda^{ab}_{cfg} = \alpha^{ab}_c\,\xi_{fg}$, then $\lambda^{db}_{cfg} = \alpha^{db}_c\,\xi_{fg}$. And the tensor that results from first applying $(a \to b)$ index substitution and then $(c \to d)$ index substitution to α^a_{cf} is the same as that resulting from reversing the order and applying first $(c \to d)$ index substitution and then $(a \to b)$ index substitution. It is written as α^{b}_{df}. All these facts, in a sense, are built into our notation.

Our final tensor operation, "(x, y) contraction," has the structure

$$V^{x\,a_1\dots a_r}_{y\,b_1\dots b_s} \to V^{a_1\dots a_r}_{b_1\dots b_s},$$

where the indices $x, y, a_1, \dots, a_r, b_1, \dots, b_s$ are all distinct. Consider, for example, (a, c) contraction with action $V^{ab}_{c} \to V^{b}$. Suppose

$$\alpha^{ab}_{c} = \sum_{i=1}^{n^3} \overset{i}{\mu}{}^{a}\, \overset{i}{\nu}{}^{b}\, \overset{i}{\tau}_{c}.$$

We take the result of applying (a, c) contraction to α^{ab}_{c} to be

$$\alpha^{ab}_{a} = \sum_{i=1}^{n^3} \overset{i}{\mu}{}^{a}\, \overset{i}{\tau}_{a}\, \overset{i}{\nu}{}^{b}.$$

(This could also be expressed as $\sum\limits_{i=1}^{n^3} \overset{i}{\mu}{}^{c}\, \overset{i}{\tau}_{c}\, \overset{i}{\nu}{}^{b}$ or as $\sum\limits_{i=1}^{n^3} \overset{i}{\mu}{}^{c}\, \overset{i}{\nu}{}^{b}\, \overset{i}{\tau}_{c}$ or as $\sum\limits_{i=1}^{n^3} \overset{i}{\mu}{}^{d}\, \overset{i}{\tau}_{d}\, \overset{i}{\nu}{}^{b}$, and so forth. The last of the listed possibilities is equal to the first because $\overset{i}{\mu}{}^{a}\, \overset{i}{\tau}_{a} = \overset{i}{\mu}{}^{d}\, \overset{i}{\tau}_{d}$ for all $\overset{i}{\mu}{}^{a}$ and $\overset{i}{\tau}_{a}$.) We write this result as α^{ab}_{a} (or α^{cb}_{c} or α^{db}_{d}, and so forth). It is important that contracted indices on a tensor—i.e., ones that appear in both contravariant and covariant position—play no role in determining the space in which the tensor resides. α^{ab}_{a} belongs to V^{b}, *not* some space V^{ab}_{a}. Indeed, there is no such space as we have set things up.

To prove that contraction is well defined—i.e., independent of one's choice of expansion—a simple lemma is needed.

LEMMA 1.4.1. *For all $r \geq 1$, and all $\overset{k}{\varphi}{}^{a}$ and $\overset{k}{\psi}_{c}\ (k = 1, \dots, r)$,*

$$\sum_{k=1}^{r} \overset{k}{\varphi}{}^{a}\, \overset{k}{\psi}_{c} = \mathbf{0} \qquad \Longrightarrow \qquad \sum_{k=1}^{r} \overset{k}{\varphi}{}^{a}\, \overset{k}{\psi}_{a} = 0.$$

Proof. Let $\overset{1}{\xi}{}^{a}, \overset{2}{\xi}{}^{a}, \dots, \overset{n}{\xi}{}^{a}$ be a basis for V^{a} with dual basis $\overset{1}{\alpha}_{a}, \overset{2}{\alpha}_{a}, \dots, \overset{n}{\alpha}_{a}$. Then, for each $k = 1, \dots, r$, there exist numbers c_{ki} and d_{kj} $(i, j = 1, \dots, n)$ where $\overset{k}{\varphi}{}^{a} = \sum\limits_{i=1}^{n} c_{ki}\, \overset{i}{\xi}{}^{a}$ and $\overset{k}{\psi}_{c} = \sum\limits_{j=1}^{n} d_{kj}\, \overset{j}{\alpha}_{c}$. Assume the left-side condition holds. Then for all $l = 1, \ldots, n$,

$$0 = \left(\sum_{k=1}^{r} \overset{k}{\varphi}{}^{a}\, \overset{k}{\psi}{}_{c}\right) \overset{l}{\alpha}{}_{a}\, \overset{l}{\xi}{}^{c} = \sum_{k=1}^{r} \left(\sum_{i=1}^{n} c_{ki}\, \overset{i}{\xi}{}^{a}\right)\left(\sum_{j=1}^{n} d_{kj}\, \overset{j}{\alpha}{}_{c}\right) \overset{l}{\alpha}{}_{a}\, \overset{l}{\xi}{}^{c}$$

$$= \sum_{k=1}^{r}\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ki}\, d_{kj}\, \delta_{il}\, \delta_{jl} = \sum_{k=1}^{r} c_{kl}\, d_{kl}.$$

It follows that the right-side condition holds, since

$$\sum_{k=1}^{r} \overset{k}{\varphi}{}^{a}\, \overset{k}{\psi}{}_{a} = \sum_{k=1}^{r} \left(\sum_{i=1}^{n} c_{ki}\, \overset{i}{\xi}{}^{a}\right)\left(\sum_{j=1}^{n} d_{kj}\, \overset{j}{\alpha}{}_{a}\right) = \sum_{k=1}^{r}\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ki}\, d_{kj}\, \delta_{ij}$$

$$= \sum_{i=1}^{n}\sum_{k=1}^{r} c_{ki}\, d_{ki} = 0.$$

(Each term $\sum_{k=1}^{r} c_{ki}\, d_{ki}$ in the final sum is 0 by the calculation just given.) □

PROBLEM 1.4.1. *Show that lemma 1.4.1 can also be derived as a corollary to the following fact (Herstein [32, p. 272]) about square matrices: if M is an $(r \times r)$ matrix $(r \geq 1)$ and M^2 is the zero matrix, then the trace of M is 0. (Hint: Consider the $r \times r$ matrix M with entries $M_{ij} = \overset{i}{\varphi}{}^{a}\, \overset{j}{\psi}{}_{a}$.)*

COROLLARY 1.4.2. *For all $r \geq 1$, and all $\overset{k}{\beta}{}^{a}$, $\overset{k}{\gamma}{}^{b}$, $\overset{k}{\psi}{}_{c} (k = 1, \ldots, r)$,*

$$\sum_{k=1}^{r} \overset{k}{\beta}{}^{a}\, \overset{k}{\gamma}{}^{b}\, \overset{k}{\psi}{}_{c} = \mathbf{0} \quad \Longrightarrow \quad \sum_{k=1}^{r} \left(\overset{k}{\beta}{}^{a}\, \overset{k}{\psi}{}_{a}\right) \overset{k}{\gamma}{}^{b} = \mathbf{0}.$$

Proof. It follows from the left-side condition that, for all λ_b, $\sum_{k=1}^{r} \overset{k}{\beta}{}^{a}(\overset{k}{\gamma}{}^{b}\lambda_b)\, \overset{k}{\psi}{}_{c} = \mathbf{0}$. Applying the lemma (with $\overset{k}{\varphi}{}^{a} = (\overset{k}{\gamma}{}^{b}\, \lambda_b)\overset{k}{\beta}{}^{a}$ for all $k = 1, \ldots, r$), we may infer that $\sum_{k=1}^{r} (\overset{k}{\beta}{}^{a}\, \overset{k}{\psi}{}_{a})\overset{k}{\gamma}{}^{b}\, \lambda_b = 0$. But here λ_b is arbitrary. So it must be the case that the right-side condition holds. □

It follows immediately that contraction is well defined for our tensor α_c^{ab}. For if

$$\alpha_c^{ab} = \sum_{i=1}^{n^3} \overset{i}{\mu}{}^a \, \overset{i}{\nu}{}^b \, \overset{i}{\tau}_c = \sum_{i=1}^{n^3} \overset{i}{\delta}{}^a \, \overset{i}{\chi}{}^b \, \overset{i}{\rho}_c,$$

we can apply corollary 1.4.2 to the difference $\sum_{i=1}^{n^3} \overset{i}{\mu}{}^a \, \overset{i}{\nu}{}^b \, \overset{i}{\tau}_c - \sum_{i=1}^{n^3} \overset{i}{\delta}{}^a \, \overset{i}{\chi}{}^b \, \overset{i}{\rho}_c$ (construed as a sum over $2n^3$ terms). And the corollary can be recast easily for tensors with other index structures.

The contraction operation commutes with addition, outer multiplication, index substitution, and other contractions. Note, once again, the consistency of our notation. The expression $\beta^a \gamma_a$, for example, can be construed as the action of the functional γ_a on β^a, or as the outer product of β^a with γ_b followed by (a, b) contraction, or as the outer product β^b with γ_a followed by (a, b) contraction, and so forth. There is no need to choose among these different readings. Similarly, $\alpha_c^a \lambda_a \sigma^c$ can be understood as the action of α_c^a on $\lambda_a \sigma^c$, or as the outer product of α_c^a with $\lambda_b \sigma^d$ followed by (a, b) and (c, d) contractions, or as the action of λ_a on $\alpha_c^a \sigma^c$, and so forth.

The operations we have introduced on tensors may seem a bit complex. But one quickly gets used to them and applies them almost automatically where appropriate. That is one of the virtues of the abstract index notation. One gets to manipulate tensors as easily as one manipulates components of tensors in traditional tensor analysis. One has the best of both worlds: complete basis (or coordinate) independence, and the computational convenience that comes with indices.

Two bits of special notation will be useful. First, we introduce the "delta tensor" δ_b^a. It is the element of V_b^a defined by setting $\delta_b^a \eta_a \xi^b = \eta_a \xi^a$ for all η_a and ξ^b. (Clearly, δ_b^a, so defined, *is* a tensor since it is linear in both indices.) Notice that the defining condition is equivalent to the requirement that $\delta_b^a \xi^b = \xi^a$ for all ξ^b, and also to the requirement that $\delta_b^a \eta_a = \eta_b$ for all η_a. We can think of δ_b^a as an $(a \to b)$ index substitution operator acting on covariant indices, or as a $(b \to a)$ index substitution operator acting on contravariant indices. So, for example, $\delta_b^a \alpha_d^{bc} = \alpha_d^{ac}$. To see this, suppose that $\alpha_d^{bc} = \sum_{i=1}^{n^3} \overset{i}{\mu}{}^b \, \overset{i}{\nu}{}^c \, \overset{i}{\tau}_d$. Then

$$\delta_b^a \alpha_d^{bc} = \sum_{i=1}^{n^3} \left(\delta_b^a \, \overset{i}{\mu}{}^b\right) \overset{i}{\nu}{}^c \, \overset{i}{\tau}_d = \sum_{i=1}^{n^3} \overset{i}{\mu}{}^a \overset{i}{\nu}{}^c \overset{i}{\tau}_d = \alpha_d^{ac}.$$

Given a basis $\overset{1}{\xi}{}^a, \overset{2}{\xi}{}^a, \ldots, \overset{n}{\xi}{}^a$ for V^a with dual basis $\overset{1}{\alpha}_a, \overset{2}{\alpha}_a, \ldots, \overset{n}{\alpha}_a$, δ^a_b can be expressed as $\delta^a_b = \sum_{i=1}^{n} \overset{i}{\xi}{}^a \, \overset{i}{\alpha}_b$. (This follows since the left- and right-side tensors in this equation have the same action on the basis elements $\overset{1}{\xi}{}^b, \overset{2}{\xi}{}^b, \ldots, \overset{n}{\xi}{}^b$.) It follows that $\delta^a_a = n$.

The second bit of useful notation is for "symmetrization" and "anti-symmetrization" of tensors. Consider, for example, the tensor β^{ab}. Corresponding to it is the tensor β^{ba}. One can think of the latter as arising from the former by a series of index substitutions: $\beta^{ab} \to \beta^{cb} \to \beta^{cd} \to \beta^{bd} \to \beta^{ba}$. (We have already discussed the fact that, though β^{ab} and β^{ba} belong to V^{ab}, in general it is *not* the case that $\beta^{ab} = \beta^{ba}$.) We take $\beta^{(ab)}$ and $\beta^{[ab]}$ to be the respective symmetrization and anti-symmetrization of β^{ab}:

$$\beta^{(ab)} = \frac{1}{2}(\beta^{ab} + \beta^{ba}),$$

$$\beta^{[ab]} = \frac{1}{2}(\beta^{ab} - \beta^{ba}).$$

Similarly, given a tensor γ^b_{cdg}, we set

$$\gamma^b_{(cdg)} = \frac{1}{6}(\gamma^b_{cdg} + \gamma^b_{gcd} + \gamma^b_{dgc} + \gamma^b_{cgd} + \gamma^b_{gdc} + \gamma^b_{dcg}),$$

$$\gamma^b_{[cdg]} = \frac{1}{6}(\gamma^b_{cdg} + \gamma^b_{gcd} + \gamma^b_{dgc} - \gamma^b_{cgd} - \gamma^b_{gdc} - \gamma^b_{dcg}).$$

In general, a tensor with round brackets surrounding a collection of p consecutive indices (all contravariant or all covariant) is to be understood as $\frac{1}{p!}$ times the sum of the $p!$ tensors obtained by taking the selected indices in all possible permutations. (Each permutation can be achieved by multiple index substitutions.) In the case of square brackets, the only difference is that each term in the sum receives a coefficient of $(+1)$ or (-1) depending on whether the indices in that term form a positive or negative permutation of the original sequence. The operations of symmetrization and antisymmetrization commute with addition, outer multiplication, and index substitution. So, for example, if $\beta^{ab} = \gamma^{ab} + \rho^{ab}$, then $\beta^{(ab)} = \gamma^{(ab)} + \rho^{(ab)}$. If $\gamma^b_{cdg} = \lambda_{cdg}\,\xi^b$, then $\gamma^b_{(cdg)} = \lambda_{(cdg)}\,\xi^b$. And if one applies $(c \to f)$ index substitution to γ^b_{cdg} and then symmetrizes over the indices f, d, and g, the resulting tensor is the same one obtained if one first symmetrizes over c, d, and g and then applies $(c \to f)$ index substitution.

We say that a tensor of the form $\alpha^{a_1 \ldots a_r}_{b_1 \ldots b_s}$ is *(totally) symmetric* in indices $b_1, \ldots, b_s$ if interchanging any two of these indices leaves the tensor intact, or, equivalently, if $\alpha^{a_1\ldots a_r}_{(b_1\ldots b_s)} = \alpha^{a_1\ldots a_r}_{b_1\ldots b_s}$. We say it is *(totally) anti-symmetric* in those indices if the interchange in each case has the effect of multiplying the tensor by (-1) or, equivalently, if $\alpha^{a_1\ldots a_r}_{[b_1\ldots b_s]} = \alpha^{a_1\ldots a_r}_{b_1\ldots b_s}$. (The conditions of symmetry and anti-symmetry in indices $a_1, \ldots, a_r$ are defined similarly.) The following proposition will be useful in what follows.

PROPOSITION 1.4.3. *If*

(1) $\alpha^{a_1\ldots a_r}_{b_1\ldots b_s}$ *is symmetric in indices* $b_1, \ldots, b_s$, *and*
(2) $\alpha^{a_1\ldots a_r}_{b_1\ldots b_s}\, \xi^{b_1} \ldots \xi^{b_s} = \mathbf{0}$ *for all* ξ *in* V,

then $\alpha^{a_1\ldots a_r}_{b_1\ldots b_s} = \mathbf{0}$. *(A parallel proposition holds if* $\alpha^{a_1\ldots a_r}_{b_1\ldots b_s}$ *is symmetric in indices* $a_1, \ldots, a_r$.*)*

Proof. We prove the proposition by induction on s. The case $s = 1$ is trivial. So assume $s > 1$ and assume the proposition holds for $s - 1$. For all vectors μ and ν in V, and all real numbers k, we have, by (2), $\mathbf{0} = \alpha^{a_1\ldots a_r}_{b_1\ldots b_s} (\mu + k\,\nu)^{b_1} \ldots (\mu + k\,\nu)^{b_s}$. Expanding the right side of the equation and using (1), we arrive at

$$\mathbf{0} = \alpha^{a_1\ldots a_r}_{b_1\ldots b_s}\, \mu^{b_1} \ldots \mu^{b_s} + \binom{s}{1} k\, \alpha^{a_1\ldots a_r}_{b_1\ldots b_s}\, \mu^{b_1} \ldots \mu^{b_{s-1}} \nu^{b_s} + \ldots$$
$$+ \binom{s}{s-1} k^{s-1} \alpha^{a_1\ldots a_r}_{b_1\ldots b_s}\, \mu^{b_1} \nu^{b_2} \ldots \nu^{b_s} + k^s \alpha^{a_1\ldots a_r}_{b_1\ldots b_s}\, \nu^{b_1} \ldots \nu^{b_s}.$$

But k is arbitrary here. The only way the right-side sum can be $\mathbf{0}$ for all values of k is if each of the terms in the sum (without the coefficient) is $\mathbf{0}$. In particular, $\alpha^{a_1\ldots a_r}_{b_1\ldots b_s}\, \mu^{b_1} \ldots \mu^{b_{s-1}}\, \nu^{b_s} = \mathbf{0}$. Now let $\alpha'^{a_1\ldots a_r}_{b_1\ldots b_{s-1}} = \alpha^{a_1\ldots a_r}_{b_s\ldots b_s}\, \nu^{b_s}$. The tensor $\alpha'^{a_1\ldots a_r}_{b_s\ldots b_{s-1}}$ is completely symmetric in the indices $b_1, \ldots, b_{s-1}$, and $\alpha'^{a_1\ldots a_r}_{b_1\ldots b_{s-1}}\, \mu^{b_1} \ldots \mu^{b_{s-1}} = \mathbf{0}$ for all μ in V. So, by our induction hypothesis, it must be the case that $\alpha^{a_1\ldots a_r}_{b_1\ldots b_s}\, \nu^{b_s} = \mathbf{0}$. But ν was an arbitrary vector. So $\alpha^{a_1\ldots a_r}_{b_1\ldots b_s} = \mathbf{0}$, as claimed. □

Sometimes it will be convenient to work with this proposition in a slightly more general form. Let Σ and Π be strings of indices, possibly empty, in which $a_1, \ldots, a_r, b_1, \ldots, b_s$ do not appear. Then we can say that a tensor $\alpha^{a_1\ldots a_r}_{\Sigma b_1\ldots b_s \Pi}$ is *(totally) symmetric* in indices $b_1, \ldots, b_s$ if $\alpha^{a_1\ldots a_r}_{\Sigma(b_1\ldots b_s)\Pi} = \alpha^{a_1\ldots a_r}_{\Sigma b_1\ldots b_s \Pi}$. The case of *(total) anti-symmetry* is handled similarly. It follows as a corollary to the proposition that if $\alpha^{a_1\ldots a_r}_{\Sigma b_1\ldots b_s \Pi}$ is symmetric in indices $b_1, \ldots, b_s$, and if $\alpha^{a_1\ldots a_r}_{\Sigma b_1\ldots b_s \Pi}\, \xi^{b_1} \ldots \xi^{b_s} = \mathbf{0}$ for all ξ in V, then $\alpha^{a_1\ldots a_r}_{\Sigma b_1\ldots b_s \Pi} = \mathbf{0}$. (It follows because we can always contract on all the indices in Σ and Π with arbitrary, distinct vectors and generate a tensor to which the proposition is directly applicable.) Of course, a similar generalization of the proposition is available in the case where the “extra indices” are in covariant position.

This completes our discussion of tensor algebra. We now return to manifolds. Suppose $(M, \mathcal{C})$ is an n-manifold and p is a point in M. Then M_p is an n-dimensional vector space. We can take it to be our fundamental space V and construct a hierarchy of tensor spaces over *it*. A *tensor field* on M is simply an assignment of a tensor (over M_p) to each point p in M, where the tensors all have the same index structure. So, for example, a vector field ξ^a on M (as defined in section 1.3) qualifies as a tensor field on M. The tensor operations (addition, outer multiplication, index substitution, and contraction) are all applied pointwise, and so they extend naturally to tensor fields.

We already know what it means for a scalar field or a (contravariant) vector field on M to be smooth. We now take a covariant vector field α_a on M to be *smooth* if $(\xi^a \alpha_a)$ is smooth for all smooth vector fields ξ^a on M. Quite generally, we say that a tensor field $\lambda^{a_1 \ldots a_r}_{b_1 \ldots b_s}$ on M is *smooth* if $\lambda^{a_1 \ldots a_r}_{b_1 \ldots b_s}\, \xi^{b_1} \ldots \eta^{b_s}\, \alpha_{a_1} \ldots \beta_{a_r}$ is smooth for all smooth fields $\xi^{b_1}, \ldots, \eta^{b_s}, \alpha_{a_1}, \ldots, \beta_{a_r}$ on M.

This pattern of definition is extremely common. One starts with a concept (in this case smoothness) applicable to scalar fields, then extends it to contravariant vector fields by considering their action on scalar fields, then extends it to covariant vector fields by considering their action on contravariant fields, then extends it to tensor fields of arbitrary index structure by considering their action on (appropriate combinations of) contravariant and covariant vector fields.

It follows from the definition of smoothness for tensor fields just given that the four tensor operations take smooth tensor fields to smooth tensor fields.

1.5. The Action of Smooth Maps on Tensor Fields

In this section, we consider when and how it is possible to use a smooth map between manifolds to carry tensors at a point, and tensor fields, from one manifold to the other.

We start with tensors at a point. Let $(M, \mathcal{C})$ and $(M', \mathcal{C}')$ be manifolds, not necessarily of the same dimension; let $\psi : M \to M'$ be a smooth map of M into M'; and let p be a point in M. There is no natural way to transfer *arbitrary* tensors between p and $\psi(p)$—at least, not without further assumptions in place. But it *is* possible to associate with ψ two restricted transfer maps.

Let us say that a tensor (at some point on some manifold) is *contravariant* (respectively, *covariant*) if all of its indices are in contravariant (respectively, covariant) position. The *rank* of such a tensor is the number of its indices. We allow the number to be 0; i.e., we regard scalars (real numbers) as both contravariant and covariant tensors of rank 0.

The first of our two restricted transfer maps, the "push-forward map" $(\psi_p)_*$, takes contravariant tensors at p to contravariant tensors of the same rank at $\psi(p)$. The second, the "pull-back map" $(\psi_p)^*$, takes covariant tensors at $\psi(p)$ to covariant tensors of the same rank at p. We define $(\psi_p)_*$ and $(\psi_p)^*$ in four stages. (For clarity, we mark objects defined on M' with a prime.)

(Stage 0) Given any real number c, we set $(\psi_p)_*(c) = (\psi_p)^*(c) = c$.

(Stage 1) Given a vector ξ^a at p, we define $(\psi_p)_*(\xi^a)$ at $\psi(p)$ as follows. Let $\alpha' \colon O' \to \mathbb{R}$ be an element of $\mathcal{S}\big(\psi(p)\big)$. Then $(\alpha' \circ \psi) \colon \psi^{-1}[O'] \to \mathbb{R}$ is an element of $\mathcal{S}(p)$. We need to specify what assignment $(\psi_p)_*(\xi^a)$ makes to α'. We set

$$\big((\psi_p)_*(\xi^a)\big)(\alpha') = \xi^a(\alpha' \circ \psi). \tag{1.5.1}$$

This makes sense because $(\alpha' \circ \psi)$ is an object of the sort to which ξ^a makes assignments.

(Stage 2) Next, consider a covariant tensor $\eta'_{b_1 \dots b_s}$ at $\psi(p)$. We define the pull-back tensor $(\psi_p)^*(\eta'_{b_1 \dots b_s})$ at p by specifying its action on arbitrary vectors $\overset{1}{\xi}{}^{b_1}, \dots, \overset{s}{\xi}{}^{b_s}$ there. We set

$$\big((\psi_p)^*(\eta'_{b_1 \dots b_s})\big)\, \overset{1}{\xi}{}^{b_1} \dots \overset{s}{\xi}{}^{b_s} = \eta'_{b_1 \dots b_s}\big((\psi_p)_*(\overset{1}{\xi}{}^{b_1})\big) \dots \big((\psi_p)_*(\overset{s}{\xi}{}^{b_s})\big). \tag{1.5.2}$$

Here, of course, we understand the right side because we know (from stage 1) how to push forward the vectors $\overset{i}{\xi}{}^{b_i}$.

(Stage 3) Finally, consider a contravariant tensor $\xi^{a_1 \dots a_r}$ at p with $r \geq 2$. We define the push-forward tensor $(\psi_p)_*(\xi^{a_1 \dots a_r})$ at $\psi(p)$ by specifying its action on arbitrary vectors $\overset{1}{\eta}{}'_{a_1}, \dots, \overset{r}{\eta}{}'_{a_r}$ there:

$$\big((\psi_p)_*(\xi^{a_1 \dots a_r})\big)\, \overset{1}{\eta}{}'_{a_1} \dots \overset{r}{\eta}{}'_{a_r} = \xi^{a_1 \dots a_r}\, \big((\psi_p)^*(\overset{1}{\eta}{}'_{a_1})\big) \dots \big((\psi_p)^*(\overset{r}{\eta}{}'_{a_r})\big). \tag{1.5.3}$$

This completes the definition of $(\psi_p)_*$ and $(\psi_p)^*$.

Several basic facts about them are recorded in the next proposition.

PROPOSITION 1.5.1. *Let* $\psi : M \to M'$ *be a smooth map of the manifold M into the manifold* M'. *Let p be any point in M. Then* $(\psi_p)_*$ *and* $(\psi_p)^*$ *have the following properties.*

(1) $(\psi_p)_*$ *and* $(\psi_p)^*$ *commute with addition.*
For example, $(\psi_p)_*(\xi^{abc} + \rho^{abc}) = (\psi_p)_*(\xi^{abc}) + (\psi_p)_*(\rho^{abc})$.
(2) $(\psi_p)_*$ *and* $(\psi_p)^*$ *commute with outer multiplication.*
For example, $(\psi_p)^*(\eta'_{abc}\,\mu'_{de}) = \big((\psi_p)^*(\eta'_{abc})\big)\,\big((\psi_p)^*(\mu'_{de})\big)$.
(3) $(\psi_p)_*$ *and* $(\psi_p)^*$ *commute with index substitution.*
(4) *For all tensors* $\xi^{a_1\ldots a_r c_1 \ldots c_s}$ *and* $\rho^{b_1\ldots b_r}$ *at p, and all tensors* $\eta'_{a_1\ldots a_r}$ *and* $\mu'_{b_1\ldots b_r d_1 \ldots d_s}$ *at* $\psi(p)$,

$$\big((\psi_p)_*(\xi^{a_1\ldots a_r c_1\ldots c_s})\big)\,\eta'_{a_1\ldots a_r} = (\psi_p)_*\big(\xi^{a_1\ldots a_r c_1\ldots c_s}\,((\psi_p)^*(\eta'_{a_1\ldots a_r}))\big), \tag{1.5.4}$$

$$\big((\psi_p)^*(\mu'_{b_1\ldots b_r d_1\ldots d_s})\big)\,\rho^{b_1\ldots b_r} = (\psi_p)^*\big(\mu'_{b_1\ldots b_r d_1\ldots d_s}\,((\psi_p)_*(\rho^{b_1\ldots b_r}))\big). \tag{1.5.5}$$

Note that we cannot replace clause (4) with the simpler assertion that $(\psi_p)_*$ and $(\psi_p)^*$ commute with contraction. We cannot claim, for example, that $(\psi_p)_*(\xi^{ac}\,\eta_a) = \big((\psi_p)_*(\xi^{ac})\big)\,\big((\psi_p)_*(\eta_a)\big)$, since the second term on the right side is not well formed. The push-forward map $(\psi_p)_*$ makes assignments only to contravariant vectors at p.

Note also that it follows as a special case of clause (2) that $(\psi_p)_*$ and $(\psi_p)^*$ commute with scalar multiplication. For example, $(\psi_p)_*(c\,\xi^{ab}) = ((\psi_p)_*(c))\,\big((\psi_p)_*(\xi^{ab})\big) = c\,\big((\psi_p)_*(\xi^{ab})\big)$. So, clearly, $(\psi_p)_*$ and $(\psi_p)^*$ are linear maps (when restricted to tensors of a fixed rank).

Proof. All four clauses in the proposition follow easily from the definitions of $(\psi_p)*$ and $(\psi_p)^*$. For the fourth clause, one first considers contractions involving (contravariant or covariant) vectors—i.e., $((\psi_p)_*(\xi^{a\,c_1\ldots c_s}))\,\eta'_a$ or $((\psi_p)^*(\eta'_{b\,d_1\ldots d_s}))\,\rho^b$—and then uses the fact that every tensor $\eta'_{a_1\ldots a_r}$ or $\rho^{b_1\ldots b_r}$ can be represented as a sum over products of such vectors. The desired conclusion then follows from clauses (1) and (2). By way of example, let us verify one instance of the fourth clause, say

$$\big((\psi_p)_*(\xi^{ac})\big)\,\eta'_a = (\psi_p)_*\big(\xi^{ac}\,\big((\psi_p)^*(\eta'_a)\big)\big).$$

To show that the two (right- and left-side) vectors at $\psi(p)$ are equal, it suffices to demonstrate that they have the same action on any vector μ'_c there. But this follows, since

$$(\psi_p)_*\big(\xi^{ac}\,\big((\psi_p)^*(\eta'_a)\big)\big)\,\mu'_c = \xi^{ac}\big((\psi_p)^*(\eta'_a)\big)\,\big((\psi_p)^*(\mu'_c)\big) = \big((\psi_p)_*(\xi^{ac})\big)\,\eta'_a\,\mu'_c.$$

Both equalities are instances of equation (1.5.3). The role of $\xi^{a_1 \ldots a_r}$ is played by $\left(\xi^{ac}\left((\psi_p)^*(\eta'_a)\right)\right)$ in the first and by ξ^{ac} in the second. □

Now we turn our attention to fields on M and M'. At each point p in M, we have transfer maps $(\psi_p)_*$ and $(\psi_p)^*$. The question arises whether they can be "aggregated" to carry contravariant fields on M to ones on M' or, alternatively, to carry covariant fields on M' to ones on M. Here an asymmetry arises.

Consider first a tensor field $\xi^{a_1 \ldots a_r}$ on M. For all p in M, $(\psi_p)_*\left(\xi^{a_1 \ldots a_r}(p)\right)$ is a tensor at $\psi(p)$. ($\xi^{a_1 \ldots a_r}(p)$ is the value of the field at p, and it is pushed forward by $(\psi_p)_*$.) But these individual assignments do not, in general, determine a field on M'. For one thing, if ψ is not injective, there will be distinct points p and q such that $\psi(p) = \psi(q)$, and nothing guarantees that $(\psi_p)_*\left(\xi^{a_1 \ldots a_r}(p)\right) = (\psi_q)_*\left(\xi^{a_1 \ldots a_r}(q)\right)$. Furthermore, even if ψ is injective, this prescription will not transfer a tensor to a point p' in M' unless it is in the range of ψ—i.e., unless $p' = \psi(p)$ for some p in M.

But no problems arise if we work in the other direction. Consider a field $\eta'_{b_1 \ldots b_s}$ on M'. Then at every point p, there is a well-defined pull-back tensor $(\psi_p)^*\left(\eta'_{b_1 \ldots b_s}(\psi(p))\right)$. It just does not matter whether ψ is injective or whether its range is all of M'. So we *can* aggregate the individual pull-back maps at different points to generate a map ψ^* that takes covariant tensor fields on M' to ones on M of the same rank.

In particular, ψ^* takes scalar fields $\alpha' : M' \to \mathbb{R}$ on M' to scalar fields

$$\psi^*(\alpha') = (\alpha' \circ \psi) \tag{1.5.6}$$

on M. (Think about it this way. The pull-back field $\psi^*(\alpha')$ assigns to any point p in M the same number that α' assigns to $\psi(p)$. (Recall the 0-th stage in the definition of $(\psi_p)^*$.) So, for all p in M, $\psi^*(\alpha')(p) = \alpha'(\psi(p)) = (\alpha' \circ \psi)(p)$.)

Three of the (pointwise) algebraic conditions listed in proposition 1.5.1 carry over immediately. Thus, ψ^* commutes with addition, outer multiplication, and index substitution (if these are now understood as operations on tensor fields rather than as operations on tensors at a point). The fourth condition, the one involving contraction, does not carry over because it refers to individual push-forward maps $(\psi_p)_*$ (and these, we know, cannot, in general, be aggregated). In addition, ψ^* satisfies a natural smoothness condition; namely, it takes smooth fields on M' to smooth fields on M. This is immediate for the case of scalar fields. (If $\alpha' : M' \to \mathbb{R}$ is smooth, then certainly the composed map $\psi^*(\alpha') = (\alpha' \circ \psi)$ is smooth as well.) But a short detour will be required for the other cases.

Let us temporarily put aside our map ψ between manifolds and consider a general fact about the representation of covariant vector fields on a manifold M. Given any smooth scalar field $\alpha\colon M \to \mathbb{R}$, we associate with it a smooth covariant vector field $d_a\alpha$ on M, called its "exterior derivative." (Here we partially anticipate our discussion of exterior derivative operators in section 1.7.) It is defined by the requirement that, for all p in M and all vectors ξ^a at p, $\xi^a\, d_a\alpha = \xi(\alpha)$; i.e., $\xi^a\, d_a\alpha$ is the directional derivative of α at p in the direction ξ^a. (The condition clearly defines a covariant vector—i.e., a linear functional over M_p, at each point p. And the resultant field $d_a\alpha$ is smooth since, given any smooth vector field ξ^a on M, $\xi^a\, d_a\alpha$ is a smooth scalar field on M.) The fact we need is the following.

LEMMA 1.5.2. *Let λ_a be a smooth field on an n-dimensional manifold $(M, \mathcal{C})$. Then, given any point p in M, there exists an open set O containing p, and smooth real-valued maps $\overset{1}{f}, \ldots, \overset{n}{f}, \overset{1}{g}, \ldots, \overset{n}{g}$ on O, such that $\lambda_a = \overset{1}{f}\, d_a \overset{1}{g} + \cdots + \overset{n}{f}\, d_a \overset{n}{g}$ on O.*

Proof. Let p be a point in M, let (O, φ) be a chart in $\mathcal{C}$ with $p \in O$, and let $u^1, \ldots, u^n$ be the associated coordinate maps on O. At every point q in O, the coordinate curve tangent vectors $(\overrightarrow{\gamma}_{1|q})^a, \ldots, (\overrightarrow{\gamma}_{n|q})^a$ associated with $u^1, \ldots, u^n$ form a basis for $(M_q)^a$. (Recall proposition 1.2.3.) Now consider the vector fields $d_a u^1, \ldots, d_a u^n$ on O. We claim that they determine a dual basis at every q; i.e., $(\overrightarrow{\gamma}_{i|q})^a (d_a u^j) = \delta_{ij}$ for all $i, j \in \{1, \ldots, n\}$. Indeed, this follows immediately since $(\overrightarrow{\gamma}_{i|q})^a (d_a u^j) = \overrightarrow{\gamma}_{i|q}(u^j)$ (by the definition of d_a) and $\overrightarrow{\gamma}_{i|q}(u^j) = \delta_{ij}$ (by equation (1.2.6)). So we can express λ_a in the form $\lambda_a = \overset{1}{f}\, d_a u^1 + \ldots + \overset{n}{f}\, d_a u^n$ on O, where $\overset{i}{f} = (\overrightarrow{\gamma}_i)^a \lambda_a$. The coordinate maps $u^1, \ldots, u^n$ are certainly smooth. And the maps $\overset{1}{f}, \ldots, \overset{n}{f}$ must be smooth as well since λ_a and the coordinate tangent fields $(\overrightarrow{\gamma}_1)^a, \ldots, (\overrightarrow{\gamma}_n)^a$ are so. □

With the lemma in hand, let us return to the original discussion. Again, let ψ be a smooth map from the manifold M into the manifold M'. Note that given any smooth field $\alpha'\colon M' \to \mathbb{R}$ on M', we have

$$\psi^*(d_a\alpha') = d_a\big(\psi^*(\alpha')\big). \tag{1.5.7}$$

(To see this, let p be any point in M and let ξ^a be any vector at p. Then

$$\xi^b\big(\psi^*(d_b\alpha')\big)_{|p} = \big((\psi_p)_*(\xi^b)\big)\,(d_b\alpha')_{|\psi(p)} = \big((\psi_p)_*(\xi^b)\big)(\alpha')$$
$$= \xi^b(\alpha' \circ \psi) = \xi^b\,\big(d_b(\psi^*(\alpha'))\big)_{|p}.$$

The first equality is an instance of equation (1.5.2), with $(d_b\alpha')_{|\psi(p)}$ playing the role of η'_b; the third is an instance of equation (1.5.1). The second follows from the definition of the operator d_a, and the fourth from that definition together with equation (1.5.6). So equation (1.5.7) holds at all points p in M.)

It is our goal, once again, to show that, for all smooth fields $\eta'_{b_1\ldots b_s}$ on M', the pull-back field $\psi^*(\eta'_{b_1\ldots b_s})$ on M is smooth as well. Consider the case of a smooth vector field η'_b on M'. Suppose M' has dimension n. We know from the lemma that given any point p' in M', we can find an open set O' containing p' in which η'_b admits the representation $\eta'_b = \sum_{i=1}^{n} \overset{i}{f}{}' d_b \overset{i}{g}{}'$ (with the constituent maps all smooth). Hence, we have

$$\psi^*(\eta'_b) = \psi^*\left(\sum_{i=1}^{n} \overset{i}{f}{}' d_b \overset{i}{g}{}'\right) = \sum_{i=1}^{n} \psi^*(\overset{i}{f}{}')\,\psi^*(d_b\overset{i}{g}{}') = \sum_{i=1}^{n} \psi^*(\overset{i}{f}{}')\,d_b\big(\psi^*(\overset{i}{g}{}')\big)$$

throughout $\psi^{-1}[O']$. (We get the second equality from the fact that ψ^* respects the tensor operations of addition and outer multiplication (in the sense discussed above). The third equality follows from equation (1.5.7).) But the constituent fields in the far right sum are all smooth. (We have already seen that ψ^* takes smooth scalar fields to smooth scalar fields.) So $\psi^*(\eta'_b)$ itself is smooth on $\psi^{-1}[O']$. But as p' ranges over M', the corresponding pull-back sets $\psi^{-1}[O']$ cover M. It follows that $\psi^*(\eta'_b)$ is smooth on (all of) M.

It remains to consider the general case: smooth fields on M' of the form $\eta'_{b_1\ldots b_s}$. But this case quickly reduces to the preceding one. We can express any such field, at least locally, in the form

$$\eta'_{b_1\ldots b_s} = \sum_{i=1}^{n^s} \overset{i}{\mu}{}'_{b_1} \ldots \overset{i}{\nu}{}'_{b_s},$$

where $\overset{i}{\mu}{}'_{b_1}, \ldots, \overset{i}{\nu}{}'_{b_s}$ ($i = 1, \ldots, n^s$) are all smooth fields on M'. Since the individual pull-back fields $\psi^*(\overset{i}{\mu}{}'_{b_1}), \ldots, \psi^*(\overset{i}{\nu}{}'_{b_s})$ are smooth (and since ψ^* commutes with addition and outer multiplication), it follows that $\psi^*(\eta'_{b_1\ldots b_s})$ must be smooth on M.

In summary, we have established the following.

PROPOSITION 1.5.3. *Let $\psi : M \to M'$ be a smooth map of the manifold M into the manifold M'. Then ψ^* is a map from smooth covariant tensor fields on M' to smooth covariant fields on M of the same rank that commutes with addition, outer multiplication, and index substitution and that also satisfies equation (1.5.7).*

The complications and asymmetries we have encountered all have their origin in the fact that we have only been assuming that ψ is a smooth map of M into M'. Now, finally, let us consider the case where ψ is, in fact, a diffeomorphism of the first onto the second; i.e., there is a well-defined inverse map $\psi^{-1}: M' \to M$ that is also smooth. Then, as one would expect, there is induced a natural one-to-one correspondence between smooth tensors fields of arbitrary index structure on the two manifolds, and this correspondence fully respects the four tensor operations. We already know how ψ^* acts on smooth covariant tensor fields (and scalar fields) on M'. Now we can characterize its action on a smooth field $\lambda'^{a_1\ldots a_r}_{b_1\ldots b_s}$ of unrestricted index structure on M'. We stipulate that, given any point p in M, and any smooth fields $\overset{1}{\eta}_{a_1}, \ldots, \overset{r}{\eta}_{a_r}, \overset{1}{\xi}{}^{b_1}, \ldots, \overset{s}{\xi}{}^{b_s}$ on M,

$$\begin{aligned} &\psi^*(\lambda'^{a_1\ldots a_r}_{b_1\ldots b_s})_{|p}\,(\overset{1}{\eta}_{a_1}\ldots\overset{s}{\xi}{}^{b_s})_{|p} \\ &\quad= (\lambda'^{a_1\ldots a_r}_{b_1\ldots b_s})_{|\psi(p)}\Big(\big((\psi)_*(\overset{1}{\eta}_{a_1})\big)\ldots\big((\psi)_*(\overset{s}{\xi}{}^{b_s})\big)\Big)_{|\psi(p)}. \end{aligned} \tag{1.5.8}$$

Of course, the right side makes sense only if we understand how ψ_* acts on smooth vector fields η_a and ξ^b on M. But we *do* understand (in this new context where ψ is a diffeomorphism). Here we can aggregate the individual push-forward maps $(\psi_p)_*$ to generate a map ψ_* that knows how to act on contravariant vector fields—just as previously we aggregated the maps $(\psi_p)^*$ to generate a map ψ^* that knows how to act on covariant vector fields. And we can take $\psi_*(\eta_a)$ to be $(\psi^{-1})^*(\eta_a)$. This completes the definition of ψ^*.

Notice that this general characterization of ψ^* reduces to the one given previously in the special case where it acts on a covariant field $\lambda'_{b_1\ldots b_s}$.

The way to remember equation (1.5.8) is this. A trade-off is involved. Pulling back $\lambda'^{a_1\ldots a_r}_{b_1\ldots b_s}$ from $\psi(p)$ to p and having it act there on particular vectors yields the same result as pushing those vectors forward from p to $\psi(p)$ and having $\lambda'^{a_1\ldots a_r}_{b_1\ldots b_s}$ act on them there.

We have just seen how to extend ψ^* so that it acts on smooth fields on M' of unrestricted index structure (when ψ is a diffeomorphism). Of course, we can extend ψ_* similarly. Indeed, we can take it to be $(\psi^{-1})^*$.

It is a straightforward matter to confirm that ψ^* (and so ψ_*) commutes with addition, outer multiplication, contraction, and index substitution. By way of example, we verify that, for all smooth fields $\alpha'^a_{\ b}$ and ξ'^b on M',

$$\psi^*(\alpha'^a_{\ b}\,\xi'^b) = \psi^*(\alpha'^a_{\ b})\psi^*(\xi'^b). \tag{1.5.9}$$

Let η_a be any smooth field on M. Then, invoking equation (1.5.8) and dropping explicit reference to points of evaluation, we have

$$\psi^*(\alpha'^a_{\ b})\psi^*(\xi'^b)\eta_a = \alpha'^a_{\ b}\psi_*\big(\psi^*(\xi'^b)\big)\ \psi_*(\eta_a) = \alpha'^a_{\ b}\,\xi'^b\ \psi_*(\eta_a) = \psi^*(\alpha'^a_{\ b}\,\xi'^b)\ \eta_a.$$

(For the second equality, we use the fact that $(\psi_* \circ \psi^*)$ = the identity map.) Since this holds for all smooth fields η_a on M, we have equation (1.5.9).

1.6. Lie Derivatives

Let $(M, \mathcal{C})$ be a fixed manifold, and let ξ^a be a smooth vector field on M. The Lie derivative operator $\mathcal{L}_\xi$ associated with ξ^a is a map from smooth tensor fields (on M) to smooth tensor fields (on M) of the same index structure. Roughly speaking, $\mathcal{L}_\xi\,\lambda^{a_1\ldots a_r}_{b_1\ldots b_s}$ represents the "rate of change" of the field $\lambda^{a_1\ldots a_r}_{b_1\ldots b_s}$ relative to a standard of constancy determined by ξ^a. We now have the tools in place to make this precise. (It is not important, but we write "$\mathcal{L}_\xi$" rather than "$\mathcal{L}_{\xi^a}$" to avoid the impression that the operator adds a new index. There is no chance for confusion since the object X in $\mathcal{L}_X$ is always a contravariant vector field and the index it carries makes no difference.)

Let $\lambda^{a_1\ldots a_r}_{b_1\ldots b_s}$ be a smooth field on M, and let p be a point in M. Further, let $\{\Gamma_t\colon U \to \Gamma_t[U]\}_{t\in I}$ be a local one–parameter group of diffeomorphisms generated by ξ^a with $p \in U$. Here I is an open interval of $\mathbb{R}$, U is an open subset of M, and the maps $\Gamma_t\colon U \to \Gamma_t[U] \subseteq M$ satisfy conditions (1)–(3) at the close of section 1.3. We set

$$\big(\mathcal{L}_\xi\,\lambda^{a_1\ldots a_r}_{b_1\ldots b_s}\big)_{|p} = \lim_{t\to 0}\frac{1}{t}\Big[\big((\Gamma_t)^*(\lambda^{a_1\ldots a_r}_{b_1\ldots b_s})\big)_{|p} - \lambda^{a_1\ldots a_r}_{b_1\ldots b_s\,|p}\Big].$$

The right-side limit is to be understood this way. We start with the tensor $(\lambda^{a_1\ldots a_r}_{b_1\ldots b_s})_{|\Gamma_t(p)}$ at $\Gamma_t(p)$, carry it back to p with the pull-back map $(\Gamma_t)^*$, subtract $(\lambda^{a_1\ldots a_r}_{b_1\ldots b_s})_{|p}$, divide by t, and then take the limit as t goes to 0. (That the limit exists, and that the resultant field $\big(\mathcal{L}_\xi\,\lambda^{a_1\ldots a_r}_{b_1\ldots b_s}\big)$ on M is smooth, follows from proposition 1.3.3.) Note that we *need* to carry $(\lambda^{a_1\ldots a_r}_{b_1\ldots b_s})_{|\Gamma_t(p)}$ back to p before comparing it with $(\lambda^{a_1\ldots a_r}_{b_1\ldots b_s})_{|p}$ because the two tensors live in different spaces. The expression $\big[(\lambda^{a_1\ldots a_r}_{b_1\ldots b_s})_{|\Gamma(p)} - \lambda^{a_1\ldots a_r}_{b_1\ldots b_s\,|p}\big]$ is not well formed.

The following proposition lists several basic properties of Lie derivatives. (The proof is straightforward.)

PROPOSITION 1.6.1. *The operator $\mathcal{L}_\xi$ has the following properties.*

(1) *It commutes with addition.*
For example, $\mathcal{L}_\xi(\alpha_c^{ab} + \beta_c^{ab}) = \mathcal{L}_\xi(\alpha_c^{ab}) + \mathcal{L}_\xi(\beta_c^{ab})$.
(2) *It satisfies the Leibniz rule with respect to outer multiplication.*
For example, $\mathcal{L}_\xi(\alpha_c^{ab}\,\beta_{df}) = \alpha_c^{ab}\,\mathcal{L}_\xi\,\beta_{df} + \beta_{df}\,\mathcal{L}_\xi\,\alpha_c^{ab}$.
(3) *It commutes with the operation of index substitution.*
(4) *It commutes with the operation of contraction.*

PROBLEM 1.6.1. *Show that $\mathcal{L}_\xi\,\delta_a^b = \mathbf{0}$. (Hint: Recall that δ_a^b can be thought of as an index substitution operator, and make use of proposition 1.6.1.)*

PROBLEM 1.6.2. *Let η^a be a smooth, non-vanishing field on M. Show that if $\mathcal{L}_\xi(\eta^a\,\eta^b) = \mathbf{0}$, then $\mathcal{L}_\xi\,\eta^a = \mathbf{0}$.*

Two cases are of special interest, namely Lie derivatives of scalar fields and of contravariant vector fields. We consider them in order.

PROPOSITION 1.6.2. *Let ξ^a and α be smooth fields on M. Then $\mathcal{L}_\xi(\alpha) = \xi(\alpha)$; i.e., at every point in M, $\mathcal{L}_\xi\,\alpha$ is just the ordinary directional derivative of α in the direction ξ^a.*

Proof. Let p be any point in M, and let $\{\Gamma_t\colon U \to \Gamma_t[U]\}_{t\in I}$ be a local one–parameter group of diffeomorphisms generated by ξ^a with $p \in U$. Since the curve $\gamma : I \to M$ defined by $\gamma(t) = \Gamma_t(p)$ is an integral curve of ξ^a with initial point p, we have

$$\xi_{|p}(\alpha) = \overrightarrow{\gamma}_{|p}(\alpha) = \frac{d}{dt}(\alpha \circ \gamma)(0) = \frac{d}{dt}\big[(\alpha \circ \Gamma_t)(p)\big]_{|t=0}.$$

But $(\Gamma_t)^*(\alpha) = (\alpha \circ \Gamma_t)$ for all $t \in I$. (Recall equation (1.5.6).) So we also have

$$\begin{aligned}
\big(\mathcal{L}_\xi\,\alpha\big)_{|p} &= \lim_{t\to 0}\frac{1}{t}\Big[\big((\Gamma_t)^*(\alpha)\big)_{|p} - \alpha_{|p}\Big] = \lim_{t\to 0}\frac{1}{t}\Big[\big(\alpha \circ \Gamma_t\big)(p) - \big(\alpha \circ \Gamma_0\big)(p)\Big] \\
&= \frac{d}{dt}\big[(\alpha \circ \Gamma_t)(p)\big]_{|t=0}.
\end{aligned}$$

So $(\mathcal{L}_\xi\,\alpha)_{|p} = \xi_{|p}(\alpha)$ at all points p in M. □

We need a lemma for the second special case (Lie derivatives of contravariant vector fields).

LEMMA 1.6.3. *Let ξ^a be a smooth vector field on M, let p be a point in M, and, once again, let $\{\Gamma_t \colon U \to \Gamma_t[U]\}_{t\in I}$ be a local one–parameter group of diffeomorphisms generated by ξ^a with $p \in U$. Then, given any smooth scalar field $\alpha \colon M \to \mathbb{R}$, there is a one–parameter family of smooth scalar fields $\{\varphi_t\}_{t\in I}$ on U such that*

(1) $\alpha \circ \Gamma_t = \alpha + t \cdot \varphi_t$ *for all t in I, and*
(2) $\varphi_0 = \xi(\alpha)$.

Proof. Consider the family of smooth scalar fields $\{\varphi_t\}_{t\in I}$ on U defined by setting

$$\varphi_t(q) = \int_0^1 \frac{d}{du}\left[\left(\alpha \circ \Gamma_u\right)(q)\right]_{|u=tx} dx$$

for all t in I and q in U. We claim that it satisfies conditions (1) and (2). First, for all t in I,

$$\begin{aligned} t \cdot \varphi_t(q) &= \int_0^1 \frac{d}{du}\left[\left(\alpha \circ \Gamma_u\right)(q)\right]_{|u=tx}\, t\, dx \\ &= \int_0^1 \frac{d}{dx}\left[\left(\alpha \circ \Gamma_{tx}(q)\right)\right] dx \\ &= \left(\alpha \circ \Gamma_t\right)(q) - \left(\alpha \circ \Gamma_0\right)(q). \end{aligned}$$

But $\Gamma_0(q) = q$ for all q in U. So we have condition (1). Next, differentiation with respect to t yields

$$t \cdot \frac{d}{dt}\varphi_t(q) + \varphi_t(q) = \frac{d}{dt}\left[\left(\alpha \circ \Gamma_t\right)(q)\right].$$

Evaluating both sides at $t = 0$ gives us

$$\varphi_0(q) = \frac{d}{dt}\left[\left(\alpha \circ \Gamma_t\right)(q)\right]_{|t=0}.$$

But now, since $\{\Gamma_t \colon U \to \Gamma_t[U]\}_{t\in I}$ is a local one–parameter group of diffeomorphisms generated by ξ^a, the curve $\gamma \colon I \to O$ defined by $\gamma(t) = \Gamma_t(q)$ is an integral curve of ξ^a with initial value q. Thus,

$$\xi_{|q}(\alpha) = \overrightarrow{\gamma}_{|q}(\alpha) = \frac{d}{dt}(\alpha \circ \gamma)_{|\,t=0} = \frac{d}{dt}\left[\left(\alpha \circ \Gamma_t\right)(q)\right]_{|\,t=0}.$$

So $\varphi_0(q) = \xi_{|q}(\alpha)$ for arbitrary q in U. This is just condition (2). $\square$

PROPOSITION 1.6.4. *Let ξ^a and λ^a be smooth vector fields on M. Then $\pounds_\xi(\lambda^a) = [\xi, \lambda]^a$, where $[\xi, \lambda]^a$ is the smooth ("commutator") vector field on M whose action* *on a smooth scalar field $\alpha : M \rightarrow \mathbb{R}$ is given by*

$$[\xi, \lambda](\alpha) = \xi(\lambda(\alpha)) - \lambda(\xi(\alpha)).$$

(Another remark about notation. One must make some decision about how to handle abstract indices when dealing with commutator vector fields. Depending on context, we shall write, for example, either "$[\xi, \lambda]^a$" or "$[\xi, \lambda]$" or "$[\xi^a, \lambda^a]$"—but never "$[\xi^a, \lambda^a]^a$." Nothing of importance turns on this decision.)

Proof. Let p be any point in M, and let $\{\Gamma_t : U \rightarrow \Gamma_t[U]\}_{t\in I}$ be a local one–parameter group of diffeomorphisms generated by ξ^a with $p \in U$. Given a smooth scalar field $\alpha : M \rightarrow \mathbb{R}$, let $\{\varphi_t\}_{t\in I}$ be a one–parameter family of smooth scalar fields on U satisfying conditions (1) and (2) in the lemma. For all t such that both t and $-t$ are in I, we have

$$[(\Gamma_t)^*(\lambda^a)]_{|p}(\alpha) = \lambda^a{}_{|\Gamma_t(p)}(\alpha \circ \Gamma_{-t}) = \lambda^a{}_{|\Gamma_t(p)}(\alpha - t \cdot \varphi_{-t}).$$

The first equality follows directly from equation (1.5.1) and the fact that $(\Gamma_t)^* = (\Gamma_{-t})_*$. The second follows from condition (1) of the lemma (with t replaced by $-t$). So

$$\begin{aligned}
\left(\mathcal{L}_\xi\, \lambda^a\right)_{|p}(\alpha) &= \lim_{t\to 0} \frac{1}{t}\left[\left((\Gamma_t)^*(\lambda^a)\right)_{|p}(\alpha) - \lambda^a{}_{|p}(\alpha)\right] \\
&= \lim_{t\to 0} \frac{1}{t}\left[\lambda^a{}_{|\Gamma_t(p)}(\alpha) - \lambda^a{}_{|p}(\alpha)\right] - \lim_{t\to 0} \lambda^a{}_{|\Gamma_t(p)}(\varphi_{-t}) \\
&= \lim_{t\to 0} \frac{1}{t}\left[(\lambda^a(\alpha))_{|\Gamma_t(p)} - (\lambda^a(\alpha))_{|p}\right] - \lambda^a{}_{|p}(\varphi_0) \\
&= \frac{d}{dt}\left[\left(\lambda^a(\alpha)\circ\Gamma_t\right)(p)\right]_{|t=0} - \lambda^a{}_{|p}(\varphi_0).
\end{aligned}$$

Now the first term on the right side of the final line is equal to $\xi_{|p}\big(\lambda(\alpha)\big)$. (The argument is the same as used in the final stage of the proof of the lemma.) And $\varphi_0 = \xi(\alpha)$, by condition (2) of the lemma. So

$$\left(\mathcal{L}_\xi\, \lambda^a\right)_{|p}(\alpha) = \xi^a{}_{|p}(\lambda(\alpha)) - \lambda^a{}_{|p}(\xi(\alpha)).$$

Since p and α are arbitrary, this establishes our claim. □

PROBLEM 1.6.3. *Show that the set of smooth contravariant vector fields on M forms a "Lie algebra" under the bracket operation (defined in the preceding proposition); i.e., show that for all smooth vector fields ξ, η, λ on M,*

$$[\xi, \eta] = -[\eta, \xi] \quad \text{and} \quad \big[\lambda, [\xi, \eta]\big] + \big[\eta, [\lambda, \xi]\big] + \big[\xi, [\eta, \lambda]\big] = \mathbf{0}.$$

PROBLEM 1.6.4. *Show that for all smooth vector fields ξ^a, η^a on M, and all smooth scalar fields α on M,*

$$\mathcal{L}_{(\alpha\,\xi)}\eta^a = \alpha\big(\mathcal{L}_\xi\,\eta^a\big) - \big(\mathcal{L}_\eta\,\alpha\big)\xi^a.$$

PROBLEM 1.6.5. *One might be tempted to take a smooth tensor field to be "constant" if its Lie derivatives with respect to all smooth vector fields are zero. But this idea does not work. Any contravariant vector field that was constant in this sense would have to vanish everywhere. Prove this.*

PROBLEM 1.6.6. *Show that for all smooth vector fields ξ^a, η^a, and all smooth tensor fields $\alpha^{a\ldots b}_{c\ldots d}$,*

$$\big(\mathcal{L}_\xi\,\mathcal{L}_\eta - \mathcal{L}_\eta\,\mathcal{L}_\xi\big)\alpha^{a\ldots b}_{c\ldots d} = \mathcal{L}_\theta\,\alpha^{a\ldots b}_{c\ldots d},$$

where θ^a is the field $\mathcal{L}_\xi\,\eta^a$. It follows that $\mathcal{L}_\xi$ and $\mathcal{L}_\eta$ commute iff $[\xi,\eta] = \mathbf{0}$. (Hint: First prove the assertion, in order, for scalar fields α and contravariant fields α^a. It will then be clear how to continue with covariant fields α_a and arbitrary tensor fields $\alpha^{a\ldots b}_{c\ldots d}$.)

Although it is important to know how Lie derivatives are defined, in practice one rarely makes direct reference to the definition. Instead, one invokes propositions 1.6.1, 1.6.2, and 1.6.4. In fact, Lie derivatives can be fully characterized in terms of the properties listed there.

PROPOSITION 1.6.5. *Let ξ^a be a smooth vector field on M. Let $\mathcal{D}$ be an operator taking smooth tensor fields on M to smooth tensor fields on M of the same index structure that satisfies the following three conditions.*

(1) *For all smooth scalar fields α on M, $\mathcal{D}(\alpha) = \xi(\alpha)$.*
(2) *For all smooth vector fields λ^a on M, $\mathcal{D}(\lambda^a) = [\xi,\lambda]^a$.*
(3) *$\mathcal{D}$ commutes with the operations of addition, index substitution, and contraction; it further satisfies the Leibniz rule with respect to tensor multiplication.*

Then $\mathcal{D} = \mathcal{L}_\xi$; i.e., $\mathcal{D}$ and $\mathcal{L}_\xi$ have the same action on all smooth tensor fields.

Proof. We are assuming outright that $\mathcal{D}$ and $\mathcal{L}_\xi$ have the same action on scalar field and contravariant vector fields. We must show that (3) induces agreement on tensor fields of all other index structures. Consider, first, the case of a field γ_a. Given any smooth field λ^a on M, we must have $\mathcal{D}(\gamma_a\,\lambda^a) = \xi(\gamma_a\,\lambda^a) = \pounds_\xi(\gamma_a\,\lambda^a)$ by (1). Hence, by (3),

$$\gamma_a\,\mathcal{D}(\lambda^a) + \mathcal{D}(\gamma_a)\lambda^a = \gamma_a\,\pounds_\xi(\lambda^a) + \pounds_\xi(\gamma_a)\lambda^a.$$

But $\mathcal{D}(\lambda^a) = \pounds_\xi(\lambda^a)$ by (2). So, for arbitrary smooth fields λ^a on M, $\big(\mathcal{D}(\gamma_a) - \pounds_\xi(\gamma_a)\big)\lambda^a = 0$. Thus, $\mathcal{D}(\gamma_a) = \pounds_\xi(\gamma_a)$.

We can now jump to the general case of a smooth tensor field $\lambda^{a_1\ldots a_r}_{b_1\ldots b_s}$ on M. We do so with an argument that is much like the one just used to handle the case of covariant vector fields. Let $\lambda^{b_1}, \ldots, \rho^{b_s}, \mu_{a_1}, \ldots, \nu_{a_r}$ be arbitrary smooth fields on M, and consider the scalar field $\alpha = \lambda^{a_1\ldots a_r}_{b_1\ldots b_s}\,\lambda^{b_1}\,\ldots\,\rho^{b_s}\,\mu_{a_1}\,\ldots\,\nu_{a_r}$. By (1), $\mathcal{D}(\alpha) = \pounds_\xi(\alpha)$. We can expand the terms $\mathcal{D}(\alpha)$ and $\pounds_\xi(\alpha)$ using the fact that both operators, $\mathcal{D}$ and $\pounds_\xi$, satisfy the Leibniz rule. The result will be an equation with $r+s+1$ terms on each side. The terms will agree completely, except that where $\mathcal{D}$ appears on the left, $\pounds_\xi$ will appear on the right. In $r+s$ terms, the operator ($\mathcal{D}$ or $\pounds_\xi$) will act on a vector field. So all these terms will cancel since $\mathcal{D}$ and $\pounds_\xi$ agree in their action on contravariant and covariant vector fields. For example, the terms

$$\lambda^{a_1\ldots a_r}_{b_1\ldots b_s}\,\mathcal{D}(\lambda^{b_1})\,\ldots\,\rho^{b_s}\,\mu_{a_1}\,\ldots\,\nu_{a_r} \quad \text{and} \quad \lambda^{a_1\ldots a_r}_{b_1\ldots b_s}\big(\pounds_\xi\,\lambda^{b_1}\big)\,\ldots\,\rho^{b_s}\,\mu_{a_1}\,\ldots\,\nu_{a_r}$$

will cancel since $\mathcal{D}(\lambda^{b_1}) = \pounds_\xi\,\lambda^{b_1}$. So we may conclude that

$$\Big[\mathcal{D}\big(\lambda^{a_1\ldots a_r}_{b_1\ldots b_s}\big) - \pounds_\xi\big(\lambda^{a_1\ldots a_r}_{b_1\ldots b_s}\big)\Big]\,\lambda^{b_1}\ldots\rho^{b_s}\,\mu_{a_1}\ldots\nu_{a_r} = \mathbf{0}$$

for all smooth fields $\lambda^{b_1}, \ldots, \rho^{b_s}, \mu_{a_1}, \ldots, \nu_{a_r}$ on M. Thus $\mathcal{D}$ and $\pounds_\xi$ agree in their action on $\lambda^{a_1\ldots a_r}_{b_1\ldots b_s}$. □

We record one more fact for future reference. For any smooth field $\lambda^{a_1\ldots a_r}_{b_1\ldots b_s}$, $\pounds_\xi\,\lambda^{a_1\ldots a_r}_{b_1\ldots b_s}$ is supposed to represent the "rate of change" of the field $\lambda^{a_1\ldots a_r}_{b_1\ldots b_s}$ relative to a standard of constancy determined by (the flow maps associated with) ξ^a. So one would expect that $\pounds_\xi\,\lambda^{a_1\ldots a_r}_{b_1\ldots b_s}$ vanishes (everywhere) iff those flow maps preserve $\lambda^{a_1\ldots a_r}_{b_1\ldots b_s}$. We make the claim precise in the following proposition. The only slightly delicate matter is the need to keep track of the domains of definition of the local flow maps.

PROPOSITION 1.6.6. *Let ξ^a and $\lambda^{a_1\ldots a_r}_{b_1\ldots b_s}$ be smooth fields on M. Then the following conditions are equivalent.*

(1) $\pounds_\xi\,\lambda^{a_1\ldots a_r}_{b_1\ldots b_s} = \mathbf{0}$ *(everywhere on M).*
(2) *For all local one–parameter groups of diffeomorphisms $\{\Gamma_t\colon U \to \Gamma_t[U]\}_{t\in I}$ generated by ξ^a, and all $t \in I$, $(\Gamma_t)^*(\lambda^{a_1\ldots a_r}_{b_1\ldots b_s}) = \lambda^{a_1\ldots a_r}_{b_1\ldots b_s}$.*

Proof. The proof is essentially the same no matter what the index structure of the field under consideration. So, for convenience, we work with a field λ^a_b. One direction ((2) $\Rightarrow$ (1)) is immediate. Let $\{\Gamma_t : U \to M\}_{t\in I}$ be any local one–parameter group of diffeomorphisms determined by ξ^a, and let p be any point in U. If (2) holds, then, in particular, $\big((\Gamma_t)^*(\lambda^a_b)\big)_{|p} = \lambda^a_{b|p}$ for all $t \in I$. Hence

$$\big(\pounds_\lambda \lambda^a_b\big)_{|p} = \lim_{t\to 0} \frac{1}{t}\Big[\big((\Gamma_t)^*(\lambda^a_b)\big)_{|p} - \lambda^a_{b|p}\Big] = \lim_{t\to 0}\frac{1}{t}[\lambda^a_{b|p} - \lambda^a_{b|p}] = \mathbf{0}.$$

The converse requires just a bit more work. Suppose that (1) holds. Let $\{\Gamma_t : U \to M\}_{t\in I}$ be any local one–parameter group of diffeomorphisms determined by ξ^a, and let p be any point in U. Further, let η_a and ρ^b and be any two vectors at p, and let $f : I \to \mathbb{R}$ be the smooth map defined by

$$f(t) = \big((\Gamma_t)^*(\lambda^a_b)\big)_{|p}\, \eta_a\, \rho^b.$$

We show that $f'(t) = 0$ for all $t \in I$. This will suffice. For then it will follow that f is constant; i.e., $\big[(\Gamma_t)^*(\lambda^a_b)\big]_{|p}\, \eta_a\, \rho^b = \big[(\Gamma_0)^*(\lambda^a_b)\big]_{|p}\, \eta_a\, \rho^b = \lambda^a_{b|p}\, \eta_a\, \rho^b$ for all $t \in I$. Hence, since η_a and ρ^a are arbitrary vectors at p, it will follow that $\big((\Gamma_t)^*(\lambda^a_b)\big)_{|p} = \lambda^a_{b|p}$ for all $t \in I$, as needed.

So let t be any number in I. Then we have

$$f'(t) = \lim_{s\to 0}\frac{1}{s}\Big[\big((\Gamma_{t+s})^*(\lambda^a_b)\big)_{|p}\, \eta_a\, \rho^b - \big((\Gamma_t)^*(\lambda^a_b)\big)_{|p}\, \eta_a\, \rho^b\Big].$$

Now suppose s is sufficiently small in absolute value that $\{s, t+s\} \subseteq I$ and $\Gamma_s(p) \in U$. Then $\Gamma_{t+s}(p) = (\Gamma_t \circ \Gamma_s)(p)$. (Recall condition (2) in the final paragraph of section 1.3.) Hence, for all such s, we have

$$\big((\Gamma_t)^*(\lambda^a_b)\big)_{|p}\, \eta_a\, \rho^b = \lambda^a_{b|\Gamma_t(p)}\, \big((\Gamma_t)_*(\eta_a)\big)_{|\Gamma_t(p)}\, \big((\Gamma_t)_*(\rho^b)\big)_{|\Gamma_t(p)}$$

and

$$\big((\Gamma_{t+s})^*(\lambda^a_b)\big)_{|p}\, \eta_a\, \rho^b = \big((\Gamma_s)^*(\lambda^a_b)\big)_{|\Gamma_t(p)}\, \big((\Gamma_t)_*(\eta_a)\big)_{|\Gamma_t(p)}\, \big((\Gamma_t)_*(\rho^b)\big)_{|\Gamma_t(p)}.$$

So, substituting into our expression for $f'(t)$, we have

$$\begin{aligned} f'(t) &= \Big[\lim_{s\to 0}\frac{1}{s}\Big(\big((\Gamma_s)^*(\lambda^a_b)\big)_{|\Gamma_t(p)} - \lambda^a_{b|\Gamma_t(p)}\Big)\Big]\, \big((\Gamma_t)_*(\eta_a)\big)_{|\Gamma_t(p)}\, \big((\Gamma_t)_*(\rho^b)\big)_{|\Gamma_t(p)} \\ &= \big(\pounds_\xi\, \lambda^a_b\big)_{|\Gamma_t(p)}\, \big((\Gamma_t)_*(\eta_a)\big)_{|\Gamma_t(p)}\, \big((\Gamma_t)_*(\rho^b)\big)_{|\Gamma_t(p)}. \end{aligned}$$

Since $\pounds_\xi\, \lambda^a_b = \mathbf{0}$ everywhere, we may conclude that $f'(t) = 0$. □

1.7. Derivative Operators and Geodesics

We have already introduced one kind of derivative operator, namely $\mathcal{L}_\lambda$, associated with a smooth contravariant vector field λ^a. In this section, we discuss a different kind. It is, in a sense, a generalization of the gradient operator ∇ that one encounters in standard vector analysis on $\mathbb{R}^n$.

Let M be a manifold, and let ∇ be a map that acts on pairs $(c, \alpha^{a_1\ldots a_r}_{b_1\ldots b_s})$, where the second is a smooth tensor field on M and the first is an abstract index distinct from $a_1 \ldots, a_r,\ b_1, \ldots, b_s$, and associates with them a smooth tensor field $\nabla_c \alpha^{a_1\ldots a_r}_{b_1\ldots b_s}$ on M in which c appears as a covariant index. (Given any one index c, we understand ∇_c to be the operator that takes the field $\alpha^{a_1\ldots a_r}_{b_1\ldots b_s}$ to the field $\nabla_c \alpha^{a_1\ldots a_r}_{b_1\ldots b_s}$.) We say that ∇ is a *(covariant) derivative operator* on M if it satisfies the following conditions.

(DO1) ∇ commutes with addition on tensor fields.
For example, $\nabla_n\big(\alpha^{ab}_c + \beta^{ba}_c\big) = \nabla_n \alpha^{ab}_c + \nabla_n \beta^{ba}_c$.

(DO2) ∇ satisfies the Leibniz rule with respect to tensor multiplication.
For example, $\nabla_n\big(\alpha^{ab}_c\, \xi_{fd}\big) = \alpha^{ab}_c \nabla_n \xi_{fd} + \big(\nabla_n \alpha^{ab}_c\big)\xi_{fd}$.

(DO3) ∇ commutes with index substitution.
For example, the result of applying $(a \to d)$ index substitution to α^{ab}_c and applying ∇_n is the same as that arising from applying $(a \to d)$ substitution to $\nabla_n \alpha^{ab}_c$. Furthermore, the result of applying $(n \to m)$ index substitution to $\nabla_n \alpha^{ab}_c$ is the same as that arising from applying ∇_m to α^{ab}_c.

(DO4) ∇ commutes with contraction.
For example, the result of applying (a, c) contraction to $\nabla_n \alpha^{ab}_c$ is the same as that arising from applying ∇_n to α^{ab}_a.

(DO5) For all smooth scalar fields α and all smooth vector fields ξ^n, $\xi^n \nabla_n \alpha = \xi(\alpha)$.

(DO6) For all (distinct) indices a and b, and for all smooth scalar fields α, $\nabla_a \nabla_b \alpha = \nabla_b \nabla_a \alpha$.

The first four conditions should seem relatively innocuous. Condition (DO5) is suggested by the situation in ordinary vector analysis on $\mathbb{R}^n$. There the directional derivative of α in the direction ξ is given by $\xi \cdot \nabla\alpha$. (Recall equation (1.2.1).) We want to interpret $\xi^n \nabla_n \alpha$ as the analog of $\xi \cdot \nabla \alpha$. So we set $\xi^n \nabla_n \alpha$ equal to the (generalized) directional derivative $\xi(\alpha)$. Condition (DO6) is a bit more delicate. One can imagine strengthening the condition to require that ∇_a and ∇_b commute on all tensor fields. This leads to the class of "flat" derivative operators and is far too restrictive for our purposes. One can also imagine dropping the condition altogether. This leads to the larger class of "derivative operators with torsion." It will be clear later why we have included (DO6). (The derivative operators determined by metrics are necessarily torsion free.)

Some authors refer to the associated maps ∇_a as "derivative operators" rather than reserving that term for the map ∇ itself. We shall do so as well, on occasion.

Having defined derivative operators, we can now pose the question of their existence, and uniqueness on manifolds. Concerning existence, one has the following basic result (Geroch [23, appendix]).

PROPOSITION 1.7.1. *A connected manifold admits a derivative operator iff a countable subset of the manifold's charts suffice to cover it.*

The restriction to connected manifolds here is harmless since, clearly, a manifold admits a derivative operator iff each of its components does. Practically all the manifolds one ever deals with in differential geometry satisfy the stated countable cover condition. Indeed, one has to work hard to find a manifold that does not. So proposition 1.7.1 has the force of a strong existence theorem. (And, of course, it implies that all manifolds admit derivative operators locally.)

The question of uniqueness is easier to deal with, and we give a complete answer. But first a lemma is needed.

LEMMA 1.7.2. *Let ∇ be a derivative operator on the n-dimensional manifold M, and let ξ_b be a co-vector at the point p. Then there is a smooth scalar field α in $\mathcal{S}(p)$ such that $\xi_b = (\nabla_b \alpha)_{|p}$.*

Proof. Here we use coordinates as in section 1.2. Suppose (U, φ) is a chart on M with $p \in U$, and $u^1, \ldots, u^n$ are the corresponding coordinate maps on U. The coordinate curve tangent vectors $\overrightarrow{\gamma}_{1|p}, \ldots, \overrightarrow{\gamma}_{n|p}$ form a basis for M_p. Let $\{\overset{1}{\beta}, \overset{2}{\beta}, \ldots, \overset{n}{\beta}\}$ be a dual basis. Then $(\overrightarrow{\gamma}_{i\,|p})\,\overset{j}{\beta} = \delta_{ij}$ for all i and j in $\{1, \ldots, n\}$, and there exist real numbers $\overset{1}{c}, \ldots, \overset{n}{c}$ such that $\xi_b = \sum_{i=1}^{n} \overset{i}{c}\, \overset{i}{\beta}_b$. Now we define a smooth scalar field $\alpha \colon U \to \mathbb{R}$ by setting $\alpha(q) = \sum_{i=1}^{n} \overset{i}{c}\, u^i(q)$. We claim $\xi_b = (\nabla_b \alpha)_{|p}$.

We must show that $\eta^b\, \xi_b = \eta^b\, (\nabla_b \alpha)_{|p}$ holds for arbitrary vectors η^b at p. Let $\eta^b = \sum_{i=1}^{n} \overset{i}{d}\, (\overrightarrow{\gamma}_{i\,|p})^b$ be one such. Then, by (DO5), and the fact that $\overrightarrow{\gamma}_{i\,|p}(u^j) = \delta_{ij}$ (recall equation (1.2.6)), we have

$$\eta^b\,(\nabla_b\,\alpha)_{|p} = \eta(\alpha)_{|p} = \left(\sum_{i=1}^{n} \overset{i}{d}\,(\overset{\rightarrow}{\gamma}_{i\,|p})\right)\left(\sum_{j=1}^{n} \overset{j}{c}\,u^j\right) = \sum_{i=1}^{n} \overset{i}{d}\,\overset{i}{c}.$$

Since we also have

$$\eta^b\,\xi_b = \left(\sum_{i=1}^{n} \overset{i}{d}\,(\overset{\rightarrow}{\gamma}_{i\,|p})^b\right)\left(\sum_{j=1}^{n} \overset{j}{c}\,\overset{j}{\beta}_b\right) = \sum_{i=1}^{n} \overset{i}{d}\,\overset{i}{c},$$

we are done. □

PROPOSITION 1.7.3. *Let ∇ and ∇' be derivative operators on the manifold M. Then there exists a smooth symmetric tensor field C^a_{bc} on M that satisfies the following condition for all smooth tensor fields $\alpha^{a_1\ldots a_r}_{b_1\ldots b_s}$ on M:*

$$(\nabla'_m - \nabla_m)\alpha^{a_1\ldots a_r}_{b_1\ldots b_s} = \alpha^{a_1\ldots a_r}_{n b_2\ldots b_s}\, C^n_{m b_1} + \ldots + \alpha^{a_1\ldots a_r}_{b_1\ldots b_{s-1} n}\, C^n_{m b_s} - \alpha^{n a_2\ldots a_r}_{b_1\ldots b_s}\, C^{a_1}_{mn} - \ldots - \alpha^{a_1\ldots a_{r-1} n}_{b_1\ldots b_s}\, C^{a_r}_{mn}. \tag{1.7.1}$$

Conversely, given any derivative operator ∇ on M and any smooth symmetric tensor field C^m_{ab} on M, if ∇' is defined by equation (1.7.1), then ∇' is also a derivative operator on M. (To get a grip on equation (1.7.1), note that for each index in $\alpha^{a_1\ldots a_r}_{b_s\ldots b_s}$ there is a corresponding term on the right. That term carries a $+$ or $-$ depending on whether the index is a subscript or superscript. In that term, the index is contracted into C^a_{bc}.)

Proof. Let ∇ and ∇' be derivative operators on M. Note first that given any smooth scalar field α on M, $\nabla'_a\,\alpha = \nabla_a\,\alpha$. (This follows from the fact that given any vector ξ^a at any point in M, $\xi^a\,\nabla'_a\,\alpha = \xi(\alpha) = \xi^a\,\nabla_a\,\alpha$.)

Next we claim that given any smooth co-vector field γ_b on M, if $\gamma_b = \mathbf{0}$ at a point p, then $\nabla'_a\,\gamma_b = \nabla_a\,\gamma_b$ at p. To see this, let ξ^b be any smooth field on M and consider the scalar field $\gamma_b\,\xi^b$. We have $\mathbf{0} = (\nabla'_a - \nabla_a)(\gamma_b\,\xi^b) = \gamma_b(\nabla'_a - \nabla_a)\xi^b + \xi^b(\nabla'_a - \nabla_a)\gamma_b$ everywhere. So, in particular, we have $\mathbf{0} = \xi^b(\nabla'_a - \nabla_a)\gamma_b$ at p. Since this is true for arbitrary ξ^a, it must be the case that $(\nabla'_a - \nabla_a)\gamma_b = \mathbf{0}$ as claimed.

It follows from the claim that given any smooth field α_b on M, the value of $(\nabla'_m - \nabla_m)\alpha_b$ at a point p is determined solely by the value of α_b itself at p. (For suppose that $\overset{1}{\alpha}_b$ and $\overset{2}{\alpha}_b$ agree at p. Then the claim is applicable to $\overset{1}{\alpha}_b - \overset{2}{\alpha}_b$, and therefore $(\nabla'_m - \nabla_m)\overset{1}{\alpha}_b = (\nabla'_m - \nabla_m)\overset{2}{\alpha}_b$ at p.)

Now we define a tensor field C^a_{bc}. Given any point p and a vector $\overset{0}{\alpha}_a$ at p, we set

$$C^n_{mb}\,\overset{0}{\alpha}_n = \left(\nabla'_m - \nabla_m\right)\alpha_b$$

where α_b is *any* smooth field on M that assumes the value $\overset{0}{\alpha}_b$ at p. (Our preliminary work shows that the choice of α_b makes no difference.) It follows immediately that C^n_{mb} satisfies $C^n_{mb}\,\alpha_n = \left(\nabla'_m - \nabla_m\right)\alpha_b$ for all smooth fields α_b and, therefore, is smooth itself.

C^a_{bc} is symmetric. To see this, consider any smooth scalar field α on M. Since $\nabla'_n\,\alpha = \nabla_n\,\alpha$, it follows that $C^n_{mb}\,\nabla_n\,\alpha = \left(\nabla'_m - \nabla_m\right)\nabla_b\,\alpha = \nabla'_m\nabla_b\,\alpha - \nabla_m\nabla_b\,\alpha = \nabla'_m\nabla'_b\,\alpha - \nabla_m\nabla_b\,\alpha$. So, by (DO6), we may conclude that $C^n_{mb}\,\nabla_n\,\alpha = C^n_{bm}\,\nabla_n\,\alpha$. But by our lemma, *all* covariant vectors at a point can be realized in the form $\nabla_n\,\alpha$ for some scalar field α. So we have $C^a_{bc} = C^a_{cb}$.

Next we show that C^a_{bc} satisfies condition (1.7.1). This involves a now familiar sequential form of argument—from scalar fields, to vector fields, to arbitrary tensor fields. We have already seen that all derivative operators agree on scalar fields. And it follows directly from our definition of C^a_{bc} that (1.7.1) holds for covariant vector fields. So let ξ^a be an arbitrary smooth contravariant field on M. Then, given any smooth field γ_a on M,

$$\begin{aligned}
\mathbf{0} &= \left(\nabla'_a - \nabla_a\right)\left(\xi^b\,\gamma_b\right) = \xi^b\left(\nabla'_a - \nabla_a\right)\gamma_b + \gamma_b\left(\nabla'_a - \nabla_a\right)\xi^b \\
&= \xi^b\,C^d_{ab}\,\gamma_d + \gamma_b\left(\nabla'_a - \nabla_a\right)\xi^b \\
&= \left[\xi^b\,C^d_{ab} + \left(\nabla'_a - \nabla_a\right)\xi^d\right]\gamma_d.
\end{aligned}$$

Since this holds for all smooth fields γ_a, it follows that $\left(\nabla'_a - \nabla_a\right)\xi^d = -\xi^b\,C^d_{ab}$, as required by (1.7.1).

To check equation (1.7.1) for tensor fields α^a_{bc}, one expands $\mathbf{0} = \left(\nabla'_m - \nabla_m\right)\left(\alpha^a_{bc}\,\xi^b\lambda^c\,\eta_a\right)$ for arbitrary fields ξ^b, λ^c, η_a and uses the known expressions for $\left(\nabla'_m - \nabla_m\right)\xi^b$, $\left(\nabla'_m - \nabla_m\right)\lambda^c$, and $\left(\nabla'_m - \nabla_m\right)\eta_a$. The calculation is straightforward. Tensor fields of arbitrary index structure can be handled similarly.

The second half of the proposition is also straightforward. □

It is worth noting that condition (DO6) entered the proof only in the demonstration that C^a_{bc} must be symmetric. If in the statement of the proposition one drops the requirement of symmetry on C^a_{bc}, then one has the appropriate formulation for derivative operators with torsion.

In what follows, if ∇' and ∇ are derivative operators on a manifold that, together with the field C^a_{bc}, satisfy condition (1.7.1), then we shall write $\nabla' = (\nabla, C^a_{bc})$. Clearly this is equivalent to $\nabla = (\nabla', -C^a_{bc})$.

We have introduced two kinds of derivative operators. The next proposition shows how the action of one can be expressed in terms of the other.

PROPOSITION 1.7.4. *Suppose ∇ is a derivative operator on the manifold M, and λ^a is a smooth vector field on M. Then for all smooth fields $\alpha^{a_1\ldots a_r}_{b_1\ldots b_s}$ on M, we have*

$$\begin{aligned}\mathcal{L}_\lambda\, \alpha^{a_1\ldots a_r}_{b_1\ldots b_s} &= \lambda^n\, \nabla_n\, \alpha^{a_1\ldots a_r}_{b_1\ldots b_s} + \alpha^{a_1\ldots a_r}_{n b_2\ldots b_s}\, \nabla_{b_1}\, \lambda^n \\ &\quad + \ldots + \alpha^{a_1\ldots a_r}_{b_1\ldots b_{s-1} n}\, \nabla_{b_s}\, \lambda^n \\ &\quad - \alpha^{n a_2\ldots a_r}_{b_1\ldots b_s}\, \nabla_n\, \lambda^{a_1} - \ldots - \alpha^{a_1\ldots a_{r-1} n}_{b_1\ldots b_s}\, \nabla_n\, \lambda^{a_r}.\end{aligned} \tag{1.7.2}$$

(Condition (1.7.2), of course, resembles (1.7.1). The difference $\mathcal{L}_\lambda\, \alpha^{a_1\ldots a_r}_{b_1\ldots b_s} - \lambda^n\, \nabla_n\, \alpha^{a_1\ldots a_r}_{b_1\ldots b_s}$ is a sum of terms, one for each index in $\alpha^{a_1\ldots a_r}_{b_1\ldots b_s}$. The terms carry a $+$ or $-$ depending on whether the associated index is a subscript or a superscript. Each term is contracted with $\nabla_a\, \lambda^b$.)

Proof. The proof is another simple sequential argument, like the one used in the preceding proof. (Note that we shall not need to invoke the definition of Lie derivatives. It will suffice to make use of the properties collected in propositions 1.6.1 and 1.6.4.)

First of all, trivially, if α is a smooth scalar field, then $\mathcal{L}_\lambda\, \alpha = \lambda(\alpha) = \lambda^n\, \nabla_n\, \alpha$. Next, suppose ξ^a is a smooth vector field. Then for arbitrary smooth scalar fields α, we have, by proposition 1.6.4),

$$\begin{aligned}\big(\mathcal{L}_\lambda\, \xi\big)(\alpha) &= \lambda\big(\xi(\alpha)\big) - \xi\big(\lambda(\alpha)\big) = \lambda\big(\xi^a\, \nabla_a\, \alpha\big) - \xi\big(\lambda^a\, \nabla_a\, \alpha\big) \\ &= \lambda^b\, \nabla_b\big(\xi^a\, \nabla_a\, \alpha\big) - \xi^b\, \nabla_b\big(\lambda^a\, \nabla_a\, \alpha\big) \\ &= \lambda^b\, \xi^a\, \nabla_b \nabla_a\, \alpha + \big(\lambda^b \nabla_b\, \xi^a\big)\nabla_a\, \alpha - \xi^b\, \lambda^a\, \nabla_b \nabla_a\, \alpha - \big(\xi^b \nabla_b\, \lambda^a\big)\nabla_a\, \alpha.\end{aligned}$$

The first and third term of the last line cancel each other by (DO6). So we have

$$\big(\mathcal{L}_\lambda\, \xi\big)(\alpha) = \big(\lambda^b\, \nabla_b\, \xi^a - \xi^b\, \nabla_b\, \lambda^a\big)\nabla_a\, \alpha = \big(\lambda^b\, \nabla_b\, \xi^a - \xi^b\, \nabla_b\, \lambda^a\big)(\alpha).$$

Since α is arbitrary, it follows that

$$\mathcal{L}_\lambda\, \xi^a = \lambda^b\, \nabla_b\, \xi^a - \xi^b\, \nabla_b\, \lambda^a,$$

as required by equation (1.7.2).

Next let α_a be a smooth covariant vector field. Then for arbitrary smooth fields ξ^a,

$$\begin{aligned}\mathcal{L}_\lambda (\alpha_a \xi^a) &= \alpha_a \mathcal{L}_\lambda \xi^a + \xi^a \mathcal{L}_\lambda \alpha_a \\ &= \alpha_a (\lambda^b \nabla_b \xi^a - \xi^b \nabla_b \lambda^a) + \xi^a \mathcal{L}_\lambda \alpha_a.\end{aligned}$$

Here we have used both the fact that $\mathcal{L}_\lambda$ satisfies the Leibniz condition and our previous expression for $\mathcal{L}_\lambda \xi^a$. But we also have

$$\mathcal{L}_\lambda (\alpha_a \xi^a) = \lambda^b \nabla_b (\alpha_a \xi^a) = \lambda^b \alpha_a \nabla_b \xi^a + \lambda^b \xi^a \nabla_b \alpha_a.$$

Therefore,

$$\xi^a \mathcal{L}_\lambda \alpha_a = \xi^a (\lambda^b \nabla_b \alpha_a + \alpha_b \nabla_a \lambda^b).$$

Since ξ^a is arbitrary, we have

$$\mathcal{L}_\lambda \alpha_a = \lambda^b \nabla_b \alpha_a + \alpha_b \nabla_a \lambda^b,$$

as required by equation (1.7.2).

Continuing this way, we can verify equation (1.7.2) for tensor fields of arbitrary index structure. □

PROBLEM 1.7.1. *Show that if* ∇ *is a derivative operator on a manifold, then* $\nabla_n \delta^b_a = \mathbf{0}$.

With the notion of a derivative operator in hand, we can now introduce the idea of "parallel transport" of tensors along curves.

Suppose M is a manifold with derivative operator ∇. The directional derivative of a scalar field α at p in the direction ξ^a, we know, is given by $\xi^n \nabla_n \alpha$. Generalizing now, we take the *directional derivative* of a smooth field $\alpha^{a_1 \ldots a_m}_{b_1 \ldots b_r}$ at p in the direction ξ^a (with respect to ∇) to be

$$\xi^n \nabla_n \alpha^{a_1 \ldots a_m}_{b_1 \ldots b_r}.$$

Furthermore, if $\gamma : I \to M$ is a smooth curve with tangent field ξ^a, we say that $\alpha^{a_1 \ldots a_m}_{b_1 \ldots b_r}$ is *constant* along γ (with respect to ∇) if $\xi^n \nabla_n \alpha^{a_1 \ldots a_m}_{b_1 \ldots b_r} = \mathbf{0}$.

Derivative operators are sometimes called "connections" (or "affine connections"). That is because, in a sense, they "connect" the tangent spaces of points "infinitesimally close" to one another, i.e., they provide a standard of identity for vectors at distinct, but "infinitesimally close" points.

So far, our tensor fields have always been defined over an entire manifold or—this amounts to the same thing—to open subsets of a manifold. It is useful also to consider tensor fields defined on curves. Suppose $\gamma : I \to M$ is a smooth curve on the manifold M. A *tensor field (of a given index structure) on* γ is just a map that assigns to each s in I a tensor of that index structure at $\gamma(s)$. (Note that this is not quite the same as assigning a tensor of that index structure to each point in $\gamma[I]$, since we are not here excluding the possibility that the curve may cross itself—i.e., that $\gamma(s_1) = \gamma(s_2)$ for distinct s_1 and s_2 in I. We do not want to insist in such a case that the tensor assigned to s_1 is the same as the one assigned to s_2.) So, for example, the tangent field to γ counts as a tensor field on γ.

It is clear what the appropriate criterion of smoothness is for tensor fields on γ. A scalar field on γ is just a map $\alpha : I \to \mathbb{R}$. So we certainly understand what it means for it to be smooth. We take a vector field ξ^a on γ to be smooth if, for all smooth scalar fields α on M, $\xi(\alpha)$ is a smooth scalar field on γ. Next, we take a co-vector field μ_a on γ to be smooth if, for all smooth fields ξ^a on M, $\xi^a \mu_a$ is a smooth scalar field on γ. One can continue in this way following the usual pattern. Note that the tangent vector field to any smooth curve qualifies as smooth.

Now suppose that $\gamma : I \to M$ is a smooth curve on the manifold M with tangent field ξ^a, $\alpha^{a_1 \ldots a_m}_{b_1 \ldots b_r}$ is a smooth field on γ, and ∇ is a derivative operator on M. We cannot meaningfully apply ∇ to $\alpha^{a_1 \ldots a_m}_{b_1 \ldots b_r}$. But we *can* make sense of the directional derivative field $\xi^n \nabla_n \alpha^{a_1 \ldots a_m}_{b_1 \ldots b_r}$ on γ. We can do so using the following proposition.

PROPOSITION 1.7.5. *Suppose ∇ is a derivative operator on the manifold M and $\gamma : I \to M$ is a smooth curve with tangent field ξ^a. Then there is a unique operator*

$$\alpha^{a_1 \ldots a_m}_{b_1 \ldots b_r} \mapsto \mathcal{D}\Big(\alpha^{a_1 \ldots a_m}_{b_1 \ldots b_r}\Big)$$

taking smooth tensor fields on γ to smooth tensor fields on γ of the same index structure that satisfies the following conditions.

(1) *$\mathcal{D}$ commutes with the operations of addition, index substitution, and contraction; it further satisfies the Leibniz rule with respect to tensor multiplication.*

(2) *For all smooth scalar fields $s \mapsto \alpha(s)$ on γ, $\mathcal{D}(\alpha) = \dfrac{d\alpha}{ds}$.*

(3) *Let $s \mapsto \alpha^{a_1 \ldots a_m}_{b_1 \ldots b_r}(s)$ be a smooth tensor field on γ. Suppose there is an open set O and a smooth field $\widetilde{\alpha}^{a_1 \ldots a_m}_{b_1 \ldots b_r}$ on O such that, for all s in some open interval $I' \subseteq I$, $\alpha^{a_1 \ldots a_m}_{b_1 \ldots b_r}(s) = \widetilde{\alpha}^{a_1 \ldots a_m}_{b_1 \ldots b_r}{}_{|\gamma(s)}$. Then, for all s in I', $\mathcal{D}\Big(\alpha^{a_1 \ldots a_m}_{b_1 \ldots b_r}\Big)(s) = \Big(\xi^n \nabla_n \widetilde{\alpha}^{a_1 \ldots a_m}_{b_1 \ldots b_r}\Big)_{|\gamma(s)}$.*

Proof. Suppose first that $\mathcal{D}$ satisfies the stated conditions, and suppose $\alpha^{a_1 \ldots a_m}_{b_1 \ldots b_r}$ is a smooth tensor field on γ. We shall derive an explicit expression for $\mathcal{D}(\alpha^{a_1 \ldots a_m}_{b_1 \ldots b_r})$ in terms of a local coordinate chart. This will show that there can be, at most, one $\mathcal{D}$ satisfying the stated conditions. To avoid drowning in indices, we shall work with a representative case—a smooth field α^a_b—but it will be clear how to adapt the argument to fields with other index structures.

Suppose our background manifold $(M, \mathcal{C})$ has dimension n. Let s be any point in I, and let (U, φ) be an n-chart in $\mathcal{C}$ whose domain U contains $\gamma(s)$. For all $i \in \{1, \ldots, n\}$, let $\overset{i}{\eta}{}^a$ be the smooth coordinate-curve tangent field $(\overrightarrow{\gamma}_i)^a$ on U. We know that the fields $\overset{1}{\eta}{}^a, \overset{2}{\eta}{}^a, \ldots, \overset{n}{\eta}{}^a$ form a basis for the tangent space at every point in U. Let $\overset{1}{\mu}_a, \overset{2}{\mu}_a, \ldots, \overset{n}{\mu}_a$ be corresponding smooth co-vector fields on U that form a dual basis at every point. Now let α^a_b be a smooth field on γ. We can certainly express it in terms of these basis and co-basis fields. That is, we can find an open subinterval $I' \subseteq I$ containing s, and smooth functions $\overset{ij}{\alpha} \colon I' \to \mathbb{R}$ such that, at all points s in I',

$$\alpha^a_b(s) = \sum_{i,j=1}^{n} \overset{ij}{\alpha}(s) (\overset{i}{\eta}{}^a)_{|\gamma(s)} (\overset{j}{\mu}_b)_{|\gamma(s)}.$$

Here, of course, $\overset{ij}{\alpha} = \alpha^a_b \, \overset{i}{\eta}{}^b \, \overset{j}{\mu}_a$. We can construe the restrictions of $\overset{i}{\eta}{}^a$ and $\overset{j}{\mu}_b$ to $\gamma[I']$ as smooth fields on (a restricted segment of) γ. It follows, therefore, that at all points in I',

$$\mathcal{D}(\alpha^a_b) = \sum_{i,j=1}^{n} \mathcal{D}\left(\overset{ij}{\alpha} \overset{i}{\eta}{}^a \, \overset{j}{\mu}_b\right) = \sum_{i,j=1}^{n} \left[\mathcal{D}(\overset{ij}{\alpha})\, (\overset{i}{\eta}{}^a \, \overset{j}{\mu}_b) + \overset{ij}{\alpha} \mathcal{D}(\overset{i}{\eta}{}^a \, \overset{j}{\mu}_b)\right].$$

Here we have just used the fact that $\mathcal{D}$ commutes with tensor addition and satisfies the Leibniz rule (and suppressed explicit reference to the evaluation point s). But now it follows from conditions (2) and (3), respectively, that $\mathcal{D}(\overset{ij}{\alpha}) = \dfrac{d\overset{ij}{\alpha}}{ds}$, and $\mathcal{D}(\overset{i}{\eta}{}^a \, \overset{j}{\mu}_b) = \xi^n \nabla_n (\overset{i}{\eta}{}^a \, \overset{j}{\mu}_b)$. So, we have our promised explicit expression for $\mathcal{D}(\alpha^a_b)$:

$$\mathcal{D}(\alpha^a_b) = \sum_{i,j=1}^{n} \left[\frac{d\overset{ij}{\alpha}}{ds} (\overset{i}{\eta}{}^a \, \overset{j}{\mu}_b) + \overset{ij}{\alpha} \xi^n \nabla_n (\overset{i}{\eta}{}^a \, \overset{j}{\mu}_b)\right].$$

To show existence, finally, it suffices to check that the operator $\mathcal{D}$ defined by this expression (and the counterpart expressions for fields with other index structures) satisfies all three conditions in the proposition. □

Under the stated conditions of the proposition, we can now understand $\xi^n \nabla_n \alpha^{a_1 \ldots a_m}_{b_1 \ldots b_r}$ to be the smooth field on γ given by $\mathcal{D}\left(\alpha^{a_1 \ldots a_m}_{b_1 \ldots b_r}\right)$. Note that condition (3) in the proposition just makes precise the requirement that $\xi^n \nabla_n \alpha^{a_1 \ldots a_m}_{b_1 \ldots b_r}$ is "what it should be" in the case where $\alpha^{a_1 \ldots a_m}_{b_1 \ldots b_r}$ arises as the restriction to $\gamma[I]$ of some smooth tensor field defined on an open set.

We have already said what it means for a tensor field defined on an open set to be "constant" along a curve γ with tangent field ξ^a. We can now extend that notion to fields $\alpha^{a_1 \ldots a_m}_{b_1 \ldots b_r}$ defined only on γ itself. The defining condition, $\xi^n \nabla_n \alpha^{a_1 \ldots a_m}_{b_1 \ldots b_r} = \mathbf{0}$ carries over intact.

The fundamental fact about constant fields on curves is the following.

PROPOSITION 1.7.6. *Given a manifold M, a derivative operator ∇ on M, a smooth curve $\gamma : I \to M$, and a tensor $\overset{0}{\alpha}{}^{a_1 \ldots a_m}_{b_1 \ldots b_r}$ at some point $\gamma(s)$, there is a unique smooth tensor field $\alpha^{a_1 \ldots a_m}_{b_1 \ldots b_r}$ on γ that is constant (with respect to ∇) and assumes the value $\overset{0}{\alpha}{}^{a_1 \ldots a_m}_{b_1 \ldots b_r}$ at s.*

When the conditions of the proposition are realized, we say that $\alpha^{a_1 \ldots a_m}_{b_1 \ldots b_r}$ results from *parallel transport* of $\overset{0}{\alpha}{}^{a_1 \ldots a_m}_{b_1 \ldots b_r}$ along γ (with respect to ∇).

Finally, we introduce "geodesics." We say that a smooth curve $\gamma : I \to M$ is a *geodesic* (with respect to ∇) if its tangent vector field ξ^a is constant along γ—i.e., if $\xi^b \nabla_b \xi^a = \mathbf{0}$. The basic existence and uniqueness theorem for geodesics is the following. (In what follows, we shall drop the qualification "with respect to ∇" except in contexts where doing so might lead to ambiguity.)

PROPOSITION 1.7.7. *Given a manifold M, a derivative operator ∇ on M, a point p in M, and a vector ξ^a at p, there is a unique geodesic $\gamma : I \to M$ with $\gamma(0) = p$ and $\xi = \overrightarrow{\gamma}_{|p}$ that satisfies the following maximality condition: if $\gamma' : I' \to M$ is also a geodesic with $\gamma'(0) = p$ and $\overrightarrow{\gamma'}_{|p} = \xi$, then $I' \subseteq I$ and $\gamma'(s) = \gamma(s)$ for all $s \in I'$.*

To prove propositions 1.7.6 and 1.7.7, one formulates the assertions in terms of local coordinates and then invokes the fundamental existence and uniqueness theorem for ordinary differential equations.

A derivative operator determines a class of geodesics. It turns out that a derivative operator is actually fully characterized by its associated geodesics. This will be important later in our discussion of relativity theory.

PROPOSITION 1.7.8. *Suppose ∇ and ∇' are both derivative operators on the manifold M. Further suppose that ∇ and ∇' admit the same geodesics (i.e., for all smooth* *curves $\gamma : I \to M$, γ is a geodesic with respect to ∇ iff it is a geodesic with respect to ∇'). Then $\nabla' = \nabla$.*

Proof. The argument provides a good example of how proposition 1.7.3 is used. Given ∇ and ∇', there must exist a smooth symmetric field C^a_{bc} on M such that $\nabla' = (\nabla, C^a_{bc})$. It will suffice to show that C^a_{bc} vanishes everywhere.

Given an arbitrary point p and an arbitrary vector $\overset{0}{\xi}{}^a$ at p, there is a geodesic γ with respect to ∇ that passes through p and has tangent $\overset{0}{\xi}{}^a$ at p. Let ξ^a be the tangent vector field of γ. Then we have $\xi^n \nabla_n \xi^a = \mathbf{0}$. By our hypothesis, γ must also be a geodesic with respect to ∇'. So $\xi^n \nabla'_n \xi^a = \mathbf{0}$ too. Now, since $\nabla'_a = (\nabla_a, C^a_{bc})$, we have $\nabla'_a \lambda^b = \nabla_a \lambda^b - C^b_{an} \lambda^n$ for all smooth fields λ^a. So, in particular, we have

$$\mathbf{0} = \xi^a \nabla'_a \xi^b = \underbrace{\xi^a \nabla_a \xi^b}_{=\mathbf{0}} - C^b_{an} \xi^a \xi^n$$

at all points on the image of γ. So $C^b_{an} \overset{0}{\xi}{}^n \overset{0}{\xi}{}^a = \mathbf{0}$ at p. But $\overset{0}{\xi}{}^a$ and p were arbitrary, and C^b_{an} is symmetric. So, by proposition 1.4.3, C^a_{bc} must vanish everywhere. □

The property of being a geodesic is not preserved under reparametrization of curves. The situation is as follows.

PROPOSITION 1.7.9. *Suppose M is a manifold with derivative operator ∇, and $\gamma : I \to M$ is a smooth curve with tangent field ξ^a. Then γ can be reparametrized so as to be a geodesic (i.e., there is a diffeomorphism $\alpha : I' \to I$ of some interval I' onto I such that $\gamma' = \gamma \circ \alpha$ is a geodesic) iff $\xi^n \nabla_n \xi^a = f\, \xi^a$ for some smooth scalar field f on γ. Furthermore, if γ is a non-trivial geodesic (i.e., a geodesic with non-vanishing tangent field), then the reparametrized curve $\gamma' = \gamma \circ \alpha$ is a geodesic iff α is linear.*

Proof. Suppose $\alpha : I' \to I$ is a diffeomorphism and ξ' is the tangent field to $\gamma' = \gamma \circ \alpha : I' \to M$. Set $t = \alpha(s)$. By the chain rule, we have $\xi' = \xi \dfrac{d\alpha}{ds}$. (This abbreviates $\xi'_{|\gamma'(s)} = \xi_{|\gamma(\alpha(s))} \dfrac{d\alpha}{ds}(s)$. Recall equation (1.3.1) in the proof of proposition 1.3.2.) Now we can construe $\dfrac{d\alpha}{ds}$ as a smooth scalar field on γ—it assigns to s the number $\dfrac{d\alpha}{ds}(s)$ at the point $\gamma(s)$—and we can make sense of the rate of change $\xi^n \nabla_n \frac{d\alpha}{ds}$. So we have

$$\xi'^n \nabla_n \xi'^a = \left(\frac{d\alpha}{ds}\,\xi^n\right) \nabla_n \left(\frac{d\alpha}{ds}\,\xi^a\right) = \left(\frac{d\alpha}{ds}\right)^2 \xi^n \nabla_n \xi^a + \left(\frac{d\alpha}{ds}\right) \xi^a \xi^n \nabla_n \frac{d\alpha}{ds}.$$

Now $\frac{d\alpha}{ds}(s) \neq 0$ for all s in I', since α is a diffeomorphism. So, by the chain rule again,

$$\xi^n \nabla_n \frac{d\alpha}{ds} = \frac{d}{dt}\left(\frac{d\alpha}{ds}\right) = \frac{d^2\alpha}{ds^2}\left(\frac{d\alpha}{ds}\right)^{-1}.$$

It follows that

$$\xi'^n \nabla_n \xi'^a = \left(\frac{d\alpha}{ds}\right)^2 \xi^n \nabla_n \xi^a + \frac{d^2\alpha}{ds^2}\,\xi^a. \tag{1.7.3}$$

Both our claims follow from this last equation. First, γ' is a geodesic, i.e., $\xi'^n \nabla_n \xi'^a = \mathbf{0}$ iff $\xi^n \nabla_n \xi^a = f\,\xi^a$, where $f = -\frac{d^2\alpha}{ds^2}\left(\frac{d\alpha}{ds}\right)^{-2}$. Second, if γ is a geodesic (i.e., if $\xi^n \nabla_n \xi^a = \mathbf{0}$), then γ' is also a geodesic iff $\frac{d^2\alpha}{ds^2}\,\xi^a = \mathbf{0}$. On the assumption that ξ^a is non-vanishing, the latter condition holds iff $\frac{d^2\alpha}{ds^2} = 0$—i.e., α is linear. □

We know that a derivative operator is determined by its associated class of geodesics. Let us now consider a different question. Suppose one does not know which (parametrized) curves are geodesics, but only which *ordered point sets* on a manifold are the *images* of geodesics. To what extent does that partial information allow one to determine the derivative operator? We answer the question in the next proposition. Let us say that two derivative operators ∇ and ∇' on a manifold are *projectively equivalent* if they admit the same geodesics up to reparametrization (i.e., if any curve can be reparametrized as to be a geodesic with respect to ∇ iff it can be reparametrized so as to be a geodesic with respect to ∇').

PROPOSITION 1.7.10. *Suppose ∇ and ∇' are derivative operators on a manifold M and $\nabla' = (\nabla, C^a_{bc})$. Then ∇ and ∇' are projectively equivalent iff there is a smooth field φ_c on M such that*

$$C^a_{bc} = \delta^a_b\,\varphi_c + \delta^a_c\,\varphi_b.$$

Proof. Suppose first that there does exist such a field φ_c. Further suppose that γ is an arbitrary smooth curve with tangent field ξ^a. Then

$$\begin{aligned}\xi^n \nabla'_n \xi^a &= \xi^n \big(\nabla_n \xi^a - C^a_{nm}\, \xi^m\big) = \xi^n \nabla_n \xi^a - \big(\delta^a_n\, \varphi_m + \delta^a_m\, \varphi_n\big)\xi^n\, \xi^m \\ &= \xi^n \nabla_n \xi^a - 2\,\xi^a (\varphi_m\, \xi^m).\end{aligned}$$

It follows by the first part of proposition 1.7.9 that γ can be reparametrized so as to be a geodesic with respect to ∇ iff it can be reparametrized so as to be a geodesic with respect to ∇'.

Conversely, suppose that ∇ and ∇' are projectively equivalent. We show there is a smooth field φ_c on M such that $C^a_{bc} = \delta^a_b\, \varphi_c + \delta^a_c\, \varphi_b$. Let γ be an arbitrary geodesic with respect to ∇ with tangent field ξ^a. Then $\xi^n \nabla_n \xi^a = \mathbf{0}$ and $\xi^n \nabla'_n \xi^a = f\, \xi^a$ for some smooth field f on γ. (Here again we use the first part of proposition 1.7.9.) It follows that

$$f\, \xi^a = \xi^b \big(\nabla_b\, \xi^a - C^a_{bc}\, \xi^c\big) = -C^a_{bc}\, \xi^b\, \xi^c.$$

Therefore, $\big(C^a_{bc}\, \xi^d - C^d_{bc}\, \xi^a\big)\xi^b\, \xi^c = \mathbf{0}$. This can be expressed as

$$\big(C^a_{bc}\, \delta^d_r - C^d_{bc}\, \delta^a_r\big)\xi^b\, \xi^c\, \xi^r = \mathbf{0}.$$

Now let φ^{ad}_{bcr} be the field $\big(C^a_{bc}\, \delta^d_r - C^d_{bc}\, \delta^a_r\big)$. Symmetrizing on the indices b, c, r, we have

$$\varphi^{ad}_{(bcr)}\, \xi^b\, \xi^c\, \xi^r = \mathbf{0}.$$

Since this equation must hold for all choices of γ, and hence all vectors ξ (at all points), and since $\varphi^{ad}_{(bcr)}$ is symmetric in b, c, r, it follows from proposition 1.4.3 that $\varphi^{ad}_{(bcr)} = \mathbf{0}$. Therefore, using the fact that C^a_{bc} is itself symmetric,

$$C^a_{bc}\, \delta^d_r - C^d_{bc}\, \delta^a_r + C^a_{rb}\, \delta^d_c - C^d_{rb}\, \delta^a_c + C^a_{cr}\, \delta^d_b - C^d_{cr}\, \delta^a_b = \mathbf{0}.$$

Now suppose n is the dimension of our underlying manifold. Then (r, d) contraction yields

$$n\, C^a_{bc} - C^a_{bc} + C^a_{cb} - C^d_{db}\, \delta^a_c + C^a_{cb} - C^d_{cd}\, \delta^a_b = \mathbf{0}.$$

Thus, $(n+1)\, C^a_{bc} = \delta^a_b\, C^d_{cd} + \delta^a_c\, C^d_{bd}$. If we set $\varphi_c = \dfrac{1}{n+1} C^d_{cd}$, this can be expressed as

$$C^a_{bc} = \delta^a_b\, \varphi_c + \delta^a_c\, \varphi_b. \qquad \square$$

We close this section with a few remarks about the "exterior derivative operator" and about "coordinate derivative operators" (associated with particular charts on a manifold).

An *m-form* (for $m \geq 0$) on a manifold M is a tensor field on M with m covariant indices that is anti-symmetric—i.e., a tensor field of the form $\alpha_{a_1 \ldots a_m}$ where $\alpha_{a_1 \ldots a_m} = \alpha_{[a_1 \ldots a_m]}$. (Scalar fields qualify as 0-forms.)

Suppose $\alpha_{a_1 \ldots a_m}$ is a smooth m-form on M, and c is an index distinct from $a_1, \ldots, a_m$. Then, given any covariant derivative operator ∇, $\nabla_{[c}\, \alpha_{a_1 \ldots a_m]}$ qualifies as a smooth $(m+1)$-form on M. It turns out that this field is independent of the choice of derivative operator ∇. (See problem 1.7.2.) In this way, we arrive at an operator d (the *exterior derivative* operator) that acts on pairs $\alpha_{a_1 \ldots a_m}$ and c, and satisfies

$$d_c\, \alpha_{a_1 \ldots a_m} = \nabla_{[c}\, \alpha_{a_1 \ldots a_m]} \tag{1.7.4}$$

for all choices of ∇. So, in particular, we have $d_a\, \alpha = \nabla_a\, \alpha$ for all smooth scalar fields α. We have $d_b\, \alpha_a = \nabla_{[b}\, \alpha_{a]} = \frac{1}{2}(\nabla_b\, \alpha_a - \nabla_a\, \alpha_b)$ for all smooth covector fields α_a. And so forth.

One can certainly introduce the exterior derivative operator directly, without reference to covariant derivative operators. Most books do so. But there is no loss in proceeding as we have, since covariant derivative operators always exist locally on manifolds, and local existence is all that is needed for our characterization.

Officially, we are taking the exterior derivative operator d to be a map that acts on a pair of objects—an index and a smooth m-form (for some m or other). One might also use the term to refer to the associated map d_c that assigns $d_c\, \alpha_{a_1 \ldots a_m}$ to $\alpha_{a_1 \ldots a_m}$. Some authors do so, and we shall too on occasion.

PROBLEM 1.7.2. *Let ∇ and ∇' be derivative operators on a manifold, and let $\alpha_{a_1 \ldots a_n}$ be a smooth n-form on it. Show that*

$$\nabla_{[b}\, \alpha_{a_1 \ldots a_n]} = \nabla'_{[b}\, \alpha_{a_1 \ldots a_n]}.$$

(Hint: Make use of proposition 1.7.3.)

It is worth asking why we do not allow the exterior derivative operator to act on arbitrary smooth covariant tensor fields. The problem is not a failure to be well defined. (Note that given *any* smooth covariant field $\alpha_{a_1 \ldots a_n}$, and any two derivative operators ∇ and ∇', it follows from problem 1.7.2 that

$$\nabla_{[b}\, \alpha_{a_1 \ldots a_n]} = \nabla_{[b}\, \alpha_{[a_1 \ldots a_n]]} = \nabla'_{[b}\, \alpha_{[a_1 \ldots a_n]]} = \nabla'_{[b}\, \alpha_{a_1 \ldots a_n]}.)$$

Rather, the problem is that we cannot *both* extend the application of the exterior derivative operator and have it satisfy the Leibniz rule—and presumably the latter is a requirement for any derivative-like operator. Here is the argument. Let α_{ab} be any smooth symmetric field. Then (if we allow ourselves to apply d_n), $d_n\alpha_{ab} = \nabla_{[n}\alpha_{ab]} = \mathbf{0}$. Now let f be any smooth scalar field. By the same argument, we have $d_n(f\alpha_{ab}) = \mathbf{0}$. So, if the Leibniz rule obtains, we have

$$\mathbf{0} = d_n(f\alpha_{ab}) = f(d_n\alpha_{ab}) + (d_nf)\,\alpha_{ab} = (d_nf)\,\alpha_{ab}.$$

But this is impossible since, given any point p, we always choose f and α_{ab} so that neither α_{ab} nor (d_nf) vanishes at p.

We have introduced three types of derivative operator on manifolds. It is helpful to contrast them with respect to two features: the background geometric structure they presuppose (if any) and the types of tensor fields to which they can be applied. One finds a trade-off of sorts. The exterior derivative operator d_a presupposes no background structure (beyond basic manifold structure). But it is only applicable to smooth m-forms (for some m or other). In contrast, the Lie derivative operator $\mathcal{L}_\xi$ and the covariant derivative operator ∇_a can both be applied to arbitrary smooth tensor fields. But the first presupposes (i.e., is defined relative to) a smooth contravariant vector field ξ; and the latter can itself be thought of as a layer of geometric structure beyond pure manifold structure. (Another difference, of course, is that $\mathcal{L}_\xi$ leaves intact the index structure of the tensor field on which it acts, whereas d_a and ∇_a both add a as a covariant index.)

Let us now consider "coordinate differentials." Let (U, φ) be an n-chart on the n-manifold $(M, \mathcal{C})$, and let $u^i\colon U \to \mathbb{R}$ $(i = 1, \ldots, n)$ be the coordinate maps on U determined by φ. We know that the associated smooth coordinate-curve tangent fields $\overrightarrow{\gamma}_1, \ldots, \overrightarrow{\gamma}_n$ form a basis for the tangent space at every point in U. (Recall the discussion in section 1.2.) The notation

$$\left(\frac{\partial}{\partial u^1}\right), \ldots, \left(\frac{\partial}{\partial u^n}\right)$$

is often used for these fields. And give any smooth scalar field f on U, the action of $\left(\dfrac{\partial}{\partial u^i}\right)$ on f is often written, simply, as $\left(\dfrac{\partial f}{\partial u^i}\right)$. Using this notation, we have, by equation (1.2.5),

$$\left(\frac{\partial f}{\partial u^i}\right)_{|p} = \left(\frac{\partial}{\partial u^i}\right)_{|p}(f) = \overrightarrow{\gamma}_{i\,|p}(f) = \left(\frac{\partial\left(f \circ \varphi^{-1}\right)}{\partial x^i}\right)_{|\varphi(p)} \tag{1.7.5}$$

for all p in U. In particular, if we take f to be u^j, it follows from equation (1.2.6) that

$$\left(\frac{\partial u^j}{\partial u^i}\right) = \left(\frac{\partial}{\partial u^i}\right)(u^j) = \overrightarrow{\gamma}_i(u^j) = \delta_{ij} \tag{1.7.6}$$

at all points in U. Furthermore, if ∇ is a derivative operator on M, we have, by condition (DO5),

$$\left(\frac{\partial}{\partial u^i}\right)^a (d_a f) = \left(\frac{\partial}{\partial u^i}\right)^a (\nabla_a f) = \left(\frac{\partial}{\partial u^i}\right)(f) = \left(\frac{\partial f}{\partial u^i}\right). \tag{1.7.7}$$

So, taking f to be u^j once again, we have

$$\left(\frac{\partial}{\partial u^i}\right)^a (d_a u^j) = \left(\frac{\partial u^j}{\partial u^i}\right) = \delta_{ij}. \tag{1.7.8}$$

This shows that the co-vectors

$$(d_a u^1), \ldots, (d_a u^n)$$

form a dual basis to $\left(\frac{\partial}{\partial u^1}\right)^a, \ldots, \left(\frac{\partial}{\partial u^n}\right)^a$ at every point in U.

Many useful facts follow from the preceding lines. For example, it follows that the index substitution field δ^a_b can be expressed as

$$\delta^a_b = \left(\frac{\partial}{\partial u^1}\right)^a (d_b u^1) + \ldots + \left(\frac{\partial}{\partial u^n}\right)^a (d_b u^n). \tag{1.7.9}$$

And it follows that, for all smooth scalar fields f on U,

$$d_b f = \sum_{j=1}^{n} \left(\frac{\partial f}{\partial u^j}\right)(d_b u^j). \tag{1.7.10}$$

(In both cases, the left- and right-side fields must be equal since they have the same action on the basis fields $\left(\frac{\partial}{\partial u^i}\right)^b$. Consider equation (1.7.10). We know from equation (1.7.7) that contraction with $\left(\frac{\partial}{\partial u^i}\right)^b$ on the left side yields $\left(\frac{\partial f}{\partial u^i}\right)$; and we know from equation (1.7.8) that contraction with $\left(\frac{\partial}{\partial u^i}\right)^b$ on the right side yields $\left(\frac{\partial f}{\partial u^i}\right)$ as well.) If we were not using the abstract index notation, we would express equation (1.7.10) in the form

$$df = \sum_{j=1}^{n} \left(\frac{\partial f}{\partial u^j}\right) du^i.$$

Next we consider "coordinate derivative operators." The basic fact is this.

PROPOSITION 1.7.11. *Let M be an n-manifold, let (U, φ) be an n-chart with non-empty domain on M (in the atlas that defines the manifold), and let $u^i : U \to \mathbb{R}$ $(i = 1, \ldots, n)$ be the coordinate maps determined by φ. Then there is a unique derivative operator ∇ on U such that $\nabla_a \left(\frac{\partial}{\partial u^i}\right)^b = \mathbf{0}$ for all i.*[2]

Proof. Uniqueness follows easily from proposition 1.7.3. Suppose ∇ and ∇' are derivative operators on U with $\nabla' = (\nabla,\ C^a_{bc})$. Then, for all i,

$$\nabla'_a \left(\frac{\partial}{\partial u^i}\right)^b = \nabla_a \left(\frac{\partial}{\partial u^i}\right)^b - C^b_{an} \left(\frac{\partial}{\partial u^i}\right)^n.$$

So if ∇ and ∇' both satisfy the stated condition, it must be the case that $C^b_{an} \left(\frac{\partial}{\partial u^i}\right)^n = \mathbf{0}$, for all i. This, in turn, implies that $C^b_{an} = \mathbf{0}$. (Somewhat more generally, if two derivative operators agree in their action on a set of vector fields that span the tangent space at each point, the derivative operators must be equal.)

We now establish existence by explicitly exhibiting a derivative operator ∇ on U that satisfies the stated condition. First, given any smooth scalar field f on U, we take $\nabla_a f$ to be the field on the right side of equation (1.7.10). (We have no choice here, since $d_a f = \nabla_a f$ for *all* derivative operators.) Next consider any smooth tensor field on U that carries at least one abstract index. It can be expressed uniquely as a sum over the basis fields $\left(\frac{\partial}{\partial u^i}\right)^a$ and $(d_a u^j)$. Consider an example. The field γ^{ab}_c can be expressed uniquely in the form

$$\gamma^{ab}_c = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \overset{ijk}{\gamma} \left(\frac{\partial}{\partial u^i}\right)^a \left(\frac{\partial}{\partial u^j}\right)^b (d_c u^k).$$

2. Here, as usual, we have suppressed explicit reference to manifold atlases. We mean, of course, that $(M, \mathcal{C})$ is an n-manifold, (U, φ) is an n-chart in $\mathcal{C}$, and ∇ is a derivative operator on the restricted manifold $(U, \mathcal{C}_{|U})$.

We take the action of ∇_m on γ_c^{ab} to be given by

$$\nabla_m \gamma_c^{ab} = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} \left(\frac{\partial \overset{ijk}{\gamma}}{\partial u^l}\right)\left(\frac{\partial}{\partial u^i}\right)^a \left(\frac{\partial}{\partial u^j}\right)^b (d_c u^k)(d_m u^l).$$

Here we introduce a new summation variable l, take the partial derivative of the scalar field $\overset{ijk}{\gamma}$ with respect to u^l, and add $(d_m u^l)$ to the list of fields on the right. This prescription can be generalized. In *every* case, we determine the action of ∇_m on a tensor field by first expressing the field as a sum over the basis fields $\left(\frac{\partial}{\partial u^i}\right)^a$ and $(d_b u^j)$, and then generating a new sum (with m as a new covariant index) in three steps: we introduce a new summation variable ι, take the partial derivative of the scalar coefficient field with respect to u^ι, and then add $(d_m u^\iota)$ to the list of fields in the sum. One can easily check that the operator so-defined satisfies conditions (DO1) through (DO6). And it is clear that $\nabla_a \left(\frac{\partial}{\partial u^i}\right)^b = \mathbf{0}$ for all i. For when we (vacuously) represent any particular field $\left(\frac{\partial}{\partial u^{i_0}}\right)^b$ in the indicated way,

$$\left(\frac{\partial}{\partial u^{i_0}}\right)^b = \sum_{i=1}^{n} \overset{i}{\alpha} \left(\frac{\partial}{\partial u^i}\right)^b,$$

the coefficients $\overset{i}{\alpha}$ are constant (either 0 or 1), and so $\frac{\partial \overset{i}{\alpha}}{\partial u^l} = 0$ for all i and l. □

We call this derivative operator—the one identified in the proposition—the *coordinate derivative operator canonically associated with* (U, φ). Sometimes, when the there is no ambiguity about the n-chart with which it is associated, the operator is written as ∂. So $\partial_a \left(\frac{\partial}{\partial u^i}\right)^b = \mathbf{0}$ for all i. As we shall see in the next section, all coordinate derivative operators are flat; i.e., their Riemann curvature fields vanish.

PROBLEM 1.7.3. *Let ∇ be the coordinate derivative operator canonically associated with (U, φ) on the n-manifold M. Let u^i be the coordinate maps on U determined by the chart. Further, let ∇' be another derivative operator on U. We know (from proposition 1.7.3) that there is a smooth field C^a_{bc} on U such that $\nabla' = (\nabla, C^a_{bc})$. Show that if*

$$C^a_{bc} = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} \overset{ijk}{C} \left(\frac{\partial}{\partial u^i}\right)^a (d_b u^j)\,(d_c u^k),$$

then a smooth vector field $\xi^a = \sum_{i=1}^{n} \overset{i}{\xi} \left(\frac{\partial}{\partial u^i}\right)^a$ *on* U *is constant with respect to* ∇' *(i.e.,* $\nabla'_a \xi^b = \mathbf{0}$*) iff*

$$\frac{\partial \overset{i}{\xi}}{\partial u^j} = \sum_{k=1}^{n} \overset{ijk}{C} \overset{k}{\xi}$$

for all i and j. (The "Christoffel symbol" Γ^i_{jk} *is often used to designate the coefficent field* $\overset{ijk}{C}$*.)*

Next, we make use of proposition 1.7.11 to prove a useful proposition about "position fields."

PROPOSITION 1.7.12. *Let* ∇ *be the coordinate derivative operator canonically associated with* (U, φ) *on the n-manifold* M*. Let* u^i *be the coordinate maps on* U *determined by the chart, and let p be a point in* U*. Then there exists a unique smooth vector field* χ^a *on* U *such that (1)* $\nabla_a \chi^b = \delta^b_a$ *and (2)* $\chi^a = \mathbf{0}$ *at p.*

Proof. (*Existence*) Consider the field χ^a defined by

$$\chi^a = \sum_{i=1}^{n} \big(u^i - u^i(p)\big)\left(\frac{\partial}{\partial u^i}\right)^a. \tag{1.7.11}$$

Clearly it satisfies condition (2) since $\big(u^i - u^i(p)\big)_{|p} = \big(u^i(p) - u^i(p)\big) = 0$. And it satisfies (1) because

$$\nabla_a \chi^b = \sum_{i=1}^{n} \nabla_a\big(u^i - u^i(p)\big)\left(\frac{\partial}{\partial u^i}\right)^b = \sum_{i=1}^{n} (\nabla_a u^i)\left(\frac{\partial}{\partial u^i}\right)^b = \delta^b_a.$$

(The first equality follows from the fact that the basis fields $\left(\frac{\partial}{\partial u^i}\right)^b$ are constant with respect to ∇; the second equality follows from the fact that (all) derivative operators annihilate all constant scalar fields; and the third equality follows from equation (1.7.9).)

(*Uniqueness*) Assume χ'^a satisfies conditions (1) and (2) as well, and consider the difference field $(\chi'^a - \chi^a)$. It is constant with respect to ∇ (because $\nabla_a(\chi'^b - \chi^b) = \delta^b_a - \delta^b_a = \mathbf{0}$), and it is the zero vector at p. So it must be the zero vector everywhere; i.e., $\chi'^a = \chi^a$. $\square$

We refer to χ^a as the *position field relative to p (associated with the coordinate derivative* ∇*).*

In the last few paragraphs we have dealt with the derivative operator ∇ canonically associated with an arbitrary n-chart (U, φ) on an arbitrary n-manifold M. Let us now consider the special case where M is the manifold $\mathbb{R}^n$, (U, φ) is the (global) n-chart where $U = \mathbb{R}^n$, and $\varphi \colon U \to \mathbb{R}^n$ is the identity map. (So $u^i = \big(x^i \circ \varphi\big) = x^i$.) In this case, we get

$$\left(\frac{\partial f}{\partial x^i}\right) = \left(\frac{\partial}{\partial x^i}\right)(f) \tag{1.7.12}$$

from equation (1.7.5). Of course, we have already encountered the fields $\left(\frac{\partial}{\partial x^i}\right)$. They were the first examples of vector fields that we considered in section 1.3. (There we used equation (1.7.12) to characterize the fields.)

Many familiar textbook assertions about "differentials" fall out as consequences of the claims we have listed. For example, the equation

$$df = \sum_{j=1}^{n} \left(\frac{\partial f}{\partial x^j}\right) dx^j$$

comes out in our notation as

$$d_b f = \sum_{j=1}^{n} \left(\frac{\partial f}{\partial x^j}\right) d_b x^j,$$

and the latter is just an instance of equation (1.7.10).

The coordinate derivative operator ∇ canonically associated with the coordinates $x^1, \ldots, x^n$ is defined on the entire manifold $\mathbb{R}^n$ (because the coordinates are). So, too, the associated positions fields χ^a (relative to particular points) are defined on the entire manifold. Note that we have encountered these position fields before as well. Suppose we take p to be the origin (i.e., suppose $x^i(p) = 0$ for all i). Then, recalling equation (1.7.11), we have

$$\chi^a = \sum_{i=1}^{n} x^i \left(\frac{\partial}{\partial x^i}\right)^a.$$

In the case $n = 2$, the right-side field is precisely what we called the "radial expansion" field in section 1.3. We can picture it as follows. Given any point q in $\mathbb{R}^2$, there is a natural isomorphism between the vector space $\mathbb{R}^2$ and the tangent space to the manifold $\mathbb{R}^2$ at q defined by

$$(x^1, x^2) \mapsto x^1 \left(\frac{\partial}{\partial x^1}\right)^a_{|q} + x^2 \left(\frac{\partial}{\partial x^2}\right)^a_{|q}.$$

If we identify these two, then we can think of $\chi^a{}_{|q}$ as just the "position vector" $\overrightarrow{oq}$ that runs from the origin o to q. (See figure 1.7.1.)

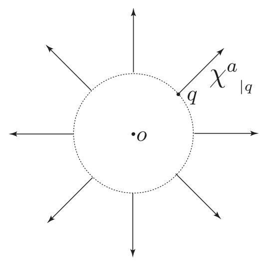

Figure 1.7.1. The position field χ^a on $\mathbb{R}^2$ (relative to point o).

1.8. Curvature

In this section we introduce the Riemann curvature tensor field R^a_{bcd} and discuss its intuitive geometric significance. We start with an existence claim.

LEMMA 1.8.1. *Suppose ∇ is a derivative operator on the manifold M. Then there is a (unique) smooth tensor field R^a_{bcd} on M such that for all smooth fields ξ^b,*

$$R^a_{bcd}\,\xi^b = -2\,\nabla_{[c}\nabla_{d]}\,\xi^a. \tag{1.8.1}$$

Proof. Uniqueness is immediate since any two fields that satisfied this condition would agree in their action on all vectors ξ^b at all points. For existence, we introduce a field R^a_{bcd} and do so in such a way that it is clear that it satisfies the required condition. Let p be any point in M and let $\overset{0}{\xi}{}^b$ be any vector at p. We define $R^a_{bcd}\,\overset{0}{\xi}{}^b$ by considering *any* smooth field ξ^b on M that assumes the value $\overset{0}{\xi}{}^b$ at p and setting $R^a_{bcd}\,\overset{0}{\xi}{}^b = -2\,\nabla_{[c}\nabla_{d]}\,\xi^a$. It suffices to verify that the choice of the field ξ^b plays no role. For this it suffices to show that if η^b is a smooth field on M that vanishes at p, then necessarily $\nabla_{[c}\nabla_{d]}\,\eta^b$ vanishes at p as well. (For then we can apply this result, taking η^b to be the difference between any two candidates for ξ^b.)

The usual argument works. Let λ_a be any smooth field on M. Then we have, by condition (DO6),

$$\begin{aligned}\mathbf{0} = \nabla_{[c}\nabla_{d]}\big(\eta^a\,\lambda_a\big) &= \big(\nabla_{[c}\,\eta^a\big)\big(\nabla_{d]}\,\lambda_a\big) + \eta^a\,\nabla_{[c}\nabla_{d]}\,\lambda_a \\ &\quad + \big(\nabla_{[c}\,\lambda_{|a|}\big)\big(\nabla_{d]}\,\eta^a\big) + \lambda_a\nabla_{[c}\nabla_{d]}\,\eta^a.\end{aligned}$$

(Note: In the third term of the final sum the vertical lines around the index indicate that it is not to be included in the anti-symmetrization.) Now the first and third terms in that sum cancel each other. And the second vanishes at

p. So we have $\mathbf{0} = \lambda_a \nabla_{[c} \nabla_{d]} \eta^a$ at p. But the field λ_a can be chosen so that it assumes any particular value at p. So $\nabla_{[c} \nabla_{d]} \eta^a = \mathbf{0}$ at p, as claimed. □

R^a_{bcd} is called the *Riemann curvature tensor field* (associated with ∇). It codes information about the degree to which the operators ∇_c and ∇_d fail to commute. Several basic properties of R^a_{bcd} are collected in the next proposition.

PROPOSITION 1.8.2. *Suppose ∇ is a derivative operator on the manifold M. Then the curvature tensor field R^a_{bcd} associated with ∇ satisfies the following conditions:*

(1) *For all smooth tensor fields $\alpha^{a_1 \ldots a_r}_{b_1 \ldots b_s}$ on M,*

$$2 \nabla_{[c} \nabla_{d]} \alpha^{a_1 \ldots a_r}_{b_1 \ldots b_s} = \alpha^{a_1 \ldots a_r}_{n b_2 \ldots b_s} R^n_{b_1 cd} + \ldots + \alpha^{a_1 \ldots a_r}_{b_1 \ldots b_{s-1} n} R^n_{b_s cd}$$
$$- \alpha^{n a_2 \ldots a_r}_{b_1 \ldots b_s} R^{a_1}_{ncd} - \ldots - \alpha^{a_1 \ldots a_{r-1} n}_{b_1 \ldots b_s} R^{a_r}_{ncd}.$$

(2) $R^a_{b(cd)} = \mathbf{0}$.
(3) $R^{\ a}_{[bcd]} = \mathbf{0}$.
(4) $\nabla_{[m} R^{\ a}_{|b|cd]} = \mathbf{0}$ *(Bianchi's identity).*

Proof. Condition (1) is proved in the now familiar way using (D06) and lemma 1.8.1. We proceed in two steps. First, we show that $2 \nabla_{[c} \nabla_{d]} \alpha_b = \alpha_n R^n_{bcd}$ for all fields α_b on M. To do so, we consider an arbitary smooth field ξ^a on M, expand $\mathbf{0} = \nabla_{[c} \nabla_{d]} (\xi^a \alpha_a)$, and invoke the lemma. Then we turn to the general case. We contract $\alpha^{a_1 \ldots a_r}_{b_1 \ldots b_s}$ with s smooth contravariant vector fields and r smooth covariant vector fields, apply $\nabla_{[c} \nabla_{d]}$, expand, and then use our previous results. Condition (2) follows immediately from lemma 1.8.1. For (3), notice that given any smooth scalar field α on M, we have, by (1),

$$R^a_{bcd} \nabla_a \alpha = 2 \nabla_{[c} \nabla_{d]} \nabla_b \alpha$$

and hence, by (D06),

$$R^{\ a}_{[bcd]} \nabla_a \alpha = 2 \nabla_{[c} \nabla_d \nabla_{b]} \alpha = \mathbf{0}.$$

Since any covariant vector at any point can be realized in the form $\nabla_a \alpha$ (recall lemma 1.7.2), it follows that $R^{\ a}_{[bcd]} = \mathbf{0}$ everywhere.

The argument for (4) is just a bit more complicated. Given any smooth field α_b on M, we have

$$2 \nabla_r \nabla_{[c} \nabla_{d]} \alpha_b = \nabla_r (R^a_{bcd} \alpha_a) = (\nabla_r R^a_{bcd}) \alpha_a + R^a_{bcd} \nabla_r \alpha_a.$$

But we also have, by (1),

$$2 \nabla_{[r} \nabla_{c]} \nabla_d \alpha_b = R^n_{drc} \nabla_n \alpha_b + R^n_{brc} \nabla_d \alpha_n.$$

If we anti-symmetrize these two equations in (r, c, d), then we have $2\,\nabla_{[r}\nabla_c\nabla_{d]}\,\alpha_b$ on the left side of both. So (equating their right sides),

$$\big(\nabla_{[r}\,R^{a}_{\ |b|cd]}\big)\alpha_a + R^{a}_{b[cd}\,\nabla_{r]}\,\alpha_a = R^{n}_{[drc]}\,\nabla_n\,\alpha_b + R^{n}_{b[rc}\,\nabla_{d]}\,\alpha_n.$$

The second term on the left here is equal to the second term on the right. So, by condition (3), we have

$$\big(\nabla_{[r}\,R^{a}_{\ |b|cd]}\big)\alpha_a = \mathbf{0}.$$

But α_a is arbitrary, and so we have (4). □

PROBLEM 1.8.1. *Let ∇ and ∇' be derivative operators on a manifold with $\nabla'_m = (\nabla_m, C^{a}_{\ bc})$, and let their respective curvature fields be $R^{a}_{\ bcd}$ and $R'^{a}_{\ bcd}$. Show that*

$$R'^{a}_{\ bcd} = R^{a}_{\ bcd} + 2\nabla_{[c}\,C^{a}_{\ d]b} + 2C^{n}_{\ b[c}\,C^{a}_{\ d]n}. \tag{1.8.2}$$

PROBLEM 1.8.2. *Show that the exterior derivative operator d on any manifold satisfies $d^2 = \mathbf{0}$; i.e., $d_n(d_m\,\alpha_{b_1\ldots b_p}) = \mathbf{0}$ for all smooth p-forms $\alpha_{b_1\ldots b_p}$. (Hint: Make use of proposition 1.8.2. Notice also that $\lambda_{[a\ldots[b\ldots c]\ldots d]} = \lambda_{[a\ldots b\ldots c\ldots d]}$ for all tensors $\lambda_{a\ldots b\ldots c\ldots d}$.)*

PROBLEM 1.8.3. *Show that given any smooth field ξ^a, and any derivative operator ∇ on a manifold, $\mathcal{L}_\xi$ commutes with ∇ (in its action on any tensor field) iff $\nabla_a\nabla_b\,\xi^m = R^{m}_{\ bna}\,\xi^n$. (Here, of course, $R^{m}_{\ bna}$ is the curvature field associated with ∇. If this conditions holds, we say that ξ^a is an "affine collineation" with respect to ∇. Hint: First show that if $K^{m}_{\ ab} = R^{m}_{\ bna}\,\xi^n - \nabla_a\nabla_b\,\xi^m$, then for all smooth fields $\alpha^{a_1\ldots a_r}_{\ b_1\ldots b_s}$,*

$$\begin{aligned}(\mathcal{L}_\xi\,\nabla_n - \nabla_n\,\mathcal{L}_\xi)\,\alpha^{a_1\ldots a_r}_{\ b_1\ldots b_s} &= \alpha^{a_1\ldots a_r}_{\ m\,b_2\ldots b_s}\,K^{m}_{\ n\,b_1} + \ldots + \alpha^{a_1\ldots a_r}_{\ b_1\ldots b_{s-1}\,m}\,K^{m}_{\ n\,b_s}\\ &\quad - \alpha^{m\,a_2\ldots a_r}_{\ b_1\ldots b_s}\,K^{a_1}_{\ n\,m} - \ldots - \alpha^{a_1\ldots a_{r-1}\,m}_{\ b_1\ldots b_s}\,K^{a_r}_{\ n\,m}.)\end{aligned}$$

PROBLEM 1.8.4. *Show that given any smooth field ξ^a on a manifold, the operators $\mathcal{L}_\xi$ and d_a commute in their action on all smooth p-forms. (Hint: Make use of the equation stated in the hint for problem 1.8.3.)*

It is not our purpose to attempt to develop systematically the theory of forms on a manifold, but we shall pause for one comment on the result stated in problem 1.8.2. Let $\alpha_{a_1\ldots a_n}$ be a smooth n-form on a manifold M with $n \geq 1$. We say that is *closed* if its exterior derivative vanishes. And we say that it is *exact* if there is a $(n-1)$-form on M of which it is the exterior derivative. (So, for example, the form α_{ab} is closed if $d_a\,\alpha_{bc} = \mathbf{0}$, and it is exact if there is a smooth form β_a such that $\alpha_{ab} = d_a\, \beta_b$.) It follows immediately from the problem that every exact form is closed. It turns out that the converse is true as well, at least locally, but the proof is non-trivial. We record the fact here for future reference.

PROPOSITION 1.8.3. *Let $\alpha_{a_1 \ldots a_n}$ be a smooth closed n-form on the manifold M with $n \geq 1$. Then, for all p in M, there is an open set O containing p, and an $(n-1)$-form $\beta_{a_1 \ldots a_{n-1}}$ on O such that $\alpha_{a_1 \ldots a_n} = d_{a_1}\, \beta_{a_2 \ldots a_n}$.*

Global assertions can also be made if M satisfies suitable conditions. If M is simply connected, for example, then all closed 1-forms are (globally) exact. And if M is contractible then, for all $n \geq 1$, all closed n-forms are (globally) exact. (See Spivak [57, volume I] for proofs of the two claims. Proposition 1.8.3 is a consequence of the second, since all manifolds are locally contractible.)

Suppose M is a manifold with derivative operator ∇ and associated curvature field $R^a_{\ bcd}$. We say that ∇ is *flat* (or that M is flat relative to ∇) if $R^a_{\ bcd}$ vanishes everywhere on M. The next proposition makes clear the intuitive geometric significance of flatness.

PROPOSITION 1.8.4. *Let ∇ be a derivative operator on the manifold M. If parallel transport of vectors on M relative to ∇ is path independent, then ∇ is flat. Conversely, if ∇ is flat, then, at least locally (i.e., within some open neighborhood of every point), parallel transport of vectors relative to ∇ is path independent. (If M is simply connected, the converse holds globally.)*

Proof. First assume that parallel transport of vectors on M is path independent. Let p be any point in M, and let $\overset{0}{\xi}{}^a$ be any vector at p. We extend $\overset{0}{\xi}{}^a$ to a smooth vector field ξ^a on all of M by parallel transporting $\overset{0}{\xi}{}^a$ (via *any* curve) to other points of M. The resulting field is constant in the sense that $\nabla_a\, \xi^b = \mathbf{0}$ everywhere. (This follows from the fact that all directional derivatives of ξ^b at all points vanish.) Hence, $R^a_{\ bcd}\, \xi^b = -2\, \nabla_{[c} \nabla_{d]}\, \xi^a = \mathbf{0}$ at all points. In particular, $R^a_{\ bcd}\, \overset{0}{\xi}{}^b = \mathbf{0}$ at p. Since $\overset{0}{\xi}{}^b$ was arbitrary, we have $R^a_{\ bcd} = \mathbf{0}$ at p.

Conversely, suppose that $R^a_{\ bcd}$ vanishes on M. To show that parallel transport on M is, at least locally, path independent, it will suffice to show that given any vector $\overset{0}{\xi}{}^a$ at point p, there is an extension of $\overset{0}{\xi}{}^a$ to a smooth field ξ^a on some open set O containing p that is constant; i.e., $\nabla_a\, \xi^b = \mathbf{0}$ everywhere in O. (For then, given any point $q \in O$, and any curve γ from p to q whose image falls within O, parallel transport of $\overset{0}{\xi}{}^a$ along γ must yield $\xi^a_{|q}$.) To see that a vector field satisfying $\nabla_a \xi^b = \mathbf{0}$ and $\xi^a_{|p} = \overset{0}{\xi}{}^a$ *does* exist locally, one writes out these two conditions in terms of local coordinates and generates a set of partial differential equations. These equations, it turns out, have a solution if a certain "integrability condition" is satisfied. That condition, is nothing but the equation $R^a_{bcd} = \mathbf{0}$ expressed in local coordinates. (For further details, see, for example, Spivak [57], volume 2, chapter 4.) □

We know from proposition 1.7.11, that given any n-chart (U, φ) (with non-empty domain) on an n-manifold, there is a unique derivative operator ∇ on U such that $\nabla_a \left(\frac{\partial}{\partial u^i}\right)^b = \mathbf{0}$ for all i. (Here $u^1, \ldots, u^n$ are the coordinate maps on U determined by (U, φ).) We called it the "coordinate derivative operator canonically associated with (U, φ)." It follows immediately, of course, that

$$R^a_{bcd} \left(\frac{\partial}{\partial u^i}\right)^b = -2\, \nabla_{[c}\, \nabla_{d]} \left(\frac{\partial}{\partial u^i}\right)^a = \mathbf{0}$$

for all i. This, in turn, implies that $R^a_{bcd} = \mathbf{0}$, since the fields $\left(\frac{\partial}{\partial u^1}\right)^b, \ldots, \left(\frac{\partial}{\partial u^n}\right)^b$ span the tangent space at every point. Thus we see that *coordinate derivative operators canonically associated with local charts are flat.*

The geometric significance of the curvature tensor field can also be explicated in terms of "geodesic deviation." Suppose ξ^a is a smooth vector field on the manifold M whose integral curves are geodesics with respect to ∇. (We shall say that ξ^a is a *geodesic field* with respect to ∇.) Further suppose that λ^a is a smooth field that satisfies $\mathcal{L}_\xi\, \lambda^a = \mathbf{0}$. Then we can think of the restriction of λ^a to an integral curve γ of ξ^a as a field that connects γ to an "infinitesimally close" integral curve γ'. If we do, the second derivative field $\xi^n\, \nabla_n (\xi^m\, \nabla_m\, \lambda^a)$ along γ represents the "relative acceleration" of γ' with respect to γ. The following proposition shows how this field can be expressed in terms of the Riemann curvature field.

PROPOSITION 1.8.5. *Suppose ξ^a is a geodesic field on the manifold M with respect to ∇. Further suppose λ^a is a smooth field that satisfies $\mathcal{L}_\xi\, \lambda^a = \mathbf{0}$. Then*

$$\xi^n\, \nabla_n (\xi^m\, \nabla_m \lambda^a) = R^a_{bcd}\, \xi^b\, \lambda^c\, \xi^d. \tag{1.8.3}$$

Proof. We have $\xi^n \nabla_n \xi^a = \mathbf{0}$ (since ξ^a is a geodesic field) and $\xi^n \nabla_n \lambda^a = \lambda^n \nabla_n \xi^a$ (since $\mathcal{L}_\xi \lambda^a = \mathbf{0}$). The rest is just a calculation.

$$\begin{aligned}
\xi^n \nabla_n(\xi^m \nabla_m \lambda^a) &= \xi^n \nabla_n(\lambda^m \nabla_m \xi^a) = (\xi^n \nabla_n \lambda^m)\nabla_m \xi^a + \xi^n \lambda^m \nabla_n \nabla_m \xi^a \\
&= (\xi^n \nabla_n \lambda^m)\nabla_m \xi^a + \xi^n \lambda^m \nabla_m \nabla_n \xi^a + \xi^n \lambda^m R^a_{rmn} \xi^r \\
&= (\xi^n \nabla_n \lambda^m)\nabla_m \xi^a + \lambda^m \nabla_m(\xi^n \nabla_n \xi^a) - (\lambda^m \nabla_m \xi^n)\nabla_n \xi^a \\
&\quad + \xi^n \lambda^m R^a_{rmn} \xi^r \\
&= R^a_{rmn} \xi^r \lambda^m \xi^n.
\end{aligned}$$

(The third equality follows from $R^a_{rmn} \xi^r = -2 \nabla_{[m} \nabla_{n]} \xi^a$. The final one follows from the fact that in the sum before the equality sign, the second term is $\mathbf{0}$, and the first and third terms cancel each other.) □

PROPOSITION 1.8.6. *Suppose ∇ is a derivative operator on the manifold M. Then ∇ is flat iff all geodesic deviation on M (with respect to ∇) vanishes; i.e., given any smooth geodesic field ξ^a, and any smooth field λ^a such that $\mathcal{L}_\xi \lambda^a = \mathbf{0}$, $\xi^n \nabla_n(\xi^m \nabla_m \lambda^a) = \mathbf{0}$.*

Proof. The "only if" half follows immediately from proposition 1.8.5. So suppose that all geodesic deviation vanishes. Then, given any vectors $\overset{0}{\xi}{}^a$ and $\overset{0}{\lambda}{}^a$ at a point p, it must be the case that $R^a_{bcd}\, \overset{0}{\xi}{}^b \overset{0}{\lambda}{}^c \overset{0}{\xi}{}^d = \mathbf{0}$. (We can always choose field ξ^a and λ^a on an open set containing p such that ξ^a is a geodesic field, $\mathcal{L}_\xi \lambda^a = \mathbf{0}$, and ξ^a and λ^a assume the values $\overset{0}{\xi}{}^a$ and $\overset{0}{\lambda}{}^a$ at p, respectively.) Equivalently, it must be the case that $R^a_{bcd}\, \overset{0}{\xi}{}^b \overset{0}{\xi}{}^d = \mathbf{0}$ for all vectors $\overset{0}{\xi}{}^b$ at p. Our conclusion now follows by the symmetries of the Riemann tensor recorded as conditions (2) and (3) in proposition 1.8.2. By (2), first, it follows that $R^a_{bcd}\, \overset{0}{\xi}{}^b \overset{0}{\xi}{}^c = \mathbf{0}$ for all vectors $\overset{0}{\xi}{}^b$ at p. Hence, by proposition 1.4.3,

$$R^{\ a}_{(bc)d} = \mathbf{0} \tag{1.8.4}$$

at p. Next, by (2) and (3), we have (everywhere)

$$R^a_{bcd} + R^a_{dbc} + R^a_{cdb} = \mathbf{0}.$$

But condition (2) and equation (1.8.4) jointly imply

$$R^a_{bcd} = R^a_{dbc} = R^a_{cdb}$$

at p. So $R^a_{bcd} = \mathbf{0}$ at our arbitrary point p. □

Equation (1.8.3) is called the *equation of geodesic deviation*. Notice that it must be the second derivative field $\xi^n \nabla_n(\xi^m \nabla_m \lambda^a)$ that enters the equation, and not the first derivative field $\xi^n \nabla_n \lambda^a$. The latter is unconstrained by the curvature of the manifold. It can assume any value at a point.

1.9. Metrics

A (*semi-Riemannian*) *metric* on a manifold M is a smooth field g_{ab} on M that is symmetric and invertible; i.e., there exists an (inverse) field g^{bc} on M such that $g_{ab}\, g^{bc} = \delta^c_a$.

It is easy to check that the inverse field g^{bc} of a metric g_{ab} is symmetric and unique. It is symmetric since

$$g^{cb} = g^{nb}\, \delta^c_n = g^{nb}\,(g_{nm}\, g^{mc}) = (g_{mn}\, g^{nb})\, g^{mc} = \delta^b_m\, g^{mc} = g^{bc}.$$

(Here we use the symmetry of g_{nm} for the third equality.) It is unique because if g'^{bc} is also an inverse field, then

$$g'^{bc} = g'^{nc}\, \delta^b_n = g'^{nc}\,(g_{nm}\, g^{mb}) = (g_{mn}\, g'^{nc})\, g^{mb} = \delta^c_m\, g^{mb} = g^{cb} = g^{bc}.$$

(Here again we use the symmetry of g_{nm} for the third equality; and we use the symmetry of g^{cb} for the final equality.) One can also check that the inverse field g^{bc} of a metric g_{ab} is smooth. This follows, essentially, because given any invertible square matrix A (over $\mathbb{R}$), the components of the inverse matrix A^{-1} depend smoothly on the components of A.

The requirement that a metric be invertible can be given a second formulation. Indeed, given any field g_{ab} on the manifold M (not necessarily symmetric and not necessarily smooth), the following conditions are equivalent.

(1) There is a tensor field g^{bc} on M such that $g_{ab}\, g^{bc} = \delta^c_a$.
(2) For all p in M, and all vectors ξ^a at p, if $g_{ab}\, \xi^a = \mathbf{0}$, then $\xi^a = \mathbf{0}$.

(When the conditions obtain, we say that g_{ab} is *non-degenerate*.) To see this, assume first that (1) holds. Then given any vector ξ^a at any point p, if $g_{ab}\, \xi^a = \mathbf{0}$, it follows that $\xi^c = \delta^c_a\, \xi^a = g^{bc}\, g_{ab}\, \xi^a = \mathbf{0}$. Conversely, suppose that (2) holds. Then at any point p, the map from $(M_p)^a$ to $(M_p)_b$ defined by $\xi^a \mapsto g_{ab}\, \xi^a$ is an injective linear map. Since $(M_p)^a$ and $(M_p)_b$ have the same dimension, it must be surjective as well. So the map must have an inverse g^{bc} defined by $g^{bc}(g_{ab}\, \xi^a) = \xi^c$ or $g^{bc}\, g_{ab} = \delta^c_a$.

In the presence of a metric g_{ab}, it is customary to adopt a notation convention for "lowering and raising indices." Consider first the case of vectors. Given a contravariant vector ξ^a at some point, we write $g_{ab}\, \xi^a$ as ξ_b; and given a covariant vector η_b, we write $g^{bc}\,\eta_b$ as η^c. The notation is evidently consistent in the sense that first lowering and then raising the index of a vector (or vice versa) leaves the vector intact.

One would like to extend this notational convention to tensors with more complex index structure. But now one confronts a problem. (It was mentioned in passing in section 1.4.) Given a tensor α^{ab}_c at a point, for example, how should we write $g^{mc}\,\alpha^{ab}_c$? As α^{mab}? Or as α^{amb}? Or as α^{abm}? In general, these three tensors will not be equal. To get around the problem, we introduce a new convention. In any context where we may want to lower or raise indices, we shall write indices, whether contravariant or covariant, in a particular sequence. So, for example, we shall write $\alpha^{ab}{}_c$ or $\alpha^a{}_c{}^b$ or $\alpha_c{}^{ab}$. (These tensors may be equal—they belong to the same vector space—but they need not be.) Clearly this convention solves our problem. We write $g^{mc}\,\alpha^{ab}{}_c$ as α^{abm}; $g^{mc}\,\alpha^a{}_c{}^b$ as α^{amb}; and so forth. No ambiguity arises. (And it is still the case that if we first lower an index on a tensor and then raise it (or vice versa), the result is to leave the tensor intact.)

We claimed in the preceding paragraph that the tensors $\alpha^{ab}{}_c$ and $\alpha^a{}_c{}^b$ (at some point) need not be equal. Here is an example. (It is just a variant of the one used in section 1.4 to show that the tensors α^{ab} and α^{ba} need not be equal.) Suppose $\overset{1}{\xi}{}^a, \overset{2}{\xi}{}^a, \ldots, \overset{n}{\xi}{}^a$ is a basis for the tangent space at a point p. Further suppose $\alpha^{abc} = \overset{i}{\xi}{}^a\,\overset{j}{\xi}{}^b\,\overset{k}{\xi}{}^c$ at the point. Then $\alpha^{acb} = \overset{i}{\xi}{}^a\,\overset{j}{\xi}{}^c\,\overset{k}{\xi}{}^b$ there. Hence, lowering indices, we have $\alpha^{ab}{}_c = \overset{i}{\xi}{}^a\,\overset{j}{\xi}{}^b\,\overset{k}{\xi}{}_c$ but $\alpha^a{}_c{}^b = \overset{i}{\xi}{}^a\,\overset{j}{\xi}{}_c\,\overset{k}{\xi}{}^b$ at p. These two will not be equal unless $j = k$.

We have reserved special notation for two tensor fields: the index substitution field δ^a_b and the Riemann curvature field R^a_{bcd} (associated with some derivative operator). Our convention will be to write these as $\delta^a{}_b$ and $R^a{}_{bcd}$—i.e., with contravariant indices before covariant ones. As it turns out, the order does not matter in the case of the first since $\delta^a{}_b = \delta_b{}^a$. (It *does* matter with the second.) To verify the equality, it suffices to observe that the two fields have the same action on an arbitrary field α^b:

$$\delta_b{}^a\,\alpha^b = (g_{bn}\,g^{am}\,\delta^n{}_m)\,\alpha^b = g_{bn}\,g^{an}\,\alpha^b = g_{bn}\,g^{na}\,\alpha^b = \delta^a{}_b\,\alpha^b.$$

Similarly we can verify (if we are raising and lowering indices with g_{ab}) that $\delta^a{}_b = g^a{}_b$ and $\delta_{ab} = g_{ab}$. (We shall take these different equalities for granted in what follows.)

Now suppose g_{ab} is a metric on the n-dimensional manifold M and p is a point in M. Then there exists an m, with $0 \leq m \leq n$, and a basis $\overset{1}{\xi}{}^a, \overset{2}{\xi}{}^a, \ldots, \overset{n}{\xi}{}^a$

for the tangent space at p such that

$$\begin{aligned} g_{ab}\overset{i}{\xi}{}^a\overset{i}{\xi}{}^b &= +1 \qquad \text{if} \quad 1 \le i \le m,\\ g_{ab}\overset{i}{\xi}{}^a\overset{i}{\xi}{}^b &= -1 \qquad \text{if} \quad m < i \le n,\\ g_{ab}\overset{i}{\xi}{}^a\overset{j}{\xi}{}^b &= 0 \qquad \text{if} \quad i \ne j. \end{aligned}$$

Such a basis is called *orthonormal*. Orthonormal bases at p are not unique, but all have the same associated number m. We call the pair $(m, n-m)$ the *signature* of g_{ab} at p. (The existence of orthonormal bases and the invariance of the associated number m are basic facts of linear algebraic life. See, for example, Lang [36].) A simple continuity argument shows that any connected manifold must have the same signature at each point. In what follows we shall restrict attention to connected manifolds and refer simply to the "signature of g_{ab}."

A metric with signature $(n, 0)$ is said to be *positive definite*. With signature $(0, n)$, it is said to be *negative definite*. With any other signature it is said to be *indefinite*. One case will be of special interest to us later. A *Lorentzian metric* is a metric with signature $(1, n-1)$. The mathematics of relativity theory is, to some degree, just a chapter in the theory of four-dimensional manifolds with Lorentzian metrics.

Suppose g_{ab} has signature $(m, n-m)$, and $\overset{1}{\xi}{}^a, \overset{2}{\xi}{}^a, \ldots, \overset{n}{\xi}{}^a$ is an orthonormal basis at a point. Further, suppose μ^a and ν^a are vectors there. If $\mu^a = \sum_{i=1}^n \overset{i}{\mu}\,\overset{i}{\xi}{}^a$ and $\nu^a = \sum_{i=1}^n \overset{i}{\nu}\,\overset{i}{\xi}{}^a$, then it follows from the linearity of g_{ab} that

$$g_{ab}\,\mu^a\,\nu^b = \overset{1}{\mu}\,\overset{1}{\nu} + \ldots + \overset{m}{\mu}\,\overset{m}{\nu} - \overset{m+1}{\mu}\,\overset{m+1}{\nu} - \ldots - \overset{n}{\mu}\,\overset{n}{\nu}. \tag{1.9.1}$$

In the special case where the metric is positive definite, this comes to

$$g_{ab}\,\mu^a\,\nu^b = \overset{1}{\mu}\,\overset{1}{\nu} + \ldots + \overset{n}{\mu}\,\overset{n}{\nu}. \tag{1.9.2}$$

And where it is Lorentzian,

$$g_{ab}\,\mu^a\,\nu^b = \overset{1}{\mu}\,\overset{1}{\nu} - \overset{2}{\mu}\,\overset{2}{\nu} - \ldots - \overset{n}{\mu}\,\overset{n}{\nu}. \tag{1.9.3}$$

So far we have introduced metrics and derivative operators as independent objects. But, in a quite natural sense, a metric determines a unique derivative operator.

Suppose g_{ab} and ∇ are both defined on the manifold M. Further suppose $\gamma : I \to M$ is a smooth curve on M with tangent field ξ^a and λ^a is a smooth field on γ. Both ∇ and g_{ab} determine a criterion of "constancy" for λ^a. λ^a is constant with respect to ∇ if $\xi^n\,\nabla_n\lambda^a = \mathbf{0}$ and is constant with respect to g_{ab} if $g_{ab}\,\lambda^a\,\lambda^b$ is constant along γ—i.e., if $\xi^n\,\nabla_n\big(g_{ab}\,\lambda^a\,\lambda^b\big) = 0$. It seems natural to consider pairs g_{ab} and ∇ for which the first condition of constancy implies the second.

Let us say that ∇ is *compatible* with g_{ab} if, for all γ and λ^a as above, λ^a is constant with respect to g_{ab} whenever it is constant with respect to ∇. The next lemma gives the condition a more economical formulation.

LEMMA 1.9.1. *Suppose ∇ is a derivative operator, and g_{ab} is a metric, on the manifold M. Then ∇ is compatible with g_{ab} iff $\nabla_a g_{bc} = \mathbf{0}$.*

Proof. Suppose γ is an arbitrary smooth curve with tangent field ξ^a and λ^a is an arbitrary smooth field on γ satisfying $\xi^n \nabla_n \lambda^a = \mathbf{0}$. Then

$$\begin{aligned}\xi^n \nabla_n (g_{ab} \lambda^a \lambda^b) &= g_{ab} \lambda^a \underbrace{\xi^n \nabla_n \lambda^b}_{=\mathbf{0}} + g_{ab} \lambda^b \underbrace{\xi^n \nabla_n \lambda^a}_{=\mathbf{0}} + \lambda^a \lambda^b \xi^n \nabla_n g_{ab} \\ &= \lambda^a \lambda^b \xi^n \nabla_n g_{ab}.\end{aligned}$$

Suppose first that $\nabla_n g_{ab} = \mathbf{0}$. Then it follows immediately that $\xi^n \nabla_n (g_{ab} \lambda^a \lambda^b) = 0$. So ∇ is compatible with g_{ab}. Suppose next that ∇ is compatible with g_{ab}. Then for *all* choices of γ and λ^a (satisfying $\xi^n \nabla_n \lambda^a = \mathbf{0}$), we have $\lambda^a \lambda^b \xi^n \nabla_n g_{ab} = 0$. Since the choice of λ^a (at any particular point) is arbitrary and g_{ab} is symmetric, it follows (by proposition 1.4.3) that $\xi^n \nabla_n g_{ab} = \mathbf{0}$. But this must be true for arbitrary ξ^a (at any particular point), and so we have $\nabla_n g_{ab} = \mathbf{0}$. □

Note that the condition of compatibility is also equivalent to $\nabla_a g^{bc} = \mathbf{0}$. To see this, recall (problem 1.7.1) that $\nabla_a \delta^m{}_n = \mathbf{0}$. Hence,

$$\begin{aligned}\mathbf{0} = g^{bn} \nabla_a \delta^c{}_n &= g^{bn} \nabla_a (g_{nr} g^{rc}) = g^{bn} g_{nr} \nabla_a g^{rc} + g^{bn} g^{rc} \nabla_a g_{nr} \\ &= \delta^b{}_r \nabla_a g^{rc} + g^{bn} g^{rc} \nabla_a g_{nr} = \nabla_a g^{bc} + g^{bn} g^{rc} \nabla_a g_{nr}.\end{aligned}$$

So if $\nabla_a g_{bc} = \mathbf{0}$, it follows immediately that $\nabla_a g^{bc} = \mathbf{0}$. Conversely, if $\nabla_a g^{bc} = \mathbf{0}$, then $g^{bn} g^{rc} \nabla_a g_{nr} = \mathbf{0}$. And therefore,

$$\mathbf{0} = g_{pb}\, g_{sc}\, g^{bn} g^{rc} \nabla_a g_{nr} = \delta^n{}_p \delta^r{}_s \nabla_a g_{nr} = \nabla_a g_{ps}.$$

The basic fact about compatible derivative operators is the following.

PROPOSITION 1.9.2. *Suppose g_{ab} is a metric on the manifold M. Then there is a unique derivative operator on M that is compatible with g_{ab}.*

—1
0
+1

Proof. To prove that M admits any derivative operator *at all* is a bit involved, and we skip the argument. (See Geroch [23]. It turns out that if a manifold admits a metric, then it necessarily satisfies the countable cover condition (M5) that we considered in section 1.1. And the latter, as noted in proposition 1.7.1, guarantees the existence of a derivative operator.) We *do* prove that if M admits a derivative operator ∇, then it admits exactly one ∇' that is compatible with g_{ab}.

Every derivative operator ∇' on M can be realized as $\nabla' = (\nabla, C^a{}_{bc})$, where $C^a{}_{bc}$ is a smooth, symmetric field on M. Now

$$\nabla'_a g_{bc} = \nabla_a g_{bc} + g_{nc}\, C^n{}_{ab} + g_{bn}\, C^n{}_{ac} = \nabla_a g_{bc} + C_{cab} + C_{bac}.$$

So ∇' will be compatible with g_{ab} (i.e., $\nabla'_a g_{bc} = \mathbf{0}$) iff

$$\nabla_a g_{bc} = -C_{cab} - C_{bac}. \tag{1.9.4}$$

Thus it suffices for us to prove that there exists a unique smooth, symmetric field $C^a{}_{bc}$ on M satisfying equation (1.9.4). To do so, we write equation (1.9.4) twice more after permuting the indices:

$$\nabla_c g_{ab} = -C_{bca} - C_{acb},$$

$$\nabla_b g_{ac} = -C_{cba} - C_{abc}.$$

If we subtract these two from the first equation, and use the fact that C_{abc} is symmetric in (b, c), we get

$$C_{abc} = \frac{1}{2}\big(\nabla_a g_{bc} - \nabla_b g_{ac} - \nabla_c g_{ab}\big), \tag{1.9.5}$$

and, therefore,

$$C^a{}_{bc} = \frac{1}{2}\, g^{an}\,\big(\nabla_n g_{bc} - \nabla_b g_{nc} - \nabla_c g_{nb}\big). \tag{1.9.6}$$

This establishes uniqueness. But clearly the field $C^a{}_{bc}$ defined by equation (1.9.6) is smooth, symmetric, and satisfies equation (1.9.4). So we have existence as well. □

In the case of positive definite metrics, there is another way to capture the significance of compatibility of derivative operators with metrics. Suppose the metric g_{ab} on M is positive definite and $\gamma : [s_1, s_2] \to M$ is a smooth curve on M.[3] We associate with γ a *length*

3. Officially (in section 1.2), we have taken a "smooth curve on M" to be a smooth map of the form $\gamma : I \to M$ where I is an open (possibly infinite or half infinite) interval in $\mathbb{R}$. Let us now agree to extend the definition and allow for the possibility that the interval I is not open. In this case, we take γ

$$|\gamma| = \int_{s_1}^{s_2} \left(g_{ab}\,\xi^a\,\xi^b\right)^{\frac{1}{2}}\,ds,$$

where ξ^a is the tangent field to γ. This assigned length is invariant under reparametrization. For suppose $\sigma : [t_1, t_2] \to [s_1, s_2]$ is a diffeomorphism (we shall write $s = \sigma(t)$) and ξ'^a is the tangent field of $\gamma' = \gamma \circ \sigma : [t_1, t_2] \to M$. Then $\xi'^a = \xi^a \dfrac{ds}{dt}$. (Recall equation (1.3.1) in the proof of proposition 1.3.2.) We may as well require that the reparametrization preserve the orientation of the original curve—i.e., require that $\sigma(t_1) = s_1$ and $\sigma(t_2) = s_2$. In this case, $\dfrac{ds}{dt} > 0$ everywhere. (Only small changes are needed if we allow the reparametrization to reverse the orientation of the curve. In that case, $\dfrac{ds}{dt} < 0$ everywhere.) It follows that

$$\begin{aligned}|\gamma'| &= \int_{t_1}^{t_2} \left(g_{ab}\,\xi'^a\,\xi'^b\right)^{\frac{1}{2}} dt = \int_{t_1}^{t_2} \left(g_{ab}\,\xi^a\,\xi^b\right)^{\frac{1}{2}}\,\frac{ds}{dt}dt \\ &= \int_{s_1}^{s_2} \left(g_{ab}\,\xi^a\,\xi^b\right)^{\frac{1}{2}}\,ds = |\gamma|.\end{aligned}$$

Let us say that $\gamma : I \to M$ is a *curve from p to q* if I is of the form $[s_1, s_2]$, $p = \gamma(s_1)$, and $q = \gamma(s_2)$. In this (positive definite) case, we take the *distance* from p to q to be

$$d(p, q) = \text{g.l.b.}\{|\gamma| : \gamma \text{ is a smooth curve from } p \text{ to } q\}.$$

Further, we say that a curve $\gamma : I \to M$ is *minimal* if, for all $s \in I$, there exists an $\varepsilon > 0$ such that, for all $s_1, s_2 \in I$ with $s_1 \leq s \leq s_2$, if $s_2 - s_1 < \varepsilon$ and if $\gamma' = \gamma_{|[s_1,s_2]}$ (the restriction of γ to $[s_1, s_2]$), then $|\gamma'| = d(\gamma(s_1), \gamma(s_2))$. Intuitively, minimal curves are "locally shortest curves." Certainly they need not be "shortest curves" outright. (Consider, for example, two points on the "equator" of a two-sphere that are not antipodal to one another. An equatorial curve running from one to the other the "long way" qualifies as a minimal curve.)

One can characterize the unique derivative operator compatible with a positive definite metric g_{ab} in terms of the latter's associated minimal curves. But in doing so, one has to pay attention to parametrization.

---

to be smooth if there is an open interval $I' \subseteq \mathbb{R}$, with $I \subseteq I'$, and a smooth map $\gamma' : I' \to M$, such that $\gamma'(s) = \gamma(s)$ for all $s \in I$. And in this case, of course, we obtain the "tangent field of γ" by restricting that of γ' to I. Furthermore, if $\sigma : I' \to I$ is a bijection betweeen (not necessarily open) intervals in $\mathbb{R}$, we understand it to be a diffeomorphism if σ and σ^{-1} are both smooth in the sense just given.

Let us say that a smooth curve $\gamma : I \to M$ with tangent field ξ^a is *parametrized by arc length* if for all ξ^a, $g_{ab}\,\xi^a\,\xi^b = 1$. In this case, if $I = [s_1, s_2]$, then

$$|\gamma| = \int_{s_1}^{s_2} \left(g_{ab}\,\xi^a\,\xi^b\right)^{\frac{1}{2}} ds = \int_{s_1}^{s_2} 1\; ds = s_2 - s_1.$$

(Any non-trivial smooth curve can always be reparametrized by arc length.) Our characterization theorem is the following.

PROPOSITION 1.9.3. *Suppose g_{ab} is a positive definite metric on the manifold M and ∇ is a derivative operator on M. Then ∇ is compatible with g_{ab} iff for all smooth curves γ parametrized by arc length, γ is a geodesic with respect to ∇ iff it is minimal with respect to g_{ab}.*

Note that the proposition would be false if the qualification "parametrized by arc length" were dropped. The class of minimal curves is invariant under reparametrization. The class of geodesics (determined by a derivative operator) is not.

We skip the proof of proposition of 1.9.3, which involves ideas from the calculus of variations. And we assert, without further discussion at this stage, that more complicated versions of the theorem are available when the metric g_{ab} under consideration is not positive definite. (We shall later consider the Lorentzian case.)

We have already demonstrated (proposition 1.8.2) that the Riemann tensor field associated with any derivative operator exhibits several index symmetries. When the derivative operator is determined by a metric, yet further symmetries are present.

PROPOSITION 1.9.4. *Suppose g_{ab} is a metric on a manifold M, ∇ is the derivative operator on M compatible with g_{ab}, and $R^a{}_{bcd}$ is associated with ∇. Then R_{abcd} $(= g_{am}\,R^m{}_{bcd})$ satisfies the following conditions.*

(1) $R_{ab(cd)} = \mathbf{0}$.
(2) $R_{a[bcd]} = \mathbf{0}$.
(3) $R_{(ab)cd} = \mathbf{0}$.
(4) $R_{abcd} = R_{cdab}$.

Proof. Conditions (1) and (2) follow directly from clauses (2) and (3) of proposition 1.8.2. And by clause (1) of that proposition, we have, since $\nabla_a g_{bc} = \mathbf{0}$,

$$\mathbf{0} = 2\,\nabla_{[c}\nabla_{d]}\,g_{ab} = g_{nb}\,R^n{}_{acd} + g_{an}\,R^n{}_{bcd} = R_{bacd} + R_{abcd}.$$

That gives us (3). So it will suffice for us to show that clauses (1)–(3) jointly imply (4). Note first that

$$\begin{aligned} \mathbf{0} &= R_{abcd} + R_{adbc} + R_{acdb} \\ &= R_{abcd} - R_{dabc} - R_{acbd}. \end{aligned}$$

(The first equality follows from (2), and the second from (1) and (3).) So anti-symmetrization over (a, b, c) yields

$$\mathbf{0} = R_{[abc]d} - R_{d[abc]} - R_{[acb]d}.$$

The second term is $\mathbf{0}$ by clause (2) again, and $R_{[abc]d} = -R_{[acb]d}$. So we have an intermediate result:

$$R_{[abc]d} = \mathbf{0}. \tag{1.9.7}$$

Now consider the octahedron in figure 1.9.1. Using conditions (1)–(3) and equation (1.9.7), one can easily verify that the sum of the terms corresponding to each triangular face vanishes. For example, the shaded face determines the sum

$$R_{abcd} + R_{bdca} + R_{adbc} = -R_{abdc} - R_{bdac} - R_{dabc} = -3\, R_{[abd]c} = \mathbf{0}.$$

So if we add the sums corresponding to the four upper faces, and subtract the sums corresponding to the four lower faces, we get (since "equatorial" terms cancel),

$$4\, R_{abcd} - 4\, R_{cdab} = \mathbf{0}.$$

This gives us (4). □

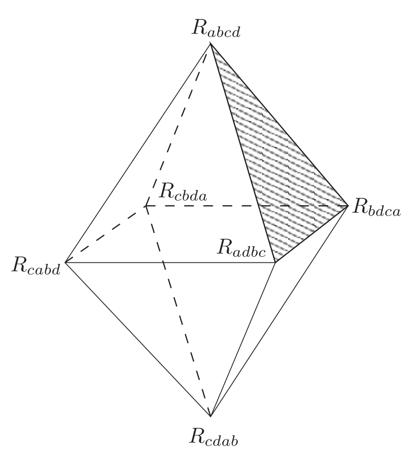

Figure 1.9.1. Symmetries of the Riemann tensor field R_{abcd}.

We say that two metrics g_{ab} and g'_{ab} on a manifold M are *projectively equivalent* if their respective associated derivative operators are projectively equivalent—i.e., if their associated derivative operators admit the same geodesics up to reparametrization. (Recall our discussion in section 1.7.) In contrast, we say that they are *conformally equivalent* if there is a map $\Omega : M \to \mathbb{R}$ such that

$$g'_{ab} = \Omega^2 \, g_{ab}.$$

Ω is called a *conformal factor*. (If such a map exists, it must be smooth and non-vanishing since both g_{ab} and g'_{ab} are.) Notice that if g_{ab} and g'_{ab} are conformally equivalent, then, given any point p and any vectors ξ^a and η^a at p, they agree on the *ratio* of their assignments to the two; i.e.,

$$\frac{g'_{ab}\, \xi^a\, \xi^a}{g'_{ab}\, \eta^a\, \eta^b} = \frac{g_{ab}\, \xi^a\, \xi^b}{g_{ab}\, \eta^a\, \eta^b}$$

(if the denominators are non-zero).

If two metrics are conformally equivalent with conformal factor Ω, then the connecting tensor field $C^a{}_{bc}$ that links their associated derivative operators can be expressed as a function of Ω.

PROPOSITION 1.9.5. *If g_{ab} and $g'_{ab} = \Omega^2 \, g_{ab}$ are metrics on the manifold M, and $\nabla' = \left(\nabla, C^a{}_{bc}\right)$, then*

$$C^a{}_{bc} = -\frac{1}{2\Omega^2}\left[\delta^a{}_b \, \nabla_c \, \Omega^2 + \delta^a{}_c \, \nabla_b \, \Omega^2 - g_{bc}\, g^{ar} \, \nabla_r \, \Omega^2\right]. \tag{1.9.8}$$

Proof. Since ∇' is compatible with g'_{ab}, it follows that

$$g'_{dr}\, C^r{}_{bc} = \frac{1}{2}\left[\nabla_d \, g'_{bc} - \nabla_b \, g'_{dc} - \nabla_c \, g'_{db}\right].$$

(Recall equation (1.9.5) in the proof of proposition 1.9.2.) If we substitute $\Omega^2 \, g_{ab}$ for g'_{ab} and use the fact that ∇ is compatible with g_{ab}, this gives us

$$\Omega^2 \, g_{dr}\, C^r{}_{bc} = \frac{1}{2}\left[g_{bc} \, \nabla_d \, \Omega^2 - g_{dc} \, \nabla_b \, \Omega^2 - g_{db} \, \nabla_c \, \Omega^2\right].$$

Contracting both sides with g^{ad} yields

$$\Omega^2 \, C^a{}_{bc} = \frac{1}{2}\left[g_{bc}\, g^{ad} \, \nabla_d \, \Omega^2 - \delta^a{}_c \, \nabla_b \, \Omega^2 - \delta^a{}_b \nabla_c \, \Omega^2\right],$$

as claimed. □

The next proposition asserts that if metrics are *both* projectively and conformally equivalent, then they can differ by—at most—a multiplicative constant.

(The converse implication is immediate.) The result will later (in section 2.1) be of crucial importance in our discussion of the physical signficance of the spacetime metric.

PROPOSITION 1.9.6. *Suppose the hypotheses of proposition 1.9.5 obtain and, in addition, g_{ab} and g'_{ab} are projectively equivalent. Further suppose that the dimension of M is at least 2. Then Ω is constant on M.*

Proof. Let the dimension of M be $n \geq 2$. We know that $C^a{}_{bc}$ must satisfy equation (1.9.8). But by proposition 1.7.10, we also have

$$C^a{}_{bc} = \delta^a{}_b\, \varphi_c + \delta^a{}_c\, \varphi_b \tag{1.9.9}$$

for some smooth field φ_c. The proof proceeds by playing off equations (1.9.8) and (1.9.9) against each other. Contracting the two equations (and using the fact that $\delta^a{}_a = n$), we get

$$C^a{}_{ba} = -\frac{1}{2\Omega^2}\left[\nabla_b\, \Omega^2 + n\, \nabla_b\, \Omega^2 - \nabla_b\, \Omega^2\right] = -\frac{n}{2\Omega^2}\nabla_b\, \Omega^2,$$

$$C^a{}_{ba} = \varphi_b + n\, \varphi_b = (n+1)\varphi_b.$$

So

$$-\frac{1}{2\Omega^2}\nabla_b\, \Omega^2 = \frac{n+1}{n}\varphi_b. \tag{1.9.10}$$

Substituting into equation (1.9.8), this yields

$$C^a{}_{bc} = \frac{n+1}{n}\left[\delta^a{}_b\, \varphi_c + \delta^a{}_c\, \varphi_b - g_{bc}\, g^{ar}\, \varphi_r\right].$$

Comparing this expression for $C^a{}_{bc}$ with equation (1.9.9), we get

$$\delta^a{}_b\, \varphi_c + \delta^a{}_c\, \varphi_b = (n+1) g_{bc}\, g^{ar}\, \varphi_r.$$

If we contract both sides with g^{bc}, we are left with

$$\varphi^a + \varphi^a = (n+1)\, n\, \varphi^a.$$

Hence, since $n \geq 2$, $\varphi^a = \mathbf{0}$. So $\nabla_b\, \Omega^2 = \mathbf{0}$, by equation (1.9.10). □

Note that in one-dimensional manifolds, *all* metrics are projectively equivalent. (All smooth curves are geodesics up to reparametrization with respect to all derivative operators.) For this reason the proposition fails if $n = 1$.

In the case where a derivative operator ∇ is determined by a metric g_{ab}, the Riemann tensor field $R^a{}_{bcd}$ associated with the former admits an instructive decomposition. Consider first the *Ricci* tensor field R_{ab} and *scalar curvature* field R defined by

$$R_{ab} = R^c{}_{abc}$$
$$R = R^a{}_a (= g^{ar} R_{ra}).$$

The first is symmetric since, by conditions (1), (3), and (4) of proposition 1.9.4,

$$R_{ab} = g^{cd} R_{dabc} = g^{cd} R_{cbad} = R_{ba}.$$

It also follows from the symmetries listed in proposition 1.9.4 that these are, up to sign, the only fields that can be obtained by contraction from $R^a{}_{bcd}$. (Contraction on any two indices yields either the zero field or $\pm R_{ab}$ and, therefore, contraction on all four indices [two at a time] yields either the zero field or $\pm R$.)

PROBLEM 1.9.1. *Let ∇ be a derivative operator on a manifold M compatible with the metric g_{ab}. Use the Bianchi identity (in proposition 1.8.2) to show that*

$$\nabla_a \left(R^{ab} - \frac{1}{2} g^{ab} R\right) = \mathbf{0}.$$

(This equation will figure later in our discussion of Einstein's equation.)

The *Weyl* (or *conformal*) tensor field C_{abcd} is defined by

(1.9.11)
$$C_{abcd} = R_{abcd} - \frac{2}{n-2}\left[g_{a[d} R_{c]b} + g_{b[c} R_{d]a}\right] - \frac{2}{(n-1)(n-2)} R\, g_{a[c}\, g_{d]b}$$

(if the dimension n of the underlying manifold is at least 3). The second and third terms on the right side exhibit symmetries (1)–(4) from proposition 1.9.4. Therefore, C_{abcd} does so as well. Furthermore, as is easily checked, $C^a{}_{bca} = \mathbf{0}$. So *all* contractions of C_{abcd} vanish. Thus equation (1.9.11) provides a decomposition of R_{abcd} in terms of R_{ab}, R, and that part of R_{abcd} whose contractions all vanish. Later we shall see that Einstein's equation in relativity theory correlates R_{ab} and R with the presence of mass-energy but does not constrain C_{abcd}. So, in a sense, the Weyl field is that part of the full Riemann curvature field that is left free by the dynamical constraints of the theory.

It turns out that the Weyl field is conformally invariant; i.e., we have the following basic result.

PROPOSITION 1.9.7. *Let g_{ab} and $g'_{ab} = \Omega^2 g_{ab}$ be metrics on a manifold with respective Weyl fields C_{abcd} and C'_{abcd}. Then $C'^{a}{}_{bcd} = C^{a}{}_{bcd}$.*

One can prove this with a laborious but straightforward calculation using problem 1.8.1 and proposition 1.9.5. (See Wald [60, pp. 446–467].)

We have said that a metric g_{ab} is *flat* if its associated Riemann tensor field R_{abcd} vanishes everywhere. In parallel, we say that it is *conformally flat* if its Weyl tensor field C_{abcd} vanishes everywhere. It follows immediately from proposition 1.9.7 (and the definition of C_{abcd}) that if a metric is conformally equivalent to a flat metric, then it is conformally flat. It turns out that the converse is true as well in manifolds of dimension at least 4. (In dimension 3, C_{abcd} vanishes automatically.)

Our next topic is "isometries" and "Killing vector fields." Given two manifolds with a metric, (M, g_{ab}) and (M', g'_{ab}), we say that a smooth map $\varphi: M \to M'$ is an *isometry* if $\varphi^*(g'_{ab}) = g_{ab}$. (Recall our discussion of "pull-back maps" in section 1.5.) This condition captures the requirement that φ preserve inner products. To see this, consider any point p in M and any two vectors ξ^a and ρ^a at p. The two have an inner product $g_{ab|p}\,\xi^a \rho^b$ at p. The push-forward map $(\varphi_p)_*$ carries them to vectors $\big((\varphi_p)_*(\xi^a)\big)$ and $\big((\varphi_p)_*(\rho^a)\big)$ at $\varphi(p)$, whose inner product there is $g'_{ab|\varphi(p)}\big((\varphi_p)_*(\xi^a)\big)\big((\varphi_p)_*(\rho^b)\big)$. In general, there is no reason why these two inner products should be equal. But they will be if $\varphi^*(g'_{ab}) = g_{ab}$, for then

$$g_{ab|p}\,\xi^a \rho^b = \big(\varphi^*(g'_{ab})\big)_{|p}\,\xi^a \rho^b = g'_{ab|\varphi(p)}\big((\varphi_p)_*(\xi^a)\big)\big((\varphi_p)_*(\rho^b)\big).$$

The second equality is just an instance of the condition (equation 1.5.2) that defines $\varphi^*(g'_{ab})$.

Now suppose λ^a is a smooth (not necessarily complete) vector field on M. We say that λ^a is a *Killing field* (with respect to g_{ab}) if $\mathcal{L}_\lambda\, g_{ab} = \mathbf{0}$ or, equivalently, if it satisfies "Killing's equation"

$$\nabla_{(a}\,\lambda_{b)} = \mathbf{0}. \tag{1.9.12}$$

(Here ∇ is understood to be the derivative operator on M compatible with g_{ab}.) Equivalence here follows from proposition 1.7.4:

$$\mathcal{L}_\lambda\, g_{ab} = \lambda^n\,\nabla_n\, g_{ab} + g_{nb}\,\nabla_a\,\lambda^n + g_{an}\,\nabla_b\,\lambda^n = \nabla_a\,\lambda_b + \nabla_b\,\lambda_a.$$

Note that

λ^a is a Killing field $\Longleftrightarrow$ the (local) flow maps determined by λ^a are isometries.

This assertion is just a special case of proposition 1.6.6, and it explains the classical description of Killing fields as "infinitesimal isometries."

The following proposition is useful when one undertakes to find or classify Killing fields.

PROPOSITION 1.9.8. *Let g_{ab} be a metric on the manifold M with associated derivative operator ∇. Further, let λ^a be a Killing field on M (with respect to g_{ab}). Then*

$$\nabla_a \nabla_b \lambda_c = -R^m{}_{abc}\, \lambda_m.$$

Proof. Given *any* smooth field λ^a on M, we have

$$2\,\nabla_{[a}\nabla_{b]}\,\lambda_c = R^m{}_{cab}\,\lambda_m,$$

$$2\,\nabla_{[c}\nabla_{a]}\,\lambda_b = R^m{}_{bca}\,\lambda_m,$$

$$2\,\nabla_{[b}\nabla_{c]}\,\lambda_a = R^m{}_{abc}\lambda_m.$$

If we subtract the third equation from the sum of the first two, and then use the fact that $\nabla_{(r}\,\lambda_{s)} = \mathbf{0}$, we get

$$\begin{aligned} 2\,\nabla_a\,\nabla_b\,\lambda_c &= (R^m{}_{cab} + R^m{}_{bca} - R^m{}_{abc})\lambda_m \\ &= 3\,R^m{}_{[abc]}\,\lambda_m - 2\,R^m{}_{abc}\,\lambda_m. \end{aligned}$$

But $R^m{}_{[abc]} = \mathbf{0}$, and so our claim follows. □

In the following problems, assume that g_{ab} is a metric on a manifold M and ∇ is its associated derivative operator.

PROBLEM 1.9.2. *Let ξ^a be a smooth vector field on M. Show that*

$$\pounds_\xi\, g^{ab} = \mathbf{0} \iff \pounds_\xi\, g_{ab} = \mathbf{0}.$$

PROBLEM 1.9.3. *Show that Killing fields on M with respect to g_{ab} are affine collineations with respect to ∇. (Recall problem 1.8.3.)*

PROBLEM 1.9.4. *Show that if ξ^a is a Killing field on M with respect to g_{ab}, then the Lie derivative operator $\pounds_\xi$ annihilates the fields R_{abcd}, R_{ab}, and R (determined by g_{ab}).*

PROBLEM 1.9.5. *Show that if ξ^a and η^a are Killing fields on M (with respect to g_{ab}), and k is a real number, then $(\xi^a + \eta^a)$, $(k\xi^a)$, and the commutator $[\xi,\eta]^a = \pounds_\xi\,\eta^a$*

are all Killing fields as well. (Thus, the set of Killing fields has the structure of a Lie algebra.)

PROBLEM 1.9.6. *Let η^a be a Killing field on M with respect to g_{ab}. (i) Let γ be a geodesic with tangent field ξ^a. Show that the function $E = \xi^a\, \eta_a$ is constant on γ. (ii) Let T^{ab} be a smooth tensor field that is symmetric and divergence free (i.e., $\nabla_a T^{ab} = \mathbf{0}$), and let J^a be the field $T^{ab}\, \eta_b$. Show that $\nabla_a J^a = 0$. (Both of these assertions will be important later in connection with our discussion of conservation principles.)*

PROBLEM 1.9.7. *A smooth field η^a on M is said to be a "conformal Killing field" (with respect to g_{ab}) if $\mathcal{L}_\eta(\Omega^2 g_{ab}) = \mathbf{0}$ for some smooth scalar field Ω. Show that if η^a is a conformal Killing field, and M has dimension n, then*

$$\nabla_{(a}\, \eta_{b)} = \frac{1}{n}\,(\nabla_c\, \eta^c)\, g_{ab}.$$

The set of Killing fields on a manifold with a metric has a natural vector space structure (problem 1.9.5). It turns out that if n is the dimension of the manifold and d is the dimension of this vector space, then $0 \leq d \leq \frac{1}{2}\, n\,(n+1)$. We will not prove this inequality but will show that "n-dimensional Euclidean space" does, in fact, admit $\frac{1}{2}\, n\,(n+1)$ linearly independent Killing fields.

Let ∇ be the flat derivative operator on the manifold $\mathbb{R}^n$ (with $n \geq 1$) canonically associated with the (globally defined) projection coordinate maps $x^1, \ldots, x^n$. (Recall our discussion toward the end of section 1.7.) We know that the basis fields $\left(\frac{\partial}{\partial x^1}\right)^a, \ldots, \left(\frac{\partial}{\partial x^n}\right)^a$ and co-basis fields $(d_a x^1), \ldots, (d_a x^n)$ are all constant with respect to ∇. We take the *Euclidean metric* on $\mathbb{R}^n$ to be the field

$$g_{ab} = (d_a x^1)(d_b x^1) + \ldots + (d_a x^n)(d_b x^n) \tag{1.9.13}$$

and take *n-dimensional Euclidean space* to be the pair $(\mathbb{R}^n, g_{ab})$. It follows that

$$g_{ab}\left(\frac{\partial}{\partial x^j}\right)^a\left(\frac{\partial}{\partial x^k}\right)^b = \sum_{i=1}^{n}(d_a x^i)\left(\frac{\partial}{\partial x^j}\right)^a(d_b x^i)\left(\frac{\partial}{\partial x^k}\right)^b = \sum_{i=1}^{n}\delta_{ij}\,\delta_{ik} = \delta_{jk}$$

for all j and k. Thus the fields $\left(\frac{\partial}{\partial x^1}\right)^a, \ldots, \left(\frac{\partial}{\partial x^n}\right)^a$ form an orthonormal basis for g_{ab} at every point, and the signature of g_{ab} is $(n, 0)$. It also follows that $\left(\frac{\partial}{\partial x^i}\right)_a = g_{an}\left(\frac{\partial}{\partial x^i}\right)^n = (d_a x^i)$ for all i. (This does not hold in general. For example, as we shall see later, when we raise and lower indices with the Minkowski metric on $\mathbb{R}^n$, $\left(\frac{\partial}{\partial x^i}\right)_a = -(d_a x^i)$ for some choices of i.) Hence

$$(d^a x^i) = g^{an}\,(d_n x^i) = g^{an}\left(\frac{\partial}{\partial x^i}\right)_n = \left(\frac{\partial}{\partial x^i}\right)^a$$

for all i and, therefore, the inverse metric field g^{ab} can be expressed in the form

$$g^{ab} = \left(\frac{\partial}{\partial x^1}\right)^a\left(\frac{\partial}{\partial x^1}\right)^b + \ldots + \left(\frac{\partial}{\partial x^n}\right)^a\left(\frac{\partial}{\partial x^n}\right)^b. \tag{1.9.14}$$

Note also that $\nabla_a g_{bc} = \mathbf{0}$—i.e., that ∇ is the unique derivative operator compatible with g_{bc}. (This follows immediately since the scalar coefficient fields on the right side of equation (1.9.13) are all constant.)

Now we proceed to find all Killing fields in n-dimensional Euclidean space. Doing so is easy given the machinery we have developed.

PROPOSITION 1.9.9. *Let ξ^a be a Killing field in n-dimensional Euclidean space $(\mathbb{R}^n, g_{ab})$ (with $n \geq 1$), let ∇ be the flat derivative operator on $\mathbb{R}^n$ canonically associated with the projection coordinate maps $x^1, \ldots, x^n$ (which is compatible with g_{ab}), let p be any point in $\mathbb{R}^n$, and let χ^a be the position field on $\mathbb{R}^n$ determined relative to p and ∇. (Recall proposition 1.7.12.) Then the following both hold.*

(1) *There exist a unique constant, anti-symmetric field F_{ab}, and a unique constant field k^a, such that*

$$\xi_b = \chi^a F_{ab} + k_b. \tag{1.9.15}$$

(Here, of course, "constant" means "constant with respect to ∇.")

(2) *The vector space of Killing fields in $(\mathbb{R}^n, g_{ab})$ has dimension $\frac{1}{2}\,n\,(n+1)$.*

Proof. (1) (*Existence*) Consider the fields $F_{ab} = \nabla_a \xi_b$, and $k_b = \xi_b - \chi^a F_{ab}$. Since ξ^a is a Killing field, $\nabla_{(a}\xi_{b)} = \mathbf{0}$. So F_{ab} is anti-symmetric. Clearly, the two fields satisfy equation 1.9.15. So what we need to show is that they are both constant with respect to ∇. F_{ab} is, since $\nabla_n F_{ab} = \nabla_n \nabla_a \xi_b = -R^m{}_{nab}\,\xi_m = \mathbf{0}$. (The second equality follows from proposition 1.9.8, and the third from the fact that ∇ is flat.) Furthermore, k_b is constant, since

$$\nabla_n k_b = \nabla_n \xi_b - \nabla_n (\chi^a F_{ab}) = F_{nb} - F_{ab} \nabla_n \chi^a = F_{nb} - F_{ab} \delta^a_{\ n} = F_{nb} - F_{nb} = \mathbf{0}.$$

(For the second equality we use the fact that $\nabla_n F_{ab} = \mathbf{0}$, and for the third that $\nabla_n \chi^a = \delta^a_{\ n}$.)

(*Uniqueness*) Assume that the fields F'_{ab} and k'_a also satisfy the stated conditions. It follows that

$$\begin{aligned} F_{ab} &= F_{nb}\, \delta^n_{\ a} = F_{nb} \nabla_a \chi^n = \nabla_a (\chi^n F_{nb} + k_b) = \nabla_a \xi_b \\ &= \nabla_a (\chi^n F'_{nb} + k'_b) \\ &= F'_{nb} \nabla_a \chi^n = F'_{nb}\, \delta^n_{\ a} = F'_{ab} \end{aligned}$$

and, therefore, $k_b = k'_b$.

(2) Let d be the dimension of the vector space of Killing fields in $(\mathbb{R}^n, g_{ab})$. It follows from part (1) that d is of the form $d = d_1 + d_2$, where d_1 is the dimension of the vector space of constant, anti-symmetric fields F_{ab} on $(\mathbb{R}^n, g_{ab})$, and d_2 is the dimension of the vector space of constant fields k^a on the manifold. Clearly, $\left(\frac{\partial}{\partial x^1}\right)^a, \ldots, \left(\frac{\partial}{\partial x^n}\right)^a$ form a basis for the latter. So $d_2 = n$. We claim that $d_1 = \frac{n(n-1)}{2}$. This will suffice, of course, for then $d = n + \frac{n(n-1)}{2} = \frac{n(n+1)}{2}$. To verify the claim, consider the expansion of any constant, anti-symmetric field F_{ab} in terms of the co-basis fields $(d_a x^1), \ldots, (d_a x^n)$. The coefficient fields are all constant (since F_{ab} is). So they determine an $n \times n$ anti-symmetric (real) matrix. (The ij^{th} entry is the coefficient of $(d_a x^i)(d_b x^j)$ in the expansion.) Thus the problem reduces to that of determining the dimension of the vector space of all $n \times n$ anti-symmetric real matrices. Since all numbers on the diagonal must be 0, and the ij^{th} and ji^{th} entries must sum to 0, the number of independent entries is just the number of ordered pairs (i, j) where $1 \leq i < j \leq n$. And *this* number is certainly $\frac{n(n-1)}{2}$. So we are done. □

Consider, for example, the case of two-dimensional Euclidean space where there should be $3 \left(= \frac{3 \times 2}{2}\right)$ linearly independent Killing fields. Here the space of constant vector fields k^a is two-dimensional and is generated by $\left(\frac{\partial}{\partial x^1}\right)^a$ and $\left(\frac{\partial}{\partial x^2}\right)^a$. The space of constant, anti-symmetric fields F_{ab} is one-dimensional and is generated by

$$F_{ab} = (d_a x^1)\,(d_b x^2) - (d_a x^2)\,(d_b x^1).$$

So the full vector space of Killing fields is generated by the three fields

$$\overset{1}{\xi}{}^b = \left(\frac{\partial}{\partial x^1}\right)^b,$$

$$\overset{2}{\xi}{}^b = \left(\frac{\partial}{\partial x^2}\right)^b,$$

$$\overset{3}{\xi}{}^b = \chi^a F_a{}^b = \left(x^1 - x^1(p)\right)\left(\frac{\partial}{\partial x^2}\right)^b - \left(x^2 - x^2(p)\right)\left(\frac{\partial}{\partial x^1}\right)^b.$$

The expression for the third is easily derived using our expression for F_{ab} and equation (1.7.11) (in the case where $u^i = x^i$):

$$\begin{aligned}
\overset{3}{\xi}{}^b &= \chi^a F_{an} g^{nb} \\
&= \left[(x^1 - x^1(p))\left(\frac{\partial}{\partial x^1}\right)^a + (x^2 - x^2(p))\left(\frac{\partial}{\partial x^2}\right)^a\right] \\
&\qquad\qquad \left[(d_a x^1)(d_n x^2) - (d_a x^2)(d_n x^1)\right] g^{nb} \\
&= \left[(x^1 - x^1(p))\left(\frac{\partial}{\partial x^1}\right)^a + (x^2 - x^2(p))\left(\frac{\partial}{\partial x^2}\right)^a\right] \\
&\qquad\qquad \left[(d_a x^1)\left(\frac{\partial}{\partial x^2}\right)^b - (d_a x^2)\left(\frac{\partial}{\partial x^1}\right)^b\right] \\
&= (x^1 - x^1(p))\left(\frac{\partial}{\partial x^2}\right)^b - (x^2 - x^2(p))\left(\frac{\partial}{\partial x^1}\right)^b.
\end{aligned}$$

The first two are the "infinitesimal generators" of horizontal and vertical translations. (See figure 1.9.2.) The third is the generator of counterclockwise rotations centered at p. If $p = (0, 0)$, the third reduces to the field $x^1\left(\frac{\partial}{\partial x^2}\right)^b - x^2\left(\frac{\partial}{\partial x^1}\right)^b$ that we have already encountered in section 1.3.

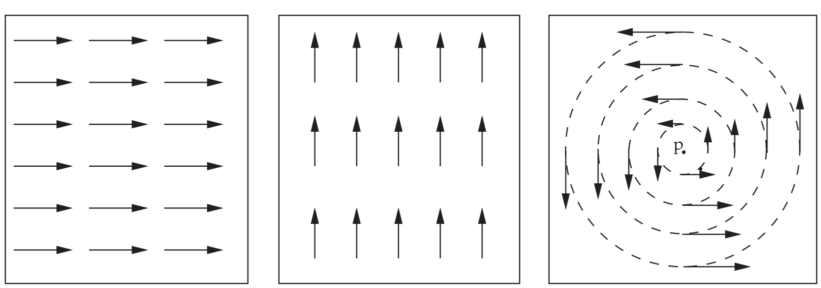

Figure 1.9.2. Killing fields in the Euclidean plane.

Finally, we briefly consider "manifolds of constant curvature," a topic that will arise when we discuss Friedmann spacetimes in section 2.11.

We say that a manifold with metric (M, g_{ab}) has *constant curvature* κ at a point in M if

$$R_{abcd} = \kappa\,(g_{ad}\,g_{bc} - g_{ac}\,g_{bd}) \tag{1.9.16}$$

holds there. (And, of course, we say that is has *constant curvature* at a point if it has constant curvature κ there for some κ.) Note that it is "possible" for equation (1.9.16) to hold only because the field $g_{abcd} = (g_{ad}\,g_{bc} - g_{ac}\,g_{bd})$ exhibits the same index symmetries as R_{abcd} (recall proposition 1.9.4):

$$R_{(ab)cd} = \mathbf{0} \qquad g_{(ab)cd} = \mathbf{0}, \tag{1.9.17}$$

$$R_{ab(cd)} = \mathbf{0} \qquad g_{ab(cd)} = \mathbf{0}, \tag{1.9.18}$$

$$R_{a[bcd]} = \mathbf{0} \qquad g_{a[bcd]} = \mathbf{0}, \tag{1.9.19}$$

$$R_{abcd} = R_{cdab} \qquad g_{abcd} = g_{cdab}. \tag{1.9.20}$$

To motivate the definition, let us temporarily assume that g_{ab} is positive-definite. (That makes things a bit easier.) Let p be a point in M and let W be a two-dimensional subspace of M_p. We take the *W-sectional curvature* of (M, g_{ab}) at p to be the number

$$\frac{R_{abcd}\,\alpha^a\,\beta^b\,\alpha^c\,\beta^d}{(g_{ad}\,g_{bc} - g_{ac}\,g_{bd})\,\alpha^a\,\beta^b\,\alpha^c\,\beta^d} \tag{1.9.21}$$

where α^a and β^a are *any* two vectors at p that span W. Note that the definition is well posed. First, the denominator cannot be 0, for that would violate our stipulation that α^a and β^a span W. (Using a more familiar notation, the point is this: if u and v are vectors such that $\langle u, v\rangle^2 = \|u\|^2\,\|v\|^2$, then u and v must be linearly dependent.) Second, the expression is independent of the choice of α^a and β^a. For suppose that $\tilde{\alpha}^a$ and $\tilde{\beta}^a$ form a basis for W as well, with $\tilde{\alpha}^a = f\,\alpha^a + g\,\beta^a$ and $\tilde{\beta}^a = h\,\alpha^a + k\,\beta^a$. Then, by equations (1.9.17) and (1.9.18),

$$R_{abcd}\,\tilde{\alpha}^a\,\tilde{\beta}^b\,\tilde{\alpha}^c\,\tilde{\beta}^d = (fk - gh)^2\,R_{abcd}\,\alpha^a\,\beta^b\,\alpha^c\,\beta^d$$

$$(g_{ad}\,g_{bc} - g_{ac}\,g_{bd})\,\tilde{\alpha}^a\,\tilde{\beta}^b\,\tilde{\alpha}^c\,\tilde{\beta}^d = (fk - gh)^2\,(g_{ad}\,g_{bc} - g_{ac}\,g_{bd})\,\alpha^a\,\beta^b\,\alpha^c\,\beta^d,$$

and the factor $(fk - gh)^2$ simply drops out.

In the special case of a smooth surface in three-dimensional Euclidean space (with the metric induced on it), the sectional curvature at any point is just what we would otherwise call the "Gaussian curvature" there. (See Spivak [57], volume 2, chapter 4.)

Now we show that constancy of curvature at a point can be understood to mean equality of sectional curvatures there.

PROPOSITION 1.9.10. *Let M be a manifold of dimension at least 2, let g_{ab} be a positive-definite metric on M, and let κ be a real number. Then*

$$R_{abcd} = \kappa\,(g_{ad}\,g_{bc} - g_{ac}\,g_{bd})$$

holds at a point iff all sectional curvatures there (i.e., all W-sectional curvatures for all two-dimensional subspaces W) are equal to κ.

Proof. The "only if" half of the assertion is immediate. For the converse, assume that all sectional curvatures are equal to κ at some point p in M. Our goal is to show that the difference tensor

$$D_{abcd} = R_{abcd} - \kappa\,(g_{ad}\,g_{bc} - g_{ac}\,g_{bd})$$

vanishes at p. Note that D_{abcd} inherits the symmetry conditions (1.9.17)–(1.9.20). Note, as well, that (i) $D_{abcd}\,\alpha^a\,\beta^b\,\alpha^c\,\beta^d = 0$ for all vectors α^a and β^a at p. For if α^a and β^a are linearly independent, the claim follows from the fact that all sectional curvatures at p are equal to κ. And if they are not linearly independent, it follows from (1.9.17) (or (1.9.18)). What we show is that D_{abcd} cannot satisfy (i) *and* the listed symmetry conditions without vanishing.

Let $\overset{1}{\mu}{}^a, \overset{2}{\mu}{}^a, \ldots, \overset{n}{\mu}{}^a$ be a basis for M_p. We claim that (ii) $D_{abcd}\,\overset{i}{\mu}{}^a\,\overset{j}{\mu}{}^b\,\overset{i}{\mu}{}^c\,\overset{k}{\mu}{}^d = 0$, for all i, j and k. This is clear, since by (i) and the symmetry (1.9.20),

$$0 = D_{abcd}\,\overset{i}{\mu}{}^a\,(\overset{j}{\mu}{}^b + \overset{k}{\mu}{}^b)\,\overset{i}{\mu}{}^c\,(\overset{j}{\mu}{}^d + \overset{k}{\mu}{}^d) = 2\,D_{abcd}\,\overset{i}{\mu}{}^a\,\overset{j}{\mu}{}^b\,\overset{i}{\mu}{}^c\,\overset{k}{\mu}{}^d.$$

We also claim that (iii) $D_{acbd} = -D_{adbc}$. For this, note that by (i) and (ii)—and by the symmetries (1.9.17), and (1.9.18), (1.9.20)—

$$\begin{aligned} 0 &= D_{abcd}\,(\overset{i}{\mu}{}^a + \overset{j}{\mu}{}^a)(\overset{k}{\mu}{}^b + \overset{l}{\mu}{}^b)(\overset{i}{\mu}{}^c + \overset{j}{\mu}{}^c)(\overset{k}{\mu}{}^d + \overset{l}{\mu}{}^d) \\ &= 2\,D_{abcd}\,(\overset{i}{\mu}{}^a\,\overset{k}{\mu}{}^b\,\overset{j}{\mu}{}^c\,\overset{l}{\mu}{}^d + \overset{i}{\mu}{}^a\,\overset{l}{\mu}{}^b\,\overset{j}{\mu}{}^c\,\overset{k}{\mu}{}^d) \\ &= 2\,(D_{acbd} + D_{adbc})\,\overset{i}{\mu}{}^a\,\overset{j}{\mu}{}^b\,\overset{k}{\mu}{}^c\,\overset{l}{\mu}{}^d. \end{aligned}$$

Since this holds for all $\overset{i}{\mu}{}^a\,\overset{j}{\mu}{}^b\,\overset{k}{\mu}{}^c\,\overset{l}{\mu}{}^d$ (and since $\overset{1}{\mu}{}^a, \overset{2}{\mu}{}^a, \ldots, \overset{n}{\mu}{}^a$ is a basis for M_p), we have (iii). Finally, it follows from (iii) and the other symmetries of D_{abcd} that

$$D_{abcd} = -D_{adcb} = D_{adbc} = -D_{acbd} = D_{adcb} + D_{abdc} = -D_{abcd} - D_{abcd}.$$

So $D_{abcd} = \mathbf{0}$. □

Now we drop our temporary assumption that we are dealing with a positive-definite metric and return to the general case.

So far, we have considered only the property of having constant curvature at a point. We say that (M, g_{ab}) has *constant curvature* if it has constant curvature at every point *and* the value of the curvature is everywhere the same. The second clause (same value at every point) needs to be added because it does not follow automatically—at least, not if M is two-dimensional. (In that special case, the property of having constant curvature at every point is vacuous and there is no reason why sectional curvatures at different points need be equal.) But, perhaps surprisingly, it *does* follow automatically if the dimension of M is at least 3.

PROPOSITION 1.9.11. (Schur's Lemma) *Let M be a manifold of dimension $n \geq 3$, and let g_{ab} be a metric on M (not necessarily positive-definite). Suppose there is a smooth scalar field κ on M such that*

$$R_{abcd} = \kappa\, (g_{ad}\, g_{bc} - g_{ac}\, g_{bd}).$$

Then κ is constant.

Proof. By Bianchi's identity (proposition 1.8.2), $\nabla_{[m} R^{ab}{}_{cd]} = \mathbf{0}$. It follows that if we apply ∇_m to $\kappa\, (\delta^a{}_d\, \delta^b{}_c - \delta^a{}_c\, \delta^b{}_d)$, and anti-symmetrize over m, c, d, we get $\mathbf{0}$. But $(\delta^a{}_d\, \delta^b{}_c - \delta^a{}_c\, \delta^b{}_d)$ is already anti-symmetric in c, d. So

$$\mathbf{0} = \nabla_{[m}\, (\kappa\, \delta^a{}_d\, \delta^b{}_{c]}) = \delta^a{}_{[d}\, \delta^b{}_c\, \nabla_{m]}\, \kappa.$$

Contracting on indices a, d and on b, c yields

$$\mathbf{0} = (n-1)\,(n-2)\,\nabla_m \kappa.$$

So (given our assumption that $n \geq 3$), we may conclude that $\nabla_m\, \kappa = \mathbf{0}$—i.e., that κ is constant on M. □

As it happens, the assertion of the proposition is also true if $n = 1$, for in that case we have (at every point) $R_{abcd} = \mathbf{0} = (g_{ad}\, g_{bc} - g_{ac}\, g_{bd})$. (Every tensor over a one-dimensional vector space vanishes if it is anti-symmetric in two indices.) The proposition fails only if $n = 2$.

Let (M, g_{ab}) and (M', g'_{ab}) be two manifolds with metric. We say they are *locally isometric* if, for all points $p \in M$ and $p' \in M'$, there exist open sets $O \subseteq$

M and $O' \subseteq M'$ containing p and p', respectively, such that the restricted manifolds $(O, g_{ab|O})$ and $(O', g'_{ab|O'})$ are isometric.

Suppose (M, g_{ab}) and (M', g'_{ab}) both have constant curvature and their respective curvature values are κ and κ'. Then, one can show, they are locally isometric iff (i) M and M' have the same dimension, (ii) g_{ab} and g'_{ab} have the same signature, and (iii) $\kappa = \kappa'$. (See Wolf [64], proposition 2.4.11.) But these conditions certainly do not guarantee that (M, g_{ab}) and (M', g'_{ab}) are (globally) isometric. (We will have more to say about this in section 2.11.)

1.10. Hypersurfaces

Let $(S, \mathcal{C}_S)$ and $(M, \mathcal{C}_M)$ be manifolds of dimension k and n, respectively, with $1 \leq k \leq n$. A smooth map $\Psi : S \to M$ is said to be an *imbedding* if it satisfies the following three conditions.

(I1) Ψ is injective.
(I2) At all points p in S, the associated (push-forward) linear map $(\Psi_p)_* : S_p \to M_{\Psi(p)}$ is injective.
(I3) For all open sets O_1 in S, $\Psi[O_1] = \Psi[S] \cap O_2$ for some open set O_2 in M. (Equivalently, the inverse map $\Psi^{-1} : \Psi[S] \to S$ is continuous with respect to the relative topology on $\Psi[S]$.)

(Recall our discussion of push-forward and pull-backward maps in section 1.5.)

Several comments about the definition are in order. First, given any point p in S, (I2) implies that $(\Psi_p)_*[S_p]$ is a k-dimensional subspace of $M_{\Psi(p)}$. So the condition *cannot* be satisfied unless $k \leq n$. Second, the three conditions are independent of one another. For example, the smooth map $\Psi : \mathbb{R} \to \mathbb{R}^2$ defined by $\Psi(s) = (\cos(s), \sin(s))$ satisfies (I2) and (I3) but is not injective. It wraps $\mathbb{R}$ round and round in a circle. On the other hand, the smooth map $\Psi : \mathbb{R} \to \mathbb{R}$ defined by $\Psi(s) = s^3$ satisfies (I1) and (I3) but is not an imbedding because $(\Psi_0)_* : \mathbb{R}_0 \to \mathbb{R}_0$ is not injective.[4] (Here $\mathbb{R}_0$ is the tangent space to the manifold $\mathbb{R}$ at the point 0). Finally, a smooth map $\Psi : S \to M$ can satisfy (I1)

4. $(\Psi_0)_*$ annihilates the vector $\dfrac{d}{dx}$ in $\mathbb{R}_0$ (and so has a non-trivial kernel). This is clear since, for any smooth real-valued function f defined on some open subset of $\mathbb{R}$ containing $\Psi(0) = 0$, we have

$$\left((\Psi_0)_*\left(\frac{d}{dx}\right)\right)(f) = \left(\frac{d}{dx}(f \circ \Psi)\right)_{|x=0} = \left(\frac{d}{dx}(f(x^3))\right)_{|x=0} = \left(f'(x^3)\,3x^2\right)_{|x=0} = 0.$$

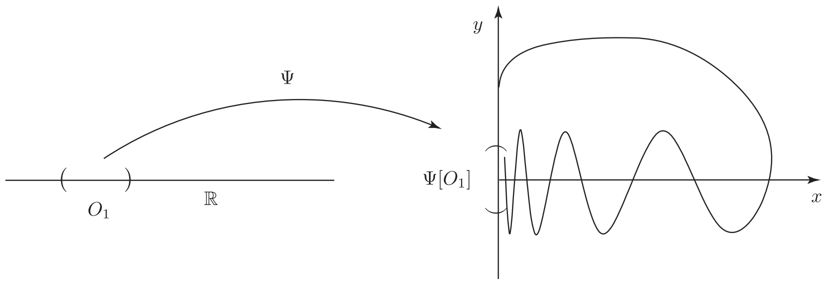

Figure 1.10.1. The map Ψ is not an imbedding, because its image bunches up on itself.

and (I2) but still have an image that "bunches up on itself." It is precisely this possibility that is ruled out by condition (I3). Consider, for example, a map $\Psi : \mathbb{R} \to \mathbb{R}^2$ whose image consists of part of the image of the curve $y = \sin(1/x)$ smoothly joined to the segment $\{(0, y) : y < 1\}$, as in figure 1.10.1. It satisfies conditions (I1) and (I2) but is not an imbedding because we can find an open interval O_1 in $\mathbb{R}$ such that given any open set O_2 in $\mathbb{R}^2$, $\Psi[O_1] \neq O_2 \cap \Psi[\mathbb{R}]$.

Suppose $(S, \mathcal{C}_S)$ and $(M, \mathcal{C}_M)$ are manifolds with $S \subseteq M$. We say that $(S, \mathcal{C}_S)$ is an *imbedded submanifold* of $(M, \mathcal{C}_M)$ if the identity map $id : S \to M$ is an imbedding. If, in addition, $k = n - 1$ (where k and n are the dimensions of the two manifolds), we say that $(S, \mathcal{C}_S)$ is a *hypersurface* in $(M, \mathcal{C}_M)$. In what follows, we first work with arbitrary imbedded submanifolds and then restrict attention to hypersurfaces. Where confusion does not arise, we suppress reference to charts.

Once and for all in this section, let $(S, \mathcal{C}_S)$ be a k-dimensional imbedded submanifold of the n-dimensional manifold $(M, \mathcal{C}_M)$, and let p be a point in S. We need to distinguish two senses in which one can speak of "tensors at p." There are tensors over the vector space S_p (call them *S-tensors at p*) and ones over the vector space M_p (call them *M-tensors at p*). So, for example, an S-vector $\tilde{\xi}^a$ at p makes assignments to maps of the form $\tilde{f} : \tilde{O} \to \mathbb{R}$ where $\tilde{O}$ *is a subset of* S that is open in the topology induced by $\mathcal{C}_S$, and $\tilde{f}$ is smooth relative to $\mathcal{C}_S$. In contrast, an M-vector ξ^a at p makes assignments to maps of the form $f : O \to \mathbb{R}$ where O *is a subset of* M that is open in the topology induced by $\mathcal{C}_M$, and f is smooth relative to $\mathcal{C}_M$.[5] Our first task is to consider the relation between S-tensors at p and M-tensors there.

5. As an aid to clarity, we shall sometimes mark S-tensors with a tilde, and sometimes we shall indicate the character of a vector ξ^a simply by indicating, explicitly, its membership in $(S_p)^a$ or $(M_p)^a$ (Co-vectors η_a shall be handled similarly.)

Let us say that $\xi^a \in (M_p)^a$ is *tangent to S* if $\xi^a \in (id_p)_*[(S_p)^a]$. (This makes sense. We know that $(id_p)_*[(S_p)^a]$ is a k-dimensional subspace of $(M_p)^a$; ξ^a either belongs to that subspace or it does not.) Let us further say that η_a in $(M_p)_a$ is *normal to S* if $\eta_a \xi^a = 0$ for all $\xi^a \in (M_p)^a$ that are tangent to S. Each of these classes of vectors has a natural vector space structure. The space of vectors $\xi^a \in (M_p)^a$ tangent to S has dimension k. The space of co-vectors $\eta_a \in (M_p)_a$ normal to S has dimension $(n-k)$ (see problem 1.10.1).

PROBLEM 1.10.1. *Let S be a k-dimensional imbedded submanifold of the n-dimensional manifold M, and let p be a point in S.*

(1) *Show that the space of co-vectors $\eta_a \in (M_p)_a$ normal to S has dimension $(n-k)$. (Hint: Consider a basis for $(M_p)^a$ containing (as a subset) k vectors tangent to S. Then consider a dual basis.)*
(2) *Show that a vector $\xi^a \in (M_p)^a$ is tangent to S iff $\eta_a \xi^a = 0$ for all co-vectors $\eta_a \in (M_p)_a$ that are normal to S.*

We note for future reference that a co-vector $\eta_a \in (M_p)_a$ is normal to S iff $(id_p)^*(\eta_a) = \mathbf{0}$. It is worth giving the argument in detail to help gain familiarity with our notation. $(id_p)^*(\eta_a)$ is the zero vector in $(S_p)_a$ iff $\left((id_p)^*(\eta_a)\right) \tilde{\xi}^a = 0$ for all $\tilde{\xi}^a \in (S_p)^a$. But (by the definition of the pull-back operation), $\left((id_p)^*(\eta_a)\right) \tilde{\xi}^a = \eta_a \left((id_p)_*(\tilde{\xi}^a)\right)$. So $(id_p)^*(\eta_a) = \mathbf{0}$ iff $\eta_a \left((id_p)_*(\tilde{\xi}^a)\right) = 0$ for all $\tilde{\xi}^a \in (S_p)^a$. But a vector $\xi^a \in (M_p)^a$ is tangent to S precisely if it is of the form $\left((id_p)_*(\tilde{\xi}^a)\right)$ for some $\tilde{\xi}^a \in (S_p)^a$. So $(id_p)^*(\eta_a) = \mathbf{0}$ iff $\eta_a \xi^a = 0$ for all vectors $\xi^a \in (M_p)^a$ that are tangent to S; i.e., η_a is normal to S.

The classification we have introduced can be extended to indices on M-tensors of higher index structure. Consider, for example, the M-tensor $\alpha^{ab}{}_{cd}$ at p. We take it to be *tangent to S* in its first contravariant index if $\eta_a \alpha^{ab}{}_{cd} = \mathbf{0}$ for all $\eta_a \in (M_p)_a$ that are normal to S. (Note that this characterization, which applies to all M-tensors with contravariant indices, is consistent with the one given initially for the special case of contravariant vectors by virtue of the second assertion in problem 1.10.1.) And we take it to be *normal* to S in its second covariant index if $\xi^d \alpha^{ab}{}_{cd} = \mathbf{0}$ for all $\xi^d \in (M_p)^d$ that are tangent to S.

So far, M-tensors at p can be tangent to S only in their contravariant indices and normal to S only in their covariant indices. But now (and henceforth in this section), let us assume that a metric g_{ab} is present on M. Then the classification can be extended. We can take take the tensor to be tangent to S in a covariant index if it is so after the index is raised with g_{ab}. And we can take it to be normal to S in a contravariant index if it is so after the index is lowered with

g_{ab}. Now we have four subspaces to consider side by side. In addition to the old k-dimensional space of contravariant M-vectors at p tangent to S, we have a new $(n-k)$-dimensional space of contravariant M-vectors at p normal to S. And in addition to the old $(n-k)$-dimensional space of covariant M-vectors at p normal to S, we have a new k-dimensional space of covariant M-vectors at p tangent to S. As one would expect, it is possible to introduce "projection tensors" that, when applied to (contravariant and covariant) M-vectors at p, yield their respective components in these four subspaces. We shall do so in a moment.

Let us say that an M-tensor at p is (fully) *tangent to S* (or *normal to S*) if it is so in each of its indices. The subspace of M-tensors $\alpha^{a_1\ldots a_r}{}_{b_1\ldots b_s}$ at p tangent to S has dimension $k^{(r+s)}$.

Nothing said so far rules out the possibility that there is a non-zero vector $\xi^a \in (M_p)^a$ that is both tangent to, and normal to, S. Such a vector would necessarily satisfy $g_{ab}\,\xi^a\,\xi^b = 0$. (Since ξ^a is tangent to S, and $g_{ab}\,\xi^b$ is normal to S, the contraction of the two must be 0.) There cannot be non-zero vectors satisfying this condition if g_{ab} is positive definite. But the possibility does arise when, for example, the metric is of Lorentzian signature.

We say that our imbedded submanifold S is a *metric submanifold* (relative to the background metric g_{ab} on M) if, for all p in S, no non-zero vector in $(M_p)^a$ is both tangent to S and normal to S. An alternative formulation is available. The pull-back field $id^*(g_{ab})$ is always a smooth, symmetric field on S. But it is non-degenerate (and so a metric) iff S is a metric submanifold (see problem 1.10.2).

PROBLEM 1.10.2. *Let S be a k-dimensional imbedded submanifold of the n-dimensional manifold M, and let g_{ab} be a metric on M. Show that S is a metric submanifold (relative to g_{ab}) iff for all p in S the pull-back tensor $(id_p)^*(g_{ab})$ at p is non-degenerate; i.e., there is no non-zero vector $\tilde{\xi}^a \in (S_p)^a$ such that $\big((id_p)^*(g_{ab})\big)\,\tilde{\xi}^a = \mathbf{0}$.*

In what follows, *we assume that S is a metric submanifold* (relative to g_{ab}). Non-metric submanifolds do arise in relativity theory. ("Null hypersurfaces," for example, are non-metric.) But they are not essential for our purposes, and it will simplify our discussion to put them aside. The assumption that S is a metric submanifold, for example, implies—and, indeed, is equivalent to the assertion that—*there is a basis for M_p consisting entirely of vectors that are either tangent to, or normal to, S (but not both)*. It is convenient to be able to work with such a basis. (It is always true (in the presence of a metric) that we can find k

linearly independent vectors at p tangent to S, and $(n-k)$ linearly independent vectors there normal to S. But the combined set of n vectors will be linearly independent iff the subspaces spanned by the two individual sets share no non-zero vector; i.e., there is no non-zero vector that is both tangent to, and normal to, S.)

The vector space of S-tensors at p of a given index structure has the same dimension as the vector space of M-tensors there that are of the same index structure and that are tangent to S. In fact, as we now show, there is a canonically defined linear map ϕ_p from the first to the second that is injective and so qualifies as an isomorphism.[6] We define this isomorphism $\tilde{\alpha}^{a_1\ldots a_r}{}_{b_1\ldots b_s} \longmapsto \phi_p(\tilde{\alpha}^{a_1\ldots a_r}{}_{b_1\ldots b_s})$ in stages, considering, in order, scalars (0^{th} order tensors), contravariant vectors, covariant vectors, and then, finally, arbitrary tensors.

For scalars α, we set $\phi_p(\alpha) = \alpha$. (We do not place a tilde over the first α because there is no distinction to be drawn here. Scalars are just scalars.) For vectors $\tilde{\xi}^a$, we set

$$\phi_p(\tilde{\xi}^a) = (id_p)_*(\tilde{\xi}^a).$$

It follows immediately from (I2)—the second condition in the definition of an imbedding—and the definition of tangency that ϕ_p determines an isomorphism between S_p and the space of contravariant M-vectors at p tangent to S. Next, we define $\phi_p(\tilde{\eta}_a)$ by specifying its action on vectors $\xi^a \in (M_p)^a$ that are either tangent to, or normal to, S. (This suffices since, as we have seen, we can always find a basis for $(M_p)^a$ consisting entirely of such vectors.)

$$\phi_p(\tilde{\eta}_a)\,\xi^a = \begin{cases} \tilde{\eta}_a\left((\phi_p)^{-1}(\xi^a)\right) & \text{if } \xi^a \text{ is tangent to } S \\ 0 & \text{if } \xi^a \text{ is normal to } S. \end{cases}$$

Clearly, $\phi_p(\tilde{\eta}_a)$ is tangent to S. That much is guaranteed by the second clause within the definition. Moreover, the action of ϕ_p on $(S_p)_a$ is injective. (Suppose $\phi_p(\tilde{\eta}_a)$ is the zero vector in $(M_p)_a$. Then $\tilde{\eta}_a\left((\phi_p)^{-1}(\xi^a)\right) = 0$, for all tangent vectors $\xi^a \in (M_p)^a$. But every vector in $(S_p)^a$ is of the form $(\phi_p)^{-1}(\xi^a)$ for some tangent vector $\xi^a \in (M_p)^a$. So $\tilde{\eta}_a$ is the zero vector in $(S_p)_a$.)

Finally, we consider the case of an S-tensor at p of higher order index structure – say $\tilde{\alpha}^{ab}{}_c$. There are no surprises. We define $\phi_p(\tilde{\alpha}^{ab}{}_c)$, once again, by specifying its action on vectors that are all tangent to, or normal to, S.

6. We are presenting a great deal of detail here. Some readers may want to skip to proposition 1.10.1.

$$\phi_p(\tilde{\alpha}^{ab}{}_c)\,\mu_a\,\nu_b\,\xi^c = \begin{cases} \tilde{\alpha}^{ab}{}_c\,\big((\phi_p)^{-1}(\mu_a)\big)\big((\phi_p)^{-1}(\nu_b)\big)\,\big((\phi_p)^{-1}(\xi^c)\big) \\ \qquad\qquad\qquad\qquad \text{if } \mu_a,\ \nu_b, \text{ and } \xi^c \text{are tangent to } S \\ 0 \quad \text{if } \mu_a, \nu_b, \text{or } \xi^c \text{ is normal to } S. \end{cases}$$

Clearly, $\phi_p(\tilde{\alpha}^{ab}{}_c)$ is tangent to S, and the argument that ϕ_p is injective in its action on $(S_p)^{ab}_c$ is very much the same as in the preceding case. This completes our definition.

We have established, so far, that for every index structure ${}^{a_1\ldots a_r}{}_{b_1\ldots b_s}$, there is an isomorphism between the vector space of S-tensors $\tilde{\alpha}^{a_1\ldots a_r}{}_{b_1\ldots b_s}$ at p and the vector space of M-tensors $\alpha^{a_1\ldots a_r}{}_{b_1\ldots b_s}$ at p that are tangent to S. If we now "aggregate" the different isomorphisms, we arrive at a map ϕ_p (we use the same notation) that commutes with all the tensor operations – addition, outer multiplication, index substitution, and contraction. It follows from our definition, for example, that $\phi_p(\tilde{\alpha}^{ab}{}_c\,\tilde{\beta}^{de}) = \phi_p(\tilde{\alpha}^{ab}{}_c)\,\phi_p(\tilde{\beta}^{de})$ and $\phi_p(\tilde{\alpha}^{ab}{}_c\,\tilde{\beta}^{ce}) = \phi_p(\tilde{\alpha}^{ab}{}_c)\,\phi_p(\tilde{\beta}^{ce})$. In summary, we have established the following.

PROPOSITION 1.10.1. *Let S be a metric submanifold of the manifold M. Then the tensor algebra of S-tensors at any point of S is isomorphic to the tensor algebra of M-tensors there that are tangent to S.*

The map ϕ_p is closely related to $(id_p)_*$. Indeed, it agrees with the latter in its action on contravariant tensors at p. But $(id_p)_*$ makes assignments *only* to contravariants tensors there, whereas ϕ_p makes assignments to all tensors. (Similarly, $(\phi_p)^{-1}$ agrees with $(id_p)^*$ in its assignment to covariant tensors at p that are tangent to S.)

Now we switch our attention to tensor *fields* on S—i.e., assignments of tensors of the same index structure to every point of S. Of course, we have to distinguish between assignments of S-tensors and assignments of M-tensors. But the isomorphisms we have been considering (defined at individual points of S) induce a correspondence $\tilde{\alpha}^{a_1\ldots a_r}{}_{b_1\ldots b_s} \longmapsto \phi(\tilde{\alpha}^{a_1\ldots a_r}{}_{b_1\ldots b_s})$ between S-fields and M-fields that are tangent to S—i.e., tangent at every point.

The correspondence respects differential structure in the following sense (in addition to algebraic structure). Let $\alpha^{a_1\ldots a_r}{}_{b_1\ldots b_s}$ be an M-field on S that is tangent to S. There are two senses in which it might be said to be "smooth." Let us say that it is *M-smooth* if, for every p in S, there is an open set $O \subseteq M$ containing p and an extension of $\alpha^{a_1\ldots a_r}{}_{b_1\ldots b_s}$ to a field $\overset{+}{\alpha}{}^{a_1\ldots a_r}{}_{b_1\ldots b_s}$ on O that is smooth relative to the charts $\mathcal{C}_M$. (This sense of smoothness applies to all M-fields on S, whether they are tangent to S or not.) Let us also say that it is *S-smooth* if the corresponding S-field $\phi^{-1}(\alpha^{a_1\ldots a_r}{}_{b_1\ldots b_s})$ is smooth relative to the charts $\mathcal{C}_S$. One would like these two senses of smoothness to agree, and in fact they do. By direct consideration of charts, one can establish the following. (We skip the proof.)

PROPOSITION 1.10.2. *Let S be a metric submanifold of the manifold M. Further, let $\alpha^{a_1\ldots a_r}{}_{b_1\ldots b_s}$ be an M-field on S that is tangent to S. Then $\alpha^{a_1\ldots a_r}{}_{b_1\ldots b_s}$ is M-smooth iff it is S-smooth.*

In what follows, we shall sometimes say that an M-field on S is smooth without further qualification. If the field is not tangent to S, this *can* only mean that it is M-smooth. If it *is* tangent to S, the proposition rules out any possibility of ambiguity.

Consider now the S-field $\tilde{h}_{ab} = id^*(g_{ab})$ on S. It is called the *induced metric* or *first fundamental form* on the manifold S.[7] (That it *is* a metric follows from our assumption that S is a metric submanifold of M. Recall problem 1.10.2.) Associated with $\tilde{h}_{ab}$ is a unique compatible derivative operator $\tilde{D}$ on S. (So it satisfies $\tilde{D}_a\, \tilde{h}_{bc} = \mathbf{0}$.) It is our goal now to show that it is possible, in a sense, to express $\tilde{D}$ in terms of the derivative operator ∇ on M that is compatible with g_{ab}. The sense involved is a bit delicate because it makes reference to the map ϕ we have been considering that takes S-fields to M-fields on S tangent to S. The idea, in effect, is to translate talk about the former into talk about the latter.

Corresponding to $\tilde{h}_{ab}$ is a smooth, symmetric M-field $h_{ab} = \phi(\tilde{h}_{ab}) = \phi(id^*(g_{ab}))$ on S that is tangent to S. (It is tangent to S because the image of *every* S-field under ϕ is so. How do we know it is smooth? Since g_{ab} is a smooth field on the manifold M, $id^*(g_{ab})$ is a smooth field on the manifold S. But $id^*(g_{ab}) = \phi^{-1}(h_{ab})$. So h_{ab} is S-smooth (and, hence, M-smooth as well).) We can characterize h_{ab} directly, without reference to $\tilde{h}_{ab}$ or ϕ, in terms of its action (at any point of S) on M-vectors that are tangent to, or normal to, S.

$$h_{ab}\,\lambda^a\,\eta^b = \begin{cases} g_{ab}\,\lambda^a\,\eta^b & \text{if } \lambda^a \text{ and } \eta^a \text{ are both tangent to } S \\ 0 & \text{if } \lambda^a \text{ or } \eta^a \text{ is normal to } S. \end{cases} \tag{1.10.1}$$

The equivalence is easy to check.[8]

Several properties of h_{ab}, as well as a companion field $k_{ab} = (g_{ab} - h_{ab})$ are listed in the following proposition. Clearly, k_{ab} is also a symmetric, smooth

7. Warning: the latter (perfectly standard) expression is potentially confusing because $\tilde{h}_{ab}$ is not a "form" in the special technical sense introduced in section 1.7; i.e., it is *not* anti-symmetric.

8. Suppose λ^a and η^a are both tangent to S. Then, by the definitions of ϕ and the pull-back map id^*, $h_{ab}\,\lambda^a\,\eta^b = \phi(id^*(g_{ab}))\,\lambda^a\,\eta^b = id^*(g_{ab})\,\phi^{-1}(\lambda^a)\,\phi^{-1}(\eta^b) = g_{ab}\,id_*(\phi^{-1}(\lambda^a))\,id_*(\phi^{-1}(\eta^b)) = g_{ab}\,\lambda^a\,\eta^b$. Alternatively, if either λ^a or η^a is normal to S, then $h_{ab}\,\lambda^a\,\eta^b = 0$ since h_{ab} is tangent to S.

M-field on S. (Here and in what follows, whenever we lower and raise indices on M-tensors, it should be understood that we do so with g_{ab}.)

PROPOSITION 1.10.3. *Let S be a metric submanifold of the manifold M (with respect to the metric g_{ab} on M). Let h_{ab} be the M-field on S defined by equation (1.10.1), and let k_{ab} be the companion M-field $(g_{ab} - h_{ab})$ on S. Then all the following hold.*

(1) *h_{ab} is tangent to S and k_{ab} is normal to S.*
(2) *For all M-vector fields α^a on S,*

(a) α^a is tangent to $S \iff h^a_{\ b}\,\alpha^b = \alpha^a \iff k^a_{\ b}\,\alpha^b = \mathbf{0}$ and
(b) α^a is normal to $S \iff k^a_{\ b}\,\alpha^b = \alpha^a \iff h^a_{\ b}\,\alpha^b = \mathbf{0}$.

(3) *$h^a_{\ b}\,h^b_{\ c} = h^a_{\ c}$ and $k^a_{\ b}\,k^b_{\ c} = k^a_{\ c}$ and $h^a_{\ b}\,k^b_{\ c} = \mathbf{0}$.*

Proof. (1) We have already given an argument to show that h_{ab} is tangent to S. (Once again, $h_{ab} = \phi(\tilde{h}_{ab})$, and the image of *every* S-field under ϕ is tangent to S.) Now let ξ^a be any M-vector tangent to S (at any point of S). Then $k_{ab}\,\xi^a = g_{ab}\,\xi^a - h_{ab}\,\xi^a$. But $g_{ab}\,\xi^a = h_{ab}\,\xi^a$, since they agree in their action on both vectors tangent to S and normal to it. So $k_{ab}\,\xi^a = \mathbf{0}$. It follows that k_{ab} is normal to S in its first index. But k_{ab} is symmetric. So it is (fully) normal to S. (2) Suppose first that $h^a_{\ b}\,\alpha^b = \alpha^a$. Then α^a is certainly tangent to S, since $h^a_{\ b}$ is tangent to S in the index a. Conversely, suppose α^a is tangent to S. Then, we claim, $h^a_{\ b}\,\alpha^b$ and α^a have the same action on any vector η_a (at any point of S) that is either tangent to, or normal to, S. In the first case, $h^a_{\ b}\,\alpha^b\,\eta_a = g_{ab}\,\eta^a\,\alpha^b = \alpha^a\,\eta_a$. In the second case, $h^a_{\ b}\,\alpha^b\,\eta_a = 0 = \alpha^a\,\eta_a$. This gives us the first equivalence in (a). The second is immediate since $k^a_{\ b}\,\alpha^b = (g^a_{\ b} - h^a_{\ b})\,\alpha^b = \alpha^a - h^a_{\ b}\,\alpha^b$. The equivalences in (b) are handled similarly. (3) It follows from (2) that $h^a_{\ b}\,h^b_{\ c}$ and $h^a_{\ c}$ have the same action on any vector ξ^c (at any point of S) that is either tangent to, or normal to, S. So $h^a_{\ b}\,h^b_{\ c} = h^a_{\ c}$. The arguments for $k^a_{\ b}\,k^b_{\ c} = k^a_{\ c}$ and $h^a_{\ b}\,k^b_{\ c} = \mathbf{0}$ are similar. □

PROBLEM 1.10.3. *Prove the following generalization of clause (2) in proposition 1.10.3. For all M-tensor fields $\alpha^{\cdots a \cdots}$ on S,*

(1) *$\alpha^{\cdots a \cdots}$ is tangent to S in the index $a \iff h^a_{\ b}\,\alpha^{\cdots b \cdots} = \alpha^{\cdots a \cdots} \iff k^a_{\ b}\,\alpha^{\cdots b \cdots} = \mathbf{0}$.*
(2) *$\alpha^{\cdots a \cdots}$ is normal to S in the index $a \iff k^a_{\ b}\,\alpha^{\cdots b \cdots} = \alpha^{\cdots a \cdots} \iff h^a_{\ b}\,\alpha^{\cdots b \cdots} = \mathbf{0}$.*

We have formulated the preceding problem in terms of contravariant M-fields on S. But, of course, this involves no essential loss of generality. For given one, instead, of form, say, $\alpha^{ab}{}_{cde}$, we can always apply the stated results to $\beta^{abcde} = \alpha^{ab}{}_{mnr} g^{mc} g^{nd} g^{re}$ and then lower indices.

We can think of $h^a{}_b$ and $k^a{}_b$ as projection operators. Given an M-field ξ^a, $h^a{}_b\,\xi^b$ is its component tangent to S, and $k^a{}_b\,\xi^b$ is its component normal to S. More generally, we can use the two operators to decompose an M-field of arbitrary index structure into a sum of component tensor fields, each of which is either tangent to S or normal to S in each of its indices (which is not to say that each of the component fields will be either [fully] tangent to S or [fully] normal to S). So, for example, in the case of a field $\alpha^a{}_b$ on S, we have the following decomposition:

$$\alpha^a{}_b = h^a{}_m\, h^n{}_b\, \alpha^m{}_n + h^a{}_m\, k^n{}_b\, \alpha^m{}_n + k^a{}_m\, h^n{}_b\, \alpha^m{}_n + k^a{}_m\, k^n{}_b\, \alpha^m{}_n.$$

(Notice that the two fields, left and right, have the same action (at any point) on any pair of vectors $\eta_a\, \xi^b$, each of which is either tangent to S or normal to S.)

We are ready to explain the sense in which the action of $\tilde{D}$ can be expressed in terms of ∇. We start with a lemma.

LEMMA 1.10.4. *Let S be a metric submanifold of the manifold M (with respect to the metric g_{ab} on M). Let h_{ab} be the M-field on S defined by equation (1.10.1). Finally, let $\overset{1}{\alpha}{}^{a_1 \ldots a_r}{}_{b_1 \ldots b_s}$ and $\overset{2}{\alpha}{}^{a_1 \ldots a_r}{}_{b_1 \ldots b_s}$ be smooth M-fields on an open set $O \subseteq M$ that agree on S. Then at all points of $S \cap O$,*

$$h^n{}_m \nabla_n\, \overset{1}{\alpha}{}^{a_1 \ldots a_r}{}_{b_1 \ldots b_s} = h^n{}_m \nabla_n\, \overset{2}{\alpha}{}^{a_1 \ldots a_r}{}_{b_1 \ldots b_s}.$$

Proof. Consider $\beta^{a_1 \ldots a_r}{}_{b_1 \ldots b_s} = \overset{1}{\alpha}{}^{a_1 \ldots a_r}{}_{b_1 \ldots b_s} - \overset{2}{\alpha}{}^{a_1 \ldots a_r}{}_{b_1 \ldots b_s}$. It vanishes on S. Let p be a point in $S \cap O$. We need to show that

$$h^n{}_m \nabla_n \beta^{a_1 \ldots a_r}{}_{b_1 \ldots b_s} = \mathbf{0}$$

at p. To do so, it suffices to show that if we contract the left side with any vector ξ^m at p that is either tangent to, or normal to, S, the result is $\mathbf{0}$. That is true automatically if ξ^m is normal to S (since $h^n{}_m$ is tangent to S). And if ξ^m is tangent to S, $h^n{}_m\, \xi^m = \xi^n$. So it suffices to show that

$$\xi^n\, \nabla_n\, \beta^{a_1 \ldots a_r}{}_{b_1 \ldots b_s} = \mathbf{0}$$

for all ξ^n at p tangent to S. The proof of this assertion is similar to other "well-definedness" arguments given before, and proceeds by considering the index structure of $\beta^{a_1\dots a_r}{}_{b_1\dots b_s}$. If β is a scalar field on S, then $\xi^n \nabla_n \beta$ is just the directional derivative $\xi(\beta)$. This has to be 0 because β is constant on S. One next proves the statement for contravariant vector fields β^a on S using the result for scalar fields together with the Leibniz rule. And so forth. □

Now suppose $\alpha^{a_1\dots a_r}{}_{b_1\dots b_s}$ is a smooth M-field on S. We cannot expect to be able to associate with it a field $\nabla_m \alpha^{a_1\dots a_r}{}_{b_1\dots b_s}$ on S. (The latter, if well defined, would encode information about how $\alpha^{a_1\dots a_r}{}_{b_1\dots b_s}$ changes as one moves away from S in arbitrary directions.) But, by the lemma, we can introduce a field $h^n{}_m \nabla_n \alpha^{a_1\dots a_r}{}_{b_1\dots b_s}$ on S. At any point of S, we simply extend $\alpha^{a_1\dots a_r}{}_{b_1\dots b_s}$ to a smooth field $\overset{1}{\alpha}{}^{a_1\dots a_r}{}_{b_1\dots b_s}$ on some open set O, and set

$$h^n{}_m \nabla_n \alpha^{a_1\dots a_r}{}_{b_1\dots b_s} = h^n{}_m \nabla_n \overset{1}{\alpha}{}^{a_1\dots a_r}{}_{b_1\dots b_s}.$$

This field need not be tangent to S even if $\alpha^{a_1\dots a_r}{}_{b_1\dots b_s}$ is. But we can "make it tangent" if we project all indices onto S with the field $h^n{}_m$. This action defines an operator D_a on the set of all smooth M-tensor fields $\alpha^{a_1\dots a_r}{}_{b_1\dots b_s}$ on S that are tangent to S:

(1.10.2) $$D_m \alpha^{a_1\dots a_r}{}_{b_1\dots b_s} = h^{a_1}{}_{c_1} \dots h^{a_r}{}_{c_s} h^{d_1}{}_{b_1} \dots h^{d_s}{}_{b_s} h^n{}_m \nabla_n \alpha^{c_1\dots c_r}{}_{d_1\dots d_s}.$$

The basic result toward which we have been working is the following.

PROPOSITION 1.10.5. *For all smooth S-fields $\tilde{\alpha}^{a_1\dots a_r}{}_{b_1\dots b_s}$,*

$$\phi(\tilde{D}_n \tilde{\alpha}^{a_1\dots a_r}{}_{b_1\dots b_s}) = D_n \phi(\tilde{\alpha}^{a_1\dots a_r}{}_{b_1\dots b_s}).$$

Proof. Let $\tilde{\tilde{D}}$ be the operator on smooth S-fields that is *defined* by the condition

$$\tilde{\tilde{D}}_n \tilde{\alpha}^{a_1\dots a_r}{}_{b_1\dots b_s} = \phi^{-1}\big(D_n \phi(\tilde{\alpha}^{a_1\dots a_r}{}_{b_1\dots b_s})\big).$$

It suffices for us to show that it is a derivative operator on S and that it is compatible with $\tilde{h}_{ab}$. For then it will follow (by proposition 1.9.2) that $\tilde{\tilde{D}} = \tilde{D}$.

Consider, first, the compatibility condition. Since $\phi(\tilde{h}_{ab}) = h_{ab}$, we have

$$\begin{aligned}\tilde{\tilde{D}}_n \tilde{h}_{ab} &= \phi^{-1}\big(D_n \phi(\tilde{h}_{ab})\big) = \phi^{-1}(D_n h_{ab}) = \phi^{-1}\big(h^r{}_a h^s{}_b h^m{}_n \nabla_m h_{rs}\big)\\ &= \phi^{-1}\big(h^s{}_b h^m{}_n[\nabla_m (h^r{}_a h_{rs}) - h_{rs} \nabla_m h^r{}_a]\big)\\ &= \phi^{-1}\big(h^s{}_b h^m{}_n \nabla_m h_{as} - h^r{}_b h^m{}_n \nabla_m h_{ra}\big) = \phi^{-1}(\mathbf{0}) = \mathbf{0}.\end{aligned}$$

Note that we have used the Leibniz rule (in reverse) to arrive at the fourth equality. We are justified in doing so because we are here working "within the shadow" of the projection operator $h^m{}_n$. We can always (locally) extend the tensor fields in question, invoke the Leibniz rule for ∇ in its standard form (where we are working with fields defined on open sets in M rather than fields defined only on S), and then invoke our lemma to show that it does not matter how we do the extension. Note also that the fifth equality follows from the third clause of proposition 1.10.3, and the sixth from the symmetry of h_{ab}.

Next we need to verify that $\tilde{\tilde{D}}$ satisfies conditions (DO1) through (DO6) (section 1.7). The first five are straightforward. The argument is very much the same in each case. Let us consider, for example, a representative instance of the Leibniz rule. We have

$$\begin{aligned}
\tilde{\tilde{D}}_n(\tilde{\alpha}^{ab}\,\tilde{\eta}_c) &= \phi^{-1}\Big(D_n[\phi(\tilde{\alpha}^{ab})\,\phi(\tilde{\eta}_c)]\Big) = \phi^{-1}\Big(h^a{}_r\,h^b{}_s\,h^q{}_c\,h^m{}_n\nabla_m\,[\phi(\tilde{\alpha}^{rs})\,\phi(\tilde{\eta}_q)]\Big)\\
&= \phi^{-1}\Big(h^a{}_r\,h^b{}_s\,h^q{}_c[\phi(\tilde{\alpha}^{rs})\,h^m{}_n\nabla_m\,\phi(\tilde{\eta}_q) + \phi(\tilde{\eta}_q)\,h^m{}_n\nabla_m\,\phi(\tilde{\alpha}^{rs})]\Big)\\
&= \phi^{-1}\Big(\phi(\tilde{\alpha}^{ab})\,h^q{}_c\,h^m{}_n\nabla_m\,\phi(\tilde{\eta}_q) + \phi(\tilde{\eta}_c)\,h^a{}_r\,h^b{}_s\,h^m{}_n\nabla_m\,\phi(\tilde{\alpha}^{rs})\Big)\\
&= \phi^{-1}\Big(\phi(\tilde{\alpha}^{ab})\,D_n\,\phi(\tilde{\eta}_c) + \phi(\tilde{\eta}_c)\,D_n\,\phi(\tilde{\alpha}^{ab})\Big)\\
&= \tilde{\alpha}^{ab}\,\phi^{-1}\Big(D_n\,\phi(\tilde{\eta}_c)\Big) + \tilde{\eta}_c\,\phi^{-1}\Big(D_n\,\phi(\tilde{\alpha}^{ab})\Big)\\
&= \tilde{\alpha}^{ab}\,\tilde{\tilde{D}}_n\,\tilde{\eta}_c + \tilde{\eta}_c\,\tilde{\tilde{D}}_n\,\tilde{\alpha}^{ab}.
\end{aligned}$$

A few steps here deserve comment. For the fourth equality, we need the fact that $h^a{}_r\,\phi(\tilde{\alpha}^{rs}) = \phi(\tilde{\alpha}^{as})$ (and a number of similar statements involving change of index). Note that this is just an instance of the assertion in problem 1.10.3, since $\phi(\tilde{\alpha}^{rs})$ is tangent to S. And the sixth equality holds because ϕ (acting at any point in S) is a tensor algebra isomorphism that commutes with the operations of addition and outer multiplication.

Let us turn, finally, to (DO6). This is the only one of the conditions that requires a bit of attention. Let α be a smooth scalar field on S. Then

$$\tilde{\tilde{D}}_a\,\tilde{\tilde{D}}_b\,\alpha = \phi^{-1}\Big(D_a\,\phi(\tilde{\tilde{D}}_b\,\alpha)\Big) = \phi^{-1}\Big(D_a D_b\,\phi(\alpha)\Big) = \phi^{-1}\Big(h^m{}_b\,h^n{}_a\nabla_n\,(h^r{}_m\nabla_r\alpha)\Big).$$

Here we have used the fact that $\phi(\alpha) = \alpha$. Now let p be any point on S. We can extend α to a smooth field $\overset{+}{\alpha}$ on an open set O in M containing p. Moreover, we can do so in such a way that $\nabla_a\,\overset{+}{\alpha}$ is tangent to S on $S \cap O$. (This can be verified with an argument involving charts. Intuitively we keep $\overset{+}{\alpha}$ constant as we move out from S in directions normal to S.) So $h^n{}_a\nabla_n\,\overset{+}{\alpha} = \nabla_a\,\overset{+}{\alpha}$ on $S \cap O$. Thus, $\nabla_a\,\overset{+}{\alpha}$ is a smooth field on O that agrees with $h^n{}_a\,\nabla_n\,\overset{+}{\alpha}$ on $S \cap O$. It follows that we can understand $h^n_{\ a} \nabla_n (h^r_{\ m} \nabla_r \alpha)$ to be $h^n_{\ a} \nabla_n \nabla_m \overset{+}{\alpha}$ on $S \cap O$, and therefore

$$\tilde{\tilde{D}}_a \, \tilde{\tilde{D}}_b \, \alpha = \phi^{-1} \Big(h^m_{\ b} \, h^n_{\ a} \nabla_n \, \nabla_m \, \overset{+}{\alpha} \Big)$$

at p. The tensor on the right side is manifestly symmetric in a and b (since ∇ satisfies condition (DO6)). Thus $\tilde{\tilde{D}}_a \, \tilde{\tilde{D}}_b \, \alpha$ is symmetric in these indices at our arbitrary point p in S. □

Up to this point we have been attentive to the distinction between S-fields and M-fields on S tangent to S, between $\tilde{h}_{ab}$ and h_{ab}, and between the operators $\tilde{D}$ and D. But it is, more or less, standard practice to be a bit casual about these distinctions or even to collapse them entirely by formally identifying the vector space S_p with the subspace of M_p whose elements are tangent to S. (The work we have done to this point—in particular, propositions 1.10.1, 1.10.2, and 1.10.5—makes clear that there is no harm in doing so.) In what follows, that will be our practice as well. We shall refer to h_{ab} as the "metric induced on S" (or the "first fundamental form on S"), refer to D as the "derivative operator induced on S," and so forth. We shall also drop the labels "S-field" and "M-field," since it is only the latter with which we shall be working.

In effect, we shall be systematically translating "S-talk" into "M-talk." Here is one more example of how this works. What should we mean by a "geodesic on S with respect to the induced metric (or induced derivative operator)"? We can certainly understand it to be a map of the form $\gamma : I \to S$ that is smooth with respect to $\mathcal{C}_S$ and whose tangent field $\tilde{\xi}^a$ satisfies $\tilde{\xi}^n \tilde{D}_n \tilde{\xi}^a = \mathbf{0}$. Instead, we shall drop explicit reference to $\mathcal{C}_S$ and $\tilde{D}$ and take it to be a map of the form $\gamma : I \to S$ that is smooth with respect to $\mathcal{C}_M$ and whose tangent field ξ^a satisfies $\xi^n D_n \xi^a = \mathbf{0}$.

We know that

$$h^m_{\ a} \, h^n_{\ b} \, h^p_{\ c} \, \nabla_m \, h_{np} = \mathbf{0} \qquad \text{(1.10.3)}$$

on S. (This is just the assertion that $D_a h_{bc} = \mathbf{0}$, and we proved it in the course of showing that $\tilde{D}_a \tilde{h}_{bc} = \mathbf{0}$. That was the first step in our proof of proposition 1.10.5.) Similarly, one can show that

$$h^m_{\ a} \, k^n_{\ b} \, k^p_{\ c} \, \nabla_m \, h_{np} = \mathbf{0} \qquad \text{(1.10.4)}$$

on S. However, the mixed projection field π_{abc} defined by

$$\pi_{abc} = h^m_{\ a} \, h^n_{\ b} \, k^p_{\ c} \, \nabla_m \, h_{np} \qquad \text{(1.10.5)}$$

need not vanish. It turns out that π_{abc} is of particular geometric interest. It is called the *extrinsic curvature* field on S.

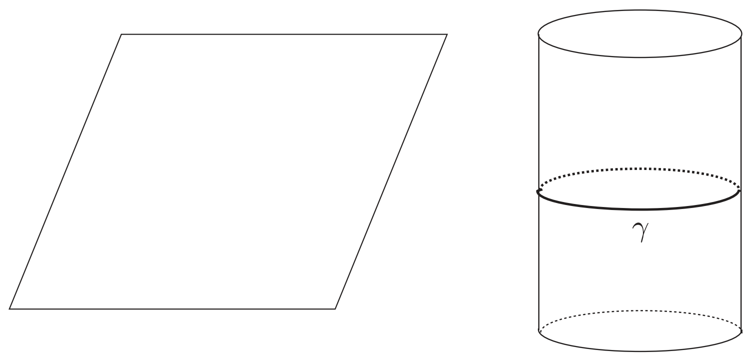

Figure 1.10.2. The cylinder and the plane (imbedded in three-dimensional Euclidean space) both have vanishing intrinsic curvature. But the cylinder, in contrast to the plane, has non-vanishing extrinsic curvature. Notice that there are curves on the cylinder—e.g., γ, that are geodesics with respect to induced derivative operator D that are not geodesics with respect to the background derivative operator ∇.

PROBLEM 1.10.4. *Prove equation (1.10.4).*

The induced metric h_{ab} and its associated derivative operator D are geometric structures "intrinsic" to S. They are not sensitive to the way S is imbedded in M. We say that (S, h_{ab}) has vanishing *intrinsic curvature* just in case D is flat. The *extrinsic curvature* of S, in contrast, *is* determined by the imbedding. Think of both a plane and a cylinder imbedded in ordinary three-dimensional Euclidean space (figure 1.10.2). They both have vanishing intrinsic curvature. But only the plane has vanishing extrinsic curvature. Notice that all geodesics of the plane are necessarily geodesics of the ambient three-dimensional space. But the corresponding statement for the cylinder is not true. There are geodesics of the cylinder (e.g., γ in figure 1.10.2) that are *not* geodesics of the larger space. This is a good way to think about extrinsic curvature. Indeed, as we shall prove (proposition 1.10.7), π_{abc} is a measure of the degree to which geodesics in (S, h_{ab}) fail to be geodesics in (M, g_{ab}). But first we need a lemma.

LEMMA 1.10.6. $\pi_{[ab]c} = \mathbf{0}$.

Proof. Consider any point p in S. If ξ^a is a vector at p tangent to S, we have $\xi^c\, k^p{}_c = \mathbf{0}$ and hence $\xi^c\, \pi_{[ab]c} = \mathbf{0}$. So it will suffice to show $\xi^c\, \pi_{[ab]c} = \mathbf{0}$ for all ξ^a at p normal to S. Since S has dimension k and M has dimension n, we can find an open set O containing p and $(n-k)$ smooth scalar fields

$\overset{i}{\alpha}(i = 1, \ldots, n-k)$ on O such that (i) $\nabla_a \overset{i}{\alpha}$ is normal to S on $S \cap O$, for all i, and (ii) the vectors $\nabla_a \overset{i}{\alpha}$ are linearly independent on $S \cap O$. (This can be verified with an argument involving charts. Indeed, the fields $\overset{i}{\alpha}$ can be local coordinates induced by a chart on M. What is required is that their associated coordinate curves all be orthogonal to S where they intersect it.) To complete the proof, it suffices to verify that $(\nabla^c \overset{i}{\alpha})\, \pi_{[ab]c} = \mathbf{0}$ at p for all i. But this follows since we have

$$\begin{aligned}
(\nabla^c \overset{i}{\alpha})\, \pi_{[ab]c} &= (\nabla^c \overset{i}{\alpha})\, h^m_{\ [a}\, h^n_{\ b]}\, k^p_{\ c}\, \nabla_m\, h_{np} = h^m_{\ [a}\, h^n_{\ b]} (\nabla^p \overset{i}{\alpha}) \nabla_m\, h_{np} \\
&= h^m_{\ [a}\, h^n_{\ b]} [\nabla_m (h_{np} \nabla^p \overset{i}{\alpha}) - h_{np}\, \nabla_m \nabla^p\, \overset{i}{\alpha}] \\
&= -\, h^m_{\ [a}\, h^p_{\ b]} \nabla_m \nabla_p\, \overset{i}{\alpha} = -h^m_{\ a}\, h^p_{\ b} \nabla_{[m} \nabla_{p]}\, \overset{i}{\alpha} = \mathbf{0}.
\end{aligned}$$

(For the fourth equality, we have used the fact that since $h_{np}\, \nabla^p\, \overset{i}{\alpha} = \mathbf{0}$ on $S \cap O$, it must be the case that $h_a^{\ m}\, \nabla_m (h_{np}\, \nabla^p\, \overset{i}{\alpha}) = \mathbf{0}$ on $S \cap O$. This follows, once again, from lemma 1.10.4.) □

Now we can give the promised geometric interpretation of π_{abc}.

PROPOSITION 1.10.7. *Let S be a metric submanifold of the manifold M (with respect to the metric g_{ab} on M). Let ∇ be the derivative operator on M determined by g_{ab}, let h_{ab} be the induced metric on S, and let π_{abc} be the extrinsic curvature field on S. Finally, let γ be a geodesic in (S, h_{ab}) with tangent field ξ^a. Then*

$$\xi^n\, \nabla_n\, \xi^c = \pi_{ab}^{\ \ c}\, \xi^a\, \xi^b. \tag{1.10.6}$$

Proof. By hypothesis, $\xi^n\, D_n\, \xi^c = \mathbf{0}$. And $\xi^n\, h^r_{\ n} = \xi^r$, since ξ^a is tangent to S. So

$$\mathbf{0} = \xi^n\, h^r_{\ n}\, h^c_{\ m} \nabla_r\, \xi^m = \xi^r (g^c_{\ m} - k^c_{\ m}) \nabla_r\, \xi^m = \xi^r\, \nabla_r\, \xi^c - k^c_{\ m}\, \xi^r\, \nabla_r\, \xi^m.$$

Therefore,

$$\begin{aligned}
\xi^r\, \nabla_r\, \xi^c &= k^c_{\ m}\, \xi^r\, \nabla_r\, \xi^m = k^c_{\ m}\, \xi^r\, \nabla_r (h^m_{\ p}\, \xi^p) \\
&= k^c_{\ m}\, \xi^r (h^m_{\ p}\, \nabla_r\, \xi^p + \xi^p\, \nabla_r\, h^m_{\ p}) = k^c_{\ m}\, \xi^r\, \xi^p\, \nabla_r\, h^m_{\ p} \\
&= k^c_{\ m} (\xi^a\, h^r_{\ a})(\xi^b\, h^p_{\ b}) \nabla_r\, h^m_{\ p} = \xi^a\, \xi^b\, h^r_{\ a}\, h^p_{\ b}\, k^{cm}\, \nabla_r\, h_{pm} = \xi^a\, \xi^b\, \pi_{ab}^{\ \ c}.
\end{aligned}$$

Here we use the fact that $k^c_{\ m}\, h^m_{\ p} = \mathbf{0}$ for the fourth equality. □

Given any point p in S, π_{abc} vanishes there iff $\pi_{abc}\,\xi^a\xi^b = \mathbf{0}$ for all vectors ξ^a at p that are tangent to S. (This follows, since π_{abc} is symmetric in its first two indices (lemma 1.10.6) and also tangent to S in them. Recall proposition 1.4.3.) But given any vector ξ^a at p tangent to S, there is a geodesic in (S, h_{ab}) that passes through p, whose tangent vector there is ξ^a. So it follows from our proposition that $\pi_{abc} = \mathbf{0}$ *iff all geodesics in* (S, h_{ab}) *are geodesics in* (M, g_{ab}). Moreover, the requirement that equation (1.10.6) hold for all geodesics in (S, h_{ab}) uniquely determines π_{abc}.

Next we consider the *Gauss-Codazzi equations.*

PROPOSITION 1.10.8. *Suppose (M, g_{ab}) and (S, h_{ab}) are as in proposition 1.10.7, and D is the derivative operator on S determined by h_{ab}. Further suppose $R^a{}_{bcd}$ is the Riemann curvature field on M associated with ∇, and $\mathcal{R}^a{}_{bcd}$ is the Riemann curvature field on S associated with D. Then*

$$\mathcal{R}^a{}_{bcd} = -2\,\pi^{a}{}_{[c}{}^{m}\,\pi_{d]bm} + h^a{}_m\, h^n{}_b\, h^p{}_c\, h^r{}_d\, R^m{}_{npr}, \tag{1.10.7}$$

$$h^m{}_{[a}\, h^n{}_{b]}\, h^p{}_c\, k^r{}_d \nabla_m \pi_{npr} = \frac{1}{2} h^m{}_a\, h^n{}_b\, h^p{}_c\, k^r{}_d\, R_{mnpr}. \tag{1.10.8}$$

Proof. The argument consists of a long computation. First, let λ^a be any smooth vector field on S tangent to S. Then $\mathcal{R}^a{}_{bcd}$ must satisfy

$$\begin{aligned} -\frac{1}{2}\mathcal{R}^a{}_{bcd}\,\lambda^b = D_{[c}\, D_{d]}\,\lambda^a &= h^p{}_{[c}\, h^r{}_{d]}\, h^a{}_s \nabla_p (h^m{}_r\, h^s{}_n \nabla_m \lambda^n) \\ &= h^p{}_{[c}\, h^r{}_{d]}\, h^a{}_s \big[(\nabla_p h^m{}_r)\, h^s{}_n\, \nabla_m \lambda^n \\ &\quad + h^m{}_r (\nabla_p h^s{}_n)\, \nabla_m \lambda^n + h^m{}_r\, h^s{}_n\, \nabla_p \nabla_m \lambda^n \big]. \end{aligned} \tag{1.10.9}$$

Now, by equations (1.10.3) and (1.10.5),

$$h^p{}_c\, h^r{}_d\, \nabla_p\, h^m{}_r = h^p{}_c\, h^r{}_d\, g^m{}_q \nabla_p\, h^q{}_r = h^p{}_c\, h^r{}_d (k^m{}_q + h^m{}_q) \nabla_p\, h^q{}_r = \pi_{cd}{}^m. \tag{1.10.10}$$

So, by lemma 1.10.6, the first term on the right side of equation (1.10.9) vanishes. The second and third terms can be simplified by using equation (1.10.10), the symmetry of h_{sn}, and the fact that $h^r{}_d\, h^m{}_r = h^m{}_d$. We have

$$\begin{aligned}
-\frac{1}{2}\mathcal{R}^a{}_{bcd}\,\lambda^b &= h^p{}_{[c}\,h^m{}_{d]}\,h^a{}_s(\nabla_p\,h^s{}_n)\,\nabla_m\,\lambda^n + h^p{}_{[c}\,h^m{}_{d]}h^a{}_n\,\nabla_p\,\nabla_m\,\lambda^n \\
&= \pi_{[c}{}^{an}\,h^m{}_{d]}\nabla_m\,\lambda_n + h^p{}_c\,h^m{}_d\,h^a{}_n\nabla_{[p}\nabla_{m]}\,\lambda^n \\
&= \pi_{[c}{}^{an}\,h^m{}_{d]}\,\nabla_m\,\lambda_n - \frac{1}{2}h^p{}_c\,h^m{}_d\,h^a{}_n\,R^n{}_{bpm}\,\lambda^b.
\end{aligned}$$

Now π_{can} is normal to S in its third index. So $\pi_c{}^{an}h_{nb} = \mathbf{0}$ and, therefore,

$$\begin{aligned}
\pi_c{}^{an}\,h^m{}_d\nabla_m\,\lambda_n &= \pi_c{}^{an}h^m{}_d\nabla_m(h_{nb}\,\lambda^b) \\
&= \pi_c{}^{an}\,h^m{}_d\,\lambda^b\,\nabla_m\,h_{nb} = \pi_c{}^{an}\,h^m{}_d\,h^b{}_r\,\lambda^r\,\nabla_m\,h_{nb} \\
&= \pi_c{}^{an}\,\pi_{drn}\,\lambda^r = \pi^a{}_c{}^n\,\pi_{drn}\lambda^r.
\end{aligned}$$

So we have, all together,

$$\Big(-\frac{1}{2}\mathcal{R}^a{}_{rcd} - \pi^a{}_{[c}{}^n\,\pi_{d]rn} + \frac{1}{2}h^a{}_n\,h^p{}_c\,h^m{}_d\,R^n{}_{rpm}\Big)\lambda^r = \mathbf{0}.$$

Now let η^b be an arbitrary smooth field on S and take $h^r{}_b\,\eta^b$ for λ^r. Then, since the first two terms are tangent to S in the index rI, we have

$$(\mathcal{R}^a{}_{bcd} + 2\,\pi^a{}_{[c}{}^n\,\pi_{d]bn} - h^a{}_n\,h^r{}_b\,h^p{}_c\,h^m{}_d\,R^n{}_{rpm})\eta^b = \mathbf{0}.$$

Since this holds for *all* smooth fields η^b on S, the field in parentheses must vanish. This gives us equation (1.10.7). The second computation is similar, and we leave it as an exercise. □

PROBLEM 1.10.5. *Derive the second Gauss-Codazzi equation (1.10.8).*

The first Gauss-Codazzi equation expresses the intrinsic Riemann curvature tensor field $\mathcal{R}^a{}_{bcd}$ in terms of the extrinsic curvature field π_{abc} and the full background Riemann curvature field $R^a{}_{bcd}$. We shall return to it later when we consider the geometric significance of Einstein's equation.

So far we have assumed only that S is a metric submanifold of M. Let us now consider the special case where S is a metric hypersurface, i.e., has dimension $k = (n-1)$. A slight simplification results. The vector space of vectors normal to S is now one-dimensional at every point of S. So it consists of multiples of some (normalized) vector ξ^a where $\xi^a\xi_a = \pm 1$. ($\xi^a\xi_a$ cannot be 0 precisely because S is a metric submanifold.) Whether the value of $\xi^a\xi_a$ is $+1$ or -1 depends on S and the signature of g_{ab}. At least if S is connnected, the value will be the same at every point of S—i.e., everywhere $+1$ or everywhere -1.

Let us assume that S and g_{ab} are such that the value is $+1$ at all points of S. (The other case is handled similarly.) So there are exactly two vectors ξ^a normal to S at every point satisfying $\xi^a \xi_a = 1$. Locally, at least, we can always make a choice so as to generate a smooth field. We say that S is *two sided* if it is possible to do so globally.

Let ξ^a be one such (local or global) smooth normal field on S satisfying $\xi^a \xi_a = 1$. Then

$$h_{ab} = (g_{ab} - \xi_a \xi_b), \tag{1.10.11}$$

$$k_{ab} = \xi_a \xi_b. \tag{1.10.12}$$

(Note that h_{ab} and $(g_{ab} - \xi_a \xi_b)$ have the same action on ξ^a and on all vectors tangent to S.) Now consider the field π_{ab} defined by

$$\pi_{ab} = -\pi_{abc}\, \xi^c. \tag{1.10.13}$$

When hypersurfaces are under discussion, it (rather than π_{abc}) is often called the *extrinsic curvature* field (relative to ξ^a). It is also called the *second fundamental form on S* (relative to ξ^a). Notice that

$$\pi_{[ab]} = \mathbf{0}, \tag{1.10.14}$$

$$\pi_{abc} = -\pi_{ab}\, \xi_c, \tag{1.10.15}$$

$$\pi_{ab} = h^m{}_a\, h^n{}_b\, \nabla_m\, \xi_n. \tag{1.10.16}$$

The first assertion follows immediately from lemma 1.10.6. For the second, it suffices to observe that π_{abc} and $-\pi_{ab}\,\xi_c$ agree in their action on ξ^c and on all vectors η^c tangent to S. For the third, we have

$$\begin{aligned}\pi_{ab} = -\pi_{abc}\,\xi^c &= -h^m{}_a\, h^n{}_b\, k^p{}_c\, \xi^c\, \nabla_m\, h_{np} = -h^m{}_a\, h^n{}_b\, \xi^p\, \nabla_m\, h_{np} \\ &= -h^m{}_a\, h^n{}_b [\nabla_m(\xi^p\, h_{np}) - h_{np}\, \nabla_m\, \xi^p] = h^m{}_a\, h^p{}_b\, \nabla_m\, \xi_p.\end{aligned}$$

Equation (1.10.16) leads to an alternative interpretation of extrinsic curvature in the case of hypersurfaces. Let $\overset{+}{\xi}{}^a$ be an extension of ξ^a to a smooth field of unit length on some open set O in M, and let $\overset{+}{h}_{ab}$ be defined by $\overset{+}{h}_{ab} = g_{ab} - \overset{+}{\xi}_a\, \overset{+}{\xi}_b$. (So $\overset{+}{h}_{ab}$ is an extension of h_{ab} and $\overset{+}{h}_{ab}\, \overset{+}{\xi}{}^a = \mathbf{0}$.) Then we have

$$\pi_{ab} = \frac{1}{2}\, \pounds_{\overset{+}{\xi}}\, \overset{+}{h}_{ab} \tag{1.10.17}$$

on $S \cap O$. To prove this, observe first that on $S \cap O$, $\mathcal{L}_{\overset{+}{\xi}} \overset{+}{h}_{ab}$ is tangent to S (in both indices). This follows since $\mathcal{L}_{\overset{+}{\xi}} \overset{+}{\xi}{}^{a} = \mathbf{0}$ and, hence,

$$\overset{+}{\xi}{}^{a} \, \mathcal{L}_{\overset{+}{\xi}} \overset{+}{h}_{ab} = \mathcal{L}_{\overset{+}{\xi}} (\overset{+}{h}_{ab} \overset{+}{\xi}{}^{a}) - \overset{+}{h}_{ab} \, \mathcal{L}_{\overset{+}{\xi}} \overset{+}{\xi}{}^{a} = \mathbf{0}.$$

Therefore on S we have

$$\begin{aligned} \mathcal{L}_{\overset{+}{\xi}} \overset{+}{h}_{ab} &= h^{r}{}_{a} \, h^{s}{}_{b} \, \mathcal{L}_{\overset{+}{\xi}} \overset{+}{h}_{rs} = h^{r}{}_{a} \, h^{s}{}_{b} \, \mathcal{L}_{\overset{+}{\xi}} (g_{rs} - \overset{+}{\xi}_{r} \overset{+}{\xi}_{s}) \\ &= h^{r}{}_{a} \, h^{s}{}_{b} \, \mathcal{L}_{\overset{+}{\xi}} g_{rs} = h^{r}{}_{a} \, h^{s}{}_{b} \, [\overset{+}{\xi}{}^{n} \nabla_{n} g_{rs} + g_{ns} \nabla_{r} \overset{+}{\xi}{}^{n} + g_{rn} \nabla_{s} \overset{+}{\xi}{}^{n}] \\ &= h^{r}{}_{a} \, h^{s}{}_{b} \, (\nabla_{r} \overset{+}{\xi}_{s} + \nabla_{s} \overset{+}{\xi}_{r}) = 2 \, h^{r}{}_{(a} \, h^{s}{}_{b)} \, \nabla_{r} \overset{+}{\xi}_{s} \\ &= 2 \, \pi_{(ab)} = 2 \, \pi_{ab}. \end{aligned}$$

(The final two equalities follow, respectively, from equations (1.10.16) and (1.10.14).)

Thus we can think of π_{ab} (up to the factor $\frac{1}{2}$) as the Lie derivative of h_{ab} in the direction ξ^{a} normal to S. This interpretation will be important later in connection with our discussion of the "initial value problem" in general relativity.

Finally let us reconsider the Gauss-Codazzi equations in the present case. Substituting $-\pi_{ab}\xi_{c}$ for π_{abc} in the equations of proposition 1.10.8 yields

(1.10.18)
$$\mathcal{R}^{a}{}_{bcd} = -2 \, \pi^{a}{}_{[c} \, \pi_{d]b} + h^{a}{}_{m} \, h^{n}{}_{b} \, h^{p}{}_{c} \, h^{r}{}_{d} \, R^{m}{}_{npr},$$
$$h^{m}{}_{[a} \, h^{n}{}_{b]} \, h^{p}{}_{c} \, \nabla_{m} \, \pi_{np} = -\frac{1}{2} h^{m}{}_{a} \, h^{n}{}_{b} \, h^{p}{}_{c} \, \xi^{r} \, R_{mnpr}.$$

The second can be expressed as

(1.10.19)
$$D_{[a} \, \pi_{b]c} = -\frac{1}{2} h^{m}{}_{a} \, h^{n}{}_{b} \, h^{p}{}_{c} \, \xi^{r} \, R_{mnpr}.$$

Contracting on equation (1.10.18) yields

$$\mathcal{R}_{bc} = -\pi^{a}{}_{c} \, \pi_{ab} + \pi^{a}{}_{a} \, \pi_{cb} + h^{r}{}_{m} \, h^{n}{}_{b} \, h^{p}{}_{c} \, R^{m}{}_{npr}.$$

Substituting $(g^{r}{}_{m} - \xi_{m} \, \xi^{r})$ for $h^{r}{}_{m}$ on the right side, we arrive at

$$\mathcal{R}_{bc} = \pi\ \pi_{bc} - \pi_{ab}\,\pi^{a}_{\ c} + h^{n}_{\ b}\,h^{p}_{\ c}\,R_{np} - R_{mbcr}\,\xi^{m}\,\xi^{r}, \tag{1.10.20}$$

where $\pi = \pi^{a}_{\ a}$. (Note that $R_{mbcr}\,\xi^{m}\,\xi^{r}$ is tangent to S, since (by proposition 1.9.4) contracting with ξ^{a} on b or c yields 0. Hence $h^{n}_{\ b}\,h^{p}_{\ c}\,R_{mbcr}\,\xi^{m}\,\xi^{r} = R_{mbcr}\,\xi^{m}\,\xi^{r}$.) Contracting once more yields

$$\begin{aligned} \mathcal{R} &= \pi^{2} - \pi_{ab}\,\pi^{ab} + h^{np}\,R_{np} - R_{mr}\,\xi^{m}\xi^{r} \\ &= \pi^{2} - \pi_{ab}\,\pi^{ab} + R - 2\,R_{nr}\,\xi^{n}\,\xi^{r}. \end{aligned} \tag{1.10.21}$$

Of course, in the special case where we are dealing with a hypersurface imbedded in a flat manifold ($R^{a}_{\ bcd} = \mathbf{0}$)—e.g., in the case of a two-dimensional surface imbedded in three-dimensional Euclidean space—our expressions for $\mathcal{R}^{a}_{\ bcd}$, $\mathcal{R}_{bc}$, and $\mathcal{R}$ simplify still further:

$$\mathcal{R}_{abcd} = \pi_{ad}\,\pi_{bc} - \pi_{ac}\,\pi_{bd}, \tag{1.10.22}$$

$$\mathcal{R}_{bc} = \pi\ \pi_{bc} - \pi_{ab}\,\pi^{a}_{\ c}, \tag{1.10.23}$$

$$\mathcal{R} = \pi^{2} - \pi_{ab}\,\pi^{ab}. \tag{1.10.24}$$

1.11. Volume Elements

In what follows, let M be an n–dimensional manifold ($n \geq 1$). As we know from section 1.7, an s-form on M ($s \geq 1$) is a covariant field $\alpha_{b_1 \ldots b_s}$ that is anti-symmetric (i.e., anti-symmetric in each pair of indices). The case where $s = n$ is of special interest.

Let $\alpha_{b_1 \ldots b_n}$ be an n-form on M. Further, let $\overset{i}{\xi}{}^{b} (i = 1, \ldots, n)$ be a basis for the tangent space at a point in M with dual basis $\overset{i}{\eta}_{b} (i = 1, \ldots, n)$. Then $\alpha_{b_1 \ldots b_n}$ can be expressed there in the form

$$\alpha_{b_1 \ldots b_n} = k\ n!\ \overset{1}{\eta}_{[b_1} \ldots \overset{n}{\eta}_{b_n]} \tag{1.11.1}$$

where

$$k = \alpha_{b_1 \ldots b_n}\,\overset{1}{\xi}{}^{b_1} \ldots \overset{n}{\xi}{}^{b_n}.$$

(To see this, observe that the two sides of equation (1.11.1) have the same action on any collection of n vectors from the set $\{\overset{1}{\xi}{}^{b}, \ldots, \overset{n}{\xi}{}^{b}\}$.) It follows that if $\alpha_{b_1 \ldots b_n}$ and $\beta_{b_1 \ldots b_n}$ are any two smooth, non-vanishing n-forms on M, then

$$\beta_{b_1 \ldots b_n} = f\,\alpha_{b_1 \ldots b_n}$$

for some smooth non-vanishing scalar field f.

Smooth, non-vanishing n-forms always exist locally on M. (Suppose (U, φ) is a chart with coordinate vector fields $(\vec{\gamma}_1)^{a}, \ldots, (\vec{\gamma}_n)^{a}$, and suppose $\overset{i}{\eta}_{b} (i =$

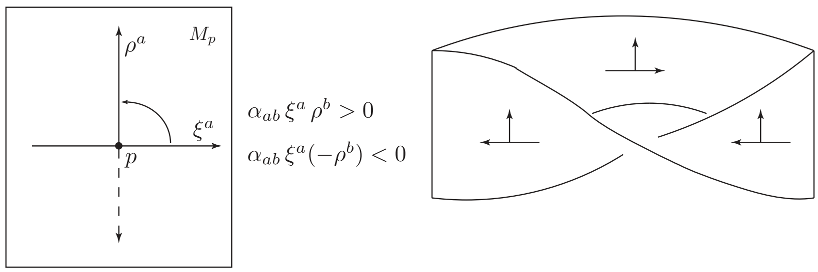

Figure 1.11.1. A 2-form α_{ab} on the Möbius strip determines a "positive direction of rotation" at every point where it is non-zero. So there cannot be a smooth, non-vanishing 2-form on the Möbius strip.

$1, \ldots, n$) are dual fields. Then $\overset{1}{\eta}_{[b_1} \ldots \overset{n}{\eta}_{b_n]}$ qualifies as a smooth, non-vanishing n-form on U.) But they do not necessarily exist globally. Suppose, for example, that M is the two-dimensional Möbius strip (see figure 1.11.1), and α_{ab} is any smooth two-form on M. We see that α_{ab} must vanish somewhere as follows.

Let p be any point on M at which $\alpha_{ab} \neq 0$, and let ξ^a be any non-zero vector at p. Consider the number $\alpha_{ab}\,\xi^a \rho^b$ as ρ^b rotates though the vectors in M_p. If $\rho^b = \pm\xi^b$, the number is zero. If $\rho^b \neq \pm\xi^b$, the number is non-zero. Therefore, as ρ^b rotates between ξ^a and $-\xi^a$, it is always positive or always negative. Thus α_{ab} determines a "positive direction of rotation" away from ξ^a on M_p. α_{ab} must vanish somewhere because one cannot continuously choose positive rotation directions over the entire Möbius strip.

M is said to be *orientable* if it admits a (globally defined) smooth, non-vanishing n-form.

So far we have made no mention of metric structure. Suppose now that our manifold M is endowed with a metric g_{ab} of signature (n^+, n^-). We take a *volume element* on M (with respect to g_{ab}) to be a smooth n-form $\epsilon_{b_1 \ldots b_n}$ that satisfies the normalization condition

(1.11.2) $$\epsilon^{b_1 \ldots b_n}\,\epsilon_{b_1 \ldots b_n} = (-1)^{n^-}\,n!.$$

Suppose $\epsilon_{b_1 \ldots b_n}$ is a volume element on M, and $\overset{i}{\xi}{}^b (i = 1, \ldots, n)$ is an orthonormal basis for the tangent space at a point in M. Then at that point we have, by equation (1.11.1),

(1.11.3) $$\epsilon_{b_1 \ldots b_n} = k\,n!\,\overset{1}{\xi}_{[b_1} \ldots \overset{n}{\xi}_{b_n]}$$

where $k = \epsilon_{b_1 \ldots b_n} \overset{1}{\xi}{}^{b_1} \ldots \overset{n}{\xi}{}^{b_n}$. Hence, by the normalization condition (1.11.2),

$$(-1)^{n^-} n! = \big(k\, n!\, \overset{1}{\xi}{}_{[b_1} \ldots \overset{n}{\xi}{}_{b_n]}\big)\big(k\, n!\, \overset{1}{\xi}{}^{[b_1} \ldots \overset{n}{\xi}{}^{b_n]}\big)$$
$$= k^2 (n!)^2 \frac{1}{n!} \big(\overset{1}{\xi}{}_{b_1} \overset{1}{\xi}{}^{b_1}\big) \ldots \big(\overset{n}{\xi}{}_{b_n} \overset{n}{\xi}{}^{b_n}\big) = k^2\, n!\, (-1)^{n^-}.$$

So $k^2 = 1$ and, therefore, equation (1.11.3) yields

$$\epsilon_{b_1 \ldots b_n} \overset{1}{\xi}{}^{b_1} \ldots \overset{n}{\xi}{}^{b_n} = \pm 1. \tag{1.11.4}$$

Clearly, if $\epsilon_{b_1 \ldots b_n}$ is a volume element on M, then so is $-\epsilon_{b_1 \ldots b_n}$. It follows from the normalization condition (1.11.4) that there cannot be any others. Thus, there are only two possibilities. Either (M, g_{ab}) admits no volume elements (at all) or it admits exactly two, and these agree up to sign.

Condition (1.11.4) also suggests where the term "volume element" comes from. Given arbitrary vectors $\overset{1}{\gamma}{}^a, \ldots, \overset{n}{\gamma}{}^a$ at a point, we can think of $\epsilon_{b_1 \ldots b_n} \overset{1}{\gamma}{}^{b_1} \ldots \overset{n}{\gamma}{}^{b_n}$ as the *volume* of the (possibly degenerate) parallelepiped determined by the vectors. Notice that, up to sign, $\epsilon_{b_1 \ldots b_n}$ is characterized by three properties.

(VE1) It is linear in each index.
(VE2) It is anti-symmetric.
(VE3) It assigns a volume V with $|V| = 1$ to each orthonormal parallelepiped.

These are conditions we would demand of any would-be volume measure (with respect to g_{ab}). If the length of one edge of a parallelepiped is multiplied by a factor k, then its volume should increase by that factor. And if a parallelepiped is sliced into two parts, with the slice parallel to one face, then its volume should be equal to the sum of the volumes of the parts. This leads to (VE1). Furthermore, if any two edges of the parallelepiped are coalligned (i.e., if it is a degenerate parallelepiped), then its volume should be zero. This leads to (VE2). (If for all vectors ξ^a, $\epsilon_{b_1 \ldots b_n} \xi^{b_1} \xi^{b_2} = \mathbf{0}$, then it must be the case that $\epsilon_{b_1 \ldots b_n}$ is anti-symmetric in indices (b_1, b_2). And similarly for all other pairs of indices.) Finally, if the edges of a parallelepiped are orthogonal, then its volume should be equal to the product of the lengths of the edges. This leads to (VE3). The only unusual thing about $\epsilon_{b_1 \ldots b_n}$ as a volume measure is that it respects orientation. If it assigns V to the ordered sequence $\overset{1}{\gamma}{}^a, \ldots, \overset{n}{\gamma}{}^a$, then it assigns $(-V)$ to $\overset{2}{\gamma}{}^a, \overset{1}{\gamma}{}^a, \overset{3}{\gamma}{}^a, \ldots, \overset{n}{\gamma}{}^a$, and so forth.

It will be helpful to collect here a few facts for subsequent calculations. Suppose $\epsilon_{a_1 \ldots a_n}$ is a volume element on M with respect to the metric g_{ab} with signature (n^+, n^-). Then

$$\epsilon^{a_1\ldots a_n}\,\epsilon_{b_1\ldots b_n} = (-1)^{n^-}\, n!\, \delta^{[a_1}_{\ b_1} \ldots \delta^{a_n]}_{\ b_n}, \tag{1.11.5}$$

$$\epsilon^{a_1\ldots a_n}\,\epsilon_{a_1 b_2\ldots b_n} = (-1)^{n^-}\, (n-1)!\, \delta^{[a_2}_{\ b_2} \ldots \delta^{a_n]}_{\ b_n}, \tag{1.11.6}$$

$$\epsilon^{a_1\ldots a_n}\,\epsilon_{a_1 a_2 b_3\ldots b_n} = (-1)^{n^-}\, 2(n-2)!\, \delta^{[a_3}_{\ b_3} \ldots \delta^{a_n]}_{\ b_n}, \tag{1.11.7}$$

$$\vdots$$

$$\epsilon^{a_1\ldots a_n}\,\epsilon_{a_1\ldots a_r b_{r+1}\ldots b_n} = (-1)^{n^-}\, r!\,(n-r)!\, \delta^{[a_{r+1}}_{\ b_{r+1}} \ldots \delta^{a_n]}_{\ b_n}. \tag{1.11.8}$$

Consider, for example, the case where $n = 3$ and $n^- = 0$—i.e., where g_{ab} is positive definite. (The general case is handled similarly.) Then equation (1.11.5) comes out as the assertion $\epsilon^{abc}\epsilon_{mnq} = 6\,\delta^{[a}_{\ m}\,\delta^{b}_{\ n}\,\delta^{c]}_{\ q}$. To see that it holds, consider any anti-symmetric tensor α^{mnq} at a point. Then $\alpha^{mnq} = k\,\epsilon^{mnq}$ for some k. So

$$\begin{aligned}\epsilon^{abc}\,\epsilon_{mnq}\,\alpha^{mnq} &= k\,\epsilon^{abc}\,\epsilon_{mnq}\,\epsilon^{mnq} = 6\,k\,\epsilon^{abc} = 6\,\alpha^{abc}\\ &= 6\,\delta^{[a}_{\ m}\,\delta^{b}_{\ n}\,\delta^{c]}_{\ q}\,\alpha^{mnq}.\end{aligned}$$

Thus for all anti-symmetric α^{mnq} at the point, we have

$$\big(\epsilon^{abc}\,\epsilon_{mnq} - 6\,\delta^{[a}_{\ m}\,\delta^{b}_{\ n}\,\delta^{c]}_{\ q}\big)\alpha^{mnq} = \mathbf{0}.$$

In particular, given arbitrary vectors λ^m, ρ^n, μ^q there,

$$\big(\epsilon^{abc}\,\epsilon_{mnq} - 6\,\delta^{[a}_{\ m}\,\delta^{b}_{\ n}\,\delta^{c]}_{\ q}\big)\lambda^{[m}\,\rho^{n}\,\mu^{q]} = \mathbf{0}.$$

But since the expression in parentheses is itself anti-symmetric in the indices (m, n, q), this condition can be expressed as

$$\big(\epsilon^{abc}\,\epsilon_{mnq} - 6\,\delta^{[a}_{\ m}\,\delta^{b}_{\ n}\,\delta^{c]}_{\ q}\big)\lambda^{m}\,\rho^{n}\,\mu^{q} = \mathbf{0}.$$

Since λ^m, ρ^n, and μ^q are arbitrary, it follows that

$$\epsilon^{abc}\,\epsilon_{mnq} - 6\,\delta^{[a}_{\ m}\,\delta^{b}_{\ n}\,\delta^{c]}_{\ q} = \mathbf{0}.$$

This gives us equation (1.11.5). Next, equation (1.11.6) follows from (1.11.5) since

$$\begin{aligned}\epsilon^{abc}\,\epsilon_{anq} &= 6\,\delta^{[a}_{\ a}\,\delta^{b}_{\ n}\,\delta^{c]}_{\ q}\\ &= 2\big(\delta^{a}_{\ a}\,\delta^{[b}_{\ n}\,\delta^{c]}_{\ q} - 2\,\delta^{[b}_{\ n}\,\delta^{c]}_{\ q}\big)\\ &= 2(3-2)\delta^{[b}_{\ n}\,\delta^{c]}_{\ q} = 2\,\delta^{[b}_{\ n}\,\delta^{c]}_{\ q}.\end{aligned}$$

Finally, equation (1.11.7) follows from (1.11.6) since

$$\epsilon^{abc}\,\epsilon_{abq} = 2\,\delta^{[b}{}_{b}\,\delta^{c]}{}_{q} = \left(\delta^{b}{}_{b} - 1\right)\delta^{c}{}_{q} = 2\,\delta^{c}{}_{q}.$$

Another fact we shall need is

$$\nabla_m\,\epsilon_{a_1\ldots a_n} = \mathbf{0} \tag{1.11.9}$$

(where ∇ is the derivative operator on M determined by g_{ab}). To see this, suppose λ^a is an arbitrary smooth field on M. Then, since $\lambda^m\,\nabla_m\,\epsilon_{b_1\ldots b_n}$ is a smooth n-form on M, we have

$$\lambda^m\,\nabla_m\,\epsilon_{b_1\ldots b_n} = \varphi\,\epsilon_{b_1\ldots b_n}$$

for some scalar field φ. But then

$$\begin{aligned}\varphi\,(-1)^{n^-}\,n! = \varphi\,\epsilon^{a_1\ldots a_n}\,\epsilon_{a_1\ldots a_n} &= \epsilon^{a_1\ldots a_n}\,\lambda^m\,\nabla_m\,\epsilon_{a_1\ldots a_n}\\ &= \frac{1}{2}\lambda^m\,\nabla_m\left(\epsilon^{a_1\ldots a_n}\,\epsilon_{a_1\ldots a_n}\right) = \frac{1}{2}\lambda^m\,\nabla_m((-1)^{n^-}\,n!) = 0.\end{aligned}$$

So $\varphi = 0$ and, hence, $\lambda^m\,\nabla_m\,\epsilon_{b_1\ldots b_n} = \mathbf{0}$. Since λ^m was arbitrary, we have equation (1.11.9).

Finally, we show how to recover ordinary vector analysis in terms of volume elements. Suppose our manifold M is $\mathbb{R}^3$, g_{ab} is the Euclidean metric defined by equation (1.9.13), ∇ is the derivative operator determined by g_{ab}, and ϵ_{abc} is a volume element on M. Then, given contravariant vectors ξ and η at some point, we define their dot and cross products as follows:

$$\xi \cdot \eta = \xi^a\,\eta_a,$$

$$\xi \times \eta = \epsilon^{abc}\,\xi_b\,\eta_c.$$

(We are deliberately not using indices on the left.) It follows immediately from the anti-symmetry of ϵ^{abc} that $\xi \times \eta = -(\eta \times \xi)$, and that $\xi \times \eta$ is orthogonal to both ξ and η. Furthermore, if we define the angular measure $\measuredangle(\xi, \eta)$ by setting

$$\cos\measuredangle(\xi, \eta) = \frac{\xi \cdot \eta}{\|\xi\|\,\|\eta\|},$$

where $\|\xi\| = (\xi \cdot \xi)^{\frac{1}{2}}$, then the magnitude of $\xi \times \eta$ is given by

$$\begin{aligned}\|\xi \times \eta\| &= \left(\epsilon^{abc}\,\xi_b\,\eta_c\,\epsilon_{amn}\,\xi^m\,\eta^n\right)^{\frac{1}{2}}\\ &= \left(2\,\delta^{[b}{}_{m}\,\delta^{c]}{}_{n}\,\xi^m\,\eta^n\,\xi_b\,\eta_c\right)^{\frac{1}{2}}\end{aligned}$$

$$= \left[(\xi^b \xi_b)(\eta^c \eta_c) - (\xi^b \eta_b)^2\right]^{\frac{1}{2}}$$

$$= \|\xi\| \, \|\eta\| \left(1 - \cos^2 \measuredangle(\xi, \eta)\right)^{\frac{1}{2}} = \|\xi\| \, \|\eta\| \sin \measuredangle(\xi, \eta).$$

Consider an example. One learns in ordinary vector analysis that, given any three vectors α, β, γ at a point,

$$\gamma \times (\alpha \times \beta) = \alpha(\gamma \cdot \beta) - \beta(\gamma \cdot \alpha).$$

In our notation, this comes out as the assertion

$$\epsilon^{abc} \gamma_b \left(\epsilon_{cmn} \alpha^m \beta^n\right) = \alpha^a (\gamma_b \beta^b) - \beta^a (\gamma_b \alpha^b),$$

and it follows easily from equation (1.11.6):

$$\begin{aligned} \epsilon^{abc} \gamma_b \epsilon_{cmn} \alpha^m \beta^n &= \epsilon^{cab} \epsilon_{cmn} \gamma_b \alpha^m \beta^n \\ &= 2 \delta^{[a}{}_m \delta^{b]}{}_n \gamma_b \alpha^m \beta^n = \alpha^a (\gamma_b \beta^b) - \beta^a (\alpha^b \gamma_b). \end{aligned}$$

Given a smooth scalar field f and a smooth contravariant vector field ξ on M, we define the following:

$$\begin{aligned} \operatorname{grad}(f) &= \nabla^a f \\ \operatorname{div}(\xi) &= \nabla_a \xi^a \\ \operatorname{curl}(\xi) &= \epsilon^{abc} \nabla_b \xi_c. \end{aligned}$$

(In the more familiar notation usually found in textbooks, these would be written as ∇f, $\nabla \cdot \xi$, and $\nabla \times \xi$.) With these definitions, we can recover all the usual formulas of vector analysis. Here are two simple examples. (Others are listed in the problems that follow.)

(1) $\operatorname{curl}(\operatorname{grad} f) = \mathbf{0}$.
(2) $\operatorname{div}(\operatorname{curl} \xi) = 0$.

The first comes out as the assertion that $\epsilon^{abc} \nabla_b \nabla_c f = \mathbf{0}$, which is immediate since $\nabla_b \nabla_c f$ is symmetric in (b, c). (For this result, flatness is not required.) The second comes out as $\nabla_a(\epsilon^{abc} \nabla_b \xi_c) = 0$. This follows from equation (1.11.9) and the fact (*now* using flatness) that $\nabla_a \nabla_b \xi_c$ is symmetric in (a, b).

PROBLEM 1.11.1. *One learns in the study of ordinary vector analysis that, for all vectors ξ, η, θ, and λ at a point, the following identities hold.*

(1) $(\xi \times \eta) \cdot (\theta \times \lambda) = (\xi \cdot \theta)(\eta \cdot \lambda) - (\xi \cdot \lambda)(\eta \cdot \theta)$.
(2) $(\xi \times (\eta \times \theta)) + (\theta \times (\xi \times \eta)) + (\eta \times (\theta \times \xi)) = \mathbf{0}$.

Reformulate these assertions in our notation and prove them.

PROBLEM 1.11.2. *Do the same for the following assertion:*

$$\mathrm{div}(\xi \times \eta) = \eta \cdot \mathrm{curl}(\xi) - \xi \cdot \mathrm{curl}(\eta).$$

(Here ξ and η are understood to be smooth vector fields.)

PROBLEM 1.11.3. *We have seen (proposition 1.9.9) that every Killing field ξ^a in n-dimensional Euclidean space ($n \geq 1$) can be expressed uniquely in the form*

$$\xi_b = \chi^a F_{ab} + k_b,$$

where F_{ab} and k_b are constant, F_{ab} is anti-symmetric, and χ^a is the position field relative to some point p. Consider the special case where $n = 3$. Let ϵ_{abc} be a volume element. Show that (in this special case) there is a unique constant field W^a such that $F_{ab} = \epsilon_{abc} W^c$. (If $W^a = \mathbf{0}$, ξ^a is the "infinitesimal generator" of a family of translations in the direction k^a. Alternatively, if $k^a = \mathbf{0}$, it generates a family of rotations about the point p with axis W^a.) (Hint: Consider $W^a = \frac{1}{2}\epsilon^{abc} F_{bc}$.)

2

CLASSICAL RELATIVITY THEORY

2.1. Relativistic Spacetimes

With the basic ideas of differential geometry now at our disposal, we turn to relativity theory.

It is helpful to think of the theory as determining a class of geometric models for the spacetime structure of our universe (and isolated subregions thereof, such as, for example, our solar system). Each represents a possible world (or world-region) compatible with the constraints of the theory. We describe these models in stages. First, we characterize a broad class of "relativistic spacetimes" and discuss their interpretation. Later, we introduce further restrictions involving global spacetime structure and Einstein's equation.

We take a *relativistic spacetime* to be a pair (M, g_{ab}), where M is a smooth, connnected, four-dimensional manifold and g_{ab} is a smooth metric on M of Lorentz signature $(1, 3)$. We interpret M as the manifold of point "events" in the world.[1] The interpretation of g_{ab} is given by a network of interconnected physical principles. We list three in this section that are relatively simple in character because they make reference only to point particles and light rays. (These objects alone suffice to determine the metric, at least up to a constant.) We list a fourth in section 2.3 that concerns the behavior of (ideal) clocks. Still other principles involving generic matter fields will come up later.

In what follows, let (M, g_{ab}) be a fixed relativistic spacetime and let ∇ be the unique derivative operator on M compatible with g_{ab}. Since g_{ab} has signature $(1, 3)$, at every point p in M, the tangent space M_p has a basis $\overset{1}{\xi}{}^a, \ldots, \overset{4}{\xi}{}^a$ such that, for all i and j in $\{1, 2, 3, 4\}$,

1. We use "event" as a neutral term here and intend no special significance. Some might prefer to speak, for example, of "equivalence classes of coincident point events" or "point event locations." We shall take this interpretation for granted in what follows and shall, for example, refer to such things as "particle worldlines in a relativistic spacetime."

—–1
— 0
—+1

$$\overset{i}{\xi}{}^a \, \overset{i}{\xi}{}_a = \begin{cases} +1 & \text{if } i = 1 \\ -1 & \text{if } i = 2,\, 3,\, 4 \end{cases}$$

and $\overset{i}{\xi}{}^a \, \overset{j}{\xi}{}_a = 0$ if $i \neq j$. It follows that given any vectors $\mu^a = \sum_{i=1}^{n} \overset{i}{\mu} \, \overset{i}{\xi}{}^a$ and $\nu^a = \sum_{i=1}^{n} \overset{i}{\nu} \, \overset{i}{\xi}{}^a$ at p,

$$\mu^a \, \nu_a = \overset{1}{\mu} \, \overset{1}{\nu} - \overset{2}{\mu} \, \overset{2}{\nu} - \overset{3}{\mu} \, \overset{3}{\nu} - \overset{4}{\mu} \, \overset{4}{\nu} \tag{2.1.1}$$

and

$$\mu^a \, \mu_a = \overset{1}{\mu} \, \overset{1}{\mu} - \overset{2}{\mu} \, \overset{2}{\mu} - \overset{3}{\mu} \, \overset{3}{\mu} - \overset{4}{\mu} \, \overset{4}{\mu}. \tag{2.1.2}$$

(Recall equation (1.9.3).)

Given a vector η^a at a point in M, we say η^a is

timelike	if	$\eta^a \eta_a > 0$,
null (*or lightlike*)	if	$\eta^a \eta_a = 0$,
causal	if	$\eta^a \eta_a \geq 0$,
spacelike	if	$\eta^a \eta_a < 0$.

In this way, g_{ab} determines a "null-cone structure" in the tangent space at every point of M. Null vectors form the boundary of the cone. Timelike vectors form its interior. Spacelike vectors fall outside the cone. Causal vectors are those that are either timelike or null.

The classification extends naturally to *curves*. We take a smooth curve $\gamma: I \to M$ to be *timelike* (respectively *null*, *causal*, *spacelike*) if its tangent vector field $\vec{\gamma}$ is of this character at every point. The property of being timelike, null, and so forth is preserved under reparametrization. So there is a clear sense in which the classification also extends to *images* of smooth curves.[2] The property of being a geodesic is not, in general, preserved under reparametrization. So it does not transfer to curve images. But, of course, the related property of being a *geodesic up to reparametrization* does carry over.

Now we can state the first three interpretive principles. For all smooth curves $\gamma : I \to M$,

(C1) γ is timelike iff $\gamma[I]$ could be the worldline of a point particle with positive mass;[3]

2. Here we are distinguishing between the map $\gamma: I \to M$ and its image $\gamma[I]$. We shall take "worldlines" to be instances of the latter—i.e., construe them as point sets rather than parametrized point sets.

3. We shall later discuss the concept of mass in relativity theory. For the moment, we take it to be just a primitive attribute of particles.

(C2) γ can be reparametrized so as to be a null geodesic iff $\gamma[I]$ could be the trajectory of a light ray;[4]

(P1) γ can be reparametrized so as to be a timelike geodesic iff $\gamma[I]$ could be the worldline of a *free*[5] point particle with positive mass.

In each case, a statement about geometric structure (on the left) is correlated with a statement about the behavior of particles or light rays (on the right).

Several comments and qualifications are called for. First, we are here working within the framework of relativity as traditionally understood and ignoring speculations about the possibility of particles that travel faster than light. (The worldlines of these so-called "tachyons" would come out as images of spacelike curves.) Second, we have restricted attention to smooth curves. So, depending on how one models collisions of point particles, one might want to restrict attention here, in parallel, to particles that do not experience collisions.

Third, the assertions require qualification because the status of "point particles" in relativity theory is a delicate matter. At issue is whether one treats a particle's own mass-energy as a source for the surrounding metric field g_{ab}—in addition to other sources that may happen to be present. (Here we anticipate our discussion of Einstein's equation.) If one does, then the curvature associated with g_{ab} may blow up as one approaches the particle's worldline. And in this case one *cannot* represent the worldline as the image of a curve in M, at least not without giving up the requirement that g_{ab} be a smooth field on M. For this reason, a more careful formulation of the principles would restrict attention to "test particles"—i.e., ones whose own mass-energy is negligible and may be ignored for the purposes at hand.

Fourth, the modal character of the assertions (i.e., the reference to possibility) is essential. It is simply not true—take the case of (C1)—that all images of smooth, timelike curves *are*, in fact, the worldlines of massive particles. The claim is that, as least so far as the laws of relativity theory are concerned, they *could* be. Of course, judgments concerning what could be the case depend on what conditions are held fixed in the background. The claim that a particular curve image could be the worldline of a massive point particle must be

4. For certain purposes, even within classical relativity theory, it is useful to think of light as constituted by streams of "photons" and to take the right-side condition here to be "$\gamma[I]$ could be the worldline of a photon." The latter formulation makes (C2) look more like (C1) and (P1) and draws attention to the fact that the distinction between particles with positive mass and those with zero mass (such as photons) has direct significance in terms of relativistic spacetime structure.

5. "Free particles" here must be understood as ones that do not experience any forces *except gravity*. It is one of the fundamental principles of relativity theory that gravity arises as a manifestation of spacetime curvature, not as an external force that deflects particles from their natural, straight (geodesic) trajectories. We shall discuss this matter further in section 2.5.

understood to mean that it could so long as there are, for example, no barriers in the way. Similarly, in (C2) there is an implicit qualification. We are considering what trajectories are available to light rays when no intervening material media are present—i.e., when we are dealing with light rays *in vacuo*.

Though these four concerns are important and raise interesting questions about the role of idealization and modality in the formulation of physical theory, they have little to do with relativity theory as such. Similar difficulties arise, for example, when one attempts to formulate corresponding principles within the framework of Newtonian gravitation theory.

It follows from the cited interpretive principles that the metric g_{ab} is determined (up to a constant) by the behavior of point particles and light rays.[6] We make this claim precise with a sequence of propositions about conformal structure and projective structure. (Recall our discussion in section 1.9.)

Let g'_{ab} be a second smooth metric of Lorentz signature on M. Clearly, if g'_{ab} is conformally equivalent to g_{ab}—i.e., if there is a smooth function $\Omega : M \to \mathbb{R}$ such that $g'_{ab} = \Omega^2 g_{ab}$—then the two agree in their classification of vectors as timelike, null, and so forth. We first verify that the converse is true as well. (Indeed, we prove something slightly stronger. To establish conformal equivalence, it suffices to require that the two metrics agree on any *one* of the four categories of vectors. If they agree on one, they agree on all.)

PROPOSITION 2.1.1. *The following conditions are equivalent.*

(1) g'_{ab} *and* g_{ab} *agree on which vectors, at arbitrary points of* M, *are timelike (or agree on which are null, or which are causal, or which are spacelike).*
(2) g'_{ab} *and* g_{ab} *are conformally equivalent.*

Proof. The equivalence of the four versions of (1) follows from the fact that the four properties in question (being timelike, null, causal, and spacelike) are interdefinable. So, for example, we can characterize null vectors in terms of timelike vectors:

> A vector η^a at p is null iff either $\eta^a = \mathbf{0}$ or, for all timelike vectors α^a at p, and all sufficiently small numbers k, of the two vectors $\eta^a + k\alpha^a$ and $\eta^a - k\alpha^a$, one is timelike and one is not.

6. This was first recognized by Hermann Weyl [62]. As he put it [63, p. 61], "it can be shown that the metrical structure of the world is already fully determined by its inertial and causal structure, that therefore measurements need not depend on clocks and rigid bodies but that light signals and mass moving under the influence of inertia alone will suffice." For more on Weyl's "causal-inertial" method of determining the spacetime metric, see Coleman and Korté [9, section 4.9].

Conversely, we can characterize timelike vectors in terms of null vectors:

> A vector η^a at p is timelike iff for all null vectors $\alpha^a \neq \mathbf{0}$ at p there is a number $k \neq 0$ and a null vector $\beta^a \neq \mathbf{0}$ at p such that $\eta^a = k\alpha^a + \beta^a$.

It follows immediately that we can also characterize causal vectors (timelike or null) and spacelike vectors (neither timelike nor null) in terms of either timelike vectors or null vectors alone. Other cases are handled similarly. (See problem 2.1.2.)

Now assume that the two metrics agree in their classification of vectors at all points of M. We show that they must be conformally equivalent. Let p be any point in M, and let ξ^a be any vector at p that is spacelike with respect to both metrics. Set

$$k = \frac{g'_{ab}\,\xi^a \xi^b}{g_{ab}\,\xi^a \xi^b}. \tag{2.1.3}$$

Since the numerator and denominator of the fraction are both negative, $k > 0$. We claim first that

$$g'_{ab}\,\eta^a \eta^b = k\, g_{ab}\,\eta^a \eta^b \tag{2.1.4}$$

for all η^a at p. If η^a is null with respect to both metrics, the assertion is trivial. So there are two cases to consider.

Case 1: η^a is timelike with respect to both metrics. Consider the following quadratic equation (in the variable x):

$$0 = g_{ab}\,(\xi^a + x\,\eta^a)(\xi^b + x\,\eta^b) = g_{ab}\,\xi^a \xi^b + 2\,x\,g_{ab}\,\xi^a \eta^b + x^2\,g_{ab}\,\eta^a \eta^b.$$

The discriminant

$$4\,(g_{ab}\,\xi^a \eta^b)^2 - 4\,(g_{ab}\,\xi^a \xi^b)(g_{ab}\,\eta^a \eta^b)$$

is positive (since $(g_{ab}\,\xi^a \xi^b) < 0$, and $(g_{ab}\,\eta^a \eta^b) > 0$). So the equation has real roots r_1 and r_2 with

$$r_1 \cdot r_2 = \frac{g_{ab}\,\xi^a \xi^b}{g_{ab}\,\eta^a \eta^b}. \tag{2.1.5}$$

Now the equation

$$0 = g'_{ab}\,(\xi^a + x\,\eta^a)(\xi^b + x\,\eta^b)$$

must have exactly the same roots as the preceding one (since the metrics agree on null vectors). So we also have

$$r_1 \cdot r_2 = \frac{g'_{ab}\,\xi^a \xi^b}{g'_{ab}\,\eta^a \eta^b}. \tag{2.1.6}$$

These two expressions for $r_1 \cdot r_2$, together with equation (2.1.3), yield equation 2.1.4).

Case 2: η^a is spacelike with respect to both metrics. Let γ^a be any vector at p that is timelike with respect to both. Repeating the argument used for case 1, with η^a now playing the role of ξ^a, we have

$$\frac{g'_{ab}\,\eta^a\eta^b}{g'_{ab}\,\gamma^a\gamma^b} = \frac{g_{ab}\,\eta^a\eta^b}{g_{ab}\,\gamma^a\gamma^b}. \tag{2.1.7}$$

But $g'_{ab}\,\gamma^a\gamma^b = k\,g_{ab}\,\gamma^a\gamma^b$, because γ^a falls under case 1. So η^a must satisfy equation (2.1.4) in this case too.

Thus, we have established our claim. Since $(g'_{ab} - k\, g_{ab})$ is symmetric, it now follows by proposition 1.4.3 that $g'_{ab} = k\, g_{ab}$ at p.

To complete the proof, we define a scalar field $\Omega : M \to \mathbb{R}$ by setting $\Omega(p) = \sqrt{k(p)}$ at each point p (where $k(p)$ is determined as above). Then $g'_{ab} = \Omega^2\, g_{ab}$, and Ω is smooth since g_{ab} and g'_{ab} are. □

It turns out that dimension plays a role in proposition 2.1.1. Our spacetimes are four-dimensional. Suppose we temporarily drop that restriction and, for any $n \geq 2$, consider "n-dimensional spacetimes" (M, g_{ab}) where M has dimension n and g_{ab} has signature $(1, n-1)$. What happens to the proposition? The proof we have given carries over intact for all $n \geq 3$. And even when $n = 2$, it carries over in part. Three versions of condition (1) are still equivalent to each other—those involving agreement on timelike, causal, or spacelike vectors—and to condition (2). But in that special case, two metrics can agree on null vectors without being conformally equivalent. (At any point p in M, a "90-degree rotation" of M_p takes null vectors to null vectors, but it takes timelike vectors to spacelike vectors.)

PROBLEM 2.1.1. *Consider our characterization of timelike vectors in terms of null vectors in the proof of proposition 2.1.1. Why does it fail if $n = 2$?*

PROBLEM 2.1.2. *(i) Show that it is possible to characterize timelike vectors (and so also null vectors and spacelike vectors) in terms of causal vectors. (ii) Show that it is possible to characterize timelike vectors (and so also null vectors and causal vectors) in terms of spacelike vectors. (Both characterizations should work for all $n \geq 2$.)*

Conformally equivalent metrics do not agree, in general, on which curves qualify as geodesics or even just as geodesics up to reparametrization. But, it turns out, they do necessarily agree on which *null* curves are geodesics up to reparametrization. Indeed, we have the following proposition. Notice that clauses (1) and (2) correspond, respectively, to interpretive principles (C1) and (C2) above.

PROPOSITION 2.1.2. *The following conditions are equivalent.*

(1) *g'_{ab} and g_{ab} agree on which smooth curves on M are timelike.*
(2) *g'_{ab} and g_{ab} agree on which smooth curves on M can be reparameterized so as to be null geodesics.*
(3) *g'_{ab} and g_{ab} are conformally equivalent.*

Proof. The implication (1) $\Rightarrow$ (3) follows immediately from the preceding proposition. So does the implication (2) $\Rightarrow$ (1). (Two metrics cannot agree on which curves are null geodesics up to reparametrization without first agreeing on which curves are null.) To complete the proof, we show that (3) implies (2). Assume that $g'_{ab} = \Omega^2\, g_{ab}$. Let γ be any smooth curve that is null (with respect to both g_{ab} and g'_{ab}), and let λ^a be its tangent field. Further, let ∇'_a be the unique derivative operator on M compatible with g'_{ab}. Then, by propositions 1.7.3 and 1.9.5,

$$\lambda^n\, \nabla'_n\, \lambda^a = \lambda^n\, (\nabla_n\, \lambda^a - C^a{}_{nm} \lambda^m)$$

where

$$C^a{}_{nm} = -\frac{1}{2\Omega^2}\big[\delta^a{}_n\, \nabla_m\, \Omega^2 + \delta^a{}_m\, \nabla_n\, \Omega^2 - g_{nm}\, g^{ar}\, \nabla_r\, \Omega^2\big].$$

Substituting for $C^a{}_{nm}$ in the first equation, and using the fact that λ^a is null, we arrive at

$$\lambda^n\, \nabla'_n\, \lambda^a = \lambda^n\, \nabla_n\, \lambda^a + \frac{1}{\Omega^2}(\lambda^n\, \nabla_n\, \Omega^2)\, \lambda^a.$$

It follows that $\lambda^n\, \nabla'_n\, \lambda^a$ is everywhere proportional to λ^a iff $\lambda^n\, \nabla_n\, \lambda^a$ is everywhere proportional to λ^a. Therefore, by proposition 1.7.9, γ can be reparametrized so as to be a geodesic with respect to g_{ab} iff it can be so reparametrized with respect to g'_{ab}. □

Question: What would go wrong if we attempted to adapt the proof to show that conformally equivalent metrics agree as to which smooth *timelike* curves are geodesics up to reparametrization?

We can understand the proposition to assert that *the spacetime metric g_{ab} is determined up to a conformal factor, independently, by the set of possible worldlines of massive point particles and by the set of possible trajectories of light rays.*

Next we turn to projective structure. Recall that g'_{ab} is said to be projectively equivalent to g_{ab} if, for all smooth curves γ on M, γ can be reparametrized so as to be geodesic with respect to g'_{ab} iff it can be so reparametrized with respect to g_{ab}. We have proved (proposition 1.9.6) that if the two metrics are both conformally *and* projectively equivalent, then the conformal factor connecting them is constant. Now, with interpretive principle P1 in mind, we prove a slightly strengthened version of the proposition that makes reference only to timelike geodesics (rather than arbitrary geodesics). To do so, we first strengthen proposition 1.4.3.

PROPOSITION 2.1.3. *Let $\alpha^{a_1 \ldots a_r}{}_{b_1 \ldots b_s}$ be a tensor at some point in M. Suppose that*

(1) *$\alpha^{a_1 \ldots a_r}{}_{b_1 \ldots b_s}$ is symmetric in indices $b_1, \ldots, b_s$, and*
(2) *$\alpha^{a_1 \ldots a_r}{}_{b_1 \ldots b_s} \xi^{b_1} \ldots \xi^{b_s} = \mathbf{0}$ for all timelike vectors ξ^a at the point.*

Then $\alpha^{a_1 \ldots a_r}{}_{b_1 \ldots b_s} = \mathbf{0}$.

Proof. Consider first the case where we are dealing with a tensor of form $\alpha_{b_1 \ldots b_s}$—i.e., one with no contravariant indices. Let ξ^a be a timelike vector at the point in question, and let η^a be an arbitrary vector there. Then there is an $\epsilon > 0$ such that, for all real numbers x, if $|x| < \epsilon$, $(\xi^a + x\, \eta^a)$ is timelike. Now consider the polynomial function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$\begin{aligned} f(x) &= \alpha_{b_1 \ldots b_s} (\xi^{b_1} + x\, \eta^{b_1}) \ldots (\xi^{b_s} + x\, \eta^{b_s}) \\ &= \alpha_{b_1 \ldots b_s} \xi^{b_1} \ldots \xi^{b_s} + \binom{s}{1} x\, \alpha_{b_1 \ldots b_s} \xi^{b_1} \ldots \xi^{b_{s-1}} \eta^{b_s} + \ldots \\ &\quad + \binom{s}{s-1} x^{s-1} \alpha_{b_1 \ldots b_s} \xi^{b_1} \eta^{b_2} \ldots \eta^{b_s} + x^s \alpha_{b_1 \ldots b_s} \eta^{b_1} \ldots \eta^{b_s}. \end{aligned}$$

By our hypothesis, $f(x) = 0$ for all x in the interval $(-\epsilon, \epsilon)$. Hence all derivatives of f vanish in the interval. So $\alpha_{b_1 \ldots b_s} \eta^{b_1} \ldots \eta^{b_s} = 0$. Since η^a was an arbitrary vector at our point, it follows, by proposition 1.4.3, that $\alpha_{b_1 \ldots b_s} = \mathbf{0}$ there. For the general case, let $\mu_{a_1} \ldots \nu_{a_r}$ be arbitrary vectors at the point. Then $\alpha^{a_1 \ldots a_r}{}_{b_1 \ldots b_s} \mu_{a_1} \ldots \nu_{a_r} = \mathbf{0}$ by the argument just given. So (since $\mu_{a_1} \ldots \nu_{a_r}$ are arbitrary vectors), $\alpha^{a_1 \ldots a_r}{}_{b_1 \ldots b_s} = \mathbf{0}$. □

Of course, a parallel proposition holds if $\alpha^{a_1 \ldots a_r}{}_{b_1 \ldots b_s}$ is symmetric in indices $a_1, \ldots, a_r$. Indeed, we can arrive at that formulation simply by lowering the a-indices and raising the b-indices, applying the proposition as proved, and then restoring the original index positions.

PROBLEM 2.1.3. *Does proposition 2.1.3 still hold if condition (1) is left intact but (2) is replaced by*

(2′) $\alpha^{a_1 \ldots a_r}{}_{b_1 \ldots b_s}\, \xi^{b_1} \ldots \xi^{b_s} = \mathbf{0}$ *for all spacelike vectors* ξ^a *at the point?*

And what if it is replaced by

(2″) $\alpha^{a_1 \ldots a_r}{}_{b_1 \ldots b_s}\, \xi^{b_1} \ldots \xi^{b_s} = \mathbf{0}$ *for all null vectors* ξ^a *at the point?*

Justify your answers.

The proposition we are after is the following.

PROPOSITION 2.1.4. *Assume* $g'_{ab} = \Omega^2 g_{ab}$*. Further, assume* g'_{ab} *and* g_{ab} *agree as to which smooth, timelike curves can be reparametrized so as to be geodesics. Then* Ω *is constant.*

Proof. Assume $\nabla' = (\nabla, C^a{}_{bc})$ where, once again, ∇' is the derivative operator associated with g'_{ab}. It suffices for us to show that $C^a{}_{bc} = \delta^a{}_b\, \varphi_c + \delta^a{}_c\, \varphi_b$ for some smooth field φ_a. For then the constancy of Ω follows *exactly* as in our proof of proposition 1.9.6.

To show that $C^a{}_{bc}$ has this form, we need only make a slight revision in our proof of proposition 1.7.10. There we started from the assumption that ∇' and ∇ agree as to which (arbitrary) smooth curves can be reparametrized so as to be geodesics. Using that assumption, we showed that the field $\varphi^{ad}{}_{bcr} = \left(C^a{}_{bc}\, \delta^d{}_r - C^d{}_{bc}\, \delta^a{}_r\right)$ satisfies the condition

$$\varphi^{ad}{}_{(bcr)}\, \xi^b\, \xi^c\, \xi^r = \mathbf{0} \tag{2.1.8}$$

for all vectors ξ^a at all points. Then we invoked proposition 1.4.3 to conclude that $\varphi^{ad}{}_{(bcr)} = \mathbf{0}$ everywhere. Arguing in exactly the same way from our weaker assumption (that the metrics agree as to which smooth, timelike curves can reparametrized so as to be geodesics), we can show that equation (2.1.8) holds for all *timelike* vectors at all points. But we know (by proposition 2.1.3) that this condition also forces the conclusion that $\varphi^{ad}{}_{(bcr)} = \mathbf{0}$ everywhere. The rest of the proof goes through exactly as in that of proposition 1.7.10. Without reference to particular types of vectors, we can show that $C^a{}_{bc} = \delta^a{}_b\, \varphi_c + \delta^a{}_c\, \varphi_b$ where $\varphi_c = \dfrac{1}{n+1}\, C^d{}_{cd}$. □

Later in this book we shall consider a few particular examples of spacetimes. But one should be mentioned immediately, namely *Minkowski spacetime*. We take it to be the pair (M, g_{ab}) where (i) M is the manifold $\mathbb{R}^4$, (ii) (M, g_{ab}) is flat—i.e., has vanishing Riemann curvature everywhere—and (iii) (M, g_{ab}) is geodesically complete—i.e., all maximally extended geodesics have domain $\mathbb{R}$.

Minkowski spacetime is very special because its structure as an affine manifold (M, ∇) is precisely the same as that of four-dimensional Euclidean space. (Here, of course, ∇ is understood to be the unique derivative operator on M compatible with g_{ab}.) In particular, given any point o in M, there is a smooth "direction field" χ^a on M that vanishes at o and satisfies the condition $\nabla_a \chi^b = \delta_a^{\ b}$. (Recall proposition 1.7.12.)

2.2. Temporal Orientation and "Causal Connectibility"

The characterization we have given of relativistic spacetimes is extremely loose. Many further conditions might be imposed. We consider one in this section, namely "temporal orientability."

First we need to review certain basic facts about Lorentzian metrics. Once again, let (M, g_{ab}) be a fixed relativistic spacetime. We start with the orthogonality relation that g_{ab} determines in the tangent space at every point of M. (Two vectors μ^a and ν^a at a point qualify as *orthogonal*, of course, if $\mu^a \nu_a = 0$.)

PROPOSITION 2.2.1. *Let μ^a and ν^a be vectors at some point p in M. Then the following both hold.*

(1) *If μ^a is timelike and ν^a is orthogonal to μ^a, then either $\nu^a = \mathbf{0}$ or ν^a is spacelike.*
(2) *If μ^a and ν^a are both null, then they are orthogonal iff they are proportional (i.e., one is a scalar multiple of the other).*

Proof. (1) Let $\overset{1}{\xi}{}^a, \ldots, \overset{4}{\xi}{}^a$ be an orthonormal basis for M_p with $\overset{1}{\xi}{}^a\, \overset{1}{\xi}{}_a = 1$, and $\overset{i}{\xi}{}^a\, \overset{i}{\xi}{}_a = -1$ for $i = 2, 3, 4$. Then we can express μ^a and ν^a in the form $\mu^a = \sum_{i=1}^{n} \overset{i}{\mu}\, \overset{i}{\xi}{}^a$ and $\nu^a = \sum_{i=1}^{n} \overset{i}{\nu}\, \overset{i}{\xi}{}^a$. Now assume μ^a is timelike, ν^a is orthogonal to μ^a, and $\nu^a \neq \mathbf{0}$. We show that ν^a is spacelike. It follows from our assumptions that

$$(\overset{1}{\mu})^2 > (\overset{2}{\mu})^2 + (\overset{3}{\mu})^2 + (\overset{4}{\mu})^2, \tag{2.2.1}$$

$$\overset{1}{\mu}\,\overset{1}{\nu} = \overset{2}{\mu}\,\overset{2}{\nu} + \overset{3}{\mu}\,\overset{3}{\nu} + \overset{4}{\mu}\,\overset{4}{\nu}, \tag{2.2.2}$$

$$\overset{1}{\mu} \neq 0, \tag{2.2.3}$$

$$(\overset{2}{\nu})^2 + (\overset{3}{\nu})^2 + (\overset{4}{\nu})^2 > 0. \tag{2.2.4}$$

(The first two assertions follow from equations (2.1.2) and (2.1.1). The third follows from the first. For the final inequality, note that if $(\overset{2}{\nu})^2 + (\overset{3}{\nu})^2 + (\overset{4}{\nu})^2 = 0$, then $\overset{2}{\nu} = \overset{3}{\nu} = \overset{4}{\nu} = 0$, and so, by equations (2.2.2) and (2.2.3), $\overset{1}{\nu} = 0$ as well. This contradicts our assumption that $\nu^a \neq \mathbf{0}$.) In turn, it now follows by the Schwarz inequality (as applied to the vectors $(\overset{2}{\mu}, \overset{3}{\mu}, \overset{4}{\mu})$ and $(\overset{2}{\nu}, \overset{3}{\nu}, \overset{4}{\nu})$) that

$$\begin{aligned}(\overset{1}{\mu})^2\,(\overset{1}{\nu})^2 &= (\overset{2}{\mu}\,\overset{2}{\nu} + \overset{3}{\mu}\,\overset{3}{\nu} + \overset{4}{\mu}\,\overset{4}{\nu})^2 \leq [(\overset{2}{\mu})^2 + (\overset{3}{\mu})^2 + (\overset{4}{\mu})^2]\,[(\overset{2}{\nu})^2 + (\overset{3}{\nu})^2 + (\overset{4}{\nu})^2]\\ &< (\overset{1}{\mu})^2\,[(\overset{2}{\nu})^2 + (\overset{3}{\nu})^2 + (\overset{4}{\nu})^2],\end{aligned}$$

and hence, by equation (2.2.3) again, that

$$(\overset{1}{\nu})^2 < (\overset{2}{\nu})^2 + (\overset{3}{\nu})^2 + (\overset{4}{\nu})^2.$$

Thus ν^a is spacelike.

(2) Assume μ^a and ν^a are both null. If they are proportional, then they are trivially orthogonal. For if, say, $\mu^a = k\,\nu^a$, then $\mu^a\nu_a = k(\nu^a\nu_a) = 0$ (since ν^a is null). Assume, conversely, that the vectors are orthogonal. Let ξ^a be a timelike vector at p. By clause (1)—since ν^a is not spacelike—either $\nu^a = \mathbf{0}$ or $\xi^a\nu_a \neq 0$. (Here ξ^a is playing role of μ^a.) In the first case, μ^a and ν^a are trivially proportional. So we may assume that $\xi^a\nu_a \neq 0$. Then there is a number k such that $k\,(\xi^a\nu_a) = \xi^a\mu_a$. Hence, $(\mu^a - k\,\nu^a)\,\xi_a = 0$. Now $(\mu^a - k\,\nu^a)$ is not spacelike. (The right side of

$$(\mu^a - k\,\nu^a)(\mu_a - k\,\nu_a) = \mu^a\mu_a - 2\,k\,(\mu^a\nu_a) + k^2\,(\nu^a\nu_a)$$

is 0 since, by assumption, μ^a and ν^a are null and $\mu^a\nu_a = 0$.) So, by clause (1) again, it must be the case that $(\mu^a - k\,\nu^a) = \mathbf{0}$; i.e., μ^a and ν^a are proportional. □

PROBLEM 2.2.1. *Let p be a point in M. Let p be a point in M. Show that there is no two-dimensional subspace of M_p all of whose elements are causal (timelike or null).*

PROBLEM 2.2.2. *Let g'_{ab} be a second metric on M (not necessarily of Lorentz signature). Show that the following conditions are equivalent.*

(1) *For all p in M, g_{ab} and g'_{ab} agree on which vectors at p are orthogonal.*
(2) *g'_{ab} is conformally equivalent to either g_{ab} or $-g_{ab}$.*

Next we consider the "lobes" of the null cone determined by g_{ab} at points of M. Let us say that two timelike vectors μ^a and ν^a at a point are *co-oriented* (or *have the same orientation*) if $\mu^a\nu_a > 0$.

PROPOSITION 2.2.2. *For all points p in M, co-orientation is an equivalence relation on the set of timelike vectors in M_p.*

Proof. Reflexivity and symmetry are immediate. For transitivity, let μ^a, ν^a, and ω^a be timelike vectors at a point, with the pairs $\{\mu^a, \nu^a\}$ and $\{\nu^a, \omega^a\}$ both co-oriented. We must show that $\{\mu^a, \omega^a\}$ is co-oriented as well. The argument is very much like that for the second clause of proposition 2.2.1.

Since $\mu^a \nu_a > 0$ and $\omega^a \nu_a > 0$, there is a real number $k > 0$ such that $\mu^a \nu_a = k\,(\omega^a \nu_a)$. Hence, $(\mu^a - k\,\omega^a)\nu_a = 0$. Since ν^a is timelike, we know from the first clause of proposition 2.2.1 that either $(\mu^a - k\,\omega^a)$ is the zero vector $\mathbf{0}$ or it is spacelike. In the first case, $\mu^a = k\,\omega^a$, and so the pair $\{\mu^a, \omega^a\}$ is certainly co-oriented ($\mu^a \omega_a = k\,(\omega^a \omega_a) > 0$). So we may assume that $(\mu^a - k\,\omega^a)$ is spacelike. But then

$$\mu^a \mu_a - 2\,k\,(\mu^a \omega_a) + k^2\,(\omega^a \omega_a) = (\mu^a - k\,\omega^a)(\mu_a - k\,\omega_a) < 0.$$

Since $\mu^a \mu_a$, $\omega^a \omega_a$, and k are all positive, it follows that $\mu^a \omega_a$ is positive as well. So, again, we are led to the conclusion that the pair $\{\mu^a, \omega^a\}$ is co-oriented. $\square$

The equivalence classes determined at each point by the co-orientation relation will be called *temporal lobes*. There must be at least two lobes at each point since, for any timelike vector μ^a there, μ^a and $-\mu^a$ are not co-oriented. There cannot be more than two since, for all timelike μ^a and ν^a at a point, ν^a is co-oriented either with μ^a or with $-\mu^a$. (Remember, two timelike vectors at a point cannot be orthogonal.) Hence there are exactly two lobes at each point. It is easy to check that each lobe is convex; i.e., if μ^a and ν^a are co-oriented at a point, and a, b are both positive real numbers, then $(a\,\mu^a + b\,\nu^a)$ is a timelike vector at the point that is co-oriented with μ^a and ν^a.

The relation of co-orientation can be extended easily to the larger set of non-zero causal (i.e., timelike or null) vectors. Given any two such vectors μ^a and ν^a at a point, we can take them to be *co-oriented* if *either* $\mu^a \nu_a > 0$ or $\nu^a = k\,\mu^a$, with $k > 0$. (The second possibility must be allowed since we want a non-zero null vector to count as being co-oriented with itself.) Once again, co-orientation turns out to be an equivalence relation with two equivalence classes that we call *causal lobes*. (Only minor changes in the proof of proposition 2.2.2 are required to establish that the extended co-orientation relation is transitive.) These lobes, too, are convex.

For future reference, we record two more facts about Lorentz metrics. (Let us agree to write $\|\mu^a\|$ for $(\mu^a \mu_a)^{\frac{1}{2}}$ when μ^a is causal, and write it for $(-\,\mu^a \mu_a)^{\frac{1}{2}}$ when μ^a is spacelike.)

PROPOSITION 2.2.3. *Let μ^a and ν^a be causal vectors at some point p in M. Then the following both hold.*

(1) *("Wrong way Schwarz inequality") $|\mu^a \nu_a| \geq \|\mu^a\| \, \|\nu^a\|$, with equality iff μ^a and ν^a are proportional.*

(2) *("Wrong way triangle inequality") If μ^a and ν^a are non-zero and co-oriented,*

$$\|\mu^a + \nu^a\| \geq \|\mu^a\| + \|\nu^a\|,$$

with equality iff μ^a and ν^a are proportional.

Proof. (1) If both μ^a and ν^a are null, the assertion follows immediately from the second assertion in proposition 2.2.1. So we may assume that one of the vectors, say μ^a, is timelike. Now we can certainly express ν^a in the form $\nu^a = k\,\mu^a + \sigma^a$, with k a real number and σ^a a vector at p orthogonal to μ^a. (It suffices to take $k = (\mu^a \nu_a)/(\mu^a \mu_a)$ and $\sigma^a = (\nu^a - k\,\mu^a)$.) Hence,

$$\mu^a \nu_a = k\,(\mu^a \mu_a),$$
$$\nu^a \nu_a = k^2\,(\mu^a \mu_a) + \sigma^a \sigma_a.$$

Since σ^a is orthogonal to μ^a, it must either be spacelike or the zero vector (by proposition 2.2.1). In either case, $(\sigma^a \sigma_a) \leq 0$. So, since $(\mu^a \mu_a) > 0$ and $(\nu^a \nu_a) \geq 0$, it follows that

$$\begin{aligned}(\mu^a \nu_a)^2 = k^2\,(\mu^a \mu_a)^2 &= [(\nu^a \nu_a) - (\sigma^a \sigma_a)]\,(\mu^a \mu_a)\\ &\geq (\nu^a \nu_a)\,(\mu^a \mu_a) = \|\mu^a\|^2\,\|\nu^a\|^2.\end{aligned}$$

Equality holds here iff $(\sigma^a \sigma_a) = 0$. But (as noted already), σ^a is either the zero vector or spacelike (in which case $(\sigma^a \sigma_a) < 0$). So equality holds iff $\sigma^a = \mathbf{0}$; i.e., $\nu^a = k\,\mu^a$.

We leave the second clause as an exercise. □

PROBLEM 2.2.3. *Prove the second clause of proposition 2.2.3.*

Now we switch our attention to considerations of global null cone structure. We say that (M, g_{ab}) is *temporally orientable* if there exists a continuous timelike vector field τ^a on M. Suppose the condition is satisfied. Then we take two such fields τ^a and τ'^a to be *co-oriented* if they are so at every point—i.e., if $\tau^a \tau'_a > 0$ holds at every point of M. Co-orientation, now understood as a relation on continuous timelike vector fields, is an equivalence relation with two equivalence classes. (It inherits this property from the original relation defined on timelike vectors at individual points.) A *temporal orientation* of (M, g_{ab}) is a choice of one of those two equivalence classes to count as the "future" one. Thus, a non-zero causal vector ξ^a at a point of M is said to be *future-directed* or *past-directed* with respect to the temporal orientation $\mathcal{T}$ depending on whether $\tau^a \xi_a > 0$ or $\tau^a \xi_a < 0$ at the point, where τ^a is any continuous timelike vector field in $\mathcal{T}$. (Remember, $\tau^a \xi_a$ cannot be 0, since no timelike vector can be orthogonal to a non-zero causal vector.) Derivatively, a smooth, causal curve $\gamma\colon I \to M$ is said to be *future-directed* (respectively *past-directed*) with respect to $\mathcal{T}$ if its tangent vector at every point is so.

Our characterization of "relativistic spacetimes" in the preceding section does not guarantee temporal orientability. But we shall take the condition for granted in what follows. We assume that our background spacetime (M, g_{ab}) is temporally orientable and that a particular temporal orientation has been specified.

Also, given points p and q in M, we shall write $p \ll q$ (resp. $p < q$) if there is a smooth, future-directed, timelike (respectively, causal) curve $\gamma : [a, b] \to M$ where $\gamma(a) = p$ and $\gamma(b) = q$. Note that $p < p$, for all points p in all spacetimes. (This is the case because the zero vector in the tangent space at any point qualifies as a null vector.) But it is not the case, in general, that $p \ll p$. The latter condition holds iff there is a smoooth, closed, future-directed timelike curve that begins and ends at p. The two relations $\ll$ and $<$ are naturally construed as relations of "causal connectibility (or accessibility)."

Appendix: Recovering Geometric Structure from the Causal Connectibility Relation

We started with a spacetime model (M, g_{ab}) exhibiting several levels of geometric structure, and used the latter to define the relations $\ll$ and $<$ on M.[7] The question now arises whether it is possible to work backward—i.e., *start* with the pair $(M, \ll)$ or $(M, <)$, with M now construed as a bare point set, and recover the geometric structure with which one began.[8] In this appendix, we briefly consider one way to make the question precise and give the answer (without proof). For convenience, we work with the relation $\ll$.

Let (M, g_{ab}) and (M', g'_{ab}) be (temporally oriented) relativistic spacetimes. We say that a bijection $\varphi\colon M \to M'$ between their underlying point sets is a *$\ll$-causal isomorphism* if, for all p and q in M,

7. The material in this appendix will play no role in what follows.
8. The question figures centrally in the "causal sets" approach to quantum gravity developed by Rafael Sorkin and co-workers. See, e.g., Sorkin [55, 56].

$$p \ll q \iff \varphi(p) \ll \varphi(q). \tag{2.2.5}$$

Then we can ask the following: *Does a $\ll$-causal isomorphism have to be a homeomorphism? A diffeomorphism? A conformal isometry?* (We know in advance that a causal isomorphism need not be a (full) isometry because conformally equivalent metrics g_{ab} and $\Omega^2 g_{ab}$ on a manifold M determine the same relation $\ll$. The best one can ask for is that it be a conformal isometry—i.e. that it be a diffeomorphism that preserves the metric up to a conformal factor.)

Without further restrictions on (M, g_{ab}) and (M', g'_{ab}), the answer is certainly "no" to all three questions. Unless the "causal structure" of a spacetime (i.e., the structure determined by $\ll$) is reasonably well behaved, it provides no useful information at all. For example, let us say that a spacetime is *causally degenerate* if $p \ll q$ for all points p and q. *Any* bijection between two causally degenerate spacetimes qualifies, trivially, as a $\ll$-causal isomorphism. But we can certainly find causally degenerate spacetimes whose underlying manifolds have different topologies. For example, we shall verify in section 3.1 that Gödel spacetime is causally degenerate. Its underlying manifold structure is $\mathbb{R}^4$. But a suitably "rolled-up" version of Minkowski spacetime is also causally degenerate, and the latter has the manifold structure $S^1 \times \mathbb{R}^3$. (Figure 2.2.1 shows a two-dimensional version.)

There is a hierarchy of "causality conditions" that is relevant here. (See, e.g., Hawking and Ellis [30, section 6.4].) They impose, with varying degrees of stringency, the requirement that there exist no closed, or "almost closed," timelike curves. Here are three.

Chronology: There do not exist smooth closed timelike curves. (Equivalently, for all p, it is *not* the case that $p \ll p$.)

Future (respectively, past) distinguishablity: For all points p, and all sufficiently small open sets O containing p, no smooth future-directed (respectively, past-directed) timelike curve that starts at p, and leaves O, ever returns to O.

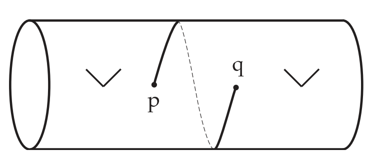

Figure 2.2.1. Two-dimensional Minkowski spacetime rolled up into a cylindrical spacetime. It is causally degenerate: $p \ll q$ for all points p and q.

Strong causality: For all points p, and all sufficiently small open sets O containing p, no smooth future-directed timelike curve that starts in O, and leaves O, ever returns to O.

It is clear that strong causality implies both future distinguishability and past distinguishability, and that each of the distinguishability conditions (alone) implies chronology. Standard examples (see Hawking and Ellis [30]) establish that the converse implications do not hold, and that neither distinguishability condition implies the other.

The names "future distinguishability" and "past distinguishability" are easily explained. For any p, let $I^+(p)$ be the set $\{q: p \ll q\}$ and let $I^-(p)$ be the set $\{q: q \ll p\}$. It turns out (see Kronheimer and Penrose [33]) that future distinguishability is equivalent to the requirement that, for all p and q,

$$I^+(p) = I^+(q) \Longrightarrow p = q.$$

And the counterpart requirement with I^+ replaced by I^- is equivalent to past distinguishability.

We mention all this because it turns out that one gets a positive answer to all three questions above if one restricts attention to spacetimes that are *both* future and past distinguishing.

PROPOSITION 2.2.4. *Let (M, g_{ab}) and (M', g'_{ab}) be (temporally oriented) relativistic spacetimes that are both future- and past-distinguishing, and let $\varphi: M \to M'$ be a $\ll$-causal isomorphism. Then φ is a diffeomorphism and preserves g_{ab} up to a conformal factor; i.e. $\varphi^\star(g'_{ab})$ is conformally equivalent to g_{ab}.*

One can prove the proposition in two stages. First one shows that, under the stated assumptions, φ must be a homeomorphism (see Malament [38]).[9] Then one invokes a result of Hawking, King, and McCarthy [29, theorem 5] that asserts, in effect, that any *continuous* $\ll$-causal isomorphism must be smooth and must preserve the metric up to a conformal factor.

The following example shows that the proposition fails if the initial restriction on causal structure is weakened to past distinguishability or to future distinguishability alone. We give the example in a two-dimensional version

9. This is a slight improvement on a well-known result. If a spacetime (M, g_{ab}) is not just past and future distinguishing, but strongly causal, then one can explicitly characterize its (manifold) topology in terms of the relation $\ll$. In this case, a subset $O \subseteq M$ is open iff, for all points p in O, there exist points q and r in O such that $q \ll p \ll r$ and $I^+(q) \cap I^-(r) \subseteq O$ (Hawking and Ellis [30, p. 196]). So a $\ll$-causal isomorphism between two strongly causal spacetimes must certainly be a homeomorphism.

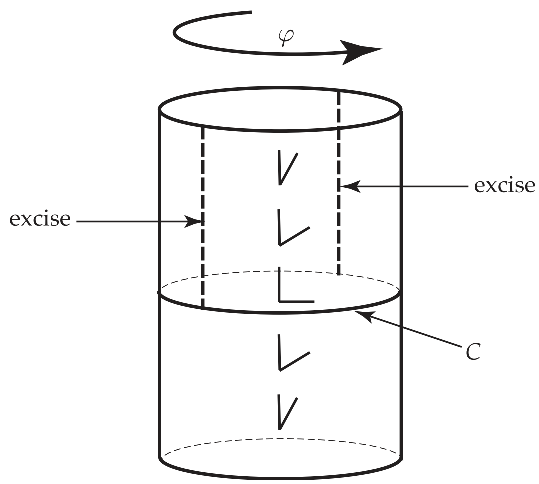

Figure 2.2.2. An example of a spacetime that is future distinguishing but not past distinguishing. Let φ be a bijection of the spacetime onto itself that leaves the lower open half below C fixed but reverses the position of the two upper slabs. It is a $\ll$-isomorphism, but it is discontinuous along C.

to simplify matters. Start with the manifold $\mathbb{R}^2$ together with the Lorentzian metric

$$g_{ab} = (d_{(a}t)(d_{b)}x) - (\sinh^2 t)(d_a x)(d_b x),$$

where t, x are global projection coordinates on $\mathbb{R}^2$. Next form a vertical cylinder by identifying the point with coordinates $(t,\ x)$ with the one having coordinates $(t,\ x+2)$. Finally, excise two closed half lines—the sets with respective coordinates $\{(t,x)\colon\ x=0 \text{ and } t \geq 0\}$ and $\{(t,x)\colon\ x=1 \text{ and } t \geq 0\}$. Figure 2.2.2 shows, roughly, what the null cones look like at every point. (The future direction at each point is taken to be the "upward one.") The exact form of the metric is not important here. All that is important is the indicated qualitative behavior of the null cones. Along the (punctured) circle C where $t = 0$, the vector fields $(\partial/\partial t)^a$ and $(\partial/\partial x)^a$ both qualify as null. But as one moves upward or downward from there, the cones close. There are no closed timelike (or null) curves in this spacetime. Indeed, it is future distinguishing because of the excisions. But it fails to be past distinguishing because $I^-(p) = I^-(q)$ for all points p and q on C. For all points p there, $I^-(p)$ is the entire region below C.

Now let φ be the bijection of the spacetime onto itself that leaves the "lower open half" fixed but reverses the position of the two upper slabs. Though φ is discontinuous along C, it *is* a $\ll$-causal isomorphism. This is the case because every point below C has all points in *both* upper slabs in its $\ll$-future.

2.3. Proper Time

So far we have discussed relativistic spacetime structure without reference to either "time" or "space." We come to them in this section and the next.

Let $\gamma : [s_1, s_2] \to M$ be a smooth, future-directed timelike curve in M with tangent field ξ^a. We associate with it an elapsed *proper time* (relative to g_{ab}) given by

$$\|\gamma\| = \int_{s_1}^{s_2} \left(g_{ab}\, \xi^a\, \xi^b\right)^{\frac{1}{2}}\, ds.$$

This elapsed proper time is invariant under reparametrization of γ and is just what we would otherwise describe as the length of (the image of) γ. The following is another basic principle of relativity theory.

(P2) Clocks record the passage of elapsed proper time along their worldlines.

Again, a number of qualifications and comments are called for. Our formulations of (C1), (C2), and (P1) were rough. The present formulation is that much more so. We have taken for granted that we know what "clocks" are. We have assumed that they have worldlines (rather than worldtubes). And we have overlooked the fact that ordinary clocks (e.g., the alarm clock on the nightstand) do not do well at all when subjected to extreme acceleration, tidal forces, and so forth. (Try smashing the alarm clock against the wall.) Again, these concerns are important and raise interesting questions about the role of idealization in the formulation of physical theory. (One might construe an "ideal clock" as a point-size test object that perfectly records the passage of proper time along its worldline, and then take (P2) to assert that real clocks are, under appropriate conditions and to varying degrees of accuracy, approximately ideal.) But they do not have much to do with relativity theory as such. Similar concerns arise when one attempts to formulate corresponding principles about clock behavior within the framework of Newtonian theory.

Now suppose that one has determined the conformal structure of spacetime, say, by using light rays. Then one can use clocks, rather than free particles, to determine the conformal factor. One has the following simple result, which should be compared with proposition 2.1.4.[10]

10. Here we not only determine the metric up to a constant, but determine the constant as well. The difference is that here, in effect, we have built in a choice of units for spacetime distance. We could obtain a more exact counterpart to proposition 2.1.4 if we worked, not with intervals of elapsed proper time, but rather with ratios of such intervals. (Note, by the way, that the condition in the second sentence of the proposition does not make sense unless the two metrics are conformally equivalent. We cannot require that they assign the same length to all timelike curves unless they first agree on which curves *are* timelike.)

PROPOSITION 2.3.1. *Let g'_{ab} be a second smooth metric on M, with $g'_{ab} = \Omega^2\, g_{ab}$. Further suppose that the two metrics assign the same lengths to timelike curves—i.e., $\|\gamma\|_{g'_{ab}} = \|\gamma\|_{g_{ab}}$ for all smooth, timelike curves $\gamma : I \to M$. Then $\Omega = 1$ everywhere. (Here $\|\gamma\|_{g_{ab}}$ is the length of γ relative to g_{ab}.)*

Proof. Let $\overset{o}{\xi}{}^a$ be an arbitrary timelike vector at an arbitrary point p in M. We can certainly find a smooth, timelike curve $\gamma : [s_1, s_2] \to M$ through p whose tangent at p is $\overset{o}{\xi}{}^a$. By our hypothesis, $\|\gamma\|_{g'_{ab}} = \|\gamma\|_{g_{ab}}$. So, if ξ^a is the tangent field to γ,

$$\int_{s_1}^{s} \left(g'_{ab}\, \xi^a\, \xi^b\right)^{\frac{1}{2}}\, ds = \int_{s_1}^{s} \left(g_{ab}\, \xi^a\, \xi^b\right)^{\frac{1}{2}}\, ds$$

for all s in $[s_1, s_2]$. It follows that $g'_{ab}\, \xi^a \xi^b = g_{ab}\, \xi^a \xi^b$ at every point on the image of γ. In particular, it follows that $(g'_{ab} - g_{ab})\, \overset{o}{\xi}{}^a\, \overset{o}{\xi}{}^b = 0$ at p. But $\overset{o}{\xi}{}^a$ was an arbitrary timelike vector at p. So, by lemma 2.1.3, $g'_{ab} = g_{ab}$ at our arbitary point p. □

(P2) gives the whole story of relativistic clock behavior (modulo the concerns noted above). In particular, it implies the path dependence of clock readings. If two clocks start at an event p and travel along different trajectories to an event q, then, in general, they will record different elapsed times for the trip. (For example, one will record an elapsed time of 3,806 seconds, the other 649 seconds.) This is true no matter how similar the clocks are. (We may stipulate that they came off the same assembly line.) This is the case because, as (P2) asserts, the elapsed time recorded by each of the clocks is just the length of the timelike curve it traverses from p to q and, in general, those lengths will be different.

Suppose we consider all future-directed timelike curves from p to q. It is natural to ask if there are any that minimize or maximize the recorded elapsed time between the events. The answer to the first question is "no." Indeed, one has the following proposition.

PROPOSITION 2.3.2. *Let p and q be events in M such that $p \ll q$. Then, for all $\epsilon > 0$, there exists a smooth, future directed timelike curve γ from p to q with $\|\gamma\| < \epsilon$. (But there is no such curve with length 0, since all timelike curves have non-zero length.)*

Though some work is required to give the proposition an honest proof (see O'Neill [46, pp. 294–295]), it should seem intuitively plausible. If there is a

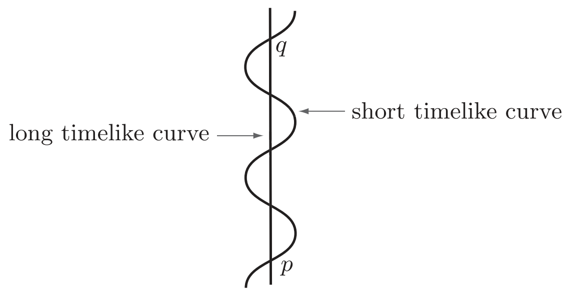

Figure 2.3.1. A long timelike curve from p to q and a very short one that approximates a broken null curve.

smooth, timelike curve connecting p and q, there is also a jointed, zig-zag null curve connecting them. It has length 0. But we can approximate the jointed null curve arbitrarily closely with smooth timelike curves that swing back and forth. So (by the continuity of the length function), we should expect that, for all $\epsilon > 0$, there is an approximating timelike curve that has length less than ϵ. (See figure 2.3.1.)

The answer to the second question ("Can one *maximize* recorded elapsed time between p and q?") is "yes" if one restricts attention to local regions of spacetime. In the case of positive definite metrics, i.e., ones with signature of form $(n, 0)$—we know geodesics are *locally shortest* curves. The corresponding result for Lorentzian metrics is that timelike geodesics are *locally longest* curves.

PROPOSITION 2.3.3. *Let $\gamma : I \to M$ be a smooth, future-directed, timelike curve. Then γ can be reparametrized so as to be a geodesic iff for all $s \in I$ there exists an open set O containing $\gamma(s)$ such that, for all $s_1, s_2 \in I$ with $s_1 \leq s \leq s_2$, if the image of $\gamma' = \gamma_{|[s_1,s_2]}$ is contained in O, then γ' (and its reparametrizations) are longer than all other timelike curves in O from $\gamma(s_1)$ to $\gamma(s_2)$. (Here $\gamma_{|[s_1,s_2]}$ is the restriction of γ to the interval* $[s_1, s_2]$.)

The proof of the proposition is very much the same as in the positive definite case. (See Hawking and Ellis [30, p. 105].) Thus, of all clocks passing locally from p to q, the one that will record the greatest elapsed time is the one that "falls freely" from p to q. To get a clock to read a smaller elapsed time than the maximal value, one will have to accelerate the clock. Now, acceleration requires fuel, and fuel is not free. So proposition 2.3.3 has the consequence that (locally) "saving time costs money." And proposition 2.3.2 may be taken to imply that "with enough money one can save as much time as one wants."

The restriction here to local regions of spacetime is essential. The connection described between clock behavior and acceleration does not, in general, hold on a global scale. In some relativistic spacetimes, one can find future-directed timelike geodesics connecting two events that have different lengths, and so clocks following the curves will record different elapsed times between the events even though *both* are in a state of free fall. Furthermore—this follows from the preceding claim by continuity considerations alone—it can be the case that of two clocks passing between the events, the one that undergoes acceleration during the trip records a *greater* elapsed time than the one that remains in a state of free fall. (A rolled-up version of two-dimensional Minkowski spacetime provides a simple example. See figure 2.3.2.)

The connection we have been considering between clock behavior and acceleration was once thought to be paradoxical. Recall the so-called "clock paradox." Suppose two clocks, A and B, pass from one event to another in a suitably small region of spacetime. Further suppose A does so in a state of free fall but B undergoes acceleration at some point along the way. Then, we know, A will record a greater elapsed time for the trip than B. This was thought paradoxical because it was believed that relativity theory denies the possibility of distinguishing "absolutely" between free-fall motion and accelerated motion. (If we are equally well entitled to think that it is clock B that is in a state of free fall and A that undergoes acceleration, then, by parity of reasoning, it should be B that records the greater elapsed time.) The resolution of the paradox, if one can call it that, is that relativity theory makes no such denial. The situations of A and B here are *not* symmetric. The distinction between accelerated motion

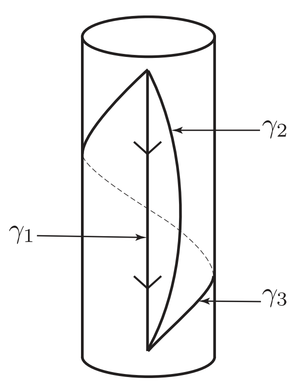

Figure 2.3.2. Two-dimensional Minkowski spacetime rolledup into a cylindrical spacetime. Three timelike curves are displayed: γ_1 and γ_3 are geodesics; γ_2 is not; γ_1 is longer than γ_2; and γ_2 is longer than γ_3.

and free fall makes every bit as much sense in relativity theory as it does in Newtonian physics.

In what follows, unless indication is given to the contrary, a "timelike curve" should be understood to be a smooth, future-directed, timelike curve parametrized by elapsed proper time—i.e., by arc length. In that case, the tangent field ξ^a of the curve has unit length ($\xi^a \xi_a = 1$). And if a particle happens to have the image of the curve as its worldline, then, at any point, ξ^a is called the particle's *four-velocity* there.

2.4. Space/Time Decomposition at a Point and Particle Dynamics

Let γ be a smooth, future-directed, timelike curve with unit tangent field ξ^a in our background spacetime (M, g_{ab}). We suppose that some massive point particle O has (the image of) this curve as its worldline. Further, let p be a point on the image of γ and let λ^a be a vector at p. Then there is a natural decomposition of λ^a into components proportional to, and orthogonal to, ξ^a:

$$\lambda^a = \underbrace{(\lambda^b \xi_b)\xi^a}_{\text{proportional to } \xi^a} + \underbrace{(\lambda^a - (\lambda^b \xi_b)\xi^a)}_{\text{orthogonal to } \xi^a}. \tag{2.4.1}$$

These are standardly interpreted, respectively, as the "temporal" and "spatial" components of λ^a relative to ξ^a (or relative to O). In particular, the three-dimensional vector space of vectors at p orthogonal to ξ^a is interpreted as the "infinitesimal" simultaneity slice of O at p.[11] If we introduce the tangent and orthogonal projection operators

$$k_{ab} = \xi_a \xi_b, \tag{2.4.2}$$

$$h_{ab} = g_{ab} - \xi_a \xi_b, \tag{2.4.3}$$

then the decomposition can be expressed in the form

$$\lambda^a = k^a{}_b \lambda^b + h^a{}_b \lambda^b. \tag{2.4.4}$$

We can think of k_{ab} and h_{ab} as the relative temporal and spatial metrics determined by ξ^a. They are symmetric and satisfy

$$k^a{}_b k^b{}_c = k^a{}_c, \tag{2.4.5}$$

$$h^a{}_b h^b{}_c = h^a{}_c. \tag{2.4.6}$$

11. Here we simply take for granted the standard identification of "relative simultaneity" with orthogonality. For discussion of how the identification is justified, see Malament [42, section 3.1] and further references cited there.

Many standard textbook assertions concerning the kinematics and dynamics of point particles can be recovered using these decomposition formulas. For example, suppose that the worldline of a second particle O' also passes through p and that its four-velocity at p is ξ'^a. (Since ξ^a and ξ'^a are both future-directed, they are co-oriented; i.e., $\xi^a \xi'_a > 0$.) We compute the speed of O' as determined by O. To do so, we take the spatial magnitude of ξ'^a relative to O and divide by its temporal magnitude relative to O:[12]

$$v = \textit{speed of } O' \textit{ relative to } O = \frac{\|h^a{}_b\, \xi'^b\|}{\|k^a{}_b\, \xi'^b\|}. \tag{2.4.7}$$

(Recall that for any vector μ^a, $\|\mu^a\|$ is $(\mu^a \mu_a)^{\frac{1}{2}}$ if μ^a is causal, and it is $(-\mu^a \mu_a)^{\frac{1}{2}}$ otherwise.) From equations (2.4.2), (2.4.3), (2.4.5), and (2.4.6), we have

$$\|k^a{}_b\, \xi'^b\| = (k^a{}_b\, \xi'^b\, k_{ac}\, \xi'^c)^{\frac{1}{2}} = (k_{bc}\, \xi'^b\, \xi'^c)^{\frac{1}{2}} = (\xi'^b\, \xi_b)$$

and

$$\|h^a{}_b\, \xi'^b\| = (-h^a{}_b\, \xi'^b\, h_{ac}\, \xi'^c)^{\frac{1}{2}} = (-h_{bc}\, \xi'^b\, \xi'^c)^{\frac{1}{2}} = ((\xi'^b\, \xi_b)^2 - 1)^{\frac{1}{2}}.$$

So

$$v = \frac{((\xi'^b\, \xi_b)^2 - 1)^{\frac{1}{2}}}{(\xi'^b\, \xi_b)} < 1. \tag{2.4.8}$$

Thus, as measured by O, no massive particle can ever attain the maximal speed 1. (A similar calculation shows that, as determined by O, light always travels with speed 1.) For future reference, we note that equation (2.4.8) implies that

$$(\xi'^b\, \xi_b) = \frac{1}{\sqrt{1 - v^2}}. \tag{2.4.9}$$

It is a basic fact of relativistic life that there is associated with every point particle, at every event on its worldline, a *four-momentum* (or *energy-momentum*) vector P^a that is tangent to its worldline there. The length $\|P^a\|$ of this vector is what we would otherwise call the *mass* (or *inertial mass* or *rest mass*) of the particle. So, in particular, if P^a is timelike, we can write it in the form $P^a = m\,\xi^a$, where $m = \|P^a\| > 0$ and ξ^a is the four-velocity of the particle. No such decomposition is possible when P^a is null and $m = \|P^a\| = 0$.

Suppose a particle O with positive mass has four-velocity ξ^a at a point, and another particle O' has four-momentum P^a there. The latter can either be a particle with positive mass or mass 0. We can recover the usual expressions

12. We are, in effect, choosing units in which $c = 1$.

for the energy and three-momentum of the second particle relative to O if we decompose P^a in terms of ξ^a. By equations (2.4.4) and (2.4.2), we have

$$P^a = \underbrace{(P^b \xi_b)}_{energy}\, \xi^a + \underbrace{h^a_{\ b} P^b}_{three-momentum} . \tag{2.4.10}$$

The *energy* relative to O is the coefficient in the first term: $E = P^b \xi_b$. If O' has positive mass and $P^a = m\, \xi'^a$, this yields, by equation (2.4.9),

$$E = m\, (\xi'^b\, \xi_b) = \frac{m}{\sqrt{1 - v^2}}. \tag{2.4.11}$$

(If we had not chosen units in which $c = 1$, the numerator in the final expression would have been mc^2 and the denominator $\sqrt{1 - (v^2/c^2)}$.) The *three-momentum* relative to O is the second term $h^a_{\ b} P^b$ in the decomposition of P^a—i.e., the component of P^a orthogonal to ξ^a. It follows from equations (2.4.8) and (2.4.9) that it has magnitude

$$p = \|h^a_{\ b}\, m\, \xi'^b\| = m\, ((\xi'^b\, \xi_b)^2 - 1)^{\frac{1}{2}} = \frac{m\, v}{\sqrt{1 - v^2}}. \tag{2.4.12}$$

Interpretive principle (P1) asserts that the worldlines of *free* particles with positive mass are the images of timelike geodesics. It can be thought of as a relativistic version of Newton's first law of motion. Now we consider acceleration and a relativistic version of the second law. Once again, let $\gamma : I \rightarrow M$ be a smooth, future-directed, timelike curve with unit tangent field ξ^a. Just as we understand ξ^a to be the four-velocity field of a massive point particle (that has the image of γ as its worldline), so we understand $\xi^n \nabla_n\, \xi^a$—the directional derivative of ξ^a in the direction ξ^a—to be its *four-acceleration field* (or just *acceleration) field*). The four-acceleration vector at any point is orthogonal to ξ^a. (This is clear, since $\xi^a\, (\xi^n \nabla_n\, \xi_a) = \frac{1}{2}\, \xi^n \nabla_n\, (\xi^a\, \xi_a) = \frac{1}{2}\, \xi^n \nabla_n\, (1) = 0$.) The magnitude $\|\xi^n \nabla_n\, \xi^a\|$ of the four-acceleration vector at a point is just what we would otherwise describe as the curvature of γ there. It is a measure of the rate at which γ "changes direction." (And γ is a geodesic precisely if its curvature vanishes everywhere.)

The notion of spacetime acceleration requires attention. Consider an example. Suppose you decide to end it all and jump off the Empire State Building. What would your acceleration history be like during your final moments? One is accustomed in such cases to think in terms of acceleration relative to the earth. So one would say that you undergo acceleration between the time of your jump and your calamitous arrival. But on the present account, that description has things backwards. Between jump and arrival, you are *not* accelerating. You are in a state of free fall and moving (approximately) along a spacetime geodesic. But before the jump, and after the arrival, you *are* accelerating. The floor of the observation deck, and then later the sidewalk, push you away from a geodesic path. The all-important idea here is that we are incorporating the "gravitational field" into the geometric structure of spacetime, and particles traverse geodesics if and only if they are acted on by no forces "except gravity."

The acceleration of our massive point particle—i.e., its deviation from a geodesic trajectory—is determined by the forces acting on it (other than "gravity"). If it has mass m, and if the vector field F^a on I represents the vector sum of the various (non-gravitational) forces acting on it, then the particle's four-acceleration $\xi^n \nabla_n \xi^a$ satisfies

$$F^a = m\, \xi^n \nabla_n \xi^a. \tag{2.4.13}$$

This is our version of Newton's second law of motion.

Consider an example. (Here we anticipate our discussion in section 2.6.) Electromagnetic fields are represented by smooth, anti-symmetric fields F_{ab}. If a particle with mass $m > 0$, charge q, and four-velocity field ξ^a is present, the force exerted by the field on the particle at a point is given by $q\, F^a{}_b\, \xi^b$. If we use this expression for the left side of equation (2.4.13), we arrive at the Lorentz law of motion for charged particles in the presence of an electromagnetic field:

$$q\, F^a{}_b\, \xi^b = m\, \xi^b\, \nabla_b\, \xi^a. \tag{2.4.14}$$

(Notice that the equation makes geometric sense. The acceleration field on the right is orthogonal to ξ^a. But so is the force field on the left, since $\xi_a(F^a{}_b\, \xi^b) = \xi^a\, \xi^b F_{ab} = \xi^a\, \xi^b F_{(ab)}$, and $F_{(ab)} = \mathbf{0}$ by the anti-symmetry of F_{ab}.)

2.5. The Energy-Momentum Field T_{ab}

In classical relativity theory, one generally takes for granted that all there is, and all that happens, can be described in terms of various "matter fields," each of which is represented by one or more smooth tensor (or spinor) fields on the spacetime manifold M.[13] The latter are assumed to satisfy particular "field equations" involving the spacetime metric g_{ab}.

For present purposes, the most important basic assumption about the matter fields is the following.

13. This being the case, the question arises as to how (or whether) one can adequately recover talk about "point particles" in terms of the matter fields. We shall briefly discuss the question later in this section.

> Associated with each matter field $\mathcal{F}$ is a symmetric smooth tensor field T_{ab} characterized by the property that, for all points p in M, and all future-directed, unit timelike vectors ξ^a at p, $T^a{}_b\,\xi^b$ is the four-momentum density of $\mathcal{F}$ at p as determined relative to ξ^a.

T_{ab} is called the *energy-momentum* field associated with $\mathcal{F}$. The four-momentum density vector $T^a{}_b\,\xi^b$ at a point can be further decomposed into its temporal and spatial components relative to ξ^a,

$$T^a{}_b\,\xi^b = \underbrace{(T_{mb}\,\xi^m\,\xi^b)}_{\textit{energy density}}\xi^a + \underbrace{T_{mb}\,h^{ma}\,\xi^b}_{\textit{three-momentum density}} ,$$

just as the four-momentum P^a of a particle was decomposed in equation (2.4.10). The coefficient of ξ^a in the first component, $T_{ab}\,\xi^a\xi^b$, is the *energy density* of $\mathcal{F}$ at the point as determined relative to ξ^a. The second component, $T_{nb}\,(g^{an} - \xi^a\,\xi^n)\,\xi^b$, is the *three-momentum density* of $\mathcal{F}$ there as determined relative to ξ^a.

A number of assumptions about matter fields can be captured as constraints on the energy-momentum tensor fields with which they are associated. Examples are the following. (Suppose T_{ab} is associated with matter field $\mathcal{F}$.)

Weak Energy Condition (WEC): Given any timelike vector ξ^a at any point in M, $T_{ab}\,\xi^a\xi^b \geq 0$.

Dominant Energy Condition (DEC): Given any timelike vector ξ^a at any point in M, $T_{ab}\,\xi^a\xi^b \geq 0$ and $T^a_b\,\xi^b$ is timelike or null.

Strengthened Dominant Energy Condition (SDEC): Given any timelike vector ξ^a at any point in M, $T_{ab}\,\xi^a\xi^b \geq 0$ and, if $T_{ab} \neq \mathbf{0}$ there, then $T^a_b\,\xi^b$ is timelike.

Conservation Condition (CC): $\nabla_a\,T^{ab} = \mathbf{0}$ at all points in M.

The WEC asserts that the energy density of $\mathcal{F}$, as determined by any observer at any point, is non-negative. The DEC adds the requirement that the four-momentum density of $\mathcal{F}$, as determined by any observer at any point, is a future-directed causal (i.e., timelike or null) vector. We can understand this second clause to assert that the energy of $\mathcal{F}$ does not propagate with superluminal velocity. The strengthened version of the DEC just changes "causal" to "timelike" in the second clause. It captures something of the flavor of (C1) in section 2.1, but avoids reference to "point particles." Each of the listed energy conditions is strictly stronger than the ones that precede it (see problem 2.5.1).

PROBLEM 2.5.1. *Give examples of each of the following.*

(1) *A smooth symmetric field* T_{ab} *that does not satisfy the WEC*
(2) *A smooth symmetric field* T_{ab} *that satisfies the WEC but not the DEC*
(3) *A smooth symmetric field* T_{ab} *that satisfies the DEC but not the SDEC*

PROBLEM 2.5.2. *Show that the DEC holds iff given any two co-oriented timelike vectors* ξ^a *and* η^a *at a point in* M, $T_{ab}\,\xi^a\,\eta^b \geq 0$.

The CC, finally, asserts that the energy-momentum carried by $\mathcal{F}$ is locally conserved. If two or more matter fields are present in the same region of spacetime, it need not be the case that each one individually satisfies the condition. Interaction may occur. But it *is* a fundamental assumption that the composite energy-momentum field formed by taking the sum of the individual ones satisfies it. Energy-momentum can be transferred from one matter field to another, but it cannot be created or destroyed.

The stated conditions have a number of consequences that support the interpretations just given. We mention two. The first requires a few preliminary definitions.

A subset S of M is said to be *achronal* if there do *not* exist points p and q in S such that $p \ll q$. Let $\gamma : I \to M$ be a smooth curve. We say that a point p in M is a *future-endpoint* of γ if, for all open sets O containing p, there exists an s_0 in I such that, for all $s \in I$, if $s \geq s_0$, then $\gamma(s) \in O$; i.e., γ eventually enters and remains in O. (*Past-endpoints* are defined similarly.) Now let S be an achronal subset of M. The *domain of dependence* $D(S)$ of S is the set of all points p in M with this property: given any smooth causal curve without (past- or future-) endpoint, if its image contains p, then it intersects S. (See figure 2.5.1.) So, in particular, $S \subseteq D(S)$.

In section 2.10, we shall make precise a sense in which "what happens on S determines what happens throughout $D(S)$." Here we consider just one aspect of that determination.

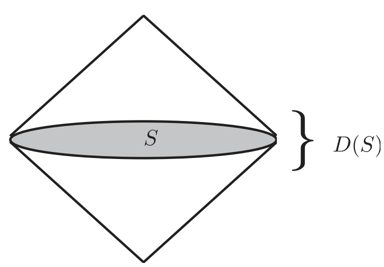

Figure 2.5.1. The domain of dependence *D(S)* of an achronal set *S*.

PROPOSITION 2.5.1. *Let S be an achronal subset of M. Further, let T_{ab} be a smooth, symmetric field on M that satisfies both the dominant energy and conservation conditions. Finally, assume $T_{ab} = \mathbf{0}$ on S. Then $T_{ab} = \mathbf{0}$ on all of $D(S)$.*

The intended interpretation of the proposition is clear. If energy-momentum cannot propagate (locally) outside the null-cone, and if it is conserved, and if it vanishes on S, then it must vanish throughout $D(S)$. After all, how could it "get to" any point in $D(S)$? Note that our formulation of the proposition does not presuppose any particular physical interpretation of the symmetric field T_{ab}. All that is required is that it satisfy the two stated conditions. (For a proof, see Hawking and Ellis [30, p. 94].)

Now recall (P1). It asserts that free massive point particles traverse (images of) timelike geodesics. The next proposition (Geroch and Jang [24]) shows that it is possible, in a sense, to capture the principle as a theorem in relativity theory. The trick is to find a way to talk about "massive point particles" in the language of extended matter fields. In effect, we shall model them as nested sequences of small, but extended, bodies that converge to a point. (See figure 2.5.2.) It turns out that if the energy-momentum content of each body in the sequence satisfies appropriate conditions, then the convergence point will necessarily traverse (the image of) a timelike geodesic.

PROPOSITION 2.5.2. *Let $\gamma: I \to M$ be smooth curve. Suppose that, given any open subset O of M containing $\gamma[I]$, there exists a smooth symmetric field T_{ab} on M such that the following conditions hold.*

(1) *T_{ab} satisfies the SDEC.*
(2) *T_{ab} satisfies the CC.*
(3) *$T_{ab} = \mathbf{0}$ outside of O.*
(4) *$T_{ab} \neq \mathbf{0}$ at some point in O.*

Then γ is timelike and can be reparametrized so as to be a geodesic.

The proposition might be paraphrased this way. Suppose that for some smooth curve γ, arbitrarily small bodies with energy-momentum satisfying conditions (1) and (2) can contain the image of γ in their worldtubes. Then γ must be a timelike geodesic (up to reparametrization). Bodies here are understood to be "free" if their internal energy-momentum is conserved (by itself). If a body is acted on by a field, it is only the composite energy-momentum of the body and field together that is conserved.

Note that our formulation of the proposition takes for granted that we can keep the background spacetime metric g_{ab} fixed while altering the fields T_{ab}

Figure 2.5.2. A non-geodesic timelike curve enclosed in a tube (as considered in propositions 2.5.2 and 2.5.3).

that live on M. This is justifiable only to the extent that we are dealing with test bodies whose effect on the background spacetime structure is negligible.[14] Note also that we do not have to assume at the outset that the curve γ is timelike. That follows from the other assumptions.

We have here a precise proposition in the language of matter fields that, at least to some degree, captures (P1). Similarly, it is possible to capture (C2), concerning the behavior of light, with a proposition about the behavior of solutions to Maxwell's equations in a limiting regime ("the optical limit") where wavelengths are small. It asserts, in effect, that when one passes to this limit, packets of electromagnetic waves are constrained to move along (images of) null geodesics. (See Wald [60, p. 71].)

It is worth noting that the Geroch-Jang result fails if condition (1) is dropped. Consider again our nested sequence of bodies converging to a point. It turns out that the CC alone imposes no restrictions whatsoever on the wordline of the convergence point. It can be a null or spacelike curve. It can also be a timelike curve that exhibits any desired pattern of large or changing acceleration or both. The next proposition, based on a suggestion of Robert Geroch (in personal communication), gives a counterexample.[15]

14. Stronger theorems have been proved (see Ehlers and Geroch [16]) where one still models a point particle as a nested sequence of extended bodies converging to a point but does not require that the perturbative effect of each body in the sequence disappear entirely. One requires only that, in a certain precise sense, it disappear in the limit.

15. It is formulated in terms of an initial curve that is timelike—the case of greatest interest—but that is not essential. The example can also be adapted to show that proposition 2.5.1 fails if the energy condition there is dropped.

PROPOSITION 2.5.3. *Let (M, g_{ab}) be Minkowski spacetime, and let $\gamma : I \to M$ be any smooth timelike curve. Then, given any open subset O of M containing $\gamma[I]$, there exists a smooth symmetric field T_{ab} on M that satisfies conditions (2), (3), and (4) in the preceding proposition. (If we want, we can also strengthen condition (4) and require that T_{ab} be non-vanishing throughout some open subset $O_1 \subseteq O$ containing $\gamma[I]$.)*

Proof. Let O be an open subset of M containing $\gamma[I]$, and let $f : M \to \mathbb{R}$ be *any* smooth scalar field on M. (Later we shall impose further restrictions on f.) Consider the fields $S^{abcd} = f(g^{ad}g^{bc} - g^{ac}g^{bd})$ and $T^{ac} = \nabla_b \nabla_d S^{abcd}$, where ∇ is the (flat) derivative operator on M compatible with g_{ab}. We have

$$T^{ac} = (g^{ad}g^{bc} - g^{ac}g^{bd}) \nabla_b \nabla_d f = \nabla^c \nabla^a f - g^{ac} (\nabla_b \nabla^b f). \tag{2.5.1}$$

So T^{ac} is clearly symmetric. It is also divergence free since

$$\nabla_a T^{ac} = \nabla_a \nabla^c \nabla^a f - \nabla^c \nabla_b \nabla^b f = \nabla^c \nabla_a \nabla^a f - \nabla^c \nabla_b \nabla^b f = \mathbf{0}.$$

(The second equality follows from the fact that ∇ is flat, and so ∇_a and ∇^c commute in their action on arbitrary tensor fields.)

To complete the proof, we now impose further restrictions on f to insure that conditions (3) and (4) are satisfied. Let O_1 be any open subset of M such that $\gamma[I] \subseteq O_1$ and $\text{cl}(O_1) \subseteq O$. (Here $\text{cl}(A)$ is the closure of A.) Our strategy will be to choose a particular f on O_1 and a particular f on $M - \text{cl}(O)$, and then fill in the buffer zone $O - \text{cl}(O_1)$ any way whatsoever (so long as the resultant field is smooth). On $M - \text{cl}(O)$, we simply take $f = 0$. This choice guarantees that, no matter how we smoothly extend f to all of M, T^{ac} will vanish outside of O.

For the other specification, let o be any point in M and let χ^a be the "position field" on M determined relative to o. So $\nabla_a \chi^b = \delta_a{}^b$ everywhere, and $\chi^a = \mathbf{0}$ at o. On O_1, we take $f = -(\chi^n \chi_n)$. With that choice, T^{ac} is non-vanishing at all points in O_1. Indeed, we have

$$\nabla_a f = -2\, \chi_n \nabla_a \chi^n = -2\, \chi_n\, \delta_a{}^n = -2\, \chi_a,$$

and, therefore,

$$\begin{aligned} T^{ac} &= \nabla^c \nabla^a f - g^{ac} (\nabla_b \nabla^b f) = -2\, \nabla^c \chi^a + 2\, g^{ac} (\nabla_b \chi^b) \\ &= -2\, g^{ca} + 2\, g^{ac}\, \delta_b{}^b = -2\, g^{ac} + 8\, g^{ac} = 6\, g^{ac} \end{aligned}$$

throughout O_1. □

One point about the proof deserves comment. As restricted to O_1 and to $M - \mathrm{cl}(O)$, the field T_{ab} that we construct *does* satisfy the SDEC. (In the first case, $T_{ab} = 6\, g_{ab}$, and in the second case, $T_{ab} = \mathbf{0}$.) But we know—from the Geroch-Jang theorem itself—that it cannot satisfy that condition everywhere. So it must fail to do so in the buffer zone $O - \mathrm{cl}(O_1)$. That shows us something. We can certainly choose f in the zone so that it smoothly joins with our choices for f on O_1 and $M - \mathrm{cl}(O)$. But, no matter how clever we are, we cannot do so in such a way that T^{ab} (as expressed in equation (2.5.1)) satisfies the SDEC.

Now we consider two examples of matter fields: perfect fluids in this section, and electromagnetic fields in the next.

"Perfect fluids" are represented by three objects: a smooth four-velocity field η^a, a smooth energy density field ρ, and a smooth isotropic pressure field p (the latter two as determined by a "co-moving" observer at rest in the fluid). In the special case where the pressure p vanishes, one speaks of a "dust field". Particular instances of perfect fluids are characterized by "equations of state" that specify p as a function of ρ. (Specifically excluded here are such complicating factors as anisotropic pressure, shear stress, and viscosity.) Though ρ is generally assumed to be non-negative, some perfect fluids (e.g., to a good approximation, water) can exert negative pressure. The energy-momentum tensor field associated with a perfect fluid is

$$T_{ab} = \rho\, \eta_a\, \eta_b - p\, (g_{ab} - \eta_a\, \eta_b). \tag{2.5.2}$$

So the energy-momentum density vector of the fluid at any point as determined by a co-moving observer (i.e., as determined relative to η^a) is $T^a_{\ b}\, \eta^b = \rho\, \eta^a$.

In the case of a perfect fluid, the WEC, DEC, and CC come out as follows.[16]

$$\begin{array}{lcl}
\text{WEC} & \iff & \rho \geq 0 \quad \text{and} \quad p \geq -\rho \\
\text{DEC} & \iff & |p| \leq \rho \\
\text{CC} & \iff & \begin{cases} (\rho + p)\, \eta^a\, \nabla_a\, \eta^b - (g^{ab} - \eta^b\, \eta^a)\, \nabla_a\, p & = \mathbf{0} \\ \eta^a\, \nabla_a\, \rho + (\rho + p)\, (\nabla_a\, \eta^a) & = 0 \end{cases}
\end{array}$$

First we verify the equivalences for the WEC and CC. (The one for the DEC is left as an exercise.) Then we make a few remarks about the physical interpretation of the two conditions jointly equivalent to CC.

(*WEC*) Clearly, the WEC holds at a point q in M iff $T_{ab}\, \xi^a \xi^b \geq 0$ for all unit timelike vectors ξ^a at q. (If the inequality holds for all unit timelike vectors, it

16. The DEC and the SDEC are not equivalent in general, as we have seen. But they are equivalent when applied, specifically, to perfect fluids. See problem 2.5.3.

holds for all timelike vectors.) It is convenient to work with the condition in this form.

If T_{ab} is given by equation (2.5.2), and ξ^a is a unit timelike vector at q, then $T_{ab}\,\xi^a\xi^b = (\rho+p)(\eta^a\xi_a)^2 - p$. So the WEC holds at q in M iff, for all such vectors ξ^a at q,

$$(\rho+p)(\eta^a\xi_a)^2 - p \geq 0. \tag{2.5.3}$$

Assume first that $(\rho+p)\geq 0$ and $\rho \geq 0$, and let ξ^a be a unit timelike vector at q. Then, by the wrong-way Schwarz inequality (proposition 2.2.3), $(\eta^a\xi_a)^2 \geq \|\eta^a\|^2\,\|\xi^a\|^2 = 1$. Hence, $(\rho+p)(\eta^a\xi_a)^2 - p \geq (\rho+p) - p = \rho \geq 0$. So we have equation (2.5.3). Conversely, assume equation (2.5.3) holds for all unit timelike vectors ξ^a at q. Then, in particular, it holds if $\xi^a = \eta^a$, and in this case we have $0 \leq (\rho+p)(\eta^a\eta_a)^2 - p = (\rho+p) - p = \rho$. Note next that there is no upper bound to the value of $(\eta^a\xi_a)^2$ as ξ^a ranges over unit timelike vectors at q. (For example, let σ^a be any unit spacelike vector at q orthogonal to η^a, and let ξ^a be of the form $\xi^a = (\cosh\theta)\,\eta^a - (\sinh\theta)\,\sigma^a$, where θ is a real number. Then ξ^a is a unit timelike vector, and $(\eta^a\xi_a)^2 = \cosh^2\theta$. The latter goes to infinity, as θ does.) So equation (2.5.3) cannot possibly hold for all unit timelike vectors at q unless $(\rho+p)\geq 0$. This gives us the stated equivalence for the WEC.

(CC) If T_{ab} is given by equation (2.5.2), then a straightforward computation shows that the conservation condition ($\nabla_a\, T^{ab} = \mathbf{0}$) holds iff

$$\begin{aligned}\rho\,(\eta^a\nabla_a\,\eta^b) + \rho\,\eta^b\,\nabla_a\,\eta^a + \eta^b(\eta^a\nabla_a\,\rho) - (\nabla_a\,p)(g^{ab} - \eta^a\eta^b)\\ + p(\eta^a\nabla_a\,\eta^b) + p\,\eta^b\,\nabla_a\,\eta^a = \mathbf{0}.\end{aligned} \tag{2.5.4}$$

Assume that equation (2.5.4) *does* hold. Then contraction with η_b yields

$$\eta^a\,\nabla_a\,\rho + (\rho+p)(\nabla_a\,\eta^a) = 0. \tag{2.5.5}$$

(Here we use the fact that the unit timelike vector field η^b is orthogonal to its associated acceleration field $\eta^a\nabla_a\,\eta^b$ and to its associated projection field $h_{ab} = (g_{ab} - \eta_a\eta_b)$.) And if we multiply equation (2.5.5) by η^b and then subtract the result from (2.5.4), we arrive at

$$(\rho+p)\,\eta^a\,\nabla_a\,\eta^b - (g^{ab} - \eta^b\,\eta^a)\,\nabla_a\,p = \mathbf{0}. \tag{2.5.6}$$

Thus equation (2.5.4) holds only if equations (2.5.5) and (2.5.6) do. And the converse is immediate. So we have our stated equivalence for the conservation condition.

PROBLEM 2.5.3. *(i) Prove the stated equivalence for the DEC. (ii) Prove that, as restricted to perfect fluids, the SDEC is equivalent to the DEC.*

Now consider the physical interpretation of the two equations jointly equivalent to the CC. Equation (2.5.6) is the equation of motion for a perfect fluid. We can think of it as a relativistic version of Euler's equation. Equation (2.5.5) is an equation of continuity (or conservation) in the sense familiar from classical fluid mechanics. It is easiest to think about the special case of a dust field ($p = 0$). In this case, the equation of motion reduces to the geodesic equation: $\eta^b \nabla_b \eta^a = \mathbf{0}$. That makes sense. In the absence of pressure, particles in the fluid are free particles. And the conservation equation reduces to $\eta^b \nabla_b \rho + \rho\, (\nabla_b \eta^b) = 0$. The first term gives the instantaneous rate of change of the fluid's energy density, as determined by a co-moving observer. The term $\nabla_b \eta^b$ gives the instantaneous rate of change of its volume, per unit volume, as determined by that observer. (We shall justify this claim in section 2.8.) In a more familiar notation, the equation might be written $\dfrac{d\rho}{ds} + \dfrac{\rho}{V}\dfrac{dV}{ds} = 0$ or, equivalently, $\dfrac{d(\rho V)}{ds} = 0$. (Here we use s for elapsed proper time.) It asserts that (in the absence of pressure, as determined by a co-moving observer) the energy contained in an (infinitesimal) fluid blob remains constant, even as its volume changes.

In the general case, the situation is more complex because the pressure in the fluid contributes to its energy (as determined relative to particular observers), and hence to what might be called its "effective mass density." (If you compress a fluid blob, it gets heavier.) In this case, the WEC comes out as the requirement that $(\rho + p) \geq 0$ in addition to $\rho \geq 0$. The equation of motion can be expressed as

$$(\rho + p)\, \eta^b \nabla_b \eta^a = h^{ab} \nabla_b p, \tag{2.5.7}$$

where h^{ab} is the projection field $(g^{ab} - \eta^a \eta^b)$. This is an instance of the "second law of motion" (see equation (2.4.13)) as applied to an (infinitesimal) blob of fluid. On the left we have "effective mass density × acceleration." On the right, we have the force acting on the blob, as determined by a co-moving observer. We can think of it as minus the gradient of the pressure (as determined by a co-moving observer). (The minus sign comes in because of our sign conventions.) Again, this makes sense. If the pressure on the left side of the blob is greater than that on the right, it will accelerate to the right.

And in the general case we are now considering—where the pressure p need not vanish—the term $(p\, \nabla_b \eta^b)$ in the conservation equation is required because the energy of the blob is *not* constant when its volume changes as a result of the pressure. The equation governs the contribution made to its energy by pressure.

2.6. Electromagnetic Fields

In this section we briefly discuss electromagnetic fields. Though our principal interest here is in the energy-momentum field T_{ab} associated with them, we mention a few fundamental ideas of classical electromagnetic theory along the way.

Electromagnetic fields are represented by smooth, anti-symmetric fields F_{ab} (on the background spacetime (M, g_{ab})). If a particle with mass $m > 0$, charge q, and four-velocity field ξ^a is present, the force exerted by the field on the particle at a point is given by $q\, F^a_{\ b}\, \xi^b$. (This condition uniquely characterizes F_{ab}.) As noted at the end of section 2.4, if we use this expression for the force term in the relativistic version of "Newton's second law" equation (2.4.13), we arrive at the Lorentz law of motion:

$$q\, F^a_{\ b}\, \xi^b = m\, \xi^b\, \nabla_b\, \xi^a. \tag{2.6.1}$$

It describes the motion of a charged particle in an electromagnetic field (at least when the contribution of the particle's own charge to the field is negligible and may be ignored). Note again that the equation makes geometric sense. The acceleration vector on the right is orthogonal to ξ^a. But so is the force vector on the left since F_{ab} is anti-symmetric.

The fundamental field equations of electromagnetic theory ("Maxwell's equations") are given by

$$\nabla_{[a}\, F_{bc]} = \mathbf{0}, \tag{2.6.2}$$

$$\nabla_a\, F^{ab} = J^b. \tag{2.6.3}$$

Here J^a is the *charge-current density* field. It is characterized by the following condition: given any background observer at a point with four-velocity ξ^a, $J^a \xi_a$ is the charge density there (arising from whatever charged matter is present) as determined by that observer. For example, in the case of a charged dust field, $J^a = \mu\, \eta^a$, where η^a is the four-velocity of the dust and μ is its charge density as measured by a co-moving observer. Thus, if equation (2.6.1) expresses the action of the electromagnetic field on a charged (test) particle, equation (2.6.3) expresses the reciprocal action of charged matter on the field. The former acts as a source for the latter.

An important constraint on the charge-curent density field J^a follows immediately from equation (2.6.3). Since F^{ab} is anti-symmetric, $\nabla_a J^a = \nabla_a \nabla_n F^{na} = \nabla_{[a} \nabla_{n]} F^{na}$. But

$$\begin{aligned} 2\,\nabla_{[a}\nabla_{n]}\,F^{na} &= -F^{ma}\,R^n{}_{man} - F^{nm}\,R^a{}_{man} = -F^{ma}\,R_{ma} + F^{nm}\,R_{mn} \\ &= -F^{mn}\,R_{mn} + F^{mn}\,R_{nm} \qquad = 0. \end{aligned}$$

(The first two equalities follow, respectively, from clauses (1) and (2) of proposition 1.8.2. The third involves nothing more than a systematic change of abstract indices. The final equality follows from the symmetry of the Ricci tensor field.) So

$$\nabla_a J^a = 0. \tag{2.6.4}$$

We can understand this as an assertion of the local conservation of charge. Notice that in the case of charged dust field with $J^a = \mu\,\eta^a$, equation (2.6.4) comes out as

$$\eta^b\,\nabla_b\,\mu + \mu\,(\nabla_b\,\eta^b) = 0. \tag{2.6.5}$$

This has exactly the same form as equation (2.5.5) in the special case where $p = 0$, and it can be analyzed in exactly the same manner. It asserts that, as determined by a co-moving observer, the total charge in an (infinitesimal) blob of charged dust remains constant, even as its volume changes.

PROBLEM 2.6.1. *Show that Maxwell's equations in the source-free case ($J^a = \mathbf{0}$) are conformally invariant; i.e., if an anti-symmetric field F_{ab} satisfies them with respect to a metric g_{ab}, then it does so as well with respect to any metric of the form $g'_{ab} = \Omega^2\,g_{ab}$. (Note: Here we need the fact that the dimension n of the background spacetime is 4. Hint: The conformal invariance of the first Maxwell equation ($\nabla_{[a}\,F_{bc]} = \mathbf{0}$) follows immediately from problem 1.7.2 and does not depend on the value of n. To establish that of the second ($\nabla_a F^{ab} = \mathbf{0}$), use proposition 1.9.5 to show that*

$$\nabla'_a\,(g'^{am}\,g'^{bn}\,F_{mn}) = \frac{1}{\Omega^4}\,(\nabla_a F^{ab}) + \frac{(n-4)}{\Omega^5}\,F^{ab}\,\nabla_a\,\Omega,$$

where $g'^{ab} = \Omega^{-2}\,g^{ab}$ is the inverse of g'_{ab}, and ∇' is the derivative operator compatible with g'_{ab}.)

The energy-momentum tensor field associated with F_{ab} is given by

$$T_{ab} = F_{am}\,F^m{}_b + \frac{1}{4}\,g_{ab}\,(F_{mn}F^{mn}). \tag{2.6.6}$$

We can gain some insight by introducing a reference observer O at a point p, with four-velocity ξ^a, and considering the decomposition of F_{ab} there into its "electric" and "magnetic" components.

Let h_{ab} be the spatial projection tensor at the point determined by ξ^a (defined by equation (2.4.3)). Further, let ϵ_{abcd} be a volume element on some open set containing p. Then we define

$$\mu = J^a \xi_a, \tag{2.6.7}$$

$$j^a = h^a_{\ b} J^b, \tag{2.6.8}$$

$$E^a = F^a_{\ b} \xi^b, \tag{2.6.9}$$

$$B^a = \frac{1}{2} \epsilon^{abcd} \xi_b F_{cd}, \tag{2.6.10}$$

$$\epsilon_{abc} = \epsilon_{abcn} \xi^n. \tag{2.6.11}$$

E^a and B^a are, respectively, the *electric* and *magnetic field* vectors at the point as determined relative to O. (Clearly, if we had chosen the other volume element, $-\epsilon_{abcd}$, we would have ended up with $-B^a$. A choice of volume element is tantamount to a choice of "right-hand rule.") μ and j^a are, respectively, the charge density and current density vectors as determined relative to O. Note that E^a, B^a, and j^a are all orthogonal to ξ^a. We can think of ϵ_{abc} as a three-dimensional volume element defined on the orthogonal subspace of ξ^a (It is anti-symmetric, it is orthogonal to ξ^a in all indices and, as one can show using equation (1.11.8), it satisfies the normalization condition $\epsilon_{abc}\epsilon^{abc} = -3!$.)

Reversing direction, we can recover F_{ab} and J^a from E^a, B^a, μ, and j^a as follows:

$$J^a = \mu \xi^a + j^a, \tag{2.6.12}$$

$$F_{ab} = 2 E_{[a} \xi_{b]} + \epsilon_{abcd} \xi^c B^d. \tag{2.6.13}$$

The first assertion is an immediate consequence of the definitions of j^a and μ. To verify the second, we substitute for B^d on the right side. By equation (1.11.8), the anti-symmetry of F_{ab}, and the definition of E^a, we have

$$\begin{aligned} 2 E_{[a} \xi_{b]} + \epsilon_{abcd} \xi^c \left(\frac{1}{2} \epsilon^{dpqr} \xi_p F_{qr}\right) &= 2 E_{[a} \xi_{b]} + \frac{1}{2} (3!) \xi^c \delta^{[p}_{\ a} \delta^{q}_{\ b} \delta^{r]}_{\ c} \xi_p F_{qr} \\ &= 2 E_{[a} \xi_{b]} + 3 \xi^c \xi_{[a} F_{bc]} \\ &= 2E_{[a}\xi_{b]} + \xi^c(\xi_a F_{bc} + \xi_c F_{ab} - \xi_b F_{ac}) = F_{ab}. \end{aligned}$$

Let us now return to our expression (2.6.6) for the energy-momentum field T_{ab}. Our observer O with four-velocity ξ^a will attribute to the electromagnetic field a four-momentum density,

$$T^a{}_b\,\xi^b = F^{am}\,F_{mb}\,\xi^b + \frac{1}{4}\,\xi^a\,(F^{mn}F_{mn}). \tag{2.6.14}$$

We can express the right side in terms of the relative electric and magnetic vectors E^a and B^a determined by O. (The computations are much like that used to prove equation (2.6.13).) We have

$$\begin{aligned} F^{am}\,F_{mb}\,\xi^b &= F^{am}\,E_m = (2\,E^{[a}\,\xi^{m]} + \epsilon^{ampr}\,\xi_p\,B_r)\,E_m \\ &= -\,\xi^a\,E^m E_m - \epsilon^{amr}\,E_m\,B_r \end{aligned} \tag{2.6.15}$$

and also

$$\begin{aligned} F^{mn}F_{mn} &= (2\,E^{[m}\,\xi^{n]} + \epsilon^{mnpq}\,\xi_p\,B_q)\,(2\,E_{[m}\,\xi_{n]} + \epsilon_{mnrs}\,\xi^r\,B^s) \\ &= 2\,E^n E_n + \epsilon_{mnrs}\,\epsilon^{mnpq}\,\xi_p\,B_q\,\xi^r\,B^s \\ &= 2\,E^n E_n - 4\,\delta^{[p}{}_r\,\delta^{q]}{}_s\,\xi_p\,B_q\,\xi^r\,B^s = 2\,(E^n E_n - B^n B_n). \end{aligned} \tag{2.6.16}$$

Hence,

$$T^a{}_b\,\xi^b = \frac{1}{2}\,(-\,E^n E_n - B^n B_n)\,\xi^a - \epsilon^{amr}\,E_m\,B_r. \tag{2.6.17}$$

The coefficient of ξ^a on the right side is the energy density of the field as determined by O. Using our notation for vector norms and temporarily dropping indices (and remembering that both E^a and B^a are spacelike [or the zero vector]), we can express it as $\frac{1}{2}\,(\|E\|^2 + \|B\|^2)$. This will be familiar as the standard textbook expression for the energy density of an electromagnetic field. The component of $T^a{}_b\,\xi^b$ orthogonal to ξ^a, namely $-\epsilon^{amr}\,E_m\,B_r$, is the three-momentum density of the electromagnetic field as determined by O. In more familiar vector notation (recall our discussion in section 1.11), it comes out as $-(E \times B)$. ($E \times B$ is called the "Poynting vector.")

Note that we can also work backward and derive equation (2.6.6), our expression for T_{ab}, from the assumption that equation (2.6.17) holds for all observers with four-velocity ξ^a. (Reversing the calculation, one shows that equation (2.6.14) or, equivalently, $(T^a{}_b - (F^{am}\,F_{mb} + \frac{1}{4}\,g^a{}_b\,F^{mn}F_{mn}))\,\xi^b = \mathbf{0}$, holds for all unit timelike vectors ξ^a. Equation (2.6.6) then follows by proposition 2.1.3.) So T_{ab} is fully determined by the requirement that it code values for $\frac{1}{2}\,(\|E\|^2 + \|B\|^2)$ and $-(E \times B)$ for all observers.

PROBLEM 2.6.2. *Textbooks standardly assert that* $(\|E\|^2 - \|B\|^2)$ *and* $E \cdot B$ *are relativistically invariant (i.e., have common values for all observers). To verify this, it suffices to note that (in our notation)*

$$(- E^a E_a + B^a B_a) = -\frac{1}{2} F^{ab} F_{ab}, \tag{2.6.18}$$

$$E^a B_a = \frac{1}{8} \epsilon^{abcd} F_{ab} F_{cd}. \tag{2.6.19}$$

We have proved the first assertion (equation (2.6.16)). Prove the second.

Now we consider our two energy conditions and the conservation condition. Given any future-directed, unit timelike vector ξ^a at a point, with corresponding electric and magnetic field vectors E^a and B^a, we have

$$T_{ab}\, \xi^a\, \xi^b = \frac{1}{2} (- E^n E_n - B^n B_n), \tag{2.6.20}$$

$$(T_{ab}\xi^b)(T^{ac}\xi_c) = \frac{1}{4} (E^n E_n - B^n B_n)^2 + (E^n B_n)^2, \tag{2.6.21}$$

$$\nabla_a T^{ab} = J_a F^{ab}. \tag{2.6.22}$$

The first follows immediately from equation (2.6.17) (and the fact that ϵ_{abc} is orthogonal to ξ^a in all indices). We leave the second as an exercise. For the third, note that

$$\begin{aligned}
\nabla_a T^{ab} &= \nabla_a (F^{am}\, F_m{}^b + \frac{1}{4} g^{ab}\, F_{mn} F^{mn}) \\
&= F^{am}\, \nabla_a F_m{}^b + F_m{}^b\, \nabla_a F^{am} + \frac{1}{2} F_{mn} \nabla^b F^{mn} \\
&= \frac{1}{2} F_{am} (\nabla^a F^{mb} - \nabla^m F^{ab}) + F_m{}^b\, \nabla_a F^{am} + \frac{1}{2} F_{ma} \nabla^b F^{ma} \\
&= -\frac{1}{2} F_{am} (\nabla^a F^{bm} + \nabla^m F^{ab} + \nabla^b F^{ma}) + F_m{}^b\, J^m = J_m F^{mb}.
\end{aligned}$$

(We get the third equality by systematically changing indices and using the anti-symmetry of F_{ab}: $F_{am}\, \nabla^a F^{mb} = F_{ma}\, \nabla^m F^{ab} = -F_{am}\, \nabla^m F^{ab}$. We get the fourth and fifth from Maxwell's equations (first $\nabla_a F^{am} = J^m$, then $\nabla^{[a} F^{bm]} = \mathbf{0}$) and, again, the anti-symmetry of F_{ab}.)

PROBLEM 2.6.3. *Prove equation (2.6.21). (It follows immediately from this result that $T_{ab}\, \xi^b$ is null iff $E^a E_a = B^a B_a$ and $E^a B_a = 0$. By problem 2.6.2, these conditions hold as determined relative to one unit timelike vector ξ^a at a point iff they hold for all such vectors there. When they do hold (at all points), we say that F_{ab} is a "null" field.)*

Maxwell's equations play no role in the proof of equations (2.6.20) and (2.6.21). So we see that for *any* anti-symmetric field F_{ab}, the corresponding energy-momentum field $T_{ab} = F_{am}\, F^{m}{}_{b} + \frac{1}{4}\, g_{ab}\, (F_{mn} F^{mn})$ satisfies both the WEC and the DEC (since E^a and B^a are always spacelike or equal to the zero vector $\mathbf{0}$). And it satisfies the SDEC except in the special case where F_{ab} is a non-vanishing null electromagnetic field (in the sense of problem 2.6.3).

The situation is different with the CC, for which Maxwell's equations are essential. Suppose that the pair $(F_{ab},\, J^a)$ satisfies them (and, therefore, that equation (2.6.22) holds). There are two cases to consider. If $J^a = \mathbf{0}$—i.e., if no sources are present—then the conservation condition $\nabla_a T^{ab} = \mathbf{0}$ is automatically satisfied. But when charged matter is present, there is the possibility of energy-momentum being transferred from the electromagnetic field to that matter. So it should not be the energy-momentum of the electromagnetic field alone that is conserved. Instead, it should be the *total* energy-momentum present (arising from both field and charged matter) that is.

By way of example, consider the case where a charged dust field serves as a source for the electromagnetic field. Suppose the dust is characterized by four-velocity field η^a, mass density ρ, and charge density μ, the latter two as determined by a co-moving observer. Then we have $J^a = \mu\, \eta^a$, and the energy-momentum field for the dust (alone) is given by $\rho\, \eta^a \eta^b$. So the total energy-momentum field in this case is given by

(2.6.23) $$T_{ab} = F_{am}\, F^{m}{}_{b} + \frac{1}{4}\, g_{ab}\, (F_{mn} F^{mn}) + \rho\, \eta_a \eta_b.$$

Hence, by equation (2.6.22),

$$\nabla_a T^{ab} = J_a\, F^{ab} + \nabla_a (\rho\, \eta^a \eta^b)$$

(2.6.24) $$= \mu\, \eta_a\, F^{ab} + \rho\, (\eta^a \nabla_a\, \eta^b) + \rho\, \eta^b\, \nabla_a\, \eta^a + \eta^b (\eta^a \nabla_a\, \rho).$$

This is the counterpart to equation (2.5.4) that we considered in our discussion of perfect fluids. Arguing much as we did there, we can verify that in the present case we have the following equivalence. (Set the right-hand side to $\mathbf{0}$, contract with η_b, and then subtract the resultant equation from the original.)

$$\text{CC} \Longleftrightarrow \begin{cases} \mu\, F^{b}{}_{a}\, \eta^a & = \quad \rho\, (\eta^a\, \nabla_a \eta^b). \\ \eta^a\, \nabla_a\, \rho + \rho\, (\nabla_a\, \eta^a) & = \quad 0. \end{cases}$$

The second equation on the right side is just equation (2.5.5) in the case where $p = 0$. It asserts that, as determined by a co-moving observer, the energy in an (infinitesimal) blob of dust remains constant, even as the volume of the blob changes. (Note that it also has exactly the same form as equation (2.6.5), which makes a corresponding assertion about charge conservation.) The first equation on the right side is an equation of motion for the dust field. It has exactly the same form as equation (2.6.1). It asserts, in a sense, that individual particles in the dust field obey the Lorentz law of motion. Thus, the energy-momentum of the electromagnetic field F_{ab} fails to be conserved only to the extent it exerts a force on those particles and causes them to accelerate.

As an afterthought, now, we recover the standard textbook formulation of Maxwell's (four) equations from our formulation. To do so, we need a bit of structure in the background. Let us temporarily assume that (M, g_{ab}) is not just any (temporally oriented) spacetime, but one that admits a future-directed, unit timelike vector field ξ^a that is *constant* ($\nabla_a \xi^b = \mathbf{0}$). Let μ, j^a, E^a, B^a, and ϵ_{abc} be as defined above. Further, let D be the derivative operator induced on hypersurfaces orthogonal to ξ^a. (Recall our discussion in section 1.10.) Then we have the following equivalences.

$$\nabla_{[a} F_{bc]} = \mathbf{0} \iff \begin{cases} D_b B^b = 0 & (\nabla \cdot \mathbf{B} = 0) \\ \epsilon^{abc} D_b E_c = -\xi^b \nabla_b B^a & \left(\nabla \times \mathbf{E} = -\dfrac{\partial \mathbf{B}}{\partial t}\right) \end{cases}$$

$$\nabla_a F^{ab} = J^b \iff \begin{cases} D_b E^b = \mu & (\nabla \cdot \mathbf{E} = \mu) \\ \epsilon^{abc} D_b B_c = \xi^b \nabla_b E^a + j^a & \left(\nabla \times \mathbf{B} = \dfrac{\partial \mathbf{E}}{\partial t} + \mathbf{j}\right) \end{cases}$$

(In each case, we have indicated how the right-side equation is formulated in standard (three-dimensional) vector notation.) We prove the first equivalence and leave the second as an exercise. Note first that by equations (2.6.13) and (1.11.8), we have

$$\begin{aligned} \epsilon^{abcd} F_{cd} &= \epsilon^{abcd} (2 E_{[c}\, \xi_{d]} + \epsilon_{cdrs}\, \xi^r B^s) \\ &= 2\, \epsilon^{abcd} E_c\, \xi_d - 4\, \delta^{[a}{}_r\, \delta^{b]}{}_s\, \xi^r B^s \\ &= 2\, \epsilon^{abcd} E_c\, \xi_d - 2\, \xi^a B^b + 2\, \xi^b B^a. \end{aligned}$$

Hence, since ξ^a is constant,

$$\epsilon^{abcd}\, \nabla_b F_{cd} = \nabla_b(\epsilon^{abcd} F_{cd}) = 2\, \epsilon^{abcd} \xi_d \nabla_b E_c - 2\, \xi^a \nabla_b B^b + 2\, \xi^b \nabla_b B^a. \tag{2.6.25}$$

And for that same reason, $\nabla_a h_{bc} = \nabla_a (g_{bc} - \xi_b \xi_c) = \mathbf{0}$. So, since $h^a{}_b E^b = E^a$ and $h^a{}_b B^b = B^a$,

$$D_b B^b = h^m{}_b\, h^b{}_n \nabla_m B^n = \nabla_m (h^m{}_b\, h^b{}_n\, B^n) = \nabla_m B^m = \nabla_b B^b \tag{2.6.26}$$

$$\begin{aligned} \epsilon^{abc} D_b E_c &= \epsilon^{abcd}\, \xi_d\, h^m{}_b\, h^n{}_c \nabla_m E_n = \epsilon^{abcd}\, \xi_d\, g^m{}_b \nabla_m (h^n{}_c E_n) \\ &= \epsilon^{abcd} \xi_d \nabla_b E_c. \end{aligned} \tag{2.6.27}$$

(For the second equality in equation (2.6.27), note that $\epsilon^{dabc}\,\xi_d\,\xi_b = \mathbf{0}$ and, hence, that $\epsilon^{dabc}\,\xi_d\,h^m{}_b = \epsilon^{dabc}\,\xi_d\,g^m{}_b$.) If we now replace $\nabla_b B^b$ and $\epsilon^{abcd}\xi_d\nabla_b E_c$ in equation (2.6.25) using equations (2.6.26) and (2.6.27), we arrive at

$$\epsilon^{abcd}\,\nabla_b F_{cd} = -2\,\xi^a (D_b B^b) + 2\,(\epsilon^{abc} D_b E_c + \xi^b \nabla_b B^a). \tag{2.6.28}$$

Now $\nabla_{[a}\,F_{bc]} = \mathbf{0}$ iff $\epsilon^{abcd}\,\nabla_b F_{cd} = \mathbf{0}$. (Why?) And the latter condition holds iff the sum on the right side of equation (2.6.28) is $\mathbf{0}$. But that sum consists of two terms, one tangent to ξ^a and one orthogonal to ξ^a. So the sum is $\mathbf{0}$ iff both terms are $\mathbf{0}$. Thus we are left with the conclusion that $\nabla_{[a}\,F_{bc]} = \mathbf{0}$ iff $D_b\,B^b = \mathbf{0}$ and $\epsilon^{abc} D_b E_c + \xi^b \nabla_b B^a = \mathbf{0}$.

PROBLEM 2.6.4. *Prove the second equivalence (for $\nabla_a\,F^{ab} = J^b$).*

2.7. Einstein's Equation

Once again, let $(M,\, g_{ab})$ be our background relativistic spacetime with a specified temporal orientation.

It is one of the fundamental ideas of relativity theory that spacetime structure is not a fixed backdrop against which the processes of physics unfold, but instead participates in that unfolding. It posits a dynamical interaction between the spacetime metric in any region and the matter fields there. The interaction is governed by *Einstein's field equation*

$$R_{ab} - \frac{1}{2}\,R\,g_{ab} = 8\,\pi\;T_{ab}, \tag{2.7.1}$$

or, equivalently,

$$R_{ab} = 8\,\pi\,(T_{ab} - \frac{1}{2}\,T\,g_{ab}). \tag{2.7.2}$$

Here $R_{ab}\,(= R^n{}_{abn})$ is the *Ricci tensor field*, $R\,(= R^a{}_a)$ is the *Riemann scalar curvature field*, and T is the contracted field $T^a{}_a$.[17] We start with four remarks about equation (2.7.1) and then consider two reformulations that provide a certain insight into the geometric significance of the equation.

(1) It is sometimes taken to be a version of "Mach's principle" that "the spacetime metric is uniquely determined by the distribution of matter." And it is sometimes proposed that the principle can be captured in the requirement that "if one first specifies the energy-momentum distribution T_{ab} on the spacetime manifold M, then there is exactly one (or at most one) Lorentzian

17. We use "geometrical units" in which the gravitational constant G and the speed of light c are 1.

metric g_{ab} on M that, together with T_{ab}, satisfies equation (2.7.1)." But there is a serious problem with the proposal. In general, one *cannot* specify the energy-momentum distribution in the absence of a spacetime metric. Indeed, in typical cases the metric enters explicitly in the expression for T_{ab}. (Recall the expression (2.5.2) for a perfect fluid.) Thus, in looking for solutions to equation (2.7.1), one must, in general, solve simultaneously for the metric *and* matter field distribution.

(2) Given any smooth metric g_{ab} on M, there certainly exists a smooth symmetric field T_{ab} on M that, together with g_{ab}, is a solution to equation (2.7.1). It suffices to define T_{ab} by the left side of the equation. But the field T_{ab} so introduced will not, in general, be the energy-momentum field associated with any known matter field. And it will not, in general, satisfy the weak energy condition discussed in section 2.5. If the latter condition is imposed as a constraint on T_{ab}, Einstein's equation is an entirely non-trivial restriction on spacetime structure.

Discussions of spacetime structure in classical relativity theory proceed on three levels according to the stringency of the constraints imposed on T_{ab}. At the first level, one considers only "exact solutions"—i.e., solutions where T_{ab} is, in fact, the aggregate energy-momentum field associated with one or more known matter fields. So, for example, one might undertake to find all perfect fluid solutions exhibiting particular symmetries. At the second level, one considers the larger class of what might be called "generic solutions"—i.e., solutions where T_{ab} satisfies one or more generic constraints (of which the weak and dominant energy conditions are examples). It is at this level, for example, that the singularity theorems of Penrose and Hawking (Hawking and Ellis [30]) are proved. Finally, at the third level, one drops all restrictions on T_{ab}, and Einstein's equation plays no role. Many results about global structure are proved at this level—e.g., the assertion that closed timelike curves exist in any relativistic spacetime (M, g_{ab}) where M is compact.

(3) We have presented Einstein's equation in its original form. He famously added a "cosmological constant" term $(-\Lambda g_{ab})$ in 1917 to allow for the possibility of a static cosmological model with a perfect fluid source, with $p = 0$ and $\rho > 0$.[18] (We shall see why the addition is necessary under those conditions at the end of section 2.11.) But Einstein was never happy with the revised equation

18. He did so for other reasons as well (see Earman [14]), but we pass over them here.

$$R_{ab} - \frac{1}{2}\,R\,g_{ab} - \Lambda\,g_{ab} = 8\,\pi\;T_{ab}, \tag{2.7.3}$$

or, equivalently,

$$R_{ab} = 8\,\pi\;\left(T_{ab} - \frac{1}{2}\,T\,g_{ab}\right) - \Lambda\,g_{ab}, \tag{2.7.4}$$

and was quick to revert to the original version after Hubble's redshift observations gave convincing evidence that the universe is, in fact, expanding. After that, he thought, there was no *need* to have a static cosmological model. (That the theory suggested the possibility of cosmic expansion *before* Hubble's observations must count as one of its great successes.) Since then the constant has often been reintroduced to help resolve discrepancies between theoretical prediction and observation, and then abandoned when the (apparent) discrepancies were resolved. (See Earman [14] for a masterful review of the history.) The story continues. Recent observations indicating an accelerating rate of cosmic expansion seem to imply that our universe is characterized by a positive value for Λ or something that mimics its effect.

In what follows, we shall continue to write Einstein's equation in the form (2.7.1) and think of the cosmological term as absorbed into the expression for the energy-momentum field T_{ab}. The magnitude and physical interpretation of this contribution to T_{ab} are topics of great importance in current physics.[19] But they will play no role in our discussion.

PROBLEM 2.7.1. *Equations (2.7.3) and (2.7.4) are equivalent only if the dimension n of the background manifold is 4. Show that in the general case (at least if $n \geq 3$), inversion of equation (2.7.3) leads to*

$$R_{ab} = 8\,\pi\;\left(T_{ab} - \frac{1}{(n-2)}\,T\,g_{ab}\right) - \frac{2}{(n-2)}\,\Lambda\,g_{ab}. \tag{2.7.5}$$

(4) It is instructive to consider the relation of Einstein's equation to Poisson's equation,

$$\nabla^2\phi = 4\,\pi\;\rho, \tag{2.7.6}$$

the field equation of Newtonian gravitation theory. Here ϕ is the Newtonian gravitational potential, and ρ is the Newtonian mass density function. In the geometrized formulation of the theory that we shall consider in chapter 4,

19. See Earman [14], once again, and references cited there.

one trades in the potential ϕ in favor of a curved derivative operator and one recovers ρ from a mass-momentum field T^{ab}. In the end, Poisson's equation comes out as

$$R_{ab} = 8\pi\,(\hat{T}_{ab} - \frac{1}{2}\,t_{ab}\,\hat{T}). \tag{2.7.7}$$

Here R_{ab} is the Ricci tensor field associated with the new curved derivative operator, t_{ab} is the temporal metric, $\hat{T}_{ab} = T^{mn}\,t_{ma}\,t_{nb}$, and $\hat{T} = T^{mn}\,t_{mn}$. (See equation (4.2.10) and the discussion that precedes it.) The resemblance to equation (2.7.2) is, of course, striking. It is particularly close in the special case where $\rho = 0$. For in this case, $T^{ab} = \mathbf{0}$ and equation (2.7.7) reduces to $R_{ab} = \mathbf{0}$. The latter is *exactly* the same as Einstein's equation (2.7.2) in the empty space case.

The geometrized formulation of Newtonian gravitation was discovered after general relativity in the 1920s. But now, after the fact, we can put ourselves in the position of a hypothetical investigator who is considering possible candidates for a relativistic field equation and who knows about the geometrized formulation of Newtonian theory. What could be more natural than to adapt equation (2.7.7) and simply replace t_{ab} with g_{ab}? This seems to me one of the nicest routes to Einstein's equation (2.7.2). Again, the route is particularly direct in the empty space case. For then one starts with the Newtonian empty space equation ($R_{ab} = \mathbf{0}$) and simply leaves it intact.

Let us now put aside the question of how one might try to motivate Einstein's equation, and consider two reformulations.

Let ξ^a be a unit timelike vector at a point p in M, and let S be a spacelike hypersurface containing p that is orthogonal to ξ^a there. (We understand a hypersurface in M to be *spacelike* if, at every point, vectors tangent to the surface are spacelike. This condition guarantees that the hypersurface is metric. (Recall our discussion in section 1.10.)) Further, let h_{ab} and π_{ab} be the first and second fundamental forms on S, and let D be the derivative operator on S determined by h_{ab}. Associated with D is a Riemann curvature field $\mathcal{R}^a{}_{bcd}$ on S. We know (recall equation (1.10.21)) that the contracted scalar field $\mathcal{R} = \mathcal{R}^a{}_{bca} h^{bc}$ satisfies

$$\mathcal{R} = \pi^2 - \pi_{ab}\,\pi^{ab} + R - 2\,R_{nr}\,\xi^n\,\xi^r \tag{2.7.8}$$

at p. In the special case where S has vanishing extrinsic curvature ($\pi_{ab} = \mathbf{0}$) at p, this can be expressed as

$$(R_{ab} - \frac{1}{2}\,g_{ab} R)\,\xi^a \xi^b = -\frac{1}{2}\,\mathcal{R}. \tag{2.7.9}$$

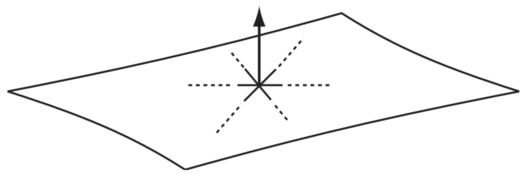

Figure 2.7.1. A "geodesic generated hypersurface" through a point is constructed by projecting geodesics in all directions orthogonal to a given timelike vector there.

If Einstein's equation holds, it therefore follows that

$$\mathcal{R} = -16\pi\ (T_{ab}\,\xi^a\xi^b).^{20} \tag{2.7.10}$$

One can also work backward. Suppose equation (2.7.10) holds for all unit timelike vectors at p and all orthogonal spacelike hypersurfaces through p with vanishing extrinsic curvature there. Then, by equation (2.7.9), it must be the case that

$$(R_{ab} - \frac{1}{2}\,R\,g_{ab})\,\xi^a\xi^b = 8\,\pi\ T_{ab}\ \xi^a\xi^b \tag{2.7.11}$$

for all unit timelike vectors ξ^a at p. This, in turn, implies Einstein's equation (by lemma 2.1.3). So we have the following equivalence.

($\star$) Einstein's equation $R_{ab} - \frac{1}{2}\,R\,g_{ab} = 8\,\pi\ T_{ab}$ holds at p iff for all unit timelike vectors ξ^a at p, and all orthogonal spacelike hypersurfaces S through p with vanishing extrinsic curvature there, the scalar curvature of S at p is given by $\mathcal{R} = -16\pi\ (T_{ab}\,\xi^a\xi^b)$.

We can give the result a somewhat more concrete formulation by casting it in terms of a particular class of spacelike hypersurfaces. Consider the set of all geodesics through p that are orthogonal to ξ^a there. The (images of these) curves, at least when restricted to a sufficiently small open set containing p, sweep out a smooth spacelike hypersurface that is orthogonal to ξ^a at p.[21] (See figure 2.7.1.) We shall call it a *geodesic generated hypersurface*. (We cannot speak of *the* geodesic generated hypersurface through p orthogonal to ξ^a because we

20. There is an issue here of sign convention that is potentially confusing. We seem to be led to the conclusion that the Riemann scalar curvature of S is less than or equal to 0—at least, if T_{ab} satisfies the weak energy condition. But it might be more natural to say that it is *greater* than or equal to 0. We are working here with $\mathcal{R}$ as determined relative to the negative definite metric h_{ab}, and a sign flip is introduced if we work, instead, with the positive definite metric $-h_{ab}$. The switch from h_{ab} to $-h_{ab}$ leaves D, $\mathcal{R}^a{}_{bcd}$, and $\mathcal{R}_{cd}$ intact but reverses the sign of $\mathcal{R} = h^{bc}\,\mathcal{R}_{bc}$.

21. More precisely, let S_p be the spacelike hyperplane in M_p orthogonal to ξ^a. Then for any sufficiently small open set O in M_p containing p, the image of $(S_p \cap O)$ under the exponential map $exp : O \to M$ is a smooth spacelike hypersurface in M containing p that is orthogonal to ξ^a there. (See, for example, Hawking and Ellis [30, p. 33].)

have left open how far the generating geodesics are extended. But given any two, their restrictions to a suitably small open set containing p coincide.)

Geodesic generated hypersurfaces are of interest in their own right, the present context aside, because they are natural candidates for a notion of "local simultaneity slice" (as determined relative to a timelike vector at a point). We can think of them as instances of *private space*. (The contrast here is with *public space*, which is determined not relative to a single timelike vector or timelike curve, but relative to a congruence of timelike curves. For more on this difference between private space and public space, see Rindler [53, 54] and Page [49].)

Now suppose S is a geodesic generated hypersurface generated from p. We claim that it has vanishing extrinsic curvature there. We can verify this with a simple calculation very much like that used to prove proposition 1.10.7. Let ξ^a be a smooth, future-directed, unit timelike field, defined on some open subset of S containing p, that is orthogonal to S. Let h_{ab} be the corresponding projection field on S. Further, let σ^a be the tangent field to a geodesic (relative to ∇) through p that is orthogonal there to ξ^a. Then along the image of the geodesic we have $\sigma^a \nabla_a \sigma^b = \mathbf{0}$ and $\sigma^a \xi_a = 0$ (or, equivalently, $h^a{}_b \sigma^b = \sigma^a$). The latter holds because the image of the geodesic is contained in S and so is everywhere orthogonal to its normal field. Hence, by equation (1.10.16), we have

$$\begin{aligned}\pi_{ab}\, \sigma^a \sigma^b &= (h^m{}_a\, h^n{}_b \nabla_m\, \xi_n)\, \sigma^a \sigma^b = \sigma^m \sigma^n \nabla_m\, \xi_n \\ &= \sigma^m \nabla_m (\sigma^n \xi_n) - \xi_n\, \sigma^m \nabla_m \sigma^n = 0\end{aligned}$$

along the image of the geodesic. In particular, the condition holds at p. But given any vector at p orthogonal to ξ^a, we can choose our initial geodesic so that it has that vector for its tangent at p. Hence, $\pi_{ab}\, \sigma^a \sigma^b = 0$ at p for *all* such orthogonal vectors. Since π_{ab} is symmetric, as well as orthogonal to the normal field ξ^a, it follows that $\pi_{ab} = \mathbf{0}$ at p.

Consider again the equivalence ($\star$). If we rerun the argument used before, but systematically cast it in terms of geodesic generated hypersurfaces, we arrive at the following alternate formulation.

PROPOSITION 2.7.1. *Let T_{ab} be a smooth symmetric field on M, and let p be a point in M. Then Einstein's equation $R_{ab} - \frac{1}{2}\, R\, g_{ab} = 8\,\pi\, T_{ab}$ holds at p iff for all unit timelike vectors ξ^a at p, and all geodesic hypersurfaces S generated from p that are orthogonal to ξ^a, the scalar curvature of S at p is given by $\mathcal{R} = -16\pi\, (T_{ab}\, \xi^a \xi^b)$.*

Our second reformulation of Einstein's equation is phrased in terms of geodesic deviation. Let ξ^a be a smooth, future-directed, unit timelike vector field whose associated integral curves are geodesics—i.e., a *geodesic reference frame*. Further, let λ^a be a vector field on one of the integral curves γ satisfying $\mathcal{L}_\xi \lambda^a = \mathbf{0}$. (So $\xi^b \nabla_b \lambda_a = \lambda^b \nabla_b \xi_a$.) Finally, assume λ^a is orthogonal to ξ^a at some point on γ. Then it must be orthogonal to the latter at all points on γ. This follows because the inner product $(\xi^a \lambda_a)$ is constant on γ:

$$\begin{aligned}\xi^b \nabla_b(\xi^a \lambda_a) &= \lambda_a\, \xi^b \nabla_b \xi^a + \xi^a \xi^b \nabla_b \lambda_a = \xi^a \xi^b \nabla_b \lambda_a = \xi^a \lambda^b \nabla_b \xi_a \\ &= \frac{1}{2}\, \lambda^b \nabla_b (\xi^a \xi_a) = \frac{1}{2}\, \lambda^b \nabla_b (1) = 0.\end{aligned}$$

We can think of λ^a as a *connecting field* that joins the image of γ to the image of another, "infinitesimally close," integral curve of ξ^a. Then the field $\xi^n \nabla_n (\xi^m \nabla_m \lambda^a)$ represents the acceleration of the latter relative to γ. We know from proposition 1.8.5 that it satisfies the "equation of geodesic deviation":

$$\xi^n \nabla_n\, (\xi^m \nabla_m \lambda^a) = R^a{}_{bcd}\, \xi^b\, \lambda^c\, \xi^d. \tag{2.7.12}$$

Now we define the "average radial acceleration" of ξ^a at a point p on γ. Let $\overset{i}{\lambda}{}^a$ $(i = 1, 2, 3)$ be any three connecting fields (as just described) such that, at p, the vectors ξ^a, $\overset{1}{\lambda}{}^a$, $\overset{2}{\lambda}{}^a$, $\overset{3}{\lambda}{}^a$ form an orthonormal set. For each i, the (outward-directed) radial component of the relative acceleration vector $\xi^n \nabla_n\, (\xi^m \nabla_m\, \overset{i}{\lambda}{}^a)$ —i.e., its component in the direction $\overset{i}{\lambda}$—has magnitude

$$-\, \overset{i}{\lambda}_a\, \xi^n \nabla_n\, (\xi^m \nabla_m\, \overset{i}{\lambda}{}^a).$$

(We need the minus sign because λ^a is spacelike.) We now take the *average radial acceleration* (ARA) of ξ^a at p to be

$$ARA = -\frac{1}{3} \sum_{i=1}^{3} \overset{i}{\lambda}_a\;\, \xi^n \nabla_n\, (\xi^m \nabla_m\, \overset{i}{\lambda}{}^a). \tag{2.7.13}$$

Of course, we need to check that the sum on the right side is independent of our initial choice of connecting fields. The orthonormality condition implies that at p we have $g_{ac} = \xi_a \xi_c - \sum_{i=1}^{3} \overset{i}{\lambda}_a\, \overset{i}{\lambda}_c$. Hence, by equation (2.7.12), we also have

$$\begin{aligned}ARA &= -\,\frac{1}{3} \sum_{i=1}^{3} \overset{i}{\lambda}_a\; R^a{}_{bcd}\, \xi^b\, \overset{i}{\lambda}{}^c\, \xi^d = -\frac{1}{3}\, R^a{}_{bcd}\, \xi^b\, \xi^d \left(\sum_{i=1}^{3} \overset{i}{\lambda}_a\, \overset{i}{\lambda}{}^c \right) \\ &= -\,\frac{1}{3}\, R^a{}_{bcd}\, \xi^b\, \xi^d\, (\xi_a \xi^c - g_a{}^c)\end{aligned}$$

at p. But $R^a{}_{bcd}\,\xi^c\xi^d = \mathbf{0}$, and $R^a{}_{bcd}\,g_a{}^c = R^a{}_{bad} = -R^a{}_{bda} = -R_{bd}$. So we may conclude that

$$ARA = -\frac{1}{3}\,R_{bd}\,\xi^b\xi^d \tag{2.7.14}$$

holds at p. Thus, as claimed, ARA is well defined.

Now if Einstein's equation holds at p, it follows that

$$ARA = -\frac{8\pi}{3}\left(T_{ab} - \frac{1}{2}\,T\,g_{ab}\right)\xi^a\xi^b \tag{2.7.15}$$

holds there as well. And conversely, if equation (2.7.15) holds at p for all geodesic reference frames, then it must be the case, by equation (2.7.14), that $R_{bd}\,\xi^b\xi^d = 8\pi\,(T_{ab} - \frac{1}{2}\,T\,g_{ab})\xi^b\xi^d$ holds for all unit timelike vectors ξ^a there. And this, in turn, implies that Einstein's equation holds at p. So we have the following equivalence.

PROPOSITION 2.7.2. *Let T_{ab} be a smooth symmetric field on M, and let p be a point in M. Then Einstein's equation $R_{ab} - \frac{1}{2}\,R\,g_{ab} = 8\pi\,T_{ab}$ holds at p iff for all geodesic reference frames ξ^a (defined on some open set containing p), the average radial acceleration of ξ^a at p is given by $ARA = -\frac{8\pi}{3}\left(T_{ab} - \frac{1}{2}\,T\,g_{ab}\right)\xi^a\xi^b$.*

We considered three energy conditions (weak, dominant, and strengthened dominant) in section 2.5. Let us now consider a fourth. Let T_{ab} be the energy-momentum field associated with a matter field $\mathcal{F}$.

Strong Energy Condition (SEC): Given any timelike vector ξ^a at any point in M,

$$\left(T_{ab} - \frac{1}{2}\,Tg_{ab}\right)\xi^a\xi^b \geq 0.$$

Equation (2.7.15) provides an interpretation. Suppose that Einstein's equation holds. Then $\mathcal{F}$ satisfies the strong energy condition iff, for all geodesic reference frames, the average (outward-directed) radial acceleration of the frame is negative or 0. This captures the claim, in a sense, that the "gravitational field" generated by $\mathcal{F}$ is "attractive".

PROBLEM 2.7.2. *Give examples of each of the following.*

(1) *A smooth symmetric field T_{ab} that satisfies the SDEC (and so also the WEC and DEC) but not the SEC*

(2) *A smooth symmetric field T_{ab} that satisfies the SEC but not the WEC (and so not the DEC or SDEC, either)*

PROBLEM 2.7.3. *Consider a perfect fluid with four-velocity η^a, energy density ρ, and pressure p. Show that it satisfies the strong energy condition iff $(\rho + p) \geq 0$ and $(\rho + 3p) \geq 0$.*

2.8. Fluid Flow

In this section, we consider fluid flow and develop the standard formalism for representing the rotation and expansion of a fluid at a point. (Later, in sections 3.2 and 3.3, we shall consider several different notions of global rotation.)

Once again, let (M, g_{ab}) be our background relativistic spacetime. We are assuming it is temporally orientable and endowed with a particular temporal orientation. Let ξ^a be a smooth, future-directed unit timelike vector field on M (or some open subset of M). We understand it to represent the four-velocity field of a fluid. Further, let h_{ab} be the spatial projection field determined by ξ^a.

The *rotation* and *expansion* fields associated with ξ^a are defined as follows:

$$\omega_{ab} = h_{[a}{}^{m}\, h_{b]}{}^{n}\, \nabla_m\, \xi_n, \tag{2.8.1}$$

$$\theta_{ab} = h_{(a}{}^{m}\, h_{b)}{}^{n}\, \nabla_m\, \xi_n. \tag{2.8.2}$$

They are smooth fields, orthogonal to ξ^a in both indices, and satisfy

$$\nabla_a\, \xi_b = \omega_{ab} + \theta_{ab} + \xi_a (\xi^m \nabla_m\, \xi_b). \tag{2.8.3}$$

(This follows since

$$\omega_{ab} + \theta_{ab} = h_a{}^m\, h_b{}^n\, \nabla_m\, \xi_n = (g_a{}^m - \xi_a\, \xi^m)\, (g_b{}^n - \xi_b\, \xi^n)\, \nabla_m\, \xi_n,$$

and $\xi^n\, \nabla_m\, \xi_n = \mathbf{0}$.) Our first task is to give the two fields a geometric interpretation and, in so doing, justify our terminology. We start with the rotation field ω_{ab}.

Let γ be an integral curve of ξ^a, and let p be a point on the image of γ. Further, let η^a be a vector field on the image of γ that is "carried along by the flow of ξ^a" (i.e., $\mathcal{L}_\xi\, \eta^a = \mathbf{0}$) and orthogonal to ξ^a at p. (It need not be orthogonal to ξ^a elsewhere.) We think of the image of γ as the worldline of a fluid element O, and think of η^a at p as a "connecting vector" that spans the distance between O and a neighboring fluid element N that is "infinitesimally close." The instantaneous velocity of N relative to O at p is given by $\xi^a\, \nabla_a\, \eta^b$. But $\xi^a\, \nabla_a\, \eta^b = \eta^a\, \nabla_a\, \xi^b$ (since $\mathcal{L}_\xi\, \eta^a = \mathbf{0}$). So, by equation (2.8.3) and the orthogonality of ξ^a with η^a at p, we have

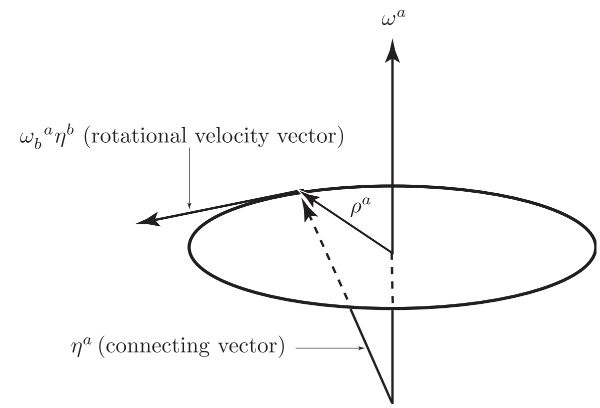

Figure 2.8.1. The angular velocity (or twist) vector ω^a. It points in the direction of the instantaneous axis of rotation of the fluid. Its magnitude $\|\omega^a\|$ is the instantaneous angular speed of the fluid about that axis. Here η^a connects the fluid element O to the "infinitesimally close" fluid element N. The rotational velocity of N relative to O is given by $\omega_b{}^a \eta^b$. The latter is orthogonal to η^a.

$$\xi^a \nabla_a \eta^b = (\omega_a{}^b + \theta_a{}^b)\, \eta^a \tag{2.8.4}$$

at the point. Here we have simply decomposed the relative velocity vector into two components. The first, $(\omega_a{}^b \eta^a)$, is orthogonal to η^a since ω_{ab} is anti-symmetric. (See figure 2.8.1.) It is naturally understood as the instantaneous *rotational velocity* of N with respect to O at p.

In support of this interpretation, consider the instantaneous rate of change of the squared length $(-\eta^b \eta_b)$ of η^a at p. It follows from equation (2.8.4) that

$$\xi^a \nabla_a (-\eta^b \eta_b) = -2\, \theta_{ab}\, \eta^a \eta^b. \tag{2.8.5}$$

Thus the rate of change depends solely on θ_{ab}. Suppose $\theta_{ab} = \mathbf{0}$. Then the instantaneous velocity of N with respect to O at p has a vanishing radial component. If $\omega_{ab} \neq \mathbf{0}$, N can still have non-zero velocity there with respect to O. But it can only be a rotational velocity. The two conditions ($\theta_{ab} = \mathbf{0}$ and $\omega_{ab} \neq \mathbf{0}$) jointly characterize "rigid rotation."

The rotation tensor ω_{ab} at a point p determines both an (instantaneous) axis of rotation there, and an (instantaneous) speed of rotation. As we shall see, both pieces of information are built into the *angular velocity* (or *twist*) vector

$$\omega^a = \frac{1}{2}\, \epsilon^{abcd}\, \xi_b\, \omega_{cd} \tag{2.8.6}$$

at p. (Here ϵ^{abcd} is a volume element defined on some open set containing p. Clearly, if we switched from the volume element ϵ_{abcd} to its negation, the result would be to replace ω^a with $-\omega^a$.)

If follows from equation (2.8.6) (and the anti-symmetry of ϵ_{abcd}) that ω^a is orthogonal to ξ^a. It further follows that

$$\omega^a = \frac{1}{2}\,\epsilon^{abcd}\,\xi_b\,\nabla_c\,\xi_d, \tag{2.8.7}$$

$$\omega_{ab} = \epsilon_{abcd}\,\xi^c\omega^d. \tag{2.8.8}$$

Hence, $\omega_{ab} = \mathbf{0}$ iff $\omega^a = \mathbf{0}$. Both equations (2.8.7) and (2.8.8) are verified with simple calculations. We do the first and leave the second as an exercise. For the first, we have

$$\begin{aligned}2\,\omega^a &= \epsilon^{abcd}\,\xi_b\,\omega_{cd} = \epsilon^{abcd}\,\xi_b\,h_{[c}{}^r\,h_{d]}{}^s\,\nabla_r\,\xi_s = \epsilon^{abcd}\,\xi_b\,h_c{}^r\,h_d^s\,\nabla_r\,\xi_s \\ &= \epsilon^{abcd}\,\xi_b\,g_c{}^r\,g_d^s\,\nabla_r\,\xi_s = \epsilon^{abcd}\,\xi_b\,\nabla_c\,\xi_d.\end{aligned}$$

(The second equality follows from the anti-symmetry of ϵ^{abcd}, and the third from the fact that $\epsilon^{abcd}\xi_b$ is orthogonal to ξ^a in all indices.) Notice that equation (2.8.6) has exactly the same form as our definition (2.6.10) of the magnetic field vector B^a determined relative to a Maxwell field F_{ab} and four-velocity vector ξ^a ($B^a = \frac{1}{2}\,\epsilon^{abcd}\,\xi_b\,F_{cd}$). It is for this reason that the magnetic field is sometimes described as the "rotational component of the electromagnetic field."

PROBLEM 2.8.1. *Prove equation (2.8.8).*

PROBLEM 2.8.2. *We have seen that the conditions (i) $\omega_{ab} = \mathbf{0}$ and (ii) $\omega^a = \mathbf{0}$ are equivalent at any point. Show that they are also equivalent (at any point) with (iii) $\xi_{[a}\nabla_b\,\xi_{c]} = \mathbf{0}$.*

We claim now that ω^a *points in the direction of the instantaneous axis of rotation (of the fluid flow associated with ξ^a)*. (See figure 2.8.1 again.) More precisely, with the connecting field η^a as above, we show that, at p,

$$\omega_b{}^a\eta^b = \mathbf{0} \Longleftrightarrow \eta^a \text{ is proportional to } \omega^a. \tag{2.8.9}$$

(Or, in the language of "infinitesimally close" fluid elements, the rotational velocity of N with respect to O vanishes iff the connecting vector from O to N is aligned with ω^a.) The implication from right to left follows immediately from equation (2.8.8) (and the anti-symmetry of ϵ_{abcd}). Conversely, suppose $\omega_b{}^a\eta^b = \mathbf{0}$. Then, by equation (2.8.8),

$$\begin{aligned}\mathbf{0} &= (\xi_n\,\omega_p\,\epsilon^{amnp})\,\omega_{ba}\,\eta^b = \xi_n\,\omega_p\,\epsilon^{amnp}\,\epsilon_{bacd}\,\xi^c\,\omega^d\,\eta^b \\ &= 3!\,\delta^{[m}{}_b\,\delta^n{}_c\,\delta^{p]}{}_d\,\eta^b\,\xi^c\,\omega^d\,\xi_n\,\omega_p = 3!\,\eta^{[m}\,\xi^n\,\omega^{p]}\,\xi_n\,\omega_p \\ &= (\eta^m\,\omega^p\,\omega_p - \omega^m\,\eta^p\,\omega_p).\end{aligned}$$

(For the final equality, here we use the fact that ξ^a is orthogonal to η^a at p and orthogonal to ω^a everywhere.) Now if $\omega^p\omega_p = \mathbf{0}$, then $\omega^a = \mathbf{0}$. (The twist vector ω^a is orthogonal to ξ^a and, by proposition 2.2.1, the only null vector orthogonal to a timelike vector is the zero vector.) And in this case, η^a is trivially aligned with ω^a. So we may assume that $\omega^p\omega_p \neq \mathbf{0}$. It then follows that $\eta^a = k\,\omega^a$, where $k = (\omega^p\eta_p)/(\omega^n\omega_n)$.

Next, we claim that *the magnitude of ω^a is the instantaneous angular speed (of the fluid flow associated with ξ^a)*. The angular speed for the connecting vector η^a is given by the ratio of the linear speed of rotation (i.e., the magnitude of $\omega_b{}^a\eta^b$) to the magnitude of the radius vector $\rho^a = \eta^a - \dfrac{\eta^b\,\omega_b}{\omega^n\,\omega_n}\,\omega^a$. (See figure 2.8.1 again.) (If $\omega^n\,\omega_n = \mathbf{0}$, then $\omega_{ab} = \mathbf{0}$, and the speed of angular rotation is 0.) It follows with a bit of calculation much like that done previously in this section that

$$\text{(angular speed)}^2 = \frac{-\omega_b{}^a\,\eta^b\,\omega_{ca}\,\eta^c}{-\rho^n\,\rho_n} = \cdots = (-\,\omega^n\omega_n); \tag{2.8.10}$$

i.e., the angular speed is $\|\omega^a\|$, as claimed.

PROBLEM 2.8.3. *Complete the calculation in equation (2.8.10). (Hint: Do not forget that we are doing the calculation at the initial point p where the connecting vector η^a is orthogonal to ξ^a.)*

The two italicized conditions concerning, respectively, the orientation and magnitude of ω^a determine it up to sign.

With the preceding remarks as motivation, we now say that our future-directed, unit timelike vector field ξ^a is *irrotational* or *twist-free* at a point if $\omega_{ab} = \mathbf{0}$ there (or, equivalently, if $\omega^a = \mathbf{0}$ or if $\xi_{[a}\nabla_b\,\xi_{c]} = \mathbf{0}$ there). It will be instructive to consider a condition that captures the requirement that ξ^a is twist-free *everywhere*. Let us say that a timelike vector field ξ^a (not necessarily of unit length) is *hypersurface orthogonal* if there exist smooth, real valued maps f and g (with the same domains of definition as ξ^a) such that, at all points, $\xi_a = f\,\nabla_a\,g$. Note that if the condition is satisfied, then the hypersurfaces of constant g value are everywhere orthogonal to ξ^a. (For if σ^a is a vector tangent to one of these hypersurfaces, $\sigma^n\nabla_n\,g = 0$. So $\sigma^n\xi_n = \sigma^n(f\,\nabla_n\,g) = 0$.) Let us further say that ξ^a is *locally hypersurface orthogonal* if the restriction of ξ^a to every sufficiently small open set is hypersurface orthogonal.

PROPOSITION 2.8.1. *Let ξ^a be a smooth, future-directed unit timelike vector field defined on M (or some open subset of M). Then the following conditions are equivalent.*

(1) $\omega_{ab} = \mathbf{0}$ *everywhere.*
(2) ξ^a *is locally hypersurface orthogonal.*

Proof. The implication from (2) to (1) is immediate. For if $\xi_a = f\, \nabla_a g$, then

$$\begin{aligned}\omega_{ab} &= h_{[a}{}^{m}\, h_{b]}{}^{n}\, \nabla_m\, \xi_n = h_{[a}{}^{m}\, h_{b]}{}^{n}\, \nabla_m\, (f\, \nabla_n\, g)\\ &= f\, h_{[a}{}^{m}\, h_{b]}{}^{n}\, \nabla_m\;\nabla_n\, g + h_{[a}{}^{m}\, h_{b]}{}^{n}\, (\nabla_m f)\, (\nabla_n\, g)\\ &= f\, h_{a}{}^{m}\, h_{b}{}^{n}\, \nabla_{[m}\;\nabla_{n]}\, g + h_{a}{}^{m}\, h_{b}{}^{n}\, (\nabla_{[m} f)\, (\nabla_{n]}\, g).\end{aligned}$$

But $\nabla_{[m}\;\nabla_{n]}\, g = \mathbf{0}$, since ∇ is torsion-free; and the second term in the final line vanishes as well since $h_{b}{}^{n}\, \nabla_n\, g = f^{-1}\, h_{b}{}^{n}\, \xi_n = \mathbf{0}$. So $\omega_{ab} = \mathbf{0}$. The converse is non-trivial. It is a special case of Frobenius' theorem (Wald [60, p. 436]). □

There is a nice picture that goes with the proposition. Think about an ordinary rope. In its natural twisted state, the rope cannot be sliced in such a way that the slice is orthogonal to all individual fibers. But if the rope is first untwisted, then such a slicing *is* possible. Thus *orthogonal sliceability is equivalent to fiber-untwistedness.* The proposition extends this intuitive equivalence to the four-dimensional "spacetime ropes" (i.e., congruences of worldlines) encountered in relativity theory. It asserts that a congruence is twist-free iff it is, at least locally, hypersurface orthogonal.

Let us now switch our attention to the expansion tensor θ_{ab} associated with ξ^a. First, we decompose it into two pieces. We set

$$\theta = \theta_{a}{}^{a} = \nabla_a\, \xi^a \tag{2.8.11}$$

$$\sigma_{ab} = \theta_{ab} - \frac{1}{3}\, h_{ab}\, \theta, \tag{2.8.12}$$

so that equation (2.8.3) can be expressed in the expanded form

$$\nabla_a\, \xi_b = \omega_{ab} + \sigma_{ab} + \frac{1}{3}\, h_{ab}\, \theta + \xi_a\, (\xi^n \nabla_n\, \xi_b). \tag{2.8.13}$$

Notice that the two expressions for θ in equation (2.8.11) are equal since $\xi^n \nabla_m \xi_n = \mathbf{0}$ and, therefore,

$$\theta_a{}^a = g^{ab}\,\theta_{ab} = g^{ab}\,h_{(a}{}^m\,h_{b)}{}^n\,\nabla_m\,\xi_n = h^{mn}\,\nabla_m\,\xi_n = (g^{mn} - \xi^m\,\xi^n)\,\nabla_m\,\xi_n = \nabla_n\,\xi^n.$$

Notice too that

$$\sigma_a{}^a = 0, \tag{2.8.14}$$

since $\sigma_a{}^a = \theta_a{}^a - \frac{1}{3}\,(g_a{}^a - \xi_a\,\xi^a)\,\theta = \theta - \theta = 0$.

θ is called the *scalar expansion field* associated with ξ^a, and σ_{ab} the *shear tensor field* associated with it. We can motivate this terminology much as we did that for ω_{ab}. We claim first that θ is a measure of the rate at which the volume of an (infinitesimal) blob of fluid increases under the flow associated with ξ^a. (It is the counterpart to the "divergence" of a vector field in ordinary three-dimensional Euclidean vector analysis.) To justify this interpretation, we do a simple calculation.

Let γ be an integral curve of ξ^a, and let p be any point on its image. Further, let $\overset{1}{\eta}{}^a$, $\overset{2}{\eta}{}^a$, $\overset{3}{\eta}{}^a$ be three vector fields on the image of γ that (i) are carried along by the flow associated with ξ^a (i.e., $\mathcal{L}_\xi\,\overset{i}{\eta}{}^a = \mathbf{0}$, for $i = 1, 2, 3$), and (ii) together with ξ^a, form an orthonormal basis at p. Then $h_{ab} = -(\overset{1}{\eta}_a\overset{1}{\eta}_b + \overset{2}{\eta}_a\overset{2}{\eta}_b + \overset{3}{\eta}_a\overset{3}{\eta}_b)$ at p. We consider the rate of change of the volume function $V = \epsilon_{abcd}\,\xi^a\overset{1}{\eta}{}^b\,\overset{2}{\eta}{}^c\,\overset{3}{\eta}{}^d$ in the direction ξ^a. It turns out that, at p,

$$\xi^n\nabla_n V = \theta\,V. \tag{2.8.15}$$

It is in this sense that θ *gives the instantaneous rate of volume increase, per unit volume, under the flow associated with* ξ^a. (This is the claim we made at the end of section 2.5.)

To verify equation (2.8.15), we compute $\xi^n\nabla_n V$. Since $\mathcal{L}_\xi\,\overset{i}{\eta}{}^a = \mathbf{0}$, we have $\xi^n\nabla_n\,\overset{i}{\eta}{}^a = \overset{i}{\eta}{}^n\nabla_n\,\xi^a$ and, hence,

$$\begin{aligned} \xi^n\nabla_n V &= \xi^n\nabla_n\,(\epsilon_{abcd}\,\xi^a\overset{1}{\eta}{}^b\,\overset{2}{\eta}{}^c\,\overset{3}{\eta}{}^d) \\ &= \epsilon_{abcd}\left[(\xi^n\nabla_n\,\xi^a)\,\overset{1}{\eta}{}^b\,\overset{2}{\eta}{}^c\,\overset{3}{\eta}{}^d + \ldots + (\overset{3}{\eta}{}^n\nabla_n\,\xi^d)\,\xi^a\,\overset{1}{\eta}{}^b\,\overset{2}{\eta}{}^c\right]. \end{aligned} \tag{2.8.16}$$

Now the vector $\epsilon_{abcd}\,\overset{1}{\eta}{}^b\,\overset{2}{\eta}{}^c\,\overset{3}{\eta}{}^d$ is orthogonal to $\overset{1}{\eta}{}^b$, $\overset{2}{\eta}{}^c$, and $\overset{3}{\eta}{}^d$. So, at p, it must be co-aligned with ξ_a. Indeed, we have $\epsilon_{abcd}\,\overset{1}{\eta}{}^b\,\overset{2}{\eta}{}^c\,\overset{3}{\eta}{}^d = (\epsilon_{nbcd}\,\xi^n\overset{1}{\eta}{}^b\,\overset{2}{\eta}{}^c\,\overset{3}{\eta}{}^d)\,\xi_a = V\,\xi_a$ there. So, $\epsilon_{abcd}(\xi^n\nabla_n\,\xi^a)\,\overset{1}{\eta}{}^b\,\overset{2}{\eta}{}^c\,\overset{3}{\eta}{}^d = \xi_a\,(\xi^n\nabla_n\,\xi^a)\,V = 0$ at p. Similarly, for example, we have

$$\epsilon_{abcd}(\overset{1}{\eta}{}^n\nabla_n\,\xi^b)\,\xi^a\,\overset{2}{\eta}{}^c\,\overset{3}{\eta}{}^d = -\overset{1}{\eta}_b\,(\overset{1}{\eta}{}^n\nabla_n\,\xi^b)\,V$$

at p. So, after handling all terms on the right side of equation (2.8.16) this way, we are left, at p, with

$$\begin{aligned}
\xi^n \nabla_n V &= -\,V\left[\overset{1}{\eta}_r\,(\overset{1}{\eta}{}^n \nabla_n\,\xi^r) + \overset{2}{\eta}_r\,(\overset{2}{\eta}{}^n \nabla_n\,\xi^r) + \overset{3}{\eta}_r\,(\overset{3}{\eta}{}^n \nabla_n\,\xi^r)\right] \\
&= -\,V(\overset{1}{\eta}_r\,\overset{1}{\eta}{}^n + \overset{2}{\eta}_r\,\overset{2}{\eta}{}^n + \overset{3}{\eta}_r\,\overset{3}{\eta}{}^n)(\nabla_n\,\xi^r) = V\,h_r{}^n\,\nabla_n\,\xi^r \\
&= V\,(g_r{}^n - \xi_r\xi^n)\,\nabla_n\,\xi^r = V\,\nabla_n\,\xi^n = V\,\theta.
\end{aligned}$$

This gives us equation (2.8.15).

Now consider σ_{ab}. It is symmetric (and orthogonal to ξ^a). So we can choose our three vector fields $\overset{1}{\eta}{}^a$, $\overset{2}{\eta}{}^a$, $\overset{3}{\eta}{}^a$ so that, in addition to being carried along by the flow of ξ^a, and (with ξ^a) forming an orthonormal basis at p, they satisfy $\sigma_{ab} = -(\overset{1}{k}\,\overset{1}{\eta}_a\overset{1}{\eta}_b + \overset{2}{k}\,\overset{2}{\eta}_a\overset{2}{\eta}_b + \overset{3}{k}\,\overset{3}{\eta}_a\overset{3}{\eta}_b)$ at p. (It is a basic fact of linear algebra that we can find an orthonormal basis at p that diagonalizes the symmetric 4×4 matrix of σ_{ab}-components.) Then $\sigma^a_{\ b}\,\overset{i}{\eta}{}^b = \overset{i}{k}\,\overset{i}{\eta}{}^a$, for each i; i.e., $\overset{i}{\eta}{}^a$ is an eigenvector of $\sigma^a_{\ b}$ with eigenvalue $\overset{i}{k}$. And the coefficients $\overset{i}{k}$ sum to 0, since $0 = \sigma_a{}^a = -(\overset{1}{k}\,\overset{1}{\eta}_a\overset{1}{\eta}{}^a + \overset{2}{k}\,\overset{2}{\eta}_a\overset{2}{\eta}{}^a + \overset{3}{k}\,\overset{3}{\eta}_a\overset{3}{\eta}{}^a) = (\overset{1}{k} + \overset{2}{k} + \overset{3}{k})$.

Suppose for the moment that $\omega_{ab} = \mathbf{0}$ and $\theta = 0$ at p. Then, by equations (2.8.4) and (2.8.12), $\xi^n \nabla_n\,\overset{i}{\eta}{}^a = \sigma_n{}^a\,\overset{i}{\eta}{}^n = \overset{i}{k}\,\overset{i}{\eta}{}^a$, for all i, at p. So, if (as above) we think of $\overset{i}{\eta}{}^a$ as a "connecting vector" pointing from an observer O to an (infinitesimally) close neighbor N, then the instaneous velocity of N relative to O is directed radially away from O at p and has magnitude $\overset{i}{k}$ there. Thus, each of the vectors $\overset{1}{\eta}{}^a$, $\overset{2}{\eta}{}^a$, $\overset{3}{\eta}{}^a$ is an axis of instantaneous expansion (or contraction) with associated magnitude $\overset{i}{k}$. Since the magnitudes sum to 0, expansion along one axis can occur only if there is contraction along another. Individual expansions and contractions so compensate each other that there is no net increase in volume. (Again, we are now considering the case where θ is 0.)

In general, the expansion factors $\overset{i}{k}$ are all different. But, for purposes of illustration, suppose that the factors on two axes are equal—say $\overset{1}{k} = \overset{2}{k}$. Further imagine that our infinitesimal blob has the shape of a sphere at p. Then there are two possibilities. If the common factor is positive, then the action of the flow flattens it into a pancake with axis $\overset{3}{\eta}{}^a$ ("pancake shear"). If it is negative, then it is elongated into a hot dog with axis $\overset{3}{\eta}{}^a$ ("hot dog shear"). The second possibility is illustrated in figure 2.8.2, where three possible actions are illustrated.

The full expansion tensor field θ_{ab} can be given another interesting geometric interpretation in the case where it is associated with a unit timelike flow ξ^a that is everywhere twist-free. In this case, by proposition 2.8.1, ξ^a is, at least

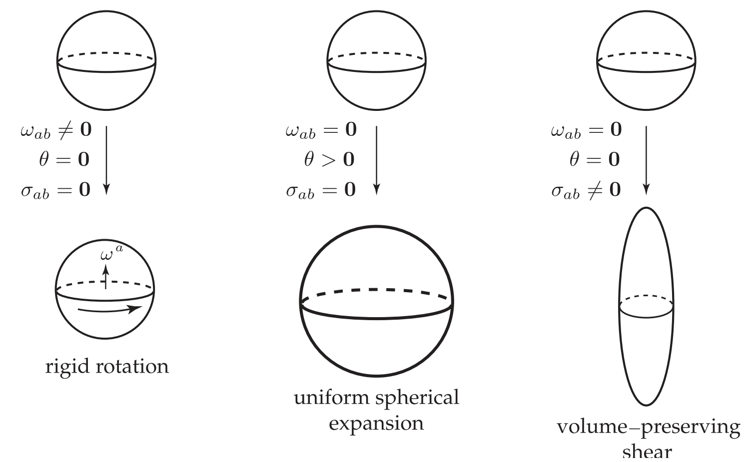

Figure 2.8.2. Rotation, expansion, shear.

locally, hypersurface orthogonal. Let S be a spacelike hypersurface to which ξ^a is orthogonal. The extrinsic curvature of S is given by $\pi_{ab} = h_a^{\ m}\, h_b^{\ n}\, \nabla_m\, \xi_n$. (Recall equation (1.10.16).) But $h_a^{\ m}\, h_b^{\ n}\, \nabla_m\, \xi_n = \omega_{ab} + \theta_{ab}$, by equations (2.8.1) and (2.8.2). So in the present case ($\omega_{ab} = \mathbf{0}$), we have $\pi_{ab} = \theta_{ab}$. Thus, *the expansion tensor field associated with a twist-free unit timelike field ξ^a is just the extrinsic curvature of the spacelike hypersurfaces to which ξ^a is orthogonal.*

This gives us another way to think about the extrinsic curvature of spacelike hypersurfaces. When $\pi_{ab} = \mathbf{0}$, normal vectors to the surface do not recede from one another. "Connecting vectors" between "infinitesimally" close surface normals do not expand. (See figure 2.8.3.) But when $\pi_{ab} \neq \mathbf{0}$, connecting vectors *do* expand.

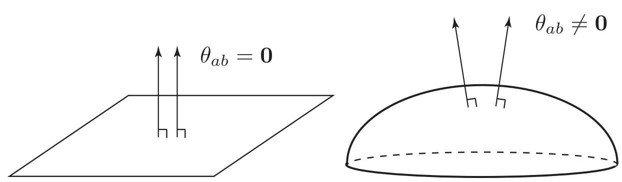

Figure 2.8.3. Expansion and extrinsic curvature.

Finally, we derive an expression for the rate of change of the scalar expansion function θ ("Raychaudhuri's equation"):

(2.8.17) $$\xi^a \nabla_a\, \theta = -R_{ab}\, \xi^a \xi^b + \omega_{ab}\, \omega^{ab} - \frac{1}{3}\,\theta^2 - \sigma_{ab}\, \sigma^{ab} + \nabla_a(\xi^n \nabla_n\, \xi^a).$$

We shall need it later in section 2.11. (Here ξ^a is still a smooth future-directed unit timelike vector field on our background spacetime (M, g_{ab}).) The derivation proceeds in two steps. First, it follows from equation (1.8.1) that

$$\begin{aligned}\xi^a \nabla_a \theta &= \xi^a \nabla_a \nabla_b \xi^b = -\xi^a R^b{}_{cab}\, \xi^c + \xi^a \nabla_b \nabla_a \xi^b \\ &= -R_{ca}\, \xi^c \xi^a + \nabla_b (\xi^a \nabla_a \xi^b) - (\nabla_b \xi^a)(\nabla_a \xi^b).\end{aligned}$$

Next, we evaluate the term $(\nabla_b \xi^a)(\nabla_a \xi^b)$ using the expansion in equation (2.8.13): $\nabla_a \xi_b = \omega_{ab} + \sigma_{ab} + \frac{1}{3}\, h_{ab}\, \theta + \xi_a\, (\xi^n \nabla_n \xi_b)$. A straightforward computation establishes that

$$(\nabla_b \xi_a)(\nabla^a \xi^b) = -\omega_{ab}\, \omega^{ab} + \frac{1}{3}\, \theta^2 + \sigma_{ab}\, \sigma^{ab}.$$

(All terms involving ξ_a or ξ_b are 0 because h_{ab}, ω_{ab}, σ_{ab}, and $\xi^n \nabla_n \xi_a$ are all orthogonal to ξ^a in all indices. The terms involving ω_{ab} together with either h^{ab} or σ^{ab} are 0 because the former is anti-symmetric whereas the latter is symmetric. The terms involving h^{ab} and σ_{ab} are 0 because $\sigma_a{}^a = 0$.) This gives us equation (2.8.17).

2.9. Killing Fields and Conserved Quantities

In relativity theory, there is a natural association between Killing fields and conserved quantities. We consider it briefly in this section.

Let κ^a be a smooth field on our background spacetime (M, g_{ab}). Recall (section 1.9) that κ^a is said to be a Killing field if its associated local flow maps Γ_s are all isometries or, equivalently, if $\pounds_\kappa\, g_{ab} = \mathbf{0}$. The latter condition can also be expressed as $\nabla_{(a}\, \kappa_{b)} = \mathbf{0}$.

Any number of standard symmetry conditions—local versions of them, at least[22]—can be cast as claims about the existence of Killing fields. Here are a few examples.

- (M, g_{ab}) is *stationary* if it has a Killing field that is everywhere timelike.
- (M, g_{ab}) is *static* if it has a Killing field that is everywhere timelike and locally hypersurface orthogonal.
- (M, g_{ab}) is *homogeneous* if its Killing fields, at every point of M, span the tangent space.

(We shall have another example in section 3.2, where we consider "stationary, axi-symmetric spacetimes.") The distinction between stationary and static

22. "Local" because Killing fields need not be complete, and their associated local flow maps need not be defined globally. (Recall our discussion at the end of section 1.3.)

spacetimes should be clear from our discussion in the preceding section. (Recall proposition 2.8.1.) Roughly speaking, in a stationary spacetime there is, at least locally, a "timelike flow" that preserves all spacetime distances. But the flow can exhibit rotation. Think of a whirlpool. It is the latter possibility that is ruled out when one passes to a static spacetime. For example, Gödel spacetime, as we shall see, is stationary but not static.

PROBLEM 2.9.1. *Let κ^a be a timelike Killing field that is locally hypersurface orthogonal ($\kappa_{[a}\nabla_b\,\kappa_{c]} = \mathbf{0}$). Further, let κ be the length of κ^a. (So $\kappa^2 = \kappa^n\kappa_n$.) Show that*

$$\kappa^2\,\nabla_a\,\kappa_b = -\kappa_{[a}\nabla_{b]}\,\kappa^2.$$

By way of example, let us find all Killing fields on Minkowski spacetime. This will be easy, as much of the work has already been prepared in sections 1.9 and 2.6.

Let κ^a be a Killing field on Minkowski spacetime (M, g_{ab}). Arguing *exactly* as in proposition 1.9.9, we can show that, given any point p in M, there is a unique constant, anti-symmetric field F_{ab} on M and a unique constant field k^a on M such that

$$\kappa_b = \chi^a\,F_{ab} + k_b, \tag{2.9.1}$$

where χ^a is the position field relative to p. (Recall that $F_{ab} = \nabla_a\,\kappa_b$, and $k_b = \kappa_b - \chi^a\,F_{ab}$.) Thus there is a one-to-one correspondence between Killing fields on Minkowski spacetime and pairs $(F_{ab},\,k_b)$ at any one point, where F_{ab} is an anti-symmetric tensor there and k^a is a vector there. It follows that the vector space of Killing fields on Minkowski spacetime has $6 + 4 = 10$ dimensions.

We can further analyze F_{ab} as in section 2.6. Let ϵ_{abcd} be a volume element on M; let ξ^a be a constant, future-directed, unit timelike field on M; and let E^a and B^a be defined as in equations (2.6.9) and (2.6.10):

$$E^a = F^a{}_b\,\xi^b,$$
$$B^a = \frac{1}{2}\,\epsilon^{abcd}\,\xi_b\,F_{cd}.$$

Then E^a and B^a are constant fields everywhere orthogonal to ξ^a. And it follows from equation (2.6.13) that we can express κ^a in the form

$$\kappa_b = \chi^a\,(2\,E_{[a}\,\xi_{b]} + \epsilon_{abcd}\,\xi^c\,B^d) + k_b. \tag{2.9.2}$$

This gives us a classification of all Killing fields (relative to an arbitrary choice of "origin" p and constant, unit timelike field ξ^a). Killing fields of the form

$\kappa^b = k^b$ generate (timelike, spacelike, or null) *translations*. Those of the form $\kappa_b = \chi^a\, \epsilon_{abcd}\, \xi^c\, B^d$ generate *spatial rotations*, based at p, with rotational axis B^a. Those of the form $\kappa_b = 2\chi^a\, E_{[a}\, \xi_{b]}$ generate *boosts*, based at p, in the plane determined by ξ^a and E^a.

PROBLEM 2.9.2. *Consider a non-trivial boost Killing field $\kappa_b = 2\chi^a E_{[a}\, \xi_{b]}$ on Minkowski spacetime (as determined relative to some point p and some constant unit timelike field ξ^a). "Non-trivial" here means that $E^a \neq \mathbf{0}$. Let η^a be a constant field on Minkowski spacetime. Show that $\pounds_\kappa\, \eta^a = \mathbf{0}$ iff η^a is orthogonal to both to ξ^a and E^a. (It follows that the boost isometries generated by κ^a leave intact all two-dimensional submanifolds orthogonal to ξ^a and E^a, but "rotate" all two-dimensional submanifolds to which ξ^a and E^a are tangent.)*

PROBLEM 2.9.3. *This time, consider a non-trivial rotational Killing field $\kappa_b = \chi^a\, \epsilon_{abcd}\, \xi^c\, B^d$ on Minkowski spacetime (with $B^a \neq \mathbf{0}$). Again, let η^a be a constant field on Minkowski spacetime. Show that $\pounds_\kappa\, \eta^a = \mathbf{0}$ iff η^a is a linear combination of ξ^a and B^a. (It follows that the isometries generated by κ^a "rotate" all two-dimensional submanifolds orthogonal to ξ^a and B^a but leave intact all two-dimensional submanifolds to which ξ^a and B^a are tangent.)*

Now we briefly consider two types of conserved quantity. One is an attribute of point particles with positive mass, the other of extended bodies. Let κ^a be a Killing field in an arbitrary spacetime (M, g_{ab}) (not necessarily Minkowski spacetime), and let $\gamma : I \to M$ be a smooth, future-directed, timelike curve, with unit tangent field ξ^a. We take its image to represent the worldline of a point particle with mass $m > 0$. Consider the quantity $J = (P^a \kappa_a)$, where $P^a = m\, \xi^a$ is the four-momentum of the particle. It certainly need not be constant on $\gamma[I]$. But it will be if γ is a geodesic. For in that case, $\xi^n \nabla_n\, \xi^a = \mathbf{0}$ and hence, by equation (1.9.12),

$$\xi^n \nabla_n J = m\, (\kappa_a\, \xi^n \nabla_n\, \xi^a + \xi^n \xi^a\, \nabla_n\, \kappa_a) = m\, \xi^n \xi^a\, \nabla_{(n}\, \kappa_{a)} = 0. \tag{2.9.3}$$

Thus, J is constant along the worldlines of *free* particles of positive mass.

We refer to J as the *conserved quantity associated with* κ^a. If κ^a is timelike, we call J the *energy* of the particle (associated with κ^a).[23] If it is spacelike,

23. Of course, one needs to ask what this notion of energy has to do with the one considered in section 2.4. There, ascriptions of energy to point particles were made relative to individual unit timelike vectors, and the value of the energy at any point was taken to be the inner product of that vector with the particle's four-momentum vector. We take the present notion of energy to be primary and the earlier one as derived. At least in the context of Minkowski spacetime, one can always extend a unit timelike vector at a point to a constant unit timelike *field* (which is, of course, a Killing field)

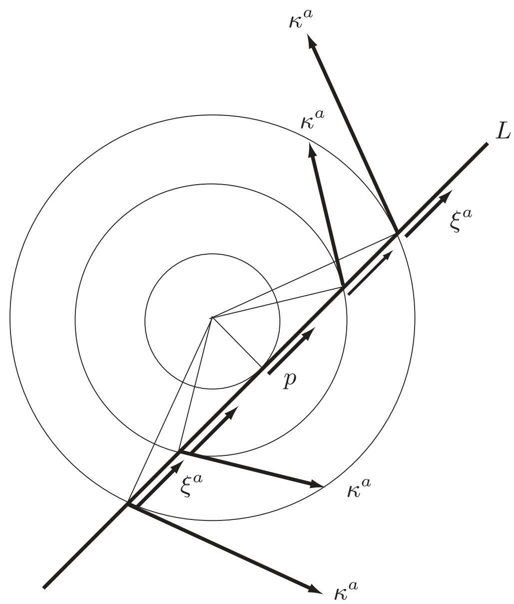

Figure 2.9.1. κ^a is a rotational Killing field. (It is everywhere orthogonal to a circle radius, and is proportional to it in length.) ξ^a is a tangent vector field of constant length on the line L. The inner product between them is constant. (Equivalently, the length of the projection of κ^a onto the line is constant.)

and if its associated flow maps resemble translations,[24] we call J the *linear momentum* of the particle (associated with κ^a). Finally, if κ^a is spacelike, and if its associated flow maps resemble rotations, then we call J the *angular momentum* of the particle (associated with κ^a).

It is useful to keep in mind a certain picture that helps one "see" why the angular momentum of free particles (to take that example) is conserved. It involves an analogue of angular momentum in Euclidean plane geometry. Figure 2.9.1 shows a rotational Killing field κ^a in the Euclidean plane, the image of a geodesic (i.e., a line) L, and the tangent field ξ^a to the geodesic. Consider the quantity $J = \xi^a \kappa_a$—i.e., the inner product of ξ^a with κ^a—along

---

and then understand relativization to the vector as relativization to the associated constant field. And perhaps the earlier usage is properly motivated only in spacetimes where individual unit timelike vectors are extendible to constant fields or, at least, to naturally distinguished Killing fields. (Similar remarks apply to components of "linear momentum" in particular directions.)

24. When one is dealing with Minkowski spacetime, one can assert without ambiguity that a Killing field generates a "translation," or a "spatial rotation," or a "boost." Things are not always so simple. Still, sometimes a Killing field in a curved spacetime resembles a Killing field in Minkowski spacetime in certain respects, and then the terminology may carry over naturally. For example, in the case of asymptotically flat spacetimes, one can classify Killing fields by their asymptotic behavior.

L. Exactly the same proof as before (of equation (2.9.3)) shows that J is constant along L.[25] But here we can better visualize the assertion.

Let us temporarily drop indices and write $\kappa \cdot \xi$ as one would in ordinary Euclidean vector calculus (rather than $\xi^a \kappa_a$). Let p be the point on L that is closest to the center point where κ vanishes. At that point, κ is parallel to ξ. As one moves away from p along L, in either direction, the length $\|\kappa\|$ of κ grows, but the angle $\angle(\kappa, \xi)$ between the vectors increases as well. It should seem at least plausible from the picture that the length of the projection of κ onto the line is constant and, hence, that the inner product $\kappa \cdot \xi = \cos(\angle(\kappa, \xi))\, \|\kappa\|\, \|\xi\|$ is constant.

That is how to think about the conservation of angular momentum for free particles in relativity theory. It does not matter that in the latter context we are dealing with a Lorentzian metric and allowing for curvature. The claim is still that a certain inner product of vector fields remains constant along a geodesic, and we can still think of that constancy as arising from a compensatory balance of two factors.

Let us now turn to the second type of conserved quantity, the one that is an attribute of extended bodies. Let κ^a be an arbitrary Killing field, and let T_{ab} be the energy-momentum field associated with some matter field. Assume it satisfies the conservation condition ($\nabla_a\, T^{ab} = \mathbf{0}$). Then ($T^{ab}\, \kappa_b$) is divergence free:

$$\nabla_a(T^{ab}\kappa_b) = \kappa_b\, \nabla_a T^{ab} + T^{ab}\nabla_a\kappa_b = T^{ab}\nabla_{(a}\kappa_{b)} = 0. \tag{2.9.4}$$

(The second equality follows from the conservation condition and the symmetry of T^{ab}; the third follows from the fact that κ^a is a Killing field.) It is natural, then, to apply Stokes's theorem to the vector field ($T^{ab}\kappa_b$). Consider a bounded system with aggregate energy-momentum field T_{ab} in an otherwise empty universe. Then there exists a (possibly huge) timelike world tube such that T_{ab} vanishes outside the tube (and vanishes on its boundary).

Let S_1 and S_2 be (non-intersecting) spacelike hypersurfaces that cut the tube as in figure 2.9.2, and let N be the segment of the tube falling between them (with boundaries included). By Stokes's theorem,[26]

25. The mass m played no special role.

26. See Wald [60, Appendix B.2] for a discussion of integration on manifolds and Stokes's theorem. We did not take the time to develop these topics in our review of differential geometry because we have so little need of them. This is the only place in this book where reference is made to integration on manifolds (except for the simple case of integration over curves).

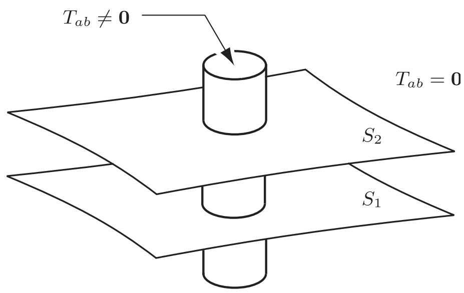

Figure 2.9.2. The integrated energy (relative to a background timelike Killing field) over the intersection of the world tube with a spacelike hypersurface is independent of the choice of hypersurface.

$$\int_{S_2} (T^{ab}\kappa_b)\, dS_a - \int_{S_1} (T^{ab}\kappa_b)\, dS_a$$
$$= \int_{S_2 \cap \partial N} (T^{ab}\kappa_b)\, dS_a - \int_{S_1 \cap \partial N} (T^{ab}\kappa_b)\, dS_a$$
$$= \int_{\partial N} (T^{ab}\kappa_b)\, dS_a = \int_N \nabla_a (T^{ab}\kappa_b)\, dV = 0.$$

Thus, the integral $\int_S (T^{ab}\kappa_b)\, dS_a$ is independent of the choice of spacelike hypersurface S intersecting the world tube, and is, in this sense, a conserved quantity (construed as an attribute of the system confined to the tube). An "early" intersection yields the same value as a "late" one. Again, the character of the background Killing field κ^a determines our description of the conserved quantity in question. If κ^a is timelike, we take $\int_S (T^{ab}\kappa_b)\, dS_a$ to be the aggregate energy of the system (associated with κ^a). And so forth.

Let us now continue the discussion that led to equation (2.9.3) and derive an inequality governing "total integrated acceleration." Once again, let κ^a be a Killing field on an arbitrary spacetime (M, g_{ab}), and let $\gamma : I \to M$ be a smooth, future-directed, timelike curve, with unit tangent field ξ^a. We take its image to represent the worldline of a point particle with mass $m > 0$. Again, we consider the quantity $J = (P^a \kappa_a)$, where $P^a = m\, \xi^a$ is the four-momentum of the particle. Even without assuming that γ is a geodesic, we have

(2.9.5) $$\xi^n \nabla_n J = m\, (\kappa_a\, \xi^n \nabla_n\, \xi^a + \xi^n \xi^a\, \nabla_n\, \kappa_a) = m\, \kappa_a\, \xi^n \nabla_n\, \xi^a.$$

Now let α be the scalar magnitude of the acceleration field; i.e., $\alpha^2 = -(\xi^n \nabla_n\, \xi^a)(\xi^m \nabla_m\, \xi_a)$. Then we have (see problem 2.9.4)

(2.9.6) $$|\xi^n \nabla_n\, J| \leq \alpha\, \sqrt{J^2 - m^2\, (\kappa^n \kappa_n)}.$$

(Of course, if γ is a geodesic—i.e., if $\alpha = 0$ everywhere—then $|\xi^n \nabla_n J|$ must vanish everywhere as well. So we recover our earlier result that J is constant in the case of geodesic motion.) If κ^a is causal (timelike or null) and future-directed everywhere, then $J = P^a \kappa_a > 0$, and it follows that

$$|\xi^n \nabla_n J| \leq \alpha\, J. \tag{2.9.7}$$

So, in this case, the *total integrated acceleration* of γ—the integral of α with respect to elapsed time—satisfies

$$TA(\gamma) = \int_\gamma \alpha\, ds \geq \int_\gamma \frac{|\xi^n \nabla_n J)|}{J}\, ds \geq \left| \int_\gamma \xi^n \nabla_n (\ln J)\, ds \right|. \tag{2.9.8}$$

Thus, if γ passes through points p_1 and p_2, the total integrated acceleration between those points is, at least, $|(\ln J)_{|p_2} - (\ln J)_{|p_1}|$. (For applications of equation (2.9.8), see Chakrabarti, Geroch, and Liang [7].)

PROBLEM 2.9.4. *Derive the inequality (2.9.6).*

2.10. The Initial Value Formulation

In this very brief section, we say a few words about the "initial value formulation" of general relativity and make precise the sense in which it is a deterministic theory. (See Hawking and Ellis [30] and Wald [60] for a proper treatment of the subject.)

Let S be a smooth, achronal, spacelike hypersurface in our background spacetime (M, g_{ab}). Recall (section 2.5) that $D(S)$, the domain of dependence of S, is the set of all points p in M with this property: given any smooth causal curve without endpoint, if its image passes through p, then it intersects S. Our goal is to explain the sense in which (at least in the empty space case) "what happens on S uniquely determines what happens on $D(S)$."

Of special interest is the case where S is a *Cauchy surface* in (M, g_{ab})—i.e., a smooth achronal spacelike hypersurface such that $D(S) = M$.

The first thing we must do is specify what is to count as "initial data" for the metric g_{ab} on S. Let ξ^a be the (unique) smooth, future-directed, unit timelike field that is everywhere orthogonal to S. (We will refer to it, simply, as the *normal field* to S.) Our first piece of initial data on S is the induced (negative definite) spatial metric $h_{ab} = g_{ab} - \xi_a \xi_b$. Our second piece is the extrinsic curvature field π_{ab} on S. We can think of the latter as the time derivative of h_{ab} in the direction ξ^a, at least up to the factor $\frac{1}{2}$, since $2\,\pi_{ab} = \pounds_\xi h_{ab}$. (Recall equation (1.10.17).)

Thus our metric initial data on S consists of the pair (h_{ab}, π_{ab}), the first and second fundamental forms on S. They correspond, respectively, to position and momentum in the initial value formulation of Newtonian particle mechanics. We know from our discussion in section 1.10 that these fields satisfy a number of constraint equations, including

$$\mathcal{R} - (\pi_a{}^a)^2 + \pi_{ab}\,\pi^{ab} = -2\,(R_{ab} - \frac{1}{2} R\, g_{ab})\,\xi^a\,\xi^b,$$
$$D_c\,\pi_a{}^c - D_a\,\pi_c{}^c = h^m{}_a\,h^{np}\,\xi^r\,R_{mnpr},$$

where D is the derivative operator induced on S, $\mathcal{R}^a{}_{bcd}$ is its associated Riemann curvature field, and $\mathcal{R}$ is the contracted scalar curvature field. (The first equation is just (1.10.21) and we get the second from (1.10.19) by contraction.) Using the symmetries of R_{mnpr}, we can re-express the right side of the second equation:

$$h^m{}_a\,h^{np}\,\xi^r\,R_{mnpr} = h^m{}_a\,(g^{np} - \xi^n\,\xi^p)\,\xi^r\,R_{nmrp} = h^m{}_a\,\xi^r\,R_{mr}$$
$$= h^m{}_a\,\xi^r\,(R_{mr} - \frac{1}{2} R\, g_{mr}).$$

And therefore, using Einstein's equation, we can express our two constraint equations as

$$\mathcal{R} - (\pi_a{}^a)^2 + \pi_{ab}\,\pi^{ab} = -16\,\pi\,T_{ab}\,\xi^a\,\xi^b, \tag{2.10.1}$$

$$D_c\,\pi_a{}^c - D_a\,\pi_c{}^c = 8\,\pi\,T_{mr}\,h^m{}_a\,\xi^r. \tag{2.10.2}$$

For simplicity, we shall restrict attention to the empty-space case—where T_{ab} vanishes and it is only the evolution of the metric field g_{ab} itself that we need to consider. In this special case, of course, the constraint equations assume the form

$$\mathcal{R} - (\pi_a{}^a)^2 + \pi_{ab}\,\pi^{ab} = 0, \tag{2.10.3}$$

$$D_c\,\pi_a{}^c - D_a\,\pi_c{}^c = \mathbf{0}. \tag{2.10.4}$$

We started with a spacetime (M, g_{ab}) and moved to an induced initial data set (h_{ab}, π_{ab}) on a smooth, achronal, spacelike hypersurface S in M satisfying particular constraint equations. Now we reverse direction.

We need a few definitions. Let us say officially that an *(empty space) initial data set* is a triple $(\Sigma, \tilde{h}_{ab}, \tilde{\pi}_{ab})$ where Σ is a smooth, connected, three-dimensional manifold, $\tilde{h}_{ab}$ is a smooth negative-definite metric on Σ, $\tilde{\pi}_{ab}$ is a smooth symmetric field on Σ, and the latter two satisfy the constraint equations (2.10.3) and (2.10.4).

A *Cauchy development* of such an initial data set $(\Sigma, \tilde{h}_{ab}, \tilde{\pi}_{ab})$ is a triple $((M, g_{ab}), S, \varphi)$ where (i) (M, g_{ab}) is a spacetime that satisfies the field equation $R_{ab} = \mathbf{0}$, (ii) S is a Cauchy surface in M, (iii) φ is a diffeomorphism of Σ onto S, and (iv) $\tilde{h}_{ab} = \phi^*(h_{ab})$ and $\tilde{\pi}_{ab} = \phi^*(\pi_{ab})$, where h_{ab} and π_{ab} are the first and second fundamental forms induced on S.

A Cauchy development $((M, g_{ab}), S, \varphi)$ of $(\Sigma, \tilde{h}_{ab}, \tilde{\pi}_{ab})$ is *maximal* if, in addition, given any other Cauchy development $((M', g'_{ab}), S', \varphi')$ of $(\Sigma, \tilde{h}_{ab}, \tilde{\pi}_{ab})$, there is an isometry ψ of M' into M that respects Σ in the sense that $\psi \circ \varphi' = \varphi$.

Our basic result (due to Choquet-Bruhat and Geroch [8]) is the following.

PROPOSITION 2.10.1. *Every empty space initial data set has a maximal Cauchy development. It is unique in the following sense. If $((M, g_{ab}), S, \varphi)$ and $((M', g'_{ab}), S', \varphi')$ are both maximal Cauchy developments of $(\Sigma, \tilde{h}_{ab}, \tilde{\pi}_{ab})$, there is a diffeomorphism $\psi : M' \to M$ such that $\psi \circ \varphi' = \varphi$ and $g'_{ab} = \psi^*(g_{ab})$.*

Proposition 2.10.1 makes precise the sense in which general relativity is a deterministic theory. But that sense is local in character because it need not be the case in an arbitrary spacetime (M, g_{ab}) that there is any one achronal spacelike hypersurface S such that $D(S) = M$; i.e., it need not be case that there is a Cauchy surface. (For example, the spacetime that arises by taking the universal covering space of anti-deSitter spacetime admits no Cauchy surface. See Hawking and Ellis [30], section 5.2.)

2.11. Friedmann Spacetimes

In this section, we briefly consider the class of Friedmann (or Friedmann-Lemaître-Robertson-Walker) spacetimes. These are the "standard models" of relativistic cosmology. (For a more complete discussion, see Wald [60] or almost any text in general relativity.) We include this section, even though we are not otherwise undertaking to survey known exact solutions to Einstein's equation, because we have a particular interest in comparing relativistic cosmology with Newtonian cosmology. We consider the latter in section 4.4.

We take a Friedmann spacetime to be one that satisfies a particular symmetry condition—"spatial homogeneity and isotropy"—together with supplemental constraints in the form of energy conditions, equations of state, or both. We start with the symmetry condition.

Roughly speaking, a spacetime is spatially homogeneous and isotropic if there is a congruence of timelike curves filling the spacetime such that "space," as determined relative to the congruence, "is the same in all directions." Here is one way to make the condition precise. (We opt for a local version of the condition. And to avoid certain distracting complications, we cast the definition directly in terms of the existence of isometries, rather than in terms of Killing fields as we did with several symmetry conditions at the beginning of section 2.9.)

Let (M, g_{ab}) be a spacetime, and let ξ^a be a smooth, future-directed, unit timelike field on M that is twist-free; i.e., $\xi_{[a}\nabla_b\,\xi_{c]} = \mathbf{0}$. (So, at least locally, it is possible to foliate M with a one-parameter family of spacelike hypersurfaces that are orthogonal to ξ^a. Recall our discussion in section 2.8. We can think of each of these hypersurfaces as constituting "space" at a given time relative to ξ^a.) We say that (M, g_{ab}) is *spatially homogeneous and isotropic relative to* ξ^a if, for all points p in M, and all unit spacelike vectors $\overset{1}{\sigma}{}^a$ and $\overset{2}{\sigma}{}^a$ at p that are orthogonal to ξ^a, there is an open set O containing p and an isometry $\varphi\colon O \to O$ that keeps p fixed, preserves the field ξ^a, and maps $\overset{1}{\sigma}{}^a$ to $\overset{2}{\sigma}{}^a$ (i.e., such that $\varphi(p) = p$, $\varphi_*(\xi^a) = \xi^a$ and $\varphi_*(\overset{1}{\sigma}{}^a) = \overset{2}{\sigma}{}^a$).[27] We further say that (M, g_{ab}) is *spatially homogeneous and isotropic* if it is so relative to *some* choice of ξ^a. The strength of the condition will become clear as we proceed.

We assume in what follows that ξ^a is as in the preceding paragraph and (M, g_{ab}) is spatially homogeneous and isotropic relative to ξ^a. We first abstract a few general principles.

(1) Given any field λ^a on M, if it is definable in terms of, or otherwise determined by, g_{ab} and ξ^a, then it must be proportional to ξ^a. (So $\lambda^a = \lambda\,\xi^a$ where $\lambda = \lambda_n \xi^n$. And if λ^a is also orthogonal to ξ^a, then $\lambda^a = \mathbf{0}$.)

This follows, for if at some point p, λ^a had a non-zero component orthogonal to ξ^a, it would determine a "preferred" orthogonal direction there and violate the isotropy condition. (Here is the argument in more detail. Since that component is determined by g_{ab} and ξ^a, it must be invariant under all maps that preserve g_{ab} and ξ^a and that leave p fixed. But, by our assumption of spatial homogeneity

27. Note, we require here that φ map the field ξ^a onto itself everywhere, not just at p. If we required only that it keep fixed the vector $\xi^a{}_{|p}$, the condition would not be strong enough for our purposes. For example, Minkowski spacetime would then qualify as spatially homogeneous and isotropic relative to *any* smooth, future-directed, unit timelike vector field ξ^a that is twist-free. It would not have to be the case, as we want it to be, that hypersurfaces orthogonal to ξ^a are manifolds of constant curvature. For the corresponding global version of the condition, we would require at the outset that ξ^a be (globally) hypersurface orthogonal and require that, for all p, $\overset{1}{\sigma}{}^a$, and $\overset{2}{\sigma}{}^a$ as specified, there is a (global) isometry $\varphi\colon M \to M$ that keeps p fixed, preserves the field ξ^a, and maps $\overset{1}{\sigma}{}^a$ to $\overset{2}{\sigma}{}^a$. We shall later consider what turns on the difference between these two (local vs. global) versions of the spatial homogeneity and isotropy condition.

and isotropy, the only vector at p, orthogonal to ξ^a, that is invariant under all such maps is the zero vector.) It follows from (1), for example, that the acceleration field $\xi^n \nabla_n \xi^a$ must vanish; i.e., ξ^a must be a geodesic field.

(2) Given any scalar field λ on M, if it is definable in terms of, or otherwise determined by, g_{ab} and ξ^a, then it must be constant on all spacelike hypersurfaces orthogonal to ξ^a. (So $\nabla_a \lambda = (\xi^n \nabla_n \lambda)\, \xi_a$.)

This is an immediate consequence of (1) as applied to $h^{ab}\, \nabla_b\, \lambda$, where h_{ab} is the spatial projection field $(g_{ab} - \xi_a\, \xi_b)$. So, for example, we have

$$\nabla_a\, \theta = (\xi^n \nabla_n\, \theta)\, \xi_a, \tag{2.11.1}$$

where $\theta = \nabla_m\, \xi^m$. (Recall section 2.8.)

(3) Given any symmetric field λ_{ab} on M, if it is definable in terms of, or otherwise determined by, g_{ab} and ξ^a, then it must be of the form $\lambda_{ab} = \alpha\, \xi_a \xi_b + \beta\, h_{ab}$ for some scalar fields α and β. (And if λ_{ab} is also orthogonal to ξ^a, then it must be of the form $\lambda_{ab} = \beta\, h_{ab}$.)

To see this, consider any point p. By (1) as applied to $\lambda^a{}_b\, \xi^b$, there is a number α such that $\lambda^a{}_b\, \xi^b = \alpha\, \xi^a$ at p. Now consider the tensor $(\lambda^{ab} - \alpha\, \xi^a\, \xi^b)$ at p. It is symmetric and orthogonal to ξ^a in both indices. So we can express it in the form

$$\lambda^{ab} - \alpha\, \xi^a\, \xi^b = -(\overset{1}{\sigma}\, \overset{1}{\sigma}{}^a\, \overset{1}{\sigma}{}^b + \overset{2}{\sigma}\, \overset{2}{\sigma}{}^a\, \overset{2}{\sigma}{}^b + \overset{3}{\sigma}\, \overset{3}{\sigma}{}^a\, \overset{3}{\sigma}{}^b), \tag{2.11.2}$$

where the vectors $\overset{1}{\sigma}{}^a, \ldots, \overset{3}{\sigma}{}^a$, together with ξ^a, form an orthonormal (eigen)basis for g_{ab} at p. But now, by the isotropy condition, the coefficients $\overset{1}{\sigma}$, $\overset{2}{\sigma}$, $\overset{3}{\sigma}$ must be equal. (For all i and j, there is an isometry that leaves p and $(\lambda^{ab} - \alpha\, \xi^a\, \xi^b)$ fixed but takes $\overset{i}{\sigma}{}^a$ to $\overset{j}{\sigma}{}^a$.) If their common value is β, then the right-side tensor in equation (2.11.2) can be expressed as $\beta\, h^{ab}$.

It follows from (3) that the shear tensor field σ_{ab} associated with ξ^a must be of the form $\sigma_{ab} = \beta\, h_{ab}$. But σ_{ab} is "trace-free," so $0 = \sigma_a{}^a = 3\, \beta$. Thus, ξ^a has vanishing shear in addition to being geodesic. And we assumed at the outset that it is twist-free. So, by equation (2.8.13),

$$\nabla_a\, \xi_b = \frac{1}{3}\, h_{ab}\, \theta. \tag{2.11.3}$$

It also follows from (3) that we can construe (M, g_{ab}) as an exact solution to Einstein's equation for a perfect fluid source with four-velocity ξ^a. For if $R_{ab} = \alpha\, \xi_a \xi_b + \beta\, h_{ab}$, then

$$R_{ab} - \frac{1}{2} R g_{ab} = 8\, \pi\, (\rho\, \xi_a \xi_b - p\, h_{ab}), \tag{2.11.4}$$

where $\rho = \dfrac{(\alpha - 3\,\beta)}{16\pi}$ and $p = \dfrac{(\alpha + \beta)}{16\pi}$. (This perfect fluid need not satisfy any of the standard energy conditions. We shall soon add one of those conditions as a supplemental constraint, but will work without it for now.) In what follows, we take T_{ab} to be the indicated energy-momentum field; i.e., we take $T_{ab} = \rho\,\xi_a\xi_b - p\,h_{ab}$. So (after inversion of equation (2.11.4)),

$$R_{ab} = 8\,\pi\,(T_{ab} - \frac{1}{2}\,T\,g_{ab}) = 4\,\pi\,(\rho + 3p)\,\xi_a\,\xi_b - 4\,\pi\,(\rho - p)h_{ab}. \tag{2.11.5}$$

Next we consider the geometry of spacelike hypersurfaces orthogonal to ξ^a. Let $\mathcal{S}$ be one such hypersurface, and let h_{ab} and π_{ab} be the first and second fundamental forms induced on $\mathcal{S}$. (Recall section 1.10.) Note that, by equations (1.10.16) and (2.11.3), the latter assumes a simple form: $\pi_{ab} = h_a{}^m\,h_b{}^n\,\nabla_m\,\xi_n = \frac{1}{3}\,h_{ab}\,\theta$. Now let $\mathcal{R}^a{}_{bcd}$ be the curvature field associated with the induced derivative operator $\mathcal{D}$. Our goal is to derive an expression for $\mathcal{R}^a{}_{bcd}$ in terms of θ, ρ, and p. We do so by first deriving one for $\mathcal{R}_{bc}$ and then invoking a general fact about the relation between the two fields that holds in the special case of three-dimensional manifolds. It follows from equation (1.10.20), our expression above for π_{ab}, and from equation (2.11.5) that

$$\begin{aligned}\mathcal{R}_{bc} &= \pi_n{}^n\,\pi_{bc} - \pi_{ab}\,\pi^{\,a}_{\ c} + h^n_{\ b}\,h^p_{\ c}\,R_{np} - R_{mbcr}\,\xi^m\,\xi^r \\ &= \frac{1}{3}\,\theta^2\,h_{bc} - \frac{1}{9}\,\theta^2\,h_{bc} - 4\,\pi\,(\rho - p)h_{bc} - R_{mbcr}\,\xi^m\,\xi^r.\end{aligned}$$

So we need only derive an expression for the fourth term on the right side. (Here and in what follows we shall use the abbreviation $\dot{\theta} = \xi^n\,\nabla_n\,\theta$.) Note that by equations (2.11.3) and (2.11.1),

$$\begin{aligned}\nabla_c\,\nabla_r\,\xi_b = \frac{1}{3}\,\nabla_c\,(h_{rb}\,\theta) &= \frac{1}{3}\,[\,h_{rb}\,\xi_c\,\dot{\theta} + \theta\,\nabla_c(g_{rb} - \xi_r\,\xi_b)\,] \\ &= \frac{1}{3}\,[\,h_{rb}\,\xi_c\,\dot{\theta} - \theta\,\xi_r\nabla_c\,\xi_b - \theta\,\xi_b\nabla_c\,\xi_r\,] \\ &= \frac{1}{3}\,[\,h_{rb}\,\xi_c\,\dot{\theta} - \frac{1}{3}\,\theta^2\,\xi_r\,h_{cb} - \frac{1}{3}\,\theta^2\,\xi_b\,h_{cr}\,].\end{aligned}$$

Hence

$$R_{mbcr}\,\xi^m = 2\,\nabla_{[c}\,\nabla_{r]}\,\xi_b = \frac{2}{3}\,\xi_{[c}\,h_{r]b}\left(\dot{\theta} + \frac{1}{3}\,\theta^2\right)$$

and, therefore,

$$R_{mbcr}\,\xi^m\,\xi^r = -\frac{1}{3}\,h_{bc}\,\left(\dot{\theta} + \frac{1}{3}\,\theta^2\right).$$

Substituting this into our expression for $\mathcal{R}_{bc}$ yields

$$\mathcal{R}_{bc} = \mathcal{K}\,h_{bc}, \tag{2.11.6}$$

where $\mathcal{K} = \frac{1}{3}(\theta^2 + \dot{\theta}) - 4\pi\,(\rho - p)$. Now we invoke our general fact. In the special case of a three-dimensional manifold with metric, we (always) have[28]

(2.11.7)
$$\mathcal{R}_{abcd} = (h_{bc}\,\mathcal{R}_{ad} + h_{ad}\,\mathcal{R}_{bc} - h_{ac}\,\mathcal{R}_{bd} - h_{bd}\,\mathcal{R}_{ac}) + \frac{1}{2}\,(h_{ac}\,h_{bd} - h_{ad}\,h_{bc})\,\mathcal{R}.$$

So it follows from equation (2.11.6) (and its contracted form $\mathcal{R} = 3\mathcal{K}$) that

(2.11.8)
$$\mathcal{R}_{abcd} = (\frac{\mathcal{R}}{2} - 2\mathcal{K})\,(h_{ac}\,h_{bd} - h_{ad}\,h_{bc}) = \frac{\mathcal{K}}{2}\,(h_{ad}\,h_{bc} - h_{ac}\,h_{bd}).$$

Thus, recalling our discussion at the end of section 1.9, we see that $(\mathcal{S}, h_{ab})$ has constant curvature $\mathcal{K}/2$. (We shall soon have a more instructive expression for $\mathcal{K}$.)

Now we turn to considerations of dynamics. We claim that

(2.11.9)
$$\dot{\theta} = -4\pi\,(\rho + 3\,p) - \frac{1}{3}\,\theta^2,$$

(2.11.10)
$$\dot{\rho} = -(\rho + p)\,\theta.$$

(We shall continue to use the dot notation. Here $\dot{\rho} = \xi^n \nabla_n\,\rho$.) We get the first from Raychaudhuri's equation (2.8.17), using equation (2.11.5) and the fact that ξ^a is geodesic, irrotational, and shear-free. The second is the continuity condition (2.5.5). Recall that the latter follows from the conservation condition $\nabla_a\,T^{ab} = \mathbf{0}$ as applied to our energy-momentum field $T^{ab} = \rho\,\xi_a\xi_b - p\,h_{ab}$. (And the conservation condition itself is a consequence of Einstein's equation.)

It is convenient and customary to introduce a new field a that we can think of as a "scaling factor." We want it to be constant on spacelike hypersurfaces orthogonal to ξ^a; i.e., $h^{mn}\nabla_m\,a = \mathbf{0}$. So we need only specify its growth along any one integral curve of ξ^a. We define it, up to a multiplicative constant, by the condition

(2.11.11)
$$\frac{1}{3}\,\theta = \frac{\dot{a}}{a}.$$

(Certainly this equation has solutions. Indeed, if the curve is parametrized by a time function t where $\xi_a = \nabla_a\,t$, then all functions of the form $a(t) = e^{f(t)}$, with $f(t) = \int_{t_0}^{t} \frac{\theta}{3}\,dt$, qualify.) The condition inherits a natural interpretation from the one we have given for θ. It concerns the rate of volume increase for

28. We shall later prove a close analogue of this result (proposition 4.1.4) in connection with our discussion of classical spacetimes. It should be clear how to adapt the proof to the present context. (We present the argument there rather than here because of added complications that arise when one is dealing with classical spacetimes.)

a fluid with four-velocity ξ^a. We saw in section 2.8 that if an ("infinitesimal") blob of the fluid has volume V, then $\dot{V} = V\,\theta$. (Recall equation (2.8.15).) If we think of the blob as a cube whose edges have length a, then $V = a^3$ and we are led immediately to equation (2.11.11). It is in this sense that a is a scaling factor. If we now express our equations for $\dot{\theta}$ and $\dot{\rho}$ above in terms of a, we have

$$3\,\frac{\ddot{a}}{a} = -4\,\pi\,(\rho + 3\,p), \tag{2.11.12}$$

$$\dot{\rho} = -3\,\frac{\dot{a}}{a}\,(\rho + p) \tag{2.11.13}$$

where, of course, $\ddot{a} = \xi^n \nabla_n (\xi^m \nabla_m\, a)$. These two jointly imply (by integration) that there is a number k such that

$$\left(\frac{\dot{a}}{a}\right)^2 - \frac{8\,\pi}{3}\,\rho = -\frac{k}{a^2}. \tag{2.11.14}$$

(This is "Friedmann's equation.") Since a was only determined initially up to a multiplicative constant, we can now normalize it so that $k = -1$ or $k = 0$ or $k = 1$.

We can use the listed equations to express several fields of interest directly in terms of the scaling factor a and k:

$$8\,\pi\,\rho = 3\,\left(\frac{\dot{a}}{a}\right)^2 + 3\,\frac{k}{a^2}, \tag{2.11.15}$$

$$8\,\pi\,p = -2\,\frac{\ddot{a}}{a} - \left(\frac{\dot{a}}{a}\right)^2 - \frac{k}{a^2}, \tag{2.11.16}$$

$$\mathcal{R}_{abcd} = -\,\frac{k}{a^2}\,(h_{ad}\,h_{bc} - h_{ac}\,h_{bd}). \tag{2.11.17}$$

Here equation (2.11.15) is just a reformulation of (2.11.14). Equation (2.11.16) follows from (2.11.12) and (2.11.15). For equation (2.11.17), recall that, by (2.11.8), $\mathcal{R}_{abcd} = \frac{\mathcal{K}}{2}\,(h_{ad}\,h_{bc} - h_{ac}\,h_{bd})$, where $\mathcal{K} = \frac{1}{3}(\theta^2 + \dot{\theta}) - 4\,\pi\,(\rho - p)$. But

$$\frac{1}{3}(\theta^2 + \dot{\theta}) = \frac{\ddot{a}}{a} + 2\left(\frac{\dot{a}}{a}\right)^2$$

by equations (2.11.9), (2.11.11), and (2.11.12). And it follows from equations (2.11.15) and (2.11.16) that

$$4\,\pi\,(\rho - p) = \frac{\ddot{a}}{a} + 2\left(\frac{\dot{a}}{a}\right)^2 + 2\,\frac{k}{a^2}.$$

So $\mathcal{K} = -2\,\frac{k}{a^2}$, as claimed.

Equation (2.11.17) tells us that $(\mathcal{S}, h_{ab})$ has constant curvature $-k/a^2$. Remember, though, that h_{ab} is negative definite, and curvature is usually reported in terms of the positive definite metric $-h_{ab}$. This introduces a sign change. (The switch from h_{ab} to $-h_{ab}$ leaves D, $\mathcal{R}^a{}_{bcd}$, and $(h_{ad}h_{bc} - h_{ac}h_{bd})$ intact, but reverses the sign of $\mathcal{R}_{abcd} = h_{an}\mathcal{R}^n{}_{bcd}$.) So we shall record our conclusion this way:

> $(\mathcal{S}, -h_{ab})$ *is a manifold of constant curvature, and the magnitude of its curvature is* $(-1/a^2)$, 0, *or* $(1/a^2)$ depending on whether k is -1, 0, or 1.

We have reached this point assuming only a local version of the spatial isotropy condition. But now suppose for a moment that the global version holds as well, and let $\mathcal{S}$ be any maximally extended spacelike hypersurface that is everywhere orthogonal to ξ^a. Then we can say more about the global structure of $(\mathcal{S}, -h_{ab})$. In this case, it follows from the way the global condition is formulated that $(\mathcal{S}, -h_{ab})$ is, itself, a homogeneous, isotropic three-manifold in this sense: for all points p in $\mathcal{S}$, and all unit vectors $\overset{1}{\sigma}{}^a$ and $\overset{2}{\sigma}{}^a$ in the tangent space to $\mathcal{S}$ at p, there is an isometry $\psi : \mathcal{S} \to \mathcal{S}$ that keeps p fixed and that maps $\overset{1}{\sigma}{}^a$ to $\overset{2}{\sigma}{}^a$. This is a very strong constraint and rules out all but a small number of possibilties (Wolf [64]). If $k = 0$, $(\mathcal{S}, -h_{ab})$ cannot be just any flat three-manifold. It must be isometric to three-dimensional Euclidean space; i.e., it must also be diffeomorphic to $\mathbb{R}^3$ and geodesically complete.[29] If $k = -1$, $(\mathcal{S}, -h_{ab})$ must be isometric to three-dimensional hyperbolic space H^3. (We shall return to consider one realization of three-dimensional hyperbolic space at the end of the section.) Finally, if $k = 1$, $(\mathcal{S}, -h_{ab})$ must be isometric either to three-dimensional spherical space S^3 or to three-dimensional elliptic space P^3. The latter arises if one identifies "antipodal points" in the former.

Let us now revert to the local version of the spatial homogeneity and isotropy condition—leaving open the global structure of maximally extended spacelike hypersurface orthogonal to ξ^a—and continue with our consideration of dynamics. The difference in strength between the two versions of the condition plays no role here.

So far, assuming only the spatial homogeneity and isotropy condition, we have established that the scaling function a must satisfy equations (2.11.15) and (2.11.16). Now for the first time, just so as to have one example, we assume

29. This should seem, at least, intuitively plausible. Consider a lower dimensional case. The Euclidean plane is not the only two-dimensional Riemannian manifold of constant 0 curvature. The cylinder and the torus also qualify. But neither of them is isotropic in the relevant sense. For, given a point in either, the only global isometry of the manifold that keeps the point fixed is the identity map.

that our perfect fluid satisfies a particular equation of state, namely $p = 0$, and consider how the latter constrains the growth of the scaling function. (We are certainly not claiming that this assumption is realistic—i.e., holds (approximately) in our universe.)

If we insert this value for p in equation (2.11.16) and multiply by $a^2\,\dot{a}$, we arrive at $2\,\ddot{a}\,\dot{a}\,a + \dot{a}^3 + k\,\dot{a} = 0$. It follows that there is a number C such that $\dot{a}^2\,a + k\,a = \frac{8\pi}{3}\rho\,a^3 = C$. (The first equality follows from equation (2.11.15).) So our task is now reduced to solving the differential equation

$$\dot{a}^2 - \frac{C}{a} + k = 0. \tag{2.11.18}$$

The solutions are the following. (It is convenient to express two of them in parametric form.)

$$k = -1 \quad \begin{cases} a(x) = \dfrac{C}{2}\,(\cosh x - 1) \\[2ex] t(x) = \dfrac{C}{2}\,(\sinh x - x) \end{cases} \qquad x \in (0,\, \infty)$$

$$k = 0 \qquad a(t) = \left(\frac{9\,C}{4}\right)^{\frac{1}{3}} t^{\frac{2}{3}} \qquad t \in (0,\, \infty)$$

$$k = +1 \quad \begin{cases} a(x) = \dfrac{C}{2}\,(1 - \cos x) \\[2ex] t(x) = \dfrac{C}{2}\,(x - \sin x). \end{cases} \qquad x \in (0,\, 2\,\pi)$$

These are maximally extended solutions for the case where θ is positive at at least one point. We get additional (time-reversed) solutions if we assume that θ is negative at at least one point.

Rough (qualitative) graphs of these solutions are given in figure 2.11.1. If $k = -1$ or $k = 0$, expansion starts at the big bang and continues forever. In both cases, the rate of expansion $\frac{da}{dt}$ decreases monotonically. But there is this difference: the rate of expansion shrinks to 0 asymptotically when $k = 0$, but has a limit value that is strictly positive when $k = -1$. (One curve is asymptotically flat; the other is not.) In contrast, if $k = 1$, expansion continues until a maximum value is reached for a (at time $t = \frac{C\,\pi}{2}$) and then a period of accelerating contraction begins that leads to a big crunch.

PROBLEM 2.11.1. *Confirm that the three stated solutions do, in fact, satisfy equation (2.11.18).*

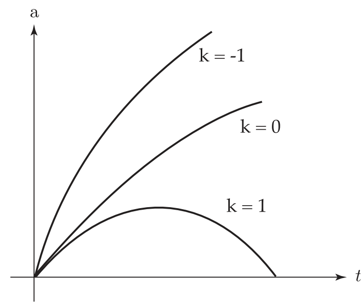

Figure 2.11.1. Rough graphs of the scaling factor a in the three cases.

PROBLEM 2.11.2. *Consider a second equation of state, namely that in which $\rho = 3\,p$. (For $T_{ab} = \rho\,\xi_a\,\xi_b - p\,h_{ab}$, this is equivalent to $T = 0$.) Show that in this case there is a number C' such that*

$$\dot{a}^2\,a^2 + k\,a^2 = \frac{8\pi}{3}\rho\,a^4 = C'.$$

(So in this case, the equation to solve is not (2.11.18), but rather

$$\dot{a}^2 - \frac{C'}{a^2} + k = 0.)$$

It will be instructive to consider an ultra-simple, degenerate Friedmann spacetime and see how some of our claims turn out in this special case. Let (M, g_{ab}) be Minkowski spacetime. Let o be any point in M, and let O be the (open) set of all points p in M such that $o \ll p$—i.e., such that there is a smooth future-directed timelike curve that runs from o to p. (See figure 2.11.2.) Further, let χ^a be the position field based at o—so χ^a vanishes at o and $\nabla_a \chi^b = \mathbf{0}$—and let ξ^a be the field

$$\xi^a = (\chi_b \chi^b)^{-\frac{1}{2}}\,\chi^a$$

as restricted to O. The latter is, clearly, a smooth, future-directed, unit timelike field on O. Moreover, it is (globally) hypersurface orthogonal; i.e., there exist smooth scalar fields f and g on O such that $\xi_a = f\,\nabla_a\,g$. Indeed, if $\chi = (\chi_a \chi^a)^{\frac{1}{2}}$, then $\chi^a = \chi\,\xi^a$, and

$$\nabla_n \chi = \frac{1}{2}\,(\chi_a \chi^a)^{-\frac{1}{2}}\,\nabla_n\,(\chi_b \chi^b) = (\chi_a \chi^a)^{-\frac{1}{2}}\,\chi_b\,\delta_n{}^b = \chi^{-1}\,\chi_n = \xi_n. \tag{2.11.19}$$

We claim that the restricted spacetime $(O, g_{ab|O})$ is spatially homogeneous and isotropic with respect to ξ^a and so qualifies as a Friedmann spacetime (with $\rho = p = 0$). Indeed, this reduces to a standard claim about the symmetries of Minkowski spacetime. Given any point p in O, and any two (distinct) unit

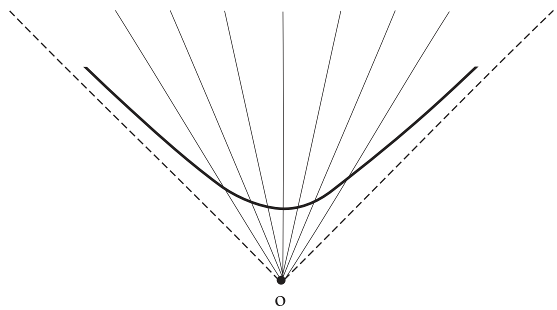

Figure 2.11.2. Minkowski spacetime (in profile) as restricted to the set of all points to the timelike future of a point o. It qualifies as a (degenerate) Friedmann spacetime with $\rho = p = 0$. A χ = constant hyperboloid is indicated. It (together with the metric induced on it) is a realization of three-dimensional hyperbolic space.

spacelike vectors $\overset{1}{\sigma}{}^a$ and $\overset{2}{\sigma}{}^a$ at p that are orthogonal to ξ^a, there is a spatial rotation that keeps p fixed, preserves the field ξ^a, and takes $\overset{1}{\sigma}{}^a$ to $\overset{2}{\sigma}{}^a$.[30]

We know from our earlier discussion that equation (2.11.3) must hold. In this special case, it is easy to check the result with a direct computation. By equation (2.11.19), we have

$$\begin{aligned} \theta &= \nabla_a \xi^a = \nabla_a (\chi^{-1} \chi^a) = \chi^{-1} (\nabla_a \chi^a) - \chi^{-2} \chi^a \nabla_a \chi \\ &= 4\chi^{-1} - \chi^{-1} = 3\chi^{-1}, \end{aligned} \tag{2.11.20}$$

and, hence,

$$\begin{aligned} \nabla_a \xi_b &= \nabla_a (\chi^{-1} \chi_b) = \chi^{-1} (\nabla_a \chi_b) - \chi^{-2} \chi_b \nabla_a \chi = \chi^{-1} g_{ab} - \chi^{-2} \chi_b \xi_a \\ &= \chi^{-1} g_{ab} - \chi^{-1} \xi_b \xi_a = \chi^{-1} h_{ab} = \frac{1}{3} \theta h_{ab}, \end{aligned}$$

as expected. Notice also, that if we take $a = \chi$, then $\dot{a} = \xi^a \nabla_a \chi = 1$ by equation (2.11.19) and

$$\frac{1}{3}\theta = \frac{\dot{a}}{a}.$$

This choice of a satisfies Friedmann's equation (2.11.14) with $\rho = 0$ and $k = -1$.

Now consider the hyperboloids in O defined by the condition χ = constant. (See figure 2.11.2 again.) Each is a spacelike hypersurface that is everywhere

---

30. Let $\overset{o}{\xi}{}^a$ be a constant unit timelike field on O that agrees with ξ^a at p, and let σ^a be a constant unit spacelike field that is orthogonal to all three vectors ξ^a, $\overset{1}{\sigma}{}^a$, and $\overset{2}{\sigma}{}^a$ at p. Then the rotation in question is generated by the Killing field $\kappa_b = \epsilon_{abcd}\, \chi^a \overset{o}{\xi}{}^c \sigma^d$ (for either choice of volume element ϵ_{abcd}). Recall equation (2.9.2).

orthogonal to ξ^a. (For if σ^a is a vector at a point of one such hypersurface $\mathcal{S}$ that is tangent to $\mathcal{S}$, then $\sigma^n \nabla_n \chi = 0$ and, therefore, by equation (2.11.19), $\sigma^n \xi_n = \sigma^n \nabla_n \chi = 0$.)

Let $\mathcal{S}$ be one such hyperboloid. Let D be the induced derivative operator on $\mathcal{S}$, and let $\mathcal{R}_{abcd}$ be its associated curvature field. We know from equation (2.11.17) that

$$\mathcal{R}_{abcd} = \frac{1}{\chi^2}\,(h_{ad}\, h_{bc} - h_{ac}\, h_{bd}),$$

since here $a = \chi$ and $k = -1$. Again, we can check this directly. To do so, we first compute the second fundamental form π_{ab} on $\mathcal{S}$. (Recall equation (1.10.16).) Since $h_b{}^n \chi_n = h_b{}^n(\chi\, \xi_n) = \mathbf{0}$, we have

$$\begin{aligned}\pi_{ab} &= h_a{}^m\, h_b{}^n\, \nabla_m\, \xi_n = h_a{}^m\, h_b{}^n\, \nabla_m\, (\chi^{-1} \chi_n) = \chi^{-1}\, h_a{}^m\, h_b{}^n\, \nabla_m\, \chi_n \\ &= \chi^{-1}\, h_a{}^m\, h_b{}^n\, g_{mn} = \chi^{-1}\, h_{ab}.\end{aligned}$$

It follows, by equation (1.10.22), that

$$\mathcal{R}_{abcd} = \pi_{ad}\, \pi_{bc} - \pi_{ac}\, \pi_{bd} = \frac{1}{\chi^2}\,(h_{ad}\, h_{bc} - h_{ac}\, h_{bd}),$$

as expected.

Thus, if $\mathcal{S}$ is characterized by the value χ, then $(\mathcal{S}, -h_{ab})$ is a three-dimensional manifold with constant curvature $-1/\chi^2$. Moreover, as we know from our discussion above, it cannot be just any such manifold, but must be, in fact, isometric to three-dimensional hyperbolic space H^3. If we had started with a three-dimensional version of Minkowski spacetime, our hyperboloid (with induced metric) would be isometric to two-dimensional hyperbolic space, otherwise known as the "Lobatchevskian plane." (For more about this "hyperboloid model" for Lobatchevskian plane geometry see, e.g., Reynolds [52].)

Finally, recall the remarks we made in section 2.7 about the cosmological constant Λ. If we include the constant in Einstein's equation—i.e., if we take the latter to be equation (2.7.4)—then Raychaudhuri's equation (2.8.17) yields

$$\dot{\theta} = -4\,\pi\,(\rho + 3\,p) - \frac{1}{3}\,\theta^2 + \Lambda \tag{2.11.21}$$

rather than equation (2.11.9). This, in turn, leads to Friedmann's equation in the form

$$\left(\frac{\dot{a}}{a}\right)^2 - \frac{8\,\pi}{3}\,\rho = -\frac{k}{a^2} + \frac{\Lambda}{3} \tag{2.11.22}$$

rather than equations (2.11.14).

Equation (2.11.21) serves to explain Einstein's introduction of the cosmological constant. He thought he needed to find a non-expanding model ($\theta = 0$) to represent the universe properly. And, in our terms, he was considering only Friedmann spacetimes and only perfect fluid sources that are pressureless ($p = 0$) and non-trivial ($\rho > 0$). It is an immediate consequence of equation (2.11.21) that these conditions can be satisfied if, but only if, $\Lambda = 4\pi\rho > 0$. And in this case, it follows from equations (2.11.22) and (2.11.11) that $k/a^2 = 4\pi\rho$. So (since k is normalized to be 1, 0, or -1), we see that the stated conditions can be satisfied iff

$$\begin{aligned} \Lambda &= 4\pi\rho > 0, \\ \frac{1}{a^2} &= 4\pi\rho, \\ k &= 1. \end{aligned}$$

(These conditions characterize *Einstein static spacetime*.)

It is also an immediate consequence of equation (2.11.21)—at least, if our universe can be represented as a Friedmann spacetime—that evidence for an accelerating rate of cosmic expansion ($\dot{\theta} > 0$) counts as evidence either for a positive value for Λ or for a violation of the strong energy condition. (Recall from problem 2.7.3 that a perfect fluid satisfies the strong energy condition iff $\rho + p \geq 0$ and $\rho + 3\,p \geq 0$.)

3

SPECIAL TOPICS

3.1. Gödel Spacetime

Kurt Gödel is, of course, best known for his work in mathematical logic and the foundations of mathematics. But in the late 1940s he made an important contribution to relativity theory by finding a new solution to Einstein's equation (Gödel [25]). It represents a possible universe with remarkable properties. For one thing, the entire material content of the Gödel universe (on a cosmological scale) is in a state of uniform, rigid rotation. For another, light rays and free test particles in it exhibit a kind of boomerang effect. Most striking of all, the Gödel universe allows for the possibility of "time travel" in a certain interesting sense.[1]

Though not a live candidate for describing our universe (the real one), Gödel's solution is of interest because of what it tells us about the possibilities allowed by relativity theory. In this section, we present the solution and establish several of its basic properties in a running list. We shall later use it as an example when we consider orbital rotation in section 3.2.

It will be helpful to keep in mind two different coordinate expressions for the Gödel metric and also a coordinate-free characterization. We start with the former. Let us officially take Gödel spacetime to be the pair (M, g_{ab}), where M is the manifold $\mathbb{R}^4$ and where

$$g_{ab} = \mu^2 \left[(d_a t)(d_b t) - (d_a x)(d_b x) + \frac{e^{2x}}{2}(d_a y)(d_b y) \right.$$
$$\left. - (d_a z)(d_b z) + 2\, e^x (d_{(a} t)(d_{b)} y) \right]. \tag{3.1.1}$$

1. In addition to finding this one new exact solution to Einstein's equation, Gödel [26] also established the existence of solutions representing universes that are rotating *and* expanding, though he did not exhibit any of the latter explicitly. For a review of Gödel's contributions to relativity theory and cosmology (and subsequent work on rotating solutions), see Ellis [18].

Here μ is an arbitrary positive number (a scale factor), and t, x, y, z are global coordinates on M.[2]

In what follows, we use the abbreviations

$$t^a = \left(\frac{\partial}{\partial t}\right)^a \quad x^a = \left(\frac{\partial}{\partial x}\right)^a \quad y^a = \left(\frac{\partial}{\partial y}\right)^a \quad z^a = \left(\frac{\partial}{\partial z}\right)^a. \tag{3.1.2}$$

To confirm that g_{ab} is a metric of signature $(1, 3)$, it suffices to check that the fields

$$\frac{t^a}{\mu}, \qquad \frac{x^a}{\mu}, \qquad \frac{\sqrt{2}}{\mu}(t^a - e^{-x} y^a), \qquad \frac{z^a}{\mu} \tag{3.1.3}$$

form an orthonormal basis (of the appropriate type) at each point. The first, in particular, is a smooth, unit timelike vector field on M. That there *exists* such a field shows us that Gödel spacetime is temporally orientable. It is also orientable since the anti-symmetrized product of the four fields in equation (3.1.3) qualifies as a volume element.

(1) *Gödel spacetime is temporally orientable and orientable.*

We shall work with the temporal orientation determined by t^a in what follows.

We note for future reference that the inverse field of g_{ab} is

$$g^{bc} = \frac{1}{\mu^2}\left[-t^b t^c - x^b x^c - 2\, e^{-2x}\, y^b y^c - z^b z^c + 4\, e^{-x}\, t^{(b} y^{c)}\right], \tag{3.1.4}$$

and that lowering indices in equation (3.1.2) with g_{ab} yields the following:

$$t_a = \mu^2(\nabla_a\, t + e^x \nabla_a\, y), \tag{3.1.5}$$

$$x_a = -\mu^2\, \nabla_a\, x, \tag{3.1.6}$$

$$y_a = \mu^2\left(\frac{e^{2x}}{2}\nabla_a\, y + e^x \nabla_a\, t\right), \tag{3.1.7}$$

$$z_a = -\mu^2\, \nabla_a\, z. \tag{3.1.8}$$

(Here ∇ is the derivative operator on M compatible with g_{ab}, and we have switched from writing, for example, "$d_a t$" to "$\nabla_a t$".)

2. More precisely, t, x, y, z are real-valued functions on M, and the composite map $\Phi\colon p \mapsto (t(p), x(p), y(p), z(p))$ is a bijection between M and $\mathbb{R}^4$ that belongs to the collection $\mathcal{C}$ of 4-charts that defines the manifold $\mathbb{R}^4$. The coordinates t, x, y, z correspond to u^1, u^2, u^3, u^4 in the notation of section 1.2. So, for example, we understand the vector $\left(\frac{\partial}{\partial t}\right)^a$ at any point p to be the tangent there to the curve $r \mapsto \Phi^{-1}(t(p) + r, x(p), y(p), z(p))$.

We claim, first, that the four fields

$$t^a, \qquad \zeta^a = (x^a - y\,y^a), \qquad y^a, \qquad z^a \tag{3.1.9}$$

are all Killing fields. They are, in fact, the generators, respectively, of one-parameter (global) isometry groups $\{\overset{t}{\Gamma}_r\}_{r\in\mathbb{R}}$, $\{\overset{\zeta}{\Gamma}_r\}_{r\in\mathbb{R}}$, $\{\overset{y}{\Gamma}_r\}_{r\in\mathbb{R}}$, $\{\overset{z}{\Gamma}_r\}_{r\in\mathbb{R}}$ on M defined by

$$\overset{t}{\Gamma}_r(p) = \Phi^{-1}\left(t(p) + r,\ x(p),\ y(p),\ z(p)\right),$$

$$\overset{\zeta}{\Gamma}_r(p) = \Phi^{-1}(t(p),\ x(p) + r,\ e^{-r}y(p),\ z(p)),$$

$$\overset{y}{\Gamma}_r(p) = \Phi^{-1}(t(p),\ x(p),\ y(p) + r,\ z(p)),$$

$$\overset{z}{\Gamma}_r(p) = \Phi^{-1}(t(p),\ x(p),\ y(p),\ z(p) + r),$$

where $\Phi\colon M \to \mathbb{R}^4$ is the chart defined by $\Phi(p) = (t(p), x(p), y(p), z(p))$. Here, of course, the group operation is composition.[3] An equivalent formulation may be more transparent. For example, we can understand $\overset{\zeta}{\Gamma}_r$ to be defined by the requirement that, for all numbers t_0, x_0, y_0, z_0,

3. There are a few things that have to be checked. First, each of these maps (for any choice of r) *is*, in fact, an isometry. This follows from basic facts we have recorded in section 1.5. Consider $\overset{\zeta}{\Gamma}_r$, for example. By equations (1.5.6) and (1.5.7), we have $(\overset{\zeta}{\Gamma}_r)^*(e^{2x}) = e^{2(x+r)}$ and $(\overset{\zeta}{\Gamma}_r)^*(d_a y) = d_a((\overset{\zeta}{\Gamma}_r)^*(y)) = d_a(e^{-r}y) = e^{-r}(d_a y)$. Hence,

$$(\overset{\zeta}{\Gamma}_r)^*(e^{2x}(d_a y)(d_b y)) = ((\overset{\zeta}{\Gamma}_r)^*(e^{2x}))\ ((\overset{\zeta}{\Gamma}_r)^*(d_a y))\ ((\overset{\zeta}{\Gamma}_r)^*(d_b y)) = e^{2x}(d_a y)(d_b y).$$

Arguing in this way, we can show that *all* the terms in g_{ab} are preserved by $(\overset{\zeta}{\Gamma}_r)^*$ and, so, $(\overset{\zeta}{\Gamma}_r)^*(g_{ab}) = g_{ab}$. Second, each of the groups *does*, in fact, have the indicated vector field as its generator. This follows from our discussion in sections 1.2 and 1.3. Consider $\{\overset{\zeta}{\Gamma}_r\}_{r\in\mathbb{R}}$, for example. Let p be a point with coordinates $\Phi(p) = (t_0, x_0, y_0, z_0)$, and let $\gamma : \mathbb{R} \to M$ be the curve through p defined by

$$\gamma(r) = \overset{\zeta}{\Gamma}_r(p) = \Phi^{-1}(t_0,\ x_0 + r,\ e^{-r}y_0,\ z_0).$$

We need to show that $\overset{\rightarrow}{\gamma}{}^a = \zeta^a$ at all points on the image of γ. Let f be any smooth field on some open set containing p. Then, by the chain rule, at all points $\gamma(r)$,

$$\overset{\rightarrow}{\gamma}{}^a(f) = \frac{d}{dr}(f\circ\gamma) = \frac{d}{dr}(f\circ\Phi^{-1})(t_0,\ x_0 + r,\ e^{-r}y_0,\ z_0) = \frac{\partial(f\circ\Phi^{-1})}{\partial x^2}\cdot 1 + \frac{\partial(f\circ\Phi^{-1})}{\partial x^3}\cdot(-e^{-r}y_0)$$
$$= \frac{\partial f}{\partial x}\cdot 1 + \frac{\partial f}{\partial y}\cdot(-e^{-r}y_0) = \frac{\partial f}{\partial x} - y\frac{\partial f}{\partial y} = \left((\frac{\partial}{\partial x})^a - y(\frac{\partial}{\partial y})^a\right)(f) = \zeta^a(f).$$

So we are done. Here x^1, x^2, x^3, x^4 are the coordinate projection functions on $\mathbb{R}^4$ that we considered in section 1.2. So, for example, $(x^3\circ\Phi)(p) = y(p)$. And the equality $\dfrac{\partial f}{\partial y} = \dfrac{\partial(f\circ\Phi^{-1})}{\partial x^3}$ is an instance of equation (1.2.7). (As mentioned in the preceding note, the coordinates t, x, y, z correspond to u^1, u^2, u^3, and u^4 in the notation of section 1.2.)

$$(\Phi \circ \overset{\zeta}{\Gamma}_r \circ \Phi^{-1})\,(t_0,\ x_0,\ y_0,\ z_0) = (t_0,\ x_0 + r,\ e^{-r} y_0,\ z_0).$$

The field x^a is not a Killing field, but it is the generator of the one-parameter group of diffeomorphisms $\{\overset{x}{\Gamma}_r\}_{r\in\mathbb{R}}$ on M given by

$$\overset{x}{\Gamma}_r\,(p) = \Phi^{-1}(t(p),\ x(p) + r,\ y(p),\ z(p)).$$

The five fields under consideration satisfy the following Lie bracket relations:

$$[t^a,\ \zeta^a] = [t^a,\ y^a] = [t^a,\ z^a] = [\zeta^a,\ z^a] = [y^a,\ z^a] = \mathbf{0}, \tag{3.1.10}$$

$$[x^a,\ t^a] = [x^a,\ \zeta^a] = [x^a,\ y^a] = [x^a,\ z^a] = \mathbf{0}, \tag{3.1.11}$$

$$[\zeta^a,\ y^a] = y^a. \tag{3.1.12}$$

There are various ways to see why these hold. For those in the first two rows, it is easiest to invoke a basic result (that we did not formulate in chapter 1).

PROPOSITION 3.1.1. *Let α^a and β^a be smooth fields on a manifold that generate one-parameter groups of diffeomorphisms $\{\overset{\alpha}{\Gamma}_r\}_{r\in\mathbb{R}}$ and $\{\overset{\beta}{\Gamma}_r\}_{r\in\mathbb{R}}$ on that manifold. Then $[\alpha^a,\ \beta^a] = \mathbf{0}$ iff $\overset{\alpha}{\Gamma}_r$ and $\overset{\beta}{\Gamma}_{r'}$ commute for all r and r'.*

(See, for example, Spivak [57, volume 1, p. 217].) It is clear in each case that the relevant commutation relations obtain; e.g., $\overset{t}{\Gamma}_r$ and $\overset{\zeta}{\Gamma}_s$ commute for all r and s.[4] For equation (3.1.12), note that

$$\begin{aligned}[\zeta^a,\ y^a] &= -[y^a,\ \zeta^a] = -\pounds_{y^a}(x^a - y\,y^a)\\ &= [x^a,\ y^a] + (\pounds_{y^a} y)\,y^a + y\,[y^a, y^a] = y^a,\end{aligned}$$

since $\pounds_{y^a} y = y^n \nabla_n y = 1$, and $[x^a,\ y^a] = [y^a, y^a] = \mathbf{0}$.

By composing the isometries $\overset{t}{\Gamma}_r$, $\overset{\zeta}{\Gamma}_r$, $\overset{y}{\Gamma}_r$, and $\overset{z}{\Gamma}_r$ (with appropriate choices for r in each case), we can go from any one point in M to any other. Moreover,

---

4. For all p, r, and s, we have

$$\begin{aligned}\overset{t}{\Gamma}_r\,(\overset{\zeta}{\Gamma}_s\,(p)) &= \Phi^{-1}\left(t(\overset{\zeta}{\Gamma}_s\,(p)) + r,\ x(\overset{\zeta}{\Gamma}_s\,(p)),\ y(\overset{\zeta}{\Gamma}_s\,(p)),\ z(\overset{\zeta}{\Gamma}_s\,(p))\right)\\ &= \Phi^{-1}\left(t(p) + r,\ x(p) + s,\ e^{-s} y(p),\ z(p)\right).\end{aligned}$$

And a similar computation shows that

$$\overset{\zeta}{\Gamma}_s(\overset{t}{\Gamma}_r\,(p)) = \Phi^{-1}\left(t(p) + r,\ x(p) + s,\ e^{-s} y(p),\ z(p)\right).$$

So $\overset{t}{\Gamma}_r\,(\overset{\zeta}{\Gamma}_s\,(p)) = \overset{\zeta}{\Gamma}_s\,(\overset{t}{\Gamma}_r\,(p))$.

each of the individual isometries, and so any composition of them, preserves the fields t^a and z^a. (This follows from propositions 1.6.6 and 1.6.4, and the fact that each of the generators t^a, ζ^a, y^a, z^a has a vanishing Lie bracket with t^a and z^a.) So we have the following homogeneity claim.

(2) *Gödel spacetime is (globally) homogeneous in this strong sense: given any two points p and q in M, there is an isometry $\psi : M \to M$ such that $\psi(p) = q$, $\psi_*(t^a) = t^a$, and $\psi_*(z^a) = z^a$.*

(The maps referred to here preserve temporal orientation automatically because they preserve t^a, and we are using that field to define temporal orientation.) We shall repeatedly invoke this strong form of homogeneity in what follows. For example, we shall prove an assertion about a particular integral curve of t^a (that makes reference only to g_{ab}, t^a, and z_a), and then claim that it necessarily holds for all integral curves of that field.

The four Killing fields t^a, ζ^a, y^a, z^a are clearly independent of each other. In fact, one can find a fifth that is independent of these four; e.g.,

$$\begin{aligned}\kappa^a &= -2\,e^{-x}\,t^a + y\,x^a + \left(e^{-2x} - \frac{1}{2}\,y^2\right) y^a \\ &= -2\,e^{-x}\,t^a + y\,\zeta^a + \left(e^{-2x} + \frac{1}{2}\,y^2\right) y^a.\end{aligned}$$

(To confirm that it is a Killing field, it suffices to expand $\nabla_a \kappa_b$ and use our expressions above for t_a, x_a, and y_b to show that its symmetric part vanishes.[5])

Now we do a bit of calculation and derive an expression for the Ricci tensor field R_{ab}. Note first that

5. We have

$$\begin{aligned}\nabla_a \kappa_b = &-2\,e^{-x}\,\nabla_a t_b + 2\,e^{-x}\,(\nabla_a x)\,t_b + y\,\nabla_a \zeta_b + (\nabla_a y)\,\zeta_b + \left(e^{-2x} + \frac{1}{2}\,y^2\right)\nabla_a y_b \\ &-2\,e^{-2x}\,(\nabla_a x)\,y_b + y\,(\nabla_a y)\,y_b \\ = &-2\,e^{-x}\,\nabla_a t_b + y\,\nabla_a \zeta_b + \left(e^{-2x} + \frac{1}{2}\,y^2\right)\nabla_a y_b \\ &+ (\nabla_a x)(2\,e^{-x}\,t_b - 2\,e^{-2x}\,y_b) + (\nabla_a y)(\zeta_b + y\,y_b).\end{aligned}$$

But $(2\,e^{-x}\,t_b - 2\,e^{-2x}\,y_b) = \mu^2\,\nabla_b\,y$ and $(\zeta_b + y\,y_b) = -\mu^2\,\nabla_b\,x$. And the first three terms have vanishing symmetric part since t^a, ζ^a, y^a are Killing fields. So $\nabla_{(a}\,\kappa_{b)} = \mathbf{0}$.

$$\nabla_a\, t_b = \mu^2\, e^x\, (\nabla_{[a}x)(\nabla_{b]}y), \tag{3.1.13}$$

$$\nabla_a\, x_b = \mu^2 \left[\frac{e^{2x}}{2} (\nabla_a y)(\nabla_b y) + e^x (\nabla_{(a} y)(\nabla_{b)} t) \right], \tag{3.1.14}$$

$$\nabla_a\, y_b = \mu^2 \left[e^{2x} (\nabla_{[a} x)(\nabla_{b]} y) + e^x (\nabla_{[a} x)(\nabla_{b]} t) \right], \tag{3.1.15}$$

$$\nabla_a\, z_b = \mathbf{0}. \tag{3.1.16}$$

These can be checked easily by using equations (3.1.5)–(3.1.8) and the fact that t^a, ζ^a, y^a, z^a are Killing fields. Since t^a is a Killing field, for example, we have $\nabla_{(a}\, t_{b)} = \mathbf{0}$ and, therefore,

$$\begin{aligned}\nabla_a\, t_b = \nabla_{[a}\, t_{b]} &= \mu^2 \left(\nabla_{[a}\nabla_{b]} t + e^x\, \nabla_{[a}\nabla_{b]} y + e^x\, (\nabla_{[a} x)(\nabla_{b]} y) \right) \\ &= \mu^2\, e^x\, (\nabla_{[a} x)(\nabla_{b]} y).\end{aligned}$$

This gives us equation (3.1.13). The other cases are handled similarly.[6] It follows immediately that t^a, x^a and z^a are all geodesic fields:

$$t^a \nabla_a\, t^b = \mathbf{0} \qquad x^a \nabla_a\, x^b = \mathbf{0} \qquad z^a \nabla_a\, z^b = \mathbf{0}. \tag{3.1.17}$$

We shall be particularly interested in the (maximally extended) integral curves of t^a. Their images are sets of the form $\{\Phi^{-1}(t, x_0, y_0, z_0) : t \in \mathbb{R}\}$, for particular choices of x_0, y_0, z_0. We shall call these curves (or their images) *t-lines*.

Now we turn to R_{ab}. We claim, first, that symmetry considerations alone establish that it must have the form

$$R_{ab} = \alpha\, \hat{t}_a \hat{t}_b + \beta\, (g_{ab} - \hat{t}_a \hat{t}_b - \hat{z}_a \hat{z}_b) \tag{3.1.18}$$

where α and β are particular numbers (to be determined), and $\hat{t}^a$ and $\hat{z}^a$ are normalized versions of t^a and z^a. (So $t^a = \mu\, \hat{t}^a$ and $z^a = \mu\, \hat{z}^a$.) The argument we use to establish this is much like that used in section 2.11 when we considered the Ricci tensor field in Friedmann spacetimes. In both cases, it turns on an isotropy condition. Shortly, when we switch to an alternate coordinate representation of the Gödel metric, it will be clear that given any t-line (through any point), there is a global isometry (a rotation) that leaves fixed every point on the line and also preserves the field z^a. In effect, we now make use of that rotational symmetry, but cast the argument in terms of Killing fields rather than of the rotations themselves.

6. For equation (3.1.14), note that since $\nabla_{[a}\, x_{b]} = -\mu^2\, \nabla_{[a}\nabla_{b]} x = \mathbf{0}$, and since $(x^a - y\, y^a)$ and y^a are Killing fields,

$$\nabla_a\, x_b = \nabla_{(a}\, x_{b)} = y\, \nabla_{(a} y_{b)} + (\nabla_{(a} y)\, y_{b)} = (\nabla_{(a} y)\, y_{b)} = \mu^2 \left[\frac{e^{2x}}{2} (\nabla_a y)(\nabla_b y) + e^x (\nabla_{(a} y)(\nabla_{b)} t) \right].$$

Since we can find an isometry that maps any one point in M to any other and preserves both t^a and z^a, it will suffice to show that equation (3.1.18) holds at one point, say $p = \Phi^{-1}(0, 0, 0, 0)$. To do so, it will suffice to show, in turn, that the two sides of equation (3.1.18) yield the same result when contracted with each of the vectors t^a, x^a, $(t^a - y^a)$, z^a. (It is convenient to work with this basis because the vectors are mutually orthogonal at p. It does not matter that they are not normalized.) So our task reduces to showing that the following all hold at p (for some values of α and β):

$$\text{(i) } R_{ab}\, t^a = \alpha\, t_b, \ \text{(ii) } R_{ab}\, x^a = \beta\, x_b,$$

$$\text{(iii) } R_{ab}\, (t^a - y^a) = \beta\, (t_b - y_b), \ \text{and (iv) } R_{ab}\, z^a = \mathbf{0}.$$

Given any Killing field λ^a in any spacetime, we have

$$R_{ab}\, \lambda^a = R^n{}_{abn}\, \lambda^a = -R^a{}_{nb}{}^n\, \lambda_a = \nabla_n\, \nabla_b\, \lambda^n. \tag{3.1.19}$$

(The second equality follows from the symmetries of the Riemann curvature tensor field, and the third follows from proposition 1.9.8.) So, in particular, applying this result to the Killing field z^a in Gödel spacetime, and recalling equation (3.1.16), we have $R_{ab}\, z^a = \nabla_n\, \nabla_b\, z^n = \mathbf{0}$. This gives us (iv).

Next, consider the field

$$\kappa'^a = -2\,(e^{-x} - 1)\, t^a + y\, x^a + \left(e^{-2x} - \frac{1}{2}\, y^2 - 1\right) y^a. \tag{3.1.20}$$

It is a linear combination of Killing fields ($\kappa'^a = \kappa^a + 2\, t^a - y^a$) and so is, itself, a Killing field. What is important about it is that it vanishes at p.[7] Notice that we have

$$[t^a,\, \kappa'^a] = [z^a,\, \kappa'^a] = \mathbf{0}, \tag{3.1.21}$$

$$[x^a,\, \kappa'^a] = 2\, e^{-x}\, t^a - 2\, e^{-2x}\, y^a, \tag{3.1.22}$$

$$[y^a,\, \kappa'^a] = x^a - y\, y^a \tag{3.1.23}$$

everywhere,[8] and so

$$\pounds_{\kappa'}\, x^a = [\kappa'^a,\, x^a] = -2(t^a - y^a), \tag{3.1.24}$$

$$\pounds_{\kappa'}\, y^a = [\kappa'^a,\, y^a] = -x^a \tag{3.1.25}$$

---

7. It is, in fact, up to a constant, just the rotational Killing field $(\partial/\partial\phi)^a$ that we shall consider below. The latter, as we shall see, generates a one-parameter group of rotations that keep fixed all points on the t-line through p (and preserve z^a).

8. These all follow easily from the Lie bracket relations that we have already established.

at p. Since κ'^a vanishes at p, we have $\pounds_{\kappa'} f = \kappa'^a \nabla_a f = 0$ at p for *all* smooth scalar fields f. So, in particular, since κ' Lie derives R_{ab} (as all Killing fields do) and Lie derives t^a (by equation (3.1.21)), we have

$$0 = \pounds_{\kappa'}(R_{ab}\, t^a y^b) = R_{ab}\, t^a\, (\pounds_{\kappa'} y^b) = -R_{ab}\, t^a x^b, \tag{3.1.26}$$

$$0 = \pounds_{\kappa'}(R_{ab}\, t^a x^b) = R_{ab}\, t^a\, (\pounds_{\kappa'} x^b) = -2\, R_{ab}\, t^a (t^b - y^b) \tag{3.1.27}$$

at p. These two, together with (iv), show that $R_{ab}\, t^a$ must be proportional to t_b at p, which is what we need for (i). Similarly, we have

$$0 = \pounds_{\kappa'}(R_{ab}\, y^a y^b) = R_{ab}\, \pounds_{\kappa'}(y^a y^b) = 2\, R_{ab}\, y^a\, (\pounds_{\kappa'} y^b) = -2\, R_{ab}\, y^a\, x^b$$

at p. This, together with equation (3.1.26) and (iv), shows that (ii) must hold for some β. Finally, (iii) follows from (ii). For if $R_{ab}\, x^a = \beta\, x_b$, then

$$\begin{aligned} -2\, R_{ab}\, (t^a - y^a) &= R_{ab}\, \pounds_{\kappa'} x^a = \pounds_{\kappa'}(R_{ab}\, x^a) = \pounds_{\kappa'}(\beta\, x_b) = \beta\, \pounds_{\kappa'} x_b \\ &= -2\, \beta\, (t_b - y_b). \end{aligned}$$

(For the final equality, we use the fact that $\pounds_{\kappa'}\, g_{ab} = \mathbf{0}$ and, so, $\pounds_{\kappa'}\, x_b = \pounds_{\kappa'}(g_{ab}\, x^a) = g_{ab}\, \pounds_{\kappa'} x^a = -2\, g_{ab}\, (t^a - y^a) = -2\, (t_b - y_b)$.)

Now it remains only to compute α and β in equation (3.1.18). It follows from equation (3.1.17)—and from equation (3.1.19) as applied to the Killing fields t^a and x^a—that

$$\begin{aligned} \alpha &= R_{ab}\, \hat{t}^a \hat{t}^b = \mu^{-2}\, R_{ab}\, t^a t^b = \mu^{-2}\, t^b\, \nabla_n\, \nabla_b\, t^n \\ &= \mu^{-2} [\nabla_n\, (t^b \nabla_b\, t^n) - (\nabla_n\, t^b)(\nabla_b\, t^n)] = -\mu^{-2}\, (\nabla_n\, t^b)(\nabla_b\, t^n) \end{aligned}$$

and (by the same argument)

$$\beta = -\mu^{-2}\, (\nabla_n\, x^b)(\nabla_b\, x^n).$$

Now, raising indices in equations (3.1.13) and (3.1.14), using equation (3.1.4), yields

$$\begin{aligned} \nabla_n\, t^b &= \frac{e^x}{2}\, x^b (\nabla_n\, y) + (-e^{-x}\, y^b + t^b)(\nabla_n\, x), \\ \nabla_n\, x^b &= \frac{e^x}{2}\, t^b (\nabla_n\, y) + (-e^{-x}\, y^b + t^b)(\nabla_n\, t). \end{aligned}$$

It follows that

$$(\nabla_n\, t^b)(\nabla_b\, t^n) = -1 \qquad (\nabla_n\, x^b)(\nabla_b\, x^n) = 0 \tag{3.1.28}$$

and, therefore, $\alpha = \mu^{-2}$ and $\beta = 0$. Thus we have

$$R_{ab} = \mu^{-2}\, \hat{t}_a\, \hat{t}_b. \tag{3.1.29}$$

So $R = \mu^{-2}$ and

$$R_{ab} - \frac{1}{2} R\, g_{ab} = \mu^{-2}\, \hat{t}_a\, \hat{t}_b - \frac{\mu^{-2}}{2} g_{ab} = \frac{\mu^{-2}}{2} \left(\hat{t}_a\, \hat{t}_b - (g_{ab} - \hat{t}_a\, \hat{t}_b) \right).$$

Therefore,

(3) *Gödel spacetime is a solution to Einstein's equation (without cosmological constant)*

$$R_{ab} - \frac{1}{2} R\, g_{ab} = 8\,\pi \left(\rho\, \hat{t}_a\, \hat{t}_b - p\, (g_{ab} - \hat{t}_a\, \hat{t}_b) \right)$$

for a perfect fluid with four-velocity $\hat{t}^a$*, mass-density* $\rho = 1/(16\,\pi\,\mu^2)$*, and pressure* $p = 1/(16\,\pi\,\mu^2)$*. (Equivalently, it is a solution to Einstein's equation with cosmological constant* $\lambda = -1/(2\,\mu^2)$

$$R_{ab} - \frac{1}{2} R\, g_{ab} - \lambda\, g_{ab} = 8\,\pi \rho'\, \hat{t}_a\, \hat{t}_b$$

for a dust field with mass-density $\rho' = 1/(8\,\pi\,\mu^2)$*.)*

Recall that a perfect fluid satisfies the dominant energy condition iff $|p| \leq \rho$. So if we construe Gödel spacetime as a perfect fluid solution to Einstein's equation without cosmological constant, the perfect fluid in question is only "borderline" for satisfying the condition.

Let us further consider the normalized field $\hat{t}^a = t^a/\mu$, which we now understand to represent the four-velocity of the background source fluid. We know that its associated expansion field θ_{ab} vanishes (because it is a Killing field), as does its acceleration (by equation (3.1.17)). Let us now compute its associated rotation field ω^a.

Let ϵ^{abcd} be a volume element on M. (We know that volume elements exist since, e.g., $t^{[a}\, x^b\, y^c\, z^{d]}$ is an anti-symmetric field on M that is everywhere non-vanishing. We need only normalize it to obtain a volume element.) The field $\nabla_a\, t_b$ is anti-symmetric, and it is orthogonal to both t^a and z^a (by equation (3.1.13)). So we can express it in the form

$$\nabla_a\, t_b = f\, \epsilon_{abcd}\, t^c\, z^d$$

for some field f. To determine f, we need only contract each side with itself and make use of equation (3.1.28):

$$\begin{aligned} 1 = (\nabla_a\, t_b)\, (\nabla^a\, t^b) &= f^2\, \epsilon_{abcd}\, t^c\, z^d\, \epsilon^{abmn}\, t_m\, z_n \\ &= -4 f^2\, \delta^{[m}{}_c\, \delta^{n]}{}_d\, t^c\, z^d\, t_m\, z_n \\ &= -2 f^2\, (t^m\, z^n - t^n\, z^m)\, t_m\, z_n = 2\,\mu^4 f^2. \end{aligned}$$

Taking f to be positive—we can always switch from the volume element ϵ_{abcd} to $-\epsilon_{abcd}$ if necessary—we have

$$\nabla_a\, t_b = \frac{1}{\sqrt{2}\,\mu^2}\,\epsilon_{abcd}\, t^c\, z^d.$$

Hence, using this volume element to compute the rotation vector field,

$$\begin{aligned}
\omega^a &= \frac{1}{2}\,\epsilon^{abcd}\,\hat{t}_b\,\nabla_c\,\hat{t}_d = \frac{1}{2\,\mu^2}\,\epsilon^{abcd}\,t_b\,\nabla_c\,t_d \\
&= \frac{1}{2\sqrt{2}\,\mu^4}\,\epsilon^{abcd}\,t_b\,\epsilon_{cdmn}\,t^m\,z^n = \frac{-4}{2\sqrt{2}\,\mu^4}\,\delta^{[a}{}_m\,\delta^{b]}{}_n\,t_b\,t^m\,z^n \\
&= \frac{1}{\sqrt{2}\,\mu^4}\,(t^b t_b)\,z^a = \frac{1}{\sqrt{2}\,\mu^2}\,z^a.
\end{aligned} \tag{3.1.30}$$

Let us record this result too.

(4) *The four-velocity $\hat{t}^a$ in Gödel spacetime is expansion free ($\theta = 0$), shear free ($\sigma_{ab} = \mathbf{0}$), and geodesic ($\hat{t}^n\nabla_n\hat{t}^a = \mathbf{0}$), but its rotation field ω^a is non-vanishing and constant ($\nabla_a\,\omega^b = \mathbf{0}$). Indeed, ω^a is just $(1/\sqrt{2}\,\mu^2)\,z^a$. The Gödel universe is thus in a state of uniform, rigid rotation.*

It turns out that there are only two homogeneous perfect fluid solutions in which (i) the mass density is non-zero, (ii) the fluid four-velocity is expansion free, shear free, and geodesic, and (iii) the underlying manifold is simply connnected,[9] namely the Einstein static universe (Hawking and Ellis [30]) and Gödel spacetime. (Gödel asserted this result, without proof, in [25]. Proofs can be found in Ozsváth [48] and Farnsworth and Kerr [19].) So Gödel spacetime itself is picked out if one adds the requirement that (iv) the rotation field of the fluid is non-vanishing.

We next want to establish the existence of closed timelike curves in Gödel spacetime and characterize its timelike and null geodesics. To do so, it will be convenient to switch to a different coordinate representation of the metric. This one, cast in terms of a cylindrical coordinate system $\tilde{t}, r, \phi, \tilde{z}$, makes manifest the rotational symmetry of Gödel spacetime about a particular axis, but hides its homogeneity:

$$\begin{aligned}
g_{ab} = 4\,\mu^2 \Big[&(d_a\tilde{t})(d_b\tilde{t}) - (d_a r)(d_b r) - (d_a\tilde{z})(d_b\tilde{z}) \\
&+ (sh^4\, r - sh\, r^2)(d_a\phi)(d_b\phi) + 2\sqrt{2}\, sh^2 r\,(d_{(a}\tilde{t})(d_{b)}\phi) \Big].
\end{aligned} \tag{3.1.31}$$

(Here we write "ch" and "sh" for "cosh" and "sinh" respectively.)

9. The third condition is needed to rule out further examples that can be generated by identifying points.

We have to be a bit careful here as to what we mean by a "coordinate system." We are not quite talking about a 4-chart in the sense of section 1.1. Here is a more precise formulation. Let A be the "axis set" consisting of all points in M of the form $\Phi^{-1}(t, 0, 0, z)$, and let M^- be the excised set $M - A$. We claim that there exist smooth maps

$$\tilde{t} \colon M \to \mathbb{R}, \quad r \colon M^- \to \mathbb{R}^+, \quad \phi \colon M^- \to S^1, \quad \tilde{z} \colon M \to \mathbb{R} \tag{3.1.32}$$

such that the composite map

$$\Delta \colon M^- \to \mathbb{R} \times \mathbb{R}^+ \times S^1 \times \mathbb{R} \tag{3.1.33}$$

determined by the rule $q \mapsto (\tilde{t}(q), r(q), \phi(q), \tilde{z}(q))$ is a diffeomorphism and equation (3.1.31) holds on M^-. (Here $\mathbb{R}^+$ is the set of reals that are strictly positive, and S^1 is identified, in the usual way, with $\mathbb{R}$ mod 2π.) Under these conditions, we can define coordinate vector fields $(\partial/\partial\tilde{t})^a$, $(\partial/\partial r)^a$, $(\partial/\partial\phi)^a$, $(\partial/\partial\tilde{z})^a$ much as we did in section 1.1.[10] We shall use the following abbreviations for them:

$$\tilde{t}^a = \left(\frac{\partial}{\partial\tilde{t}}\right)^a \quad r^a = \left(\frac{\partial}{\partial r}\right)^a \quad \phi^a = \left(\frac{\partial}{\partial\phi}\right)^a \quad \tilde{z}^a = \left(\frac{\partial}{\partial\tilde{z}}\right)^a.$$

The radial coordinate r can be extended to a map $r \colon M \to \mathbb{R}^+ \cup \{0\}$ that is, at least, continuous on the axis A.

The relation between the new coordinates and the old is given by the following conditions:

$$e^x = ch\, 2r + (cos\, \phi)(sh\, 2r), \tag{3.1.34}$$

$$y\, e^x = \sqrt{2}\,(sin\, \phi)(sh\, 2r), \tag{3.1.35}$$

$$z = 2\,\tilde{z}, \tag{3.1.36}$$

$$tan\left(\frac{\phi}{2} + \frac{t - 2\,\tilde{t}}{2\sqrt{2}}\right) = e^{-2r}\, tan\, \frac{\phi}{2} \qquad \text{where} \quad \left|\frac{t - 2\,\tilde{t}}{2\sqrt{2}}\right| < \frac{\pi}{2}. \tag{3.1.37}$$

With some work, one can show directly that these conditions do, in fact, properly define smooth maps over the domains indicated in (3.1.32)[11] and use them to derive the expression for g_{ab} given in equation (3.1.31). (The details are worked out with great care in Stein [58].) We skip this work and make just two remarks about the conditions. Later, in an appendix, following Gödel [25], we shall establish the equivalence of the two coordinate representations

10. So, for example, let q be any point in M^-. Then $s \mapsto \Delta^{-1}(\tilde{t}(q), r(q), \phi(q) + s, \tilde{z}(q))$ is a smooth curve through q. We understand $(\partial/\partial\phi)^a$ at q to be the tangent vector to the curve there.

11. Strictly speaking, the conditions define only $\tilde{t}$ on the restricted domain M^-. But it can be smoothly extended to all of M.

a somewhat different way. It will involve a direct appeal to a coordinate-free description of the metric.

First, it is clear from the first two conditions in the list why we need to restrict attention to M^-. If $x = y = 0$, they will be satisfied iff $r = 0$. But if $r = 0$, those conditions impose no constraints on ϕ (and neither do the other conditions). So ϕ is not well defined on $M - M^-$. (On the other hand, if either $x \neq 0$ or $y \neq 0$, then equations (3.1.34) and (3.1.35) determine unique values for *both* ϕ and r.)

Second, though the exact relation between t and $\tilde{t}$ is complex, their associated coordinate fields $\tilde{t}^a$ and t^a are proportional to each other; i.e, we have $\tilde{t}^a = \alpha\, t^a$ for some α. This follows from the first three conditions. For when $r, \phi, \tilde{z}$ are fixed, x, y, z are fixed as well. So every $\tilde{t}$-line (characterized by constant values of $r, \phi, \tilde{z}$) is also a t-line (characterized by constant values for x, y, z). And it follows from equation (3.1.37) that the proportionality factor must be 2.[12] So we have

$$\tilde{t}^a = 2\, t^a. \tag{3.1.38}$$

We also have

$$\tilde{z}^a = 2\, z^a \tag{3.1.39}$$

from equation (3.1.36).

Let us now accept as given the second coordinate representation of the Gödel metric (in terms of cylindrical coordinates). We shall work with it much as we did the first representation. Note that the inverse of the metric now comes out (in M^-) as

$$g^{bc} = \frac{1}{4\,\mu^2}\left[-\frac{(sh^4 r - sh^2 r)}{(sh^4 r + sh^2 r)}\,\tilde{t}^b\,\tilde{t}^c - r^b\,r^c - \tilde{z}^b\,\tilde{z}^c - \frac{1}{(sh^4 r + sh^2 r)}\,\phi^b\,\phi^c + \frac{2\sqrt{2}}{(sh^4 r + sh^2 r)}\,\tilde{t}^{(b}\,\phi^{c)} \right]. \tag{3.1.40}$$

Consider ϕ^a. Since

$$\phi_a\,\phi^a = 4\,\mu^2\,(sh^4 r - sh^2 r), \tag{3.1.41}$$

it qualifies as spacelike, null, or timelike at a point q in M^- depending on whether $r(q)$ is less than, equal to, or greater than the critical value $r_c = \ln (1 + \sqrt{2})$ where sh assumes the value 1. The angular coordinate ϕ is defined

12. It follows from equation (3.1.37), specifically, that the difference $(t - 2\,\tilde{t})$ is constant on every $\tilde{t}$-line; i.e., once r and ϕ are fixed, $(t - 2\,\tilde{t})$ is fixed as well. So $\tilde{t}^n \nabla_n (t - 2\,\tilde{t}) = 0$. It follows that $\alpha = \alpha\, t^n \nabla_n t = \tilde{t}^n \nabla_n t = \tilde{t}^n \nabla_n (2\,\tilde{t}) = 2$.

only on M^-, but we can smoothly extend ϕ^a itself to all of M by taking it to be the zero vector on $M - M^-$—i.e., on the axis A. We shall understand it that way in what follows. Where ϕ^a *is* timelike (and where it is null but non-zero) it qualifies as future-directed, because temporal orientation is determined by t^a (or, equivalently, $\tilde{t}^a$), and $\tilde{t}^a \phi_a = 4\sqrt{2}\,\mu^2\, sh^2 r$. (So $\tilde{t}^a \phi_a > 0$, unless $r = 0$.)

Both $\tilde{t}^a$ and $\tilde{z}^a$ are, of course, Killing fields. We know that from before. So is ϕ^a. It is the generator of a one-parameter family of (global) isometries $\{\overset{\phi}{\Gamma}_s\}_{s \in S^1}$ defined by

$$\overset{\phi}{\Gamma}_s\,(p) = \begin{cases} \Delta^{-1}(\tilde{t}(p),\, r(p),\, \phi(p) + s,\, \tilde{z}(p)) & \text{if } p \in M^-, \\ p & \text{if } p \in A. \end{cases}$$

The three Killing fields under consideration have vanishing Lie brackets with one another:

$$[\tilde{t}^a,\, \phi^a] = [\tilde{t}^a,\, \tilde{z}^a] = [\phi^a,\, \tilde{z}^a] = \mathbf{0}. \tag{3.1.42}$$

(Once again, these relations follow most easily from proposition 3.1.1.) Now let p be any point on the axis A. The maps $\overset{\phi}{\Gamma}_s$ all leave p fixed, and leave $\tilde{t}^a$ and $\tilde{z}^a$ fixed as well (by proposition 1.6.6). So if U is the two-dimensional subspace of M_p that is orthogonal to both $\tilde{t}^a$ and $\tilde{z}^a$, the maps $\overset{\phi}{\Gamma}_s$ induce a one-parameter family of rotations of U. And what is true here of p is true quite generally, because of homogeneity as formulated in (2). So we have the following isotropy claim.

(5) *Gödel spacetime is (globally) isotropic in the following sense: given any point p, and any two unit spacelike vectors $\overset{1}{\sigma}{}^a$ and $\overset{2}{\sigma}{}^a$ at p that are orthogonal to both $\tilde{t}^a$ and z^a, there is an isometry $\psi : M \to M$ such that $\psi(p) = p$, $\psi_*(\tilde{t}^a) = \tilde{t}^a$, $\psi_*(\tilde{z}^a) = \tilde{z}^a$, and $\psi_*(\overset{1}{\sigma}{}^a) = \overset{2}{\sigma}{}^a$.*

And now it is also clear, as announced, that Gödel spacetime admits closed timelike (and closed null) curves. Indeed, consider the set of (maximally extended) integral curves of ϕ^a. They are closed curves, characterized by constant values for $\tilde{t}$, r, and $\tilde{z}$. We shall call them (or their images) *Gödel circles*. As we have just seen, they qualify as timelike if $r > r_c$ and null if $r = r_c$. These particular curves are centered on the axis A. But by homogeneity, it follows that given *any* point in Gödel spacetime, there are closed timelike and closed null curves passing through the point. Indeed, we can make a much stronger assertion. The "causal structure" of Gödel spacetime is completely degenerate in the following sense.

(6) *Given any two points p and q in Gödel spacetime, there is a smooth, future-directed timelike curve that runs from p and q. (Hence, since we can always*

−1
0
+1

combine timelike curves that run in the two directions and smooth out the joints, there is a smooth, closed timelike curve that contains p and q.)

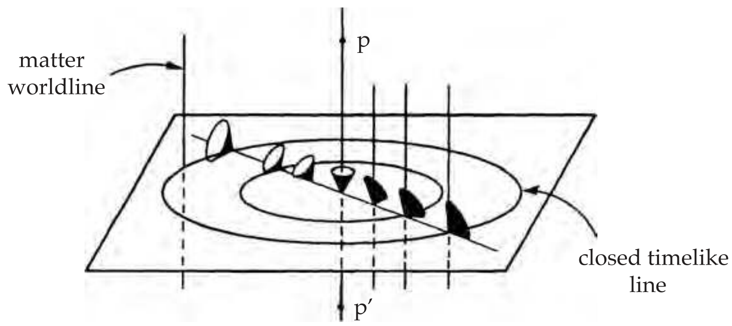

Figure 3.1.1. Gödel spacetime with one dimension (the $\tilde{z}$ dimension) suppressed.

Thus a time traveler in Gödel spacetime can start at any point p, return to that point, and stop off at any other desired point q along the way. To see why (6) holds, consider figure 3.1.1. It gives, at least, a rough, qualitative picture of Gödel spacetime with one dimension suppressed. We may as well take the central line to be the axis A and take p to be a point on A. (By homogeneity once again, there is no loss in generality in doing so.) Notice first that given any other point p' on A, no matter how "far down," there is a smooth, future-directed timelike curve that runs from p to p'. We can think of it as arising in three stages. (i) By moving "radially outward and upward" from p (i.e., along a future-directed timelike curve whose tangent vector field is of the form $\tilde{t}^a + \alpha\, r^a$, with α positive[13]), we can reach a point p_1 with coordinate value $r > r_c$. At that radius, we know, ϕ^a is timelike and future-directed. So we can find an $\epsilon > 0$ such that $(-\epsilon\, \tilde{t}^a + \phi^a)$ is also timelike and future-directed there. (ii) Now consider the maximally extended, future-directed timelike curve γ through p_1 whose tangent is *everywhere* equal to $(-\epsilon\, \tilde{t}^a + \phi^a)$ (for that value of ϵ). It is a spiral-shaped curve of fixed radius, with "downward pitch." By following γ far enough, we can teach a point p_2 that is well "below" p'. (We can overshoot as much as we might want.) Now, finally, (iii) we can reach p' by working our way upward and inward from p_2 via a curve whose tangent vector is the form $\tilde{t}^a + \alpha\, r^a$, but now with α negative. It remains only to smooth out the "joints" at intermediate points p_1 and p_2 to arrive at a smooth timelike curve that, as required, runs from p to p'.

13. Note that $\tilde{t}^a + \alpha\, r^a$ is timelike so long as $\alpha^2 < 1$.

Now consider any point q. It might not be possible to reach q from p in the same simple way we went from p to p_1—i.e., along a future-directed timelike curve that moves radially outward and upward—p might be too "high" for that. But we can get around this problem by first moving to an intermediate point p' on A sufficiently "far down"—we have established that that is possible—and then going from there to q. (This completes the argument for (6).)

Other interesting features of Gödel spacetime are closely related to the existence of closed timelike curves. So, for example, a *slice* (in any relativistic spacetime) is a spacelike hypersurface that, as a subset of the background manifold, is closed. We can think of it as a candidate for a "global simultaneity slice." It turns out that there are no slices in Gödel spacetime. More generally, given any relativistic spacetime, if it is temporally orientable and simply connected and has smooth closed timelike curves through every point, then it does not admit any slices (Hawking and Ellis [30, p. 170]).

Next we have the following basic fact.

(7) *There are no closed timelike or null geodesics in Gödel spacetime.*

We can easily confirm this, even before we characterize the class of timelike and null geodesics. It suffices (by homogeneity) to show that there are no closed timelike or closed null geodesics that pass through some particular point p on the axis A. Consider the set $C = \{q\colon r(q) < r_c\}$. We shall call it the *critical cylinder surrounding A*. We can establish our claim by showing two things: (i) all timelike geodesics that pass through p are fully contained within C, and all null geodesics that pass through p are fully contained within the closure of C; and (ii) there are no (non-trivial) closed causal curves within the closure of C.

For (i), let γ be any timelike or null geodesic that passes through p, and let λ^a be its tangent field. We may as well assume that γ is future-directed (since otherwise we can run the argument on a new curve that results from reversing the orientation of γ). Since ϕ^a is a Killing field, the quantity $\lambda^a\phi_a$ is constant on γ. (Recall problem 1.9.6.) It is equal to 0 at p, since ϕ^a is the zero vector there. So it must be 0 everywhere. Now on the boundary of C (where $r = r_c$), ϕ^a is a non-zero, future-directed null vector. So its inner product there with any future-directed timelike vector is strictly positive. It follows that if γ is timelike, it can never reach the boundary of C. (If it did, we would have $\lambda^a\phi_a > 0$ there.) It must stay within the (open) set C. Similarly, at all points outside the closure of C, ϕ^a is a future-directed timelike vector. So its inner product with all future-directed causal vectors (even null ones) is strictly positive. And therefore, if γ is null, it must remain within the closure C. (As we shall see in a moment, null geodesics through p *do* periodically intersect the boundary of C.)

For (ii), note that, by equation (3.1.40),

$$g^{ab}\,(\nabla_a\tilde{t})\,(\nabla_b\tilde{t}) = -\frac{1}{4\,\mu^2}\,\frac{(sh^4 r - sh^2 r)}{(sh^4 r + sh^2 r)}.$$

So $\nabla_a \tilde{t}$ is timelike within C and null (and non-zero) on the boundary of the set. It is future-directed both in C and on its boundary (since $\tilde{t}^a \nabla_a \tilde{t} = 1$). Now let γ be any non-trivial future-directed causal curve that passes through p, and let λ^a be its tangent field. Then (since λ^a and $\nabla_a \tilde{t}$ are co-oriented), we have $\lambda^n \nabla_n \tilde{t} > 0$ at all points in C and $\lambda^n \nabla_n \tilde{t} \geq 0$ at all points on the boundary of the set. So γ cannot possibly stay within the closure of C and still close back on itself.

Now, finally, let us characterize the set of all timelike and null geodesics in Gödel spacetime. The $\tilde{z}^a$ direction is not very interesting here, and we may as well restrict attention to curves that fall within a $\tilde{z}^a =$ constant submanifold—i.e., curves whose tangent fields are orthogonal to $\tilde{z}^a$ (or equivalently to z^a).[14]

We shall first consider certain examples that admit a particularly simple description. Then we shall argue that they are, up to isometry (and reparametrization), the only ones. A small bit of computation is involved. For that we need the following simple results that are the counterparts to ones presented earlier for the first set of coordinates. At points in M^-, where $r > 0$, we have

(3.1.43) $$\phi_b = 4\,\mu^2\,[(sh^4\, r - sh^2\, r)\nabla_b\,\phi + \sqrt{2}\,sh^2\, r\,\nabla_b\,\tilde{t}],$$

(3.1.44) $$\nabla_a\,\phi_b = 4\,\mu^2\,[\,(4\,sh^3\, r - 2\,sh\, r)(ch\, r)(\nabla_{[a}\, r)(\nabla_{b]}\,\phi) + 2\,\sqrt{2}\,(sh\, r)(ch\, r)\,(\nabla_{[a}\, r)(\nabla_{b]}\,\tilde{t})\,],$$

(3.1.45) $$\phi^a\nabla_a\,\phi_b = -4\,\mu^2\,(2\,sh^3\, r - sh\, r)(ch\, r)\,\nabla_b\, r,$$

(3.1.46) $$\phi^a\nabla_a\,\tilde{t}_b = \tilde{t}^a\nabla_a\,\phi_b = -4\,\sqrt{2}\,\mu^2\,(sh\, r)(ch\, r)\,\nabla_b\, r.$$

(For the second equation, we use the fact that ϕ^a is a Killing field and, so, $\nabla_{(a}\,\phi_{b)} = \mathbf{0}$. For the fourth, we use equation (3.1.42).)

Consider fields of the form $\tilde{t}^a + k\,\phi^a$, where k is some real number. Their integral curves are "helices" on which r and $\tilde{z}$ are constant (since $\tilde{t}^a\nabla_a\, r = \tilde{t}^a\nabla_a\,\tilde{z} = 0$, and similarly for ϕ^a). Our goal is to show that some of these helices—characterized by particular choices for k and r—are causal geodesics. Let k and r be fixed, and let γ be an integral curve of $\tilde{t}^a + k\,\phi^a$ associated with these values. Then, we have

14. Given any smooth curve $s \mapsto \Phi^{-1}(t(s), x(s), y(s), z(s))$ in Gödel spacetime, it qualifies as a geodesic iff (i) $z(s)$ is of the form $z(s) = z_0 + k\,s$, for some numbers z_0 and k, and (ii) the projected curve $s \mapsto \Phi^{-1}(t(s), x(s), y(s), z_0)$ qualifies as a geodesic. This follows because $\nabla_a\, z^b = \mathbf{0}$.

$$(\tilde{t}^a + k\,\phi^a)(\tilde{t}_a + k\,\phi_a) = 4\,\mu^2\,[\,k^2\,(sh^4\,r - sh^2\,r) + 2\,\sqrt{2}\,(sh^2\,r)\,k + 1\,] \tag{3.1.47}$$

and (by equation (3.1.45) and (3.1.46) and the fact that $\tilde{t}^a$ is a geodesic field),

$$\begin{aligned}(\tilde{t}^a + k\,\phi^a)\nabla_a\,(\tilde{t}_b + k\,\phi_b) &= 2\,k\,[-4\,\sqrt{2}\,\mu^2\,(sh\,r)(ch\,r)\,\nabla_b\,r\,] \\ &\quad + k^2\,[-4\,\mu^2\,(2\,sh^3\,r - sh\,r)(ch\,r)\,\nabla_b\,r\,] \\ &= -4\,\mu^2\,k\,(sh\,r)(ch\,r)\,[2\,\sqrt{2} \\ &\quad + k\,\,(2\,sh^2\,r - 1)]\,\nabla_b\,r.\end{aligned} \tag{3.1.48}$$

Thus γ is a geodesic iff $k = 0$ (in which case it is just an integral curve of $\tilde{t}^a$), $r = 0$ (in which case, again, it is an integral curve of $\tilde{t}^a$, now on the axis), or

$$k\,\,(2\,sh^2\,r - 1) + 2\,\sqrt{2} = 0. \tag{3.1.49}$$

It is a null geodesic iff this condition holds and the right side of equation (3.1.47) is 0. That leaves us with two equations in two unknowns. They yield

$$\gamma \text{ is a null geodesic} \iff sh^2\,r = \frac{(\sqrt{2} - 1)}{2} \quad \text{and} \quad k = 2(1 + \sqrt{2})$$

or, equivalently (since $sh\,2r = 2\,(sh\,r)(ch\,r)$),

$$\gamma \text{ is a null geodesic} \iff r = \frac{r_c}{2} \quad \text{and} \quad k = 2(1 + \sqrt{2}).$$

Similarly, after excluding the trivial cases where $k = 0$ or $r = 0$, , we have

$$\gamma \text{ is a timelike geodesic} \iff r < \frac{r_c}{2} \quad \text{and} \quad k = \frac{2\,\sqrt{2}}{(1 - 2\,sh^2\,r)}.$$

Thus, given any point q with r coordinate satisfying $0 < r < r_c/2$, there is exactly one value of k for which the helix through q with tangent field $\tilde{t}^a + k\,\phi^a$ is a timelike geodesic.

The number k here has a natural physical interpretation in terms of relative speed. Think of the tangent vector $\tilde{t}^a + k\,\phi^a$ as a (non-normalized, possibly null) velocity vector. We can extract a "speed relative to $\tilde{t}^a$" if we first decompose it into components tangent—and orthogonal to—$\tilde{t}^a$, and then divide the norm of the second by the norm of the first. With just a bit of calculation, we get

$$v = \text{speed relative to } \tilde{t}^a = \frac{k\,(sh\,r)(ch\,r)}{1 + k\,\sqrt{2}\,sh^2\,r}.$$

It follows that $k = 2\,\sqrt{2}/(1 - 2\,sh^2\,r)$ holds iff $v = \sqrt{2}\,(sh\,2r)/(ch\,2r)$. So we can reformulate our equivalence this way:

$$\gamma \text{ is a timelike geodesic} \iff r < \frac{r_c}{2} \quad \text{and} \quad v = \sqrt{2}\,\frac{sh\,2r}{ch\,2r}.$$

(Notice that $\sqrt{2}\,(sh\,2r)/(ch\,2r)$ goes to 1 as r approaches $r_c/2$.)

Here is our characterization claim.

(8) *The special geodesics we have just considered—the ones that are (maximally extended) integral curves of* $\tilde{t}^a + k\,\phi^a$ *for some* k*—are, up to isometry and reparametrization, the only maximally extended, future-directed, null and timelike geodesics in Gödel spacetime (confined to a* $\tilde{z} =$ *constant submanifold).*

Let us verify it, first, for null geodesics. Let γ_1 be any maximally extended, future-directed, null geodesic confined to a submanifold N whose points all have some particular $\tilde{z}$ value. Let q be any point in N whose r coordinate satisfies $sh^2\, r = (\sqrt{2} - 1)/2$. Pick any point on γ_1. By virtue of the homogeneity of Gödel spacetime—as recorded in (2)—we can find a (temporal orientation preserving) global isometry that maps that point to q and maps N to itself. Let γ_2 be the image of γ_1 under that isometry. We know that at q the vector $(\tilde{t}^a + k\,\phi^a)$ is null if $k = 2(1 + \sqrt{2})$. So, by virtue of the isotropy of Gödel spacetime (in the sense of (5)), we can find a global isometry that keeps q fixed, maps N to itself, and rotates γ_2 onto a new null geodesic γ_3 whose tangent vector at q is, at least, proportional to $(\tilde{t}^a + 2(1 + \sqrt{2})\,\phi^a)$, with positive proportionality factor. If, finally, we reparametrize γ_3 so that its tangent vector at q is equal to $(\tilde{t}^a + 2(1 + \sqrt{2})\,\phi^a)$, then the resultant curve *must* be a special null geodesic helix through q since (up to a uniform parameter shift) there can *be* only one (maximally extended) geodesic through q that has that tangent vector there.

The corresponding argument for timelike geodesics is almost the same. Let γ_1 this time be any maximally extended, future-directed, timelike geodesic confined to a submanifold N whose points all have some particular $\tilde{z}$ value. Let v be the speed of that curve relative to $\tilde{t}^a$. (The value as determined at any point must be constant along the curve since it is a geodesic.). Further, let q be any point in N whose r coordinate satisfies $\sqrt{2}\,(sh\,2r)/(ch\,2r) = v$. (We can certainly find such a point since $\sqrt{2}\,(sh\,2r)/(ch\,2r)$ runs through all values between 0 and 1 as r ranges between 0 and $r_c/2$.) Now we can proceed in three stages, as before. We map γ_1 to a curve that runs through q. Then we rotate that curve so that its tangent vector (at q) is aligned with $(\tilde{t}^a + k\,\phi^a)$ for the appropriate value of k, namely $k = 2\,\sqrt{2}/(1 - 2\,sh^2\, r)$. Finally, we reparametrize the rotated curve so that it has that vector itself as its tangent vector at q. That final curve must be one of our special helical geodesics by the uniqueness theorem for geodesics. (This completes the argument for (8).)

The special timelike and null geodesics we started with—the special helices centered on the axis A—exhibit various features. Some are exhibited by *all* timelike and null geodesics (confined to a $\tilde{z}$=constant submanifold); some are not. It is important to keep track of the difference. What is at issue is whether the features can or cannot be captured in terms of g_{ab}, $\tilde{t}^a$, and $\tilde{z}^a$ (or whether they make essential reference to the coordinates $\tilde{t}, r, \phi$ themselves). So, for example, if a curve is parametrized by s, one might take its vertical "pitch" (relative to $\tilde{t}$) at any point to be given by the value of $d\tilde{t}/ds$ there. Understood this way, the vertical pitch of the special helices centered on A is constant, but that of other timelike and null geodesics is not. For this reason, it is not correct to think of the latter, simply, as "translated" versions of the former. On the other hand, the following *is* true of all timelike and null geodesics (confined to a $\tilde{z}$=constant submanifold). If we project them (via $\tilde{t}^a$) onto a two-dimensional submanifold characterized by constant values for $\tilde{t}$ as well as $\tilde{z}$, the result is a circle.[15]

Here is another way to make the point. Consider any timelike or null geodesic γ (confined to a $\tilde{z}$ = constant submanifold). It certainly need not be centered on the axis A and need not have constant vertical pitch relative to $\tilde{t}$. But we can always find a (new) axis A' and a new set of cylindrical coordinates $\tilde{t}', r', \phi'$ adapted to A' such that γ qualifies as a special helical geodesic relative to *those* coordinates. In particular, it will have constant vertical pitch relative to $\tilde{t}'$.

Let us now consider all the timelike and null geodesics that pass through some point p (and are confined to a $\tilde{z}$=constant submanifold). It may as well be on the original axis A. We can better visualize the possibilities if we direct our attention to the circles that arise after projection (via $\tilde{t}^a$). Figure 3.1.2 shows a two-dimensional submanifold through p on which $\tilde{t}$ and $\tilde{z}$ are both constant. The dotted circle has radius r_c. Once again, that is the "critical radius" at which the rotational Killing field ϕ^a is null. Call this dotted circle the "critical circle." The circles that pass through p and have radius $r = r_c/2$ are projections of null geodesics.[16] Each shares exactly one point with the critical circle. In contrast, the circles of smaller radius that pass through p are the projections of timelike geodesics. The diagram captures one of the claims we made in the course of arguing for claim (7)—namely, that no timelike or null geodesic that passes through a point can "escape" to a radial distance from it greater than r_c.

We said at the beginning of this section that Gödel spacetime exhibits a "boomerang effect." It should now be clear what was intended. Suppose an individual is at rest with respect to the cosmic source fluid in Gödel spacetime (and so his worldline coincides with some $\tilde{t}$-line). If that individual shoots a

15. Notice that we can capture this projection condition in terms of g_{ab}, $\tilde{t}^a$, and $\tilde{z}^a$. It holds of a given curve γ iff there is an integral curve of $\tilde{t}^a$ such that all points on γ are the same "distance" from it, where distance is measured along geodesic segments that are orthogonal to both $\tilde{t}^a$ and $\tilde{z}^a$.

16. The assertion that a certain timelike or null geodesic has a certain "radius" can be expressed without reference to the value of a radial coordinate based on some axis. See note 15.

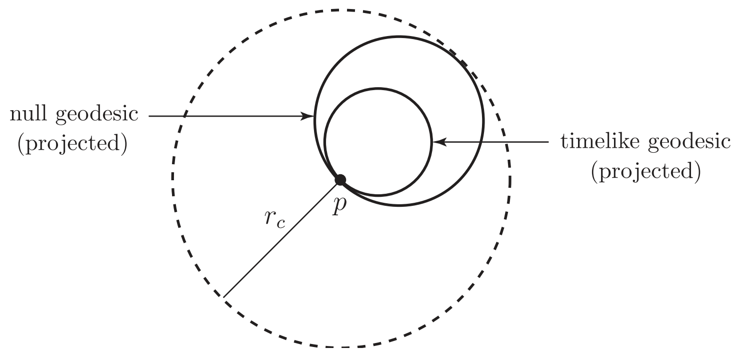

Figure 3.1.2. Projections of timelike and null geodesics in Gödel spacetime. r_c is the "critical radius" at which the rotational Killing field ϕ^a centered at p is null.

gun at some point, in any direction orthogonal to $\tilde{z}^a$, then, no matter what the muzzle speed of the gun, the bullet will eventually come back and hit him (unless it hits something else first or disintegrates). Here is a purely geometric formulation.

(9) *(Boomerang Effect) Let L be any $\tilde{t}$-line in Gödel spacetime, and let γ be any maximally extended timelike or null (but non-degenerate) geodesic on which the value of $\tilde{z}^a$ is constant. Then if γ intersects L once, it does so infinitely many times; and the temporal interval between intersection points (as measured along L) is constant.*

Appendix: A Coordinate-Free Characterization of Gödel Spacetime

Here, following Gödel [25] and [27], we characterize the geometric structure of Gödel spacetime in coordinate-free terms, and use this characterization to establish the equivalence of our two coordinate representations of the metric.[17]

First, Gödel spacetime (M, g_{ab}) can be decomposed as a metric product. One component is the manifold $\mathbb{R}$ together with the (negative-definite) metric $-\mu^2\, dz_a dz_b$. The other component is the manifold $\mathbb{R}^3$ together with a certain metric h_{ab} of signature $(1, 2)$. The latter can be expressed as

$$h_{ab} = \tilde{h}_{ab} + \tau_a\, \tau_b,$$

where

(1) $\tilde{h}_{ab}$ is a geodesically complete metric on $\mathbb{R}^3$ of signature $(1, 2)$ and constant positive-curvature $1/(4\mu^2)$, and

17. The material in this appendix is taken, with only minor changes in notation, from Malament [39].

(2) $\tau^a = \tilde{h}^{ab}\, \tau_b$ is a unit timelike Killing field with respect to $\tilde{h}_{ab}$.

(In (2), $\tilde{h}^{ab}$ is the inverse of $\tilde{h}_{ab}$; i.e., we are not using some other metric to raise indices.)

We can recover this characterization by starting with either of our two coordinate representations of the Gödel metric. Consider the first, equation (3.1.1). Here the coordinates t, x, y, z range over all of $\mathbb{R}$. We arrive at the structure $(\mathbb{R}^3, h_{ab})$ by dropping the $dz_a\, dz_b$ term and restricting the reduced metric to any submanifold of constant z value. The reduced metric assumes the form $\tilde{h}_{ab} + \tau_a\, \tau_b$ if we set

$$\tilde{h}_{ab} = \mu^2 \left[\frac{1}{2} (\nabla_a\, t)(\nabla_b\, t) + e^x\, (\nabla_{(a} t)(\nabla_{b)} y) - (\nabla_a\, x)(\nabla_b\, x) \right], \tag{3.1.50}$$

$$\tau_a = \frac{\mu}{\sqrt{2}}\, (\nabla_a\, t + e^x\, \nabla_a\, y). \tag{3.1.51}$$

So, to justify the proposed characterization, it will suffice to confirm that these two fields satisfy (1) and (2).

The inverse of $\tilde{h}_{ab}$ is

$$\tilde{h}^{bc} = \frac{1}{\mu^2} \left[4\, e^{-x}\, t^{(b} y^{c)} - x^b\, x^c - 2\, e^{-2x}\, y^b\, y^c \right], \tag{3.1.52}$$

and so τ^a comes out to be $(\sqrt{2}/\mu)\, t^a$. (We are continuing to use the abbreviations in equation (3.1.2).) The latter is a unit timelike field with respect to $\tilde{h}_{ab}$, as required. It is also a Killing field with respect to that metric. (The argument is almost exactly the same as the one used above to establish that t^a is a Killing field with respect to the original metric g_{ab}.) So we have (2). For (1), note first that $\tilde{h}_{ab}$ has signature $(1, 2)$, since the vectors $(\sqrt{2}/\mu)\, t^a$, $(\sqrt{2}/\mu)\, (t^a - e^{-x}\, y^a)$, and $(1/\mu)\, x^a$ form an orthonormal triple (of the appropriate type) at every point. Next, consider the map

$$\Psi : (t, x, y) \;\mapsto\; (u_1, u_2, u_3, u_4)$$

from $\mathbb{R}^3$ into $\mathbb{R}^4$ where

$$u_1 = 2\,\mu \left[cos\left(\frac{t}{2\sqrt{2}}\right) ch\left(\frac{x}{2}\right) - \frac{1}{2\sqrt{2}}\, y\, e^{x/2}\, sin\left(\frac{t}{2\sqrt{2}}\right) \right], \tag{3.1.53}$$

$$u_2 = 2\,\mu \left[sin\left(\frac{t}{2\sqrt{2}}\right) ch\left(\frac{x}{2}\right) + \frac{1}{2\sqrt{2}}\, y\, e^{x/2}\, cos\left(\frac{t}{2\sqrt{2}}\right) \right], \tag{3.1.54}$$

$$u_3 = 2\,\mu \left[-sin\left(\frac{t}{2\sqrt{2}}\right) sh\left(\frac{x}{2}\right) + \frac{1}{2\sqrt{2}}\, y\, e^{x/2}\, cos\left(\frac{t}{2\sqrt{2}}\right) \right], \tag{3.1.55}$$

(3.1.56) $$u_4 = 2\,\mu\left[\cos\left(\frac{t}{2\sqrt{2}}\right)\,sh\left(\frac{x}{2}\right) + \frac{1}{2\sqrt{2}}\,y\,e^{x/2}\,\sin\left(\frac{t}{2\sqrt{2}}\right)\right].$$

A straightforward computation establishes that

(3.1.57) $$(u_1)^2 + (u_2)^2 - (u_3)^2 - (u_4)^2 = 4\,\mu^2$$

and, using equation (1.5.7), that

(3.1.58) $$\Psi^*\big((\nabla_a u_1)(\nabla_b u_1) + (\nabla_a u_2)(\nabla_b u_2) - (\nabla_a u_3)(\nabla_b u_3) - (\nabla_a u_4)(\nabla_b u_4)\big)$$
$$= \mu^2\left(\frac{1}{2}(\nabla_a t)(\nabla_b t) + e^x(\nabla_{(a} t)(\nabla_{b)} y) - (\nabla_a x)(\nabla_b x)\right).$$

The map Ψ, as it stands, is not injective. It makes the same assignment to (t, x, y) and $(t + 4\sqrt{2}\,\pi, x, y)$. But it is injective if we restrict t to the interval $[0, 4\sqrt{2}\,\pi)$. Indeed, if $\sim$ is the equivalence relation on $\mathbb{R}^3$ defined by

$$(t, x, y) \sim (t', x', y') \quad \text{iff} \quad x' = x \text{ and } y' = y \text{ and } t' = t \ (\text{mod } 4\sqrt{2}\,\pi),$$

then Ψ determines a diffeomorphism between the quotient manifold $\mathbb{R}^3/\sim$ and the manifold

$$H = \left\{(u_1, u_2, u_3, u_4) \in \mathbb{R}^4 : (u_1)^2 + (u_2)^2 - (u_3)^2 - (u_4)^2 = 4\,\mu^2\right\}.$$ [18]

By equation (3.1.58), it qualifies as an isometry with respect to the metric induced on the latter by the background flat metric on $\mathbb{R}^4$ of signature $(2, 2)$. But it is a standard result that H together with this induced metric is a complete manifold of constant curvature $1/(4\,\mu^2)$. (See, for example, O'Neill [46, p. 113].) So—since $(\mathbb{R}^3, \tilde{h}_{ab})$ is an isometric covering manifold of the latter—$(\mathbb{R}^3, \tilde{h}_{ab})$ is, itself, a complete manifold of constant curvature $1/(4\,\mu^2)$. This gives us (1).

We can proceed in much the same way starting with equation (3.1.31), the second coordinate representation of the Gödel metric. This time we drop the $d\tilde{z}_a\, d\tilde{z}_b$ term and arrive at the desired decomposition of the reduced metric $(h_{ab} = \tilde{h}_{ab} + \tau_a\, \tau_b)$ if we set

(3.1.59) $$\tilde{h}_{ab} = 4\,\mu^2\left[\frac{1}{2}(\nabla_a \tilde{t})(\nabla_b \tilde{t}) - (\nabla_a r)(\nabla_b r) - sh^2 r\,(\nabla_a \phi)(\nabla_b \phi) + \sqrt{2}\,sh^2 r\,(\nabla_{(a}\tilde{t})(\nabla_{b)}\phi)\right],$$

18. Note that we can invert the restricted map and explicitly solve for t, x, y in terms of u_1, u_2, u_3, u_4. For example,

$$t = 2\sqrt{2}\,\arccos\frac{u_1 + u_4}{\sqrt{(u_1 + u_4)^2 + (u_2 - u_3)^2}}.$$

$$\tau_a = \sqrt{2}\,\mu\,(\nabla_a\,\tilde{t} + \sqrt{2}\,sh^2\,r\,\nabla_a\,\phi). \tag{3.1.60}$$

Here $\tau^a = \tilde{h}^{ab}\tau_b$ comes out as $(1/\sqrt{2}\,\mu)\,\tilde{t}^a$. So we see, once again, by equation (3.1.38), that $\tau^a = (\sqrt{2}/\mu)\,t^a$. And this time we can show that $(\mathbb{R}^3, \tilde{h}_{ab})$ is an isometric covering manifold of H (with respect to the induced metric on H) by considering the map[19]

$$\Psi' : (\tilde{t}, r, \phi) \mapsto (u_1, u_2, u_3, u_4)$$

where

$$u_1 = 2\mu\,cos\left(\frac{\tilde{t}}{\sqrt{2}}\right)\,ch\,r, \tag{3.1.61}$$

$$u_2 = 2\mu\,sin\left(\frac{\tilde{t}}{\sqrt{2}}\right)\,ch\,r, \tag{3.1.62}$$

$$u_3 = 2\mu\,sin\left(\phi - \frac{\tilde{t}}{\sqrt{2}}\right)\,sh\,r, \tag{3.1.63}$$

$$u_4 = 2\mu\,cos\left(\phi - \frac{\tilde{t}}{\sqrt{2}}\right)\,sh\,r. \tag{3.1.64}$$

One can check that equation (3.1.57) holds, once again, as does the counterpart to (3.1.58):

$$\begin{aligned}\Psi'^*\big((\nabla_a\,u_1)(\nabla_b\,u_1) + (\nabla_a\,u_2)(\nabla_b\,u_2) - (\nabla_a\,u_3)(\nabla_b\,u_3) - (\nabla_a\,u_4)(\nabla_b\,u_4)\big)\\ = 4\,\mu^2\left(\frac{1}{2}(\nabla_a\,\tilde{t})(\nabla_b\,\tilde{t}) - (\nabla_a\,r)(\nabla_b\,r) - sh^2\,r\,(\nabla_a\,\phi)(\nabla_b\,\phi)\right.\\ \left. + \sqrt{2}\,sh^2 r\,(\nabla_{(a}\tilde{t})(\nabla_{b)}\phi)\right).\end{aligned} \tag{3.1.65}$$

Here Ψ' is not injective, but it is so if we restrict $\tilde{t}$ to the interval $[0,\, 2\sqrt{2}\,\pi)$.

It should be clear now that our two coordinate expressions for the Gödel metric are fully equivalent. They are but alternate expressions for a metric on $\mathbb{R}^4$ that we have been able to characterize in a coordinate independent way.

We can gain further insight into the two maps Ψ and Ψ' if we recast them. Consider the (associative, distributive) algebra of "hyperbolic quaternions." We can construe them as elements of the form

19. As characterized here, the map is defined only where $r \neq 0$. But it can be smoothly extended to points at which $r = 0$.

$$\varphi = w_1 + w_2\mathbf{i} + w_3\mathbf{j} + w_4\mathbf{k}$$

where $w_1, \ldots, w_4$ are real numbers. Addition is defined by the rule

$$\begin{aligned}
&(w_1 + w_2\mathbf{i} + w_3\mathbf{j} + w_4\mathbf{k}) + (w_1' + w_2'\mathbf{i} + w_3'\mathbf{j} + w_4'\mathbf{k}) \\
&\quad = \big((w_1 + w_1') + (w_2 + w_2')\mathbf{i} + (w_3 + w_3')\mathbf{j} + (w_4 + w_4')\mathbf{k}\big).
\end{aligned}$$

Multiplication is defined by the requirement that (the real number) 1 serve as an identity element and by the relations

$$\begin{aligned}
\mathbf{i}\cdot\mathbf{i} &= -1, \\
\mathbf{j}\cdot\mathbf{j} &= \mathbf{k}\cdot\mathbf{k} = 1, \\
\mathbf{i}\cdot\mathbf{j} &= -\mathbf{j}\cdot\mathbf{i} = \mathbf{k}, \\
\mathbf{j}\cdot\mathbf{k} &= -\mathbf{k}\cdot\mathbf{j} = -\mathbf{i}, \\
\mathbf{k}\cdot\mathbf{i} &= -\mathbf{i}\cdot\mathbf{k} = \mathbf{j}.
\end{aligned}$$

If we define the conjugate and norm of φ by setting

$$\begin{aligned}
\overline{\varphi} &= w_1 - w_2\mathbf{i} - w_3\mathbf{j} - w_4\mathbf{k}, \\
\mathrm{norm}(\varphi) = \varphi\cdot\overline{\varphi} &= (w_1)^2 + (w_2)^2 - (w_3)^2 - (w_4)^2,
\end{aligned}$$

then it follows that $\overline{\varphi\cdot\psi} = \overline{\psi}\cdot\overline{\varphi}$ and, hence,

$$\mathrm{norm}(\varphi\cdot\psi) = \mathrm{norm}(\varphi)\ \mathrm{norm}(\psi) \tag{3.1.66}$$

for all φ and ψ. To simplify notation, we shall identify the hyperbolic quaternion $w_1 + w_2\mathbf{i} + w_3\mathbf{j} + w_4\mathbf{k}$ with the corresponding element (w_1, w_2, w_3, w_4) of $\mathbb{R}^4$. Then H is identified with the set of hyperbolic quaternions of norm $4\,\mu^2$, and it acquires a natural (Lie) group structure: given any two elements u and u' in H, we take their product to be $(1/4\,\mu^2)\,u\cdot u'$. The norm product condition (3.1.66) guarantees that the product is well defined. The element u has $\overline{u}$ for an inverse.

Notice now that for all real number t, x, y, the quadruples

$$(\cos t,\ \sin t,\ 0,\ 0) \qquad (\operatorname{ch} x,\ 0,\ 0,\ \operatorname{sh} x) \qquad (1,\ y,\ y,\ 0)$$

all have norm 1. So their product has norm 1. Straightforward computation confirms that the associated map

$$(t,\ x,\ y) \mapsto 2\,\mu\,(\cos t,\ \sin t,\ 0,\ 0)\cdot(\operatorname{ch} x,\ 0,\ 0,\ \operatorname{sh} x)\cdot(1,\ y,\ y,\ 0)$$

is essentially just the first of the two maps from $(\mathbb{R}^3, \tilde{h}_{ab})$ onto H displayed in equations (3.1.53)–(3.1.56). This is where it "comes from." Strictly speaking, to match the coefficients in that map, we need to make a small change and take the product to be

$$2\mu\left(\cos\left(\frac{t}{2\sqrt{2}}\right), \sin\left(\frac{t}{2\sqrt{2}}\right), 0, 0\right)\cdot\left(ch\left(\frac{x}{2}\right), 0, 0, sh\left(\frac{x}{2}\right)\right)\cdot\left(1, \frac{y}{2\sqrt{2}}, \frac{y}{2\sqrt{2}}, 0\right).$$

Similarly, we can recover the second of the maps from $(\mathbb{R}^3, \tilde{h}_{ab})$ onto H, the one displayed in equations (3.1.61)–(3.1.64), in the form

$$(\tilde{t}, r, \phi) \mapsto 2\mu\left(\cos\left(\frac{\tilde{t}}{\sqrt{2}}\right), \sin\left(\frac{\tilde{t}}{\sqrt{2}}\right), 0, 0\right)\cdot(ch\, r, 0, sh\, r \sin\phi, sh\, r \cos\phi).$$

3.2. Two Criteria of Orbital (Non-)Rotation

In general relativity, there is a natural and unambiguous notion of rotation at a point as it applies, for example, to a fluid. This is the notion we considered in section 2.8. If the four-velocity field of the fluid is ξ^a, then we say that the fluid is *non-rotating* at a given point if its associated rotation field ω_{ab} vanishes there or, equivalently, if $\xi_{[a}\nabla_b\,\xi_{c]} = \mathbf{0}$ there. (Recall problem 2.8.1.)

But when we consider notions of rotation that make essential reference to what happens over extended regions of spacetime, the situation changes immediately. So, for example, consider a (one-dimensional) ring centered about an axis of rotational-symmetry (figure 3.2.1). Just what does it mean to say that the ring is "not rotating" around the axis? (It will be convenient to stick with the negative formulation.) This turns out to be a subtle and interesting question in relativity theory. Various criteria for non-rotation readily come to mind. In garden-variety circumstances, they are equivalent. But the theory allows for conditions under which they come apart. It can happen that the ring is non-rotating in one perfectly natural sense but *is* rotating in another.

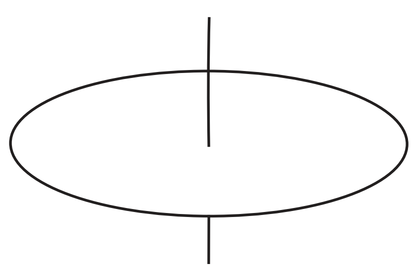

Figure 3.2.1. What does it mean to say that a ring is "not rotating" around a central axis of rotational symmetry?

In this section we consider two[20] such natural criteria for ring non-rotation: (i) the zero angular momentum (ZAM) criterion, and (ii) the compass of inertia on the ring (CIR) criterion. In each case, we give both a direct, geometric formulation and also a somewhat more intuitive, quasi-operational formulation. We verify that the ZAM and CIR criteria agree if a certain simplifying condition obtains, and we show that they do not agree in Gödel spacetime.

In the next section, we step back from these two particular criteria and formulate a no-go result[21] that applies to a large class of "generalized criteria" of ring non-rotation. We abstract three conditions that one might want a criterion of ring non-rotation to satisfy, and show that, at least in the case of some relativistic spacetime models, no generalized criterion of ring non-rotation satisfies all three. The upshot is that no notion of orbital non-rotation in relativity theory fully answers to our classical intuitions.

We need a certain amount of background structure to set things up. In what follows, let (M, g_{ab}) be a spacetime with two complete Killing fields, $\tilde{t}^a$ and ϕ^a, satisfying the following conditions: (i) $\tilde{t}^a$ is timelike; (ii) the orbits of ϕ^a are closed; (iii) ϕ^a is spacelike except at "axis points" (if there are any) where $\phi^a = \mathbf{0}$; (iv) not all points are axis points (i.e., ϕ^a does not vanish everwhere) (v) $[\tilde{t}^a, \phi^a] = \mathbf{0}$; and (vi) $\tilde{t}_{[a}\, \phi_b \nabla_c\, \tilde{t}_{d]} = \mathbf{0}$ and $\tilde{t}_{[a}\, \phi_b \nabla_c\, \phi_{d]} = \mathbf{0}$.

Gödel spacetime meets this description, at least if we restrict attention to the open set where $r < r_c$.[22] Another example is Minkowski spacetime. Yet a third—at least if we restrict attention, once again, to a certain open set—is Kerr spacetime, which we shall consider very briefly in the next section.

The stated conditions are, more or less, the usual ones defining a "stationary, axi-symmetric spacetime" (Wald [60]). For convenience, we have strengthened things a bit (compared to some formulations) by requiring that $\tilde{t}^a$ and ϕ^a be complete. The added strength is harmless. The point here is that even *with* this much structure in place, the two criteria of ring non-rotation need not agree. In what follows, when we refer to a *stationary, axi-symmetric*

20. It would be easy to assemble a longer list of criteria. For example, we could consider non-rotation as determined at "spatial infinity" (at least for the case of asymptotically flat spacetimes), non-rotation as determined relative to the compass of inertia on the axis (CIA) criterion (Malament [41]), and yet other criteria (see Page [50]). We are not attempting here a systematic account of orbital rotation in relativity theory. Our goal is to give an indication of the subject's interest and to prepare the way for a particular no-go theorem.

21. The result presented here is a variant of the one in Malament [41].

22. That condition (vi) holds in Gödel spacetime follows from equations (3.2.11) and (3.2.12) below. (We are deliberately using the same notation that we used in the preceding section for Gödel spacetime so that we can easily go back and forth between claims about the general case and claims about that one example.)

spacetime with Killing fields $\tilde{t}^a$ and ϕ^a, it should be understood that the stated conditions obtain.

The conditions themselves should be clear except, possibly, (vi). It asserts that, at least locally, there exist two-dimensional submanifolds that are orthogonal to both $\tilde{t}^a$ and ϕ^a. (This is a consequence of Frobenius's theorem. See the first part of the proof of theorem 7.1.1 in Wald [60, p. 163].) In Gödel spacetime, for example, these are submanifolds characterized by fixed values for $\tilde{t}$ and ϕ, and free values for r and $\tilde{z}$.

With this structure in place, we can represent our ring as an imbedded two-dimensional submanifold $\mathcal{R}$ that is invariant under the isometries generated by $\tilde{t}^a$ and ϕ^a (and on which $\phi^a \neq \mathbf{0}$). We call the latter an *orbit cylinder*. To represent the rotational state of the ring, we need to keep track of the motion of individual points on it. Each such point has a worldline that can be represented as a timelike curve on $\mathcal{R}$. So we are led to consider not just $\mathcal{R}$, but $\mathcal{R}$ together with a congruence of smooth timelike curves on $\mathcal{R}$ (figure 3.2.2).

We want to think of the ring as being in a state of rigid rotation, i.e., rotation with the distance between points on the ring remaining constant. So we are further led to restrict attention to just those congruences of timelike curves on $\mathcal{R}$ that are invariant under all isometries generated by $\tilde{t}^a$. Equivalently (moving from the curves themselves to their tangent fields), we are led to consider future-directed timelike vector fields on $\mathcal{R}$ of the form $(\tilde{t}^a + k\,\phi^a)$, where k is a number. We shall call the pair $(\mathcal{R}, k)$ a *striated orbit cylinder*. And, quite generally, we can take a "criterion of ring non-rotation" to be, simply, a specification, for every striated cylinder $(\mathcal{R}, k)$, whether it is to count as "non-rotating."

Officially, now, our two criteria can be formulated as follows. Let $(\mathcal{R}, k)$ be a striated cylinder. (Recall that we say a timelike vector field η^a, normalized or not, is non-rotating at a point if $\eta_{[a}\nabla_b\,\eta_{c]} = \mathbf{0}$ there.)

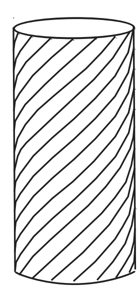

Figure 3.2.2. A "striated orbit cylinder" that represents a particular rotational (or non-rotational) state of the ring.

(1) $(\mathcal{R}, k)$ is non-rotating according to the *zero angular momentum (ZAM)* criterion if $(\tilde{t}^a + k\,\phi^a)$ is orthogonal to ϕ^a on $\mathcal{R}$; i.e., $(\tilde{t}^a + k\,\phi^a)\phi_a = 0$.

(2) $(\mathcal{R}, k)$ is non-rotating according to the *compass of inertia on the ring (CIR)* criterion if $(\tilde{t}^a + k\,\phi^a)$ is non-rotating on $\mathcal{R}$; i.e., the following condition holds on $\mathcal{R}$:

$$\tilde{t}_{[a}\nabla_b\,\tilde{t}_{c]} + k\,\tilde{t}_{[a}\nabla_b\,\phi_{c]} + k\,\phi_{[a}\nabla_b\,\tilde{t}_{c]} + k^2\,\phi_{[a}\nabla_b\,\phi_{c]} = \mathbf{0}. \tag{3.2.1}$$

The orthogonality condition in (1) just captures the requirement that every point on the ring have zero angular momentum with respect the rotational Killing field ϕ^a. (Recall our discussion in section 2.9.) So the terminology makes sense.

Let us now recast the two criteria in quasi-operational terms. Let us start with the second. Here is one way to set up an experimental test. Suppose we mount a gyroscope at some fixed point on the ring in such a way that it can rotate freely. And suppose that at some initial moment the axis of the gyroscope is oriented so as to be tangent to the ring (figure 3.2.3). Then we can consider whether it remains tangent over time. It turns out that it will do so (i.e., remain tangent to the ring) iff the ring is non-rotating according to the CIR criterion.

We shall verify this equivalence in a moment. But first, notice that the stated experimental test *does* seem to provide a natural criterion of non-rotation. Think about it. If the ring were rotating—here we are simply appealing to ordinary intuitions—we would expect that the angle between the gyroscope axis and (an oriented) tangent line would shift from 0^o to 90^o to 180^o to 270^o and back to 0^o as the ring passed through one complete rotation. The intuition here is that the tangent line changes direction as the ring rotates, but the axis of the gyroscope does not.

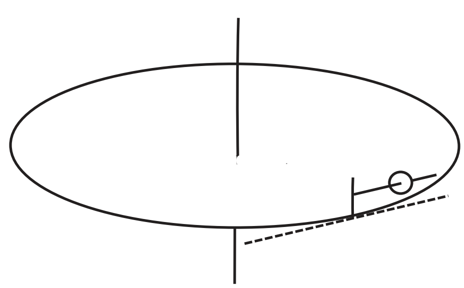

Figure 3.2.3. An experimental test to determine whether the ring is non-rotating according to the CIR criterion.

Now consider how we can capture, most directly, the stated "gyroscope remains tangent" condition. Let γ be a future-directed timelike curve that represents the worldline of the point on the ring where the gyroscope is mounted. The gyroscope there does not "change (spatial) direction as determined relative to γ." That is what makes it a gyroscope. So the "gyroscope remains tangent" condition will be satisfied iff the tangent field ϕ^a itself (now conceived as a field on γ) does not "change (spatial) direction relative to γ." We need only spell out the latter condition.

Let $\eta^a = (\tilde{t}^a + k\,\phi^a)$, let $\eta = (\eta^n \eta_n)^{1/2}$, and let $\hat{\eta}^a$ be the normalized field defined by $\eta^a = \eta\,\hat{\eta}^a$. Finally, let h_{ab} be the spatial projection field $(g_{ab} - \hat{\eta}_a\,\hat{\eta}_a)$ determined relative to $\hat{\eta}^a$. Then the spatial direction of ϕ^a as determined relative to γ is $h^b{}_n \phi^n$. And ϕ^a is "not changing (spatial) direction relative to γ" iff

$$h^a{}_b\,\hat{\eta}^m \nabla_m\,(h^b{}_n \phi^n) = \mathbf{0}. \tag{3.2.2}$$

This condition asserts that the spatial component of $\hat{\eta}^m \nabla_m\,(h^b{}_n \phi^n)$ as determined relative to γ vanishes. When it holds, we say that $h^b{}_n \phi^n$ is *Fermi transported* along γ.

We can simplify the condition slightly if we cast it in terms of $\eta^a = (\tilde{t}^a + k\,\phi^a)$ rather than the normalized field $\hat{\eta}^a$. Here and in what follows, we make repeated use of the fact that η^a is a Killing field and that η^a Lie derives ϕ^a and $\tilde{t}^a$ (since the Lie bracket of ϕ^a and $\tilde{t}^a$ vanishes); i.e., we have

$$\mathcal{L}_\eta\,\phi^a = \mathcal{L}_\eta\,\tilde{t}^a = \mathbf{0} \quad \text{and} \quad \mathcal{L}_\eta\,g_{ab} = \mathbf{0}. \tag{3.2.3}$$

Expanding h_{ab}, we see that equation (3.2.2) holds iff

$$(g^a{}_b - \hat{\eta}^a \hat{\eta}_b)\,\hat{\eta}^m \nabla_m\,[\phi^b - (\phi^n \hat{\eta}_n)\,\hat{\eta}^b] = \mathbf{0}.$$

But $\hat{\eta}^m \nabla_m\,\eta = \mathbf{0}$ and $\hat{\eta}^m \nabla_m (\phi^n \hat{\eta}_n) = \mathbf{0}$ by equation (3.2.3), and $\hat{\eta}_b\,\hat{\eta}^m \nabla_m\,\phi^b = \mathbf{0}$ since ϕ^a is a Killing field. Furthermore, $\hat{\eta}_b\,\hat{\eta}^m \nabla_m\,\hat{\eta}^b = 0$, since $\hat{\eta}^b$ is of unit length. So equation (3.2.2) holds iff

$$\eta^2\,\eta^m \nabla_m\,\phi^a = (\phi^n \eta_n)\,\eta^m \nabla_m\,\eta^a. \tag{3.2.4}$$

With all this as motivation, we have the following definition.

(2′) $(\mathcal{R}, k)$ is non-rotating according to the *gyroscope remains tangent (GRT)* criterion if $\eta^a = (\tilde{t}^a + k\,\phi^a)$ satisfies equation (3.2.4) on $\mathcal{R}$ (with $\eta = (\eta^n \eta_n)^{1/2}$).

Our earlier claim of equivalence now comes out as the following proposition.

PROPOSITION 3.2.1. *A striated orbit cylinder* $(\mathcal{R}, k)$ *qualifies as non-rotating according to the CIR criterion iff it qualifies as non-rotating according to the GRT criterion.*

Proof. One direction is easy. Assume that $\eta_{[n}\nabla_m\, \eta_{a]} = \mathbf{0}$ on $\mathcal{R}$. Then contraction with $\eta^m \phi^n$ yields

$$(\phi^n \eta_n)\, \eta^m \nabla_m\, \eta_a + \eta_a\, \eta^m \phi^n \nabla_n\, \eta_m + (\eta^m \eta_m)\, \phi^n \nabla_a\, \eta_n = \mathbf{0}$$

on $\mathcal{R}$. But the Lie bracket of ϕ^a with η^a vanishes. And ϕ^a and η^a are both Killing fields. So the second term in the sum vanishes ($\eta^m\, \phi^n \nabla_n\, \eta_m = \eta^m\, \eta^n \nabla_n\, \phi_m = 0$), and the third term is equal to

$$(\eta^m \eta_m)\, \phi^n \nabla_a\, \eta_n = -\eta^2\, \phi^n \nabla_n\, \eta_a = -\eta^2\, \eta^n \nabla_n\, \phi_a.$$

So equation (3.2.4) holds on $\mathcal{R}$.

Conversely, assume that equation (3.2.4) holds on $\mathcal{R}$. Then (once again using the fact that $\eta^m \nabla_m\, \phi_a = \phi^m \nabla_m\, \eta_a$), we have

$$[\eta^2\, \phi^m - (\phi^n \eta_n)\, \eta^m]\, \nabla_m\, \eta_a = \mathbf{0}$$

on $\mathcal{R}$. Now consider the field $\psi^m = [\eta^2 \phi^m - (\phi^n \eta_n) \eta^m]$. We have (i) $\psi^m \eta_m = 0$; (ii) $\psi^m \neq \mathbf{0}$; and (iii) $\psi^m\, \nabla_m\, \eta_a = \mathbf{0}$ on $\mathcal{R}$. (Condition (ii) holds because $\eta^2\, \phi^m$ is spacelike and $(\phi^n \eta_n)\, \eta^m$ is timelike or equal to $\mathbf{0}$.) It follows that $\psi^m \eta_{[n} \nabla_m\, \eta_{a]} = \mathbf{0}$ on $\mathcal{R}$. Now assume that $\eta_{[n}\nabla_m\, \eta_{a]} \neq \mathbf{0}$ at some point p on $\mathcal{R}$. Let ϵ_{abcd} be a volume element defined on some open set containing p. The space of anti-symmetric tensors α_{nma} at p that are orthogonal to ψ^m is one-dimensional. So at p, $\epsilon_{nmad}\, \psi^d = k_1\, \eta_{[n} \nabla_m\, \eta_{a]}$ for some k_1. Or, equivalently, $\psi^d = k_2\, \epsilon^{dnma} \eta_n \nabla_m\, \eta_a$ at p for some k_2. It follows (after expanding $\eta^a = \tilde{t}^a + k\, \phi^a$) that

$$\psi^d \phi_d = k_2\, \epsilon^{dnma} \phi_d\, \eta_n \nabla_m\, \eta_a = k_2\, \epsilon^{dnma} \phi_d\, \tilde{t}_n \nabla_m\, \tilde{t}_a + k_2\, k\, \epsilon^{dnma} \phi_d\, \tilde{t}_n \nabla_m\, \phi_a$$

at p. It now follows, by condition (vi) in our characterization of stationary axisymmetric spacetimes, that $\psi^d \phi_d = 0$ at p. So $\eta^2\, (\phi^m \phi_m) - (\phi^n \eta_n)^2 = 0$. But this is impossible, since ϕ^a is spacelike and $\eta > 0$. So we may conclude that $\eta_{[n} \nabla_m\, \eta_{a]} = \mathbf{0}$ at all points on $\mathcal{R}$. □

Now we turn to the ZAM criterion of ring non-rotation. Various experimental tests are possible. One involves the Sagnac effect. Imagine that we mount a light source at some point Q on the ring and arrange for its light pulses to travel around the ring in opposite (clockwise and counterclockwise) directions. (See Figure 3.2.4) This can be done, for example, using concave mirrors attached to

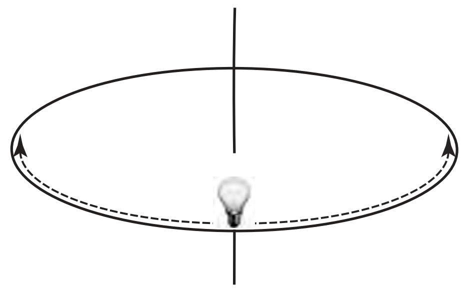

Figure 3.2.4. An experimental test to determine whether the ring is non-rotating according to the ZAM criterion.

the ring. Imagine further that we keep track of whether the pulses arrive back at Q simultaneously, using, for example, an interferometer). It turns out that this will be the case—i.e., they will arrive back simultaneously—iff the ring has zero angular momentum (with the respect to the background rotational symmetry). We shall soon verify this equivalence.

But notice, once again, that the stated experimental test *does* seem to provide a natural criterion of non-rotation. Suppose the ring is rotating in, say, a counterclockwise direction. (Here, again, we are simply appealing to ordinary intuitions about rotation.) Then the "C pulse," the one that moves in a clockwise direction, should get back to Q before completing a full circuit of the ring, because it is moving toward an approaching target. In contrast, the "CC pulse," the one moving in a counterclockwise direction, is chasing a receding target. To get back to Q it will have to traverse the entire length of the ring, and then it will have to cover the distance that Q has moved in the interim time. So one should expect, in this case, that the C pulse will arrive back at Q before the CC pulse. (Here we presume that light travels at the same speed in all directions.) Similarly, if the ring is rotating in a clockwise direction, one would expect that the CC pulse would arrive back at Q before the C pulse. Only if the ring is not rotating should they arrive simultaneously. Thus, our experimental test for whether the ring has zero angular momentum provides what would seem to be a natural criterion of non-rotation.

Let us now make precise our claim of equivalence. Let $(\mathcal{R}, k)$ be a striated orbit cylinder, let γ be any (maximally extended) integral curve of $(\tilde{t}^a + k\,\phi^a)$ on $\mathcal{R}$, and let p_0 be an arbitrary point on the image of γ. Further, let λ_1 and λ_2 be two future-directed (maximally extended) null curves on $\mathcal{R}$ that start at p_0 (figure 3.2.5). The latter represent light pulses that are emitted at p_0 and traverse the ring in opposite directions. Call them "pulse 1" and "pulse 2." Both λ_1 and λ_2 must intersect γ a second time (indeed infinitely many times);

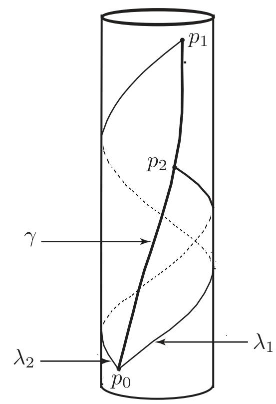

Figure 3.2.5. Sagnac effect.

i.e., the pulses must eventually return to their point of emission on the ring. (We shall soon verify this.) Let p_1 be the next intersection point of γ with λ_1, and let p_2 be the next intersection point of γ with λ_2. In general, there is no reason why p_1 and p_2 should coincide. We are interested in the case where they do. So we are led to consider the following criterian of non-rotation.

(1′) $(\mathcal{R}, k)$ is non-rotating according to the *Sagnac effect (SE)* criterion if, in the case just described, the first re-intersection points p_1 and p_2 coincide.

Note that the stated condition—agreement of first re-intersection points—will hold for one choice of initial integral curve γ and initial point p_0 iff it holds for any other. The symmetries of $(\mathcal{R}, k)$ guarantee as much. So there is no ambiguity in our formulation. Now we can verify our claim of equivalence.[23]

PROPOSITION 3.2.2. *A striated orbit cylinder $(\mathcal{R}, k)$ qualifies as non-rotating according to the ZAM criterion iff it qualifies as non-rotating according to the SE criterion.*

Proof. We have to verify that, in the case described,

$$p_1 = p_2 \iff (\tilde{t}^a + k\,\phi^a)\phi_a = 0. \tag{3.2.5}$$

The tangent field to γ is $(\tilde{t}^a + k\,\phi^a)$. The tangent fields to λ_1, and λ_2 can be rescaled so that they have the form $(\tilde{t}^a + l_1\,\phi^a)$ and $(\tilde{t}^a + l_2\,\phi^a)$. Since the first

23. Our proof proceeds by way of a "low-brow" calculation. For a more insightful argument, see Ashtekar and Magnon [3].

is timelike and the second two are null, we have $l_i \neq k$ and

$$(\tilde{t}^a + l_i\,\phi^a)(\tilde{t}_a + l_i\,\phi_a) = 0$$

for $i = 1, 2$. This equation has roots

$$l_1 = \frac{-(\tilde{t}^a\phi_a) + \sqrt{D}}{(\phi^n\phi_n)}, \tag{3.2.6}$$

$$l_2 = \frac{-(\tilde{t}^a\phi_a) - \sqrt{D}}{(\phi^n\phi_n)}, \tag{3.2.7}$$

where $D = \left[(\tilde{t}^a\phi_a)^2 - (\tilde{t}^a\tilde{t}_a)(\phi^b\phi_b)\right]$. (Clearly there is no loss in generality in choosing to label them this way.) Note that $D > (\tilde{t}^a\phi_a)^2 \geq 0$, since $\tilde{t}^a$ is timelike and ϕ^a is spacelike on $\mathcal{R}$. So $l_1 > 0$ and $l_2 < 0$. Moreover, $l_2 < k < l_1$. (Consider the quadratic function $f(x) = (\tilde{t}^a + x\,\phi^a)(\tilde{t}_a + x\,\phi_a)$. It is concave downward because $(\phi_a\,\phi^a)$ is negative. So, since $f(k) > 0$ and $f(l_1) = f(l_2) = 0$, it must be the case that k falls between l_1 and l_2.) So

$$(l_1 - k) > 0 \qquad \text{and} \qquad (l_2 - k) < 0.$$

It follows from our initial assumptions about the background spacetime (M, g_{ab}) that there there exist smooth coordinate maps $\tilde{t}\colon \mathcal{R} \to \mathbb{R}$ and $\phi\colon \mathcal{R} \to \mathbb{R}\,(\mathrm{mod}\,2\pi)$ on the orbit cylinder $\mathcal{R}$ such that $\tilde{t}^a\nabla_a\,\tilde{t} = \phi^a\nabla_a\,\phi = 1$ and $\tilde{t}^a\nabla_a\,\phi = \phi^a\nabla_a\,\tilde{t} = 0$.[24] Now consider the hybrid field $\phi'\colon \mathcal{R} \to \mathbb{R}\ (\mathrm{mod}\ 2\pi)$ defined by

$$\phi' = (\phi - k\,\tilde{t})\,(\mathrm{mod}\,2\pi).$$

It is adapted to $(\mathcal{R}, k)$ in the sense that it is constant on all integral curves of $(\tilde{t}^a + k\,\phi^a)$:

$$(\tilde{t}^n + k\,\phi^n)\nabla_n\,(\phi - k\,\tilde{t}) = \tilde{t}^n\nabla_n\,(-k\,\tilde{t}) + (k\,\phi^n)\nabla_n\,\phi = 0.$$

In particular, ϕ' is constant on γ. In contrast, ϕ' increases (respectively, decreases) uniformly with respect to elapsed parameter distance along λ_1 (respectively, λ_2) since $(\tilde{t}^n + l_i\,\phi^n)\nabla_n\,\phi' = (l_i - k)$. (It follows, as claimed that λ_1 and λ_2 *must* reintersect γ.)

---

24. We can introduce the coordinates as follows. Pick any initial point on $\mathcal{R}$ and take its coordinates to be $\tilde{t} = 0$ and $\phi = 0$. Given any other point on $\mathcal{R}$, we can "get to it" from the initial point by moving a certain (signed) parameter distance along an integral curve of $\tilde{t}^a$ and moving a certain (signed) parameter distance along an integral curve of ϕ^a. It does not matter in what order we perform the operations because the fields $\tilde{t}^a$ and ϕ^a have a vanishing Lie bracket. We take the respective parameter distances to be the $\tilde{t}$ and ϕ coordinates of the new point.

Let the points p_0, and p_1, p_2 have respective $\tilde{t}, \phi'$ coordinates $(\tilde{t}_0, \phi')$, $(\tilde{t}_1, \phi')$, and $(\tilde{t}_2, \phi')$. They share a common ϕ' coordinate since ϕ' is constant on γ. But ϕ' increases along λ_1 in the stretch between p_0 and p_1: it goes from 0 to 2π. Similarly, ϕ' decreases along λ_2 in the stretch between p_0 and p_2: it goes from 0 to -2π.

The coordinate $\tilde{t}$ increases along all three curves, γ, λ_1, and λ_2. (Indeed, we have $(\tilde{t}^n + k\,\phi^n)\nabla_n \tilde{t} = (\tilde{t}^n + l_i\,\phi^n)\nabla_n \tilde{t} = 1$.) So we can think of the curves as parametrized by $\tilde{t}$ and consider the rate of change of ϕ' with respect to $\tilde{t}$ on them. This rate of change on λ_i is (by the chain rule)

$$\frac{d\phi'}{d\tilde{t}} = \frac{(\tilde{t}^n + l_i\,\phi^n)\nabla_n\,(\phi - k\,\tilde{t})}{(\tilde{t}^n + l_i\,\phi^n)\nabla_n\,\tilde{t}} = (l_i - k).$$

So, considering the total change of ϕ' along λ_1 and λ_2, we have

$$2\pi = (\tilde{t}_1 - \tilde{t}_0)\,\frac{d\phi'}{d\tilde{t}}_{|\text{on}\,\lambda_1} = (\tilde{t}_1 - \tilde{t}_0)\,(l_1 - k),$$

$$-2\pi = (\tilde{t}_2 - \tilde{t}_0)\,\frac{d\phi'}{d\tilde{t}}_{|\text{on}\,\lambda_2} = (\tilde{t}_2 - \tilde{t}_0)\,(l_2 - k).$$

It follows that

$$\tilde{t}_1 - \tilde{t}_2 = \frac{2\,\pi\,(l_1 + l_2 - 2k)}{(l_1 - k)\,(l_2 - k)}.$$

Hence, by equations (3.2.6) and (3.2.7),

$$p_1 = p_2 \Longleftrightarrow \tilde{t}_1 = \tilde{t}_2 \Longleftrightarrow (l_1 + l_2 - 2k) = 0 \Longleftrightarrow k = -\frac{(\tilde{t}^a \phi_a)}{(\phi^n \phi_n)}.$$

This gives us equation (3.2.5). □

Now we consider the two criteria in the special case of Gödel spacetime. We start with a calculation.

PROPOSITION 3.2.3. *Let ϵ^{abcd} be a volume element on Gödel spacetime, and let η^a be the field $\tilde{t}^a + k\,\phi^c$ for some choice of k. Then*

(3.2.8) $$\epsilon^{abcd}\,\eta_{[b}\nabla_c\,\eta_{d]} = \pm\,2\left[k^2\,\sqrt{2}\,sh^4\,r + k\,(2\,sh^2\,r - 1) + \sqrt{2}\right]\tilde{z}^a$$

where, as in the previous section, $\tilde{z}^a = (\partial\,/\partial\,\tilde{z})^a$.

Note that in the special case where $k = 0$, this yields

$$\epsilon^{abcd}\,\tilde{t}_{[b}\nabla_c\,\tilde{t}_{d]} = \pm\, 2\sqrt{2}\,\tilde{z}^a.$$

If we re-express this in terms of $\hat{t}^a = t^a/\mu = \tilde{t}^a/(2\mu)$ and $z^a = \tilde{z}^a/2$, and choose a volume element so that the right side sign is $+1$, we recover equation (3.1.30); i.e.,

$$\frac{1}{2}\,\epsilon^{abcd}\,\hat{t}_b\,\nabla_c\,\hat{t}_d = \frac{1}{\sqrt{2}\,\mu^2}\,z^a.$$

Proof. As before, let A be the set of axis points in Gödel spacetime where $r = 0$, and let M^- be the complement set $M - A$. The vector fields

$$\tilde{t}^a = (\partial/\partial\,\tilde{t})^a, \quad r^a = (\partial/\partial\,r)^a, \quad \phi^a = (\partial/\partial\,\phi)^a, \quad \tilde{z}^a = (\partial/\partial\,\tilde{z})^a$$

are linearly independent on M^-. So we can express ϵ^{abcd} in the form

$$\epsilon^{abcd} = f\,\tilde{t}^{[a}\,r^b\,\phi^c\,\tilde{z}^{d]}$$

on M^-. We can determine f, up to sign, as follows. We certainly have

$$-(4!) = \epsilon^{abcd}\,\epsilon_{abcd} = f^2\,\tilde{t}^{[a}\,r^b\,\phi^c\,\tilde{z}^{d]}\,\tilde{t}_{[a}\,r_b\,\phi_c\,\tilde{z}_{d]} = f^2\,\tilde{t}^{[a}\,r^b\,\phi^c\,\tilde{z}^{d]}\,\tilde{t}_a\,r_b\,\phi_c\,\tilde{z}_d.$$

And by equation (3.1.31),

$$\begin{aligned}
\tilde{t}_a &= 4\,\mu^2\,[\sqrt{2}\,sh^2\,r\,\nabla_a\,\phi + \nabla_a\,\tilde{t}],\\
r_b &= 4\,\mu^2\,\nabla_b\,r,\\
\phi_c &= 4\,\mu^2\,[(sh^4\,r - sh^2\,r)\nabla_c\,\phi + \sqrt{2}\,sh^2\,r\,\nabla_c\,\tilde{t}],\\
\tilde{z}_d &= 4\,\mu^2\,\nabla_d\,\tilde{z}.
\end{aligned}$$

So

$$\begin{aligned}
-(4!) &= f^2\,\tilde{t}^{[a}\,r^b\,\phi^c\,\tilde{z}^{d]}\,(4\,\mu^2)^4\left[(sh^4\,r - sh^2\,r) - 2\,sh^4\,r\right](\nabla_a\tilde{t})\,(\nabla_b\,r)\,(\nabla_c\,\phi)\,(\nabla_d\,\tilde{z})\\
&= -f^2\,(4\,\mu^2)^4\,(sh^4\,r + sh^2\,r)\,\frac{1}{4\,!} = -f^2\,(4\,\mu^2)^4\,(sh^2\,r)\,(ch^2\,r)\,\frac{1}{4\,!}.
\end{aligned}$$

Thus, on M^-, we have

$$\epsilon^{abcd} = \pm\,\frac{4!}{16\,\mu^4\,(sh\,r)(ch\,r)}\,\tilde{t}^{[a}\,r^b\,\phi^c\,\tilde{z}^{d]}. \tag{3.2.9}$$

Next, we derive an expression for

$$\eta_{[b}\nabla_c\,\eta_{d]} = \tilde{t}_{[a}\nabla_b\,\tilde{t}_{c]} + k\,\tilde{t}_{[a}\nabla_b\,\phi_{c]} + k\,\phi_{[a}\nabla_b\,\tilde{t}_{c]} + k^2\,\phi_{[a}\nabla_b\,\phi_{c]} \tag{3.2.10}$$

on M^-. Note first that

$$\nabla_b \phi_c = 4\,\mu^2\,[(4\,sh^3\,r - 2\,sh\,r)(ch\,r)(\nabla_{[b}\,r)(\nabla_{c]}\,\phi) + 2\,\sqrt{2}\,(sh\,r)(ch\,r)\,(\nabla_{[b}\,r)(\nabla_{c]}\,\tilde{t})],$$

$$\nabla_b\,\tilde{t}_c = 4\,\mu^2\,[\,2\,\sqrt{2}\,(sh\,r)(ch\,r)(\nabla_{[b}\,r)(\nabla_{c]}\,\phi)\,],$$

both hold on M^-. (The first is equation (3.1.44). The second is derived similarly, using the fact that $\nabla_{(b}\,\tilde{t}_{c)} = \mathbf{0}$.) These expressions, together with the preceding ones for $\tilde{t}_a$ and ϕ_a, yield

$$\tilde{t}_{[a}\nabla_b\,\tilde{t}_{c]} = 16\,\mu^4\,2\,\sqrt{2}\,(sh\,r)(ch\,r)\,(\nabla_{[a}\,\tilde{t})(\nabla_b\,r)(\nabla_{c]}\,\phi), \tag{3.2.11}$$

$$\tilde{t}_{[a}\nabla_b\,\phi_{c]} = 16\,\mu^4\,[\,(4\,sh^3\,r - 2\,sh\,r)(ch\,r) - 4\,(sh^3\,r)(ch\,r)](\nabla_{[a}\,\tilde{t})(\nabla_b\,r)(\nabla_{c]}\,\phi), \tag{3.2.12}$$

$$\phi_{[a}\nabla_b\,\tilde{t}_{c]} = 16\,\mu^4\,4\,(sh^3\,r)(ch\,r)\,(\nabla_{[a}\,\tilde{t})(\nabla_b\,r)(\nabla_{c]}\,\phi), \tag{3.2.13}$$

$$\phi_{[a}\nabla_b\,\phi_{c]} = 16\,\mu^4\,[\,\sqrt{2}\,(sh^2\,r)(ch\,r)(4\,sh^3\,r - 2\,sh\,r) - 2\,\sqrt{2}\,(sh\,r)(ch\,r)(4\,sh^3\,r - 2\,sh\,r)\,](\nabla_{[a}\,\tilde{t})(\nabla_b\,r)(\nabla_{c]}\,\phi). \tag{3.2.14}$$

If we insert these expressions in equation (3.2.10), we arrive at

$$\eta_{[b}\nabla_c\,\eta_{d]} = 32\,\mu^4\,(sh\,r)(ch\,r)\,\left[k^2\,\sqrt{2}\,sh^4\,r + k\,(2\,sh^2\,r - 1) + \sqrt{2}\right](\nabla_{[a}\,\tilde{t})(\nabla_b\,r)(\nabla_{c]}\,\phi). \tag{3.2.15}$$

Finally, combining this result with equation (3.2.9) yields

$$\begin{aligned} \epsilon^{abcd}\,\eta_{[b}\nabla_c\,\eta_{d]} &= \pm\,\frac{4!}{16\,\mu^4\,(sh\,r)(ch\,r)}\,\tilde{t}^{[a}\,r^b\,\phi^c\,\tilde{z}^{d]}\,\eta_{[b}\nabla_c\,\eta_{d]} \\ &= \pm\,2\,(k^2\,\sqrt{2}\,sh^4\,r + k\,(2\,sh^2\,r - 1) + \sqrt{2}\,)\,\tilde{z}^a \end{aligned} \tag{3.2.16}$$

on M^-. Since both η^a and the final vector field in equation (3.2.16) are smooth (everywhere), the equation must hold on A as well. □

Our desired characterization result for Gödel spacetime follows as a corollary. (For clause (2), recall that $sh^2\,r_c = 1$.)

PROPOSITION 3.2.4. *Let $\mathcal{R}$ be a striated orbit cylinder in Gödel spacetime generated by $\tilde{t}^a$ and ϕ^a. It is characterized by particular values for r (where $0 < r < r_c$) and $\tilde{z}$. Let k be such that $\eta^a = \tilde{t}^a + k\,\phi^a$ is timelike on $\mathcal{R}$. Then the following both hold.*

(1) $(\mathcal{R}, k)$ *is non-rotating according to the ZAM criterion* $\Longleftrightarrow k = \dfrac{\sqrt{2}}{(1 - sh^2\, r)}$.

(2) $(\mathcal{R}, k)$ *is non-rotating according to the CIR criterion* $\Longleftrightarrow$

$$r < \frac{r_c}{2} \qquad \text{and} \qquad k = \frac{2\sqrt{2}}{(1 - 2\, sh^2\, r) + \sqrt{1 - sh^2\,(2r)}}.$$

Proof. Note that our assumption that $\tilde{t}^a + k\,\phi^a$ is timelike on $\mathcal{R}$ comes out as the assumption that the relation

$$k^2 (sh^4\, r - sh^2\, r) + k\, 2\sqrt{2}\, sh^2\, r + 1 > 0 \tag{3.2.17}$$

holds there. (We are making use of equation (3.1.31) here and shall do so repeatedly in what follows.)

$(\mathcal{R}, k)$ qualifies as non-rotating according to the ZAM criterion iff $(\tilde{t}^a + k\,\phi^a)\phi_a = 0$ on $\mathcal{R}$. The latter condition comes out as

$$\sqrt{2}\, sh^2\, r + k\,(sh^4\, r - sh^2\, r) = 0.$$

Moreover, as is easy to check, if $k = \sqrt{2}/(1 - sh^2\, r)$, then equation (3.2.17) is automatically satisfied; i.e., (3.2.17) imposes no further constraint on k in this case. So we have clause (1).

Next, $(\mathcal{R}, k)$ is non-rotating according to the CIR criterion iff $\eta_{[b}\nabla_c \eta_{d]} = \mathbf{0}$ on $\mathcal{R}$ or, equivalently, if $\epsilon^{abcd}\, \eta_{[b}\nabla_c\, \eta_{d]}$ vanishes there (for either choice of ϵ^{abcd}). We know from the preceding proposition that this is the case iff

$$k^2\, \sqrt{2}\, sh^4\, r + k\,(2\, sh^2\, r - 1) + \sqrt{2} = 0 \tag{3.2.18}$$

on $\mathcal{R}$. This equation has two roots:

$$k_1 = \frac{(1 - 2\, sh^2\, r) - \sqrt{1 - sh^2\,(2r)}}{2\sqrt{2}\, sh^4\, r} \quad \text{and} \quad k_2 = \frac{(1 - 2\, sh^2\, r) + \sqrt{1 - sh^2\,(2r)}}{2\sqrt{2}\, sh^4\, r}.$$

So it is a necessary condition for $(\mathcal{R}, k)$ to be non-rotating according to the CIR criterion (for any choice of k) that $sh^2\, 2r \leq 1$ or, equivalently, that $r \leq r_c/2$. So assume this condition holds. We claim that the root k_2 can be ruled out because it leads to a violation of equation (3.2.17). We also claim that k_1 *is* compatible with that inequality if we further restrict r so that $sh^2\, 2r < 1$. To see this, note that in the presence of (3.2.18), equation (3.2.17) holds iff

$$k^2\, \sqrt{2}\, sh^2\, r - k\,(2\, sh^2\, r + 1) < 0,$$

and this holds, in turn, iff

$$0 < k < \frac{2\, sh^2\, r + 1}{\sqrt{2}\, sh^2\, r}. \tag{3.2.19}$$

—–1
—– 0
—–+1

With a bit of straightforward algebra, one can easily check that k_2 violates this inequality but that k_1 satisfies it if $sh^2\, 2r < 1$. Finally, note that $X = [(1 - 2\, sh^2\, r) + \sqrt{1 - sh^2\, 2r}\,] \neq 0$. So we have

$$k_1 = k_1\, \frac{X}{X} = \frac{2\sqrt{2}}{(1 - 2\, sh^2\, r) + \sqrt{1 - sh^2\,(2r)}}.$$

This gives us (2). □

There are two regimes to consider here. If $0 < r < (r_c/2)$, there is one rotational state of the ring (i.e., one choice of k) that counts as non-rotating according to the ZAM criterion, and one that counts as non-rotating according to the CIR criterion, but the two are different. In contrast, if $(r_c/2) \leq r < r_c$, then there is still one rotational state of the ring that counts as non-rotating according to the ZAM criterion, but now there is no state whatsoever that counts as non-rotating according to the CIR criterion.

Notice that though the two criteria do not agree for any choice of r, there is a sense in which they agree "in the limit" as $r \to 0$. They have a common limiting value for k:

$$\lim_{r \to 0}\, \frac{\sqrt{2}}{(1 - sh^2\, r)} = \lim_{r \to 0}\, \frac{2\sqrt{2}}{(1 - 2\, sh^2\, r) + \sqrt{1 - sh^2\,(2r)}} = \sqrt{2}.$$

That this is so should not be surprising. We began this section by asserting that there is a robust, unambiguous notion of non-rotation at a point in relativity theory. Here, in a sense, we recover that notion as we pass to the limit of "infinitesimally small rings." Notice that $\sqrt{2}$ is the unique value of k for which $\eta^a = \tilde{t}^a + k\,\phi^c$ is non-rotating (i.e., satisfies $\eta_{[a}\nabla_b\, \eta_{d]} = \mathbf{0}$) at points on the axis where $r = 0$. (This follows immediately from proposition 3.2.3.) It is *that* value of k that we recover in the limit as $r \to 0$. This will be important in what follows.

Let us now leave Gödel spacetime behind and return to the general case with which we started (where we are dealing with an arbitrary stationary, axi-symmetric spacetime). We claimed earlier on the section that the two criteria of ring non-rotation *do* agree if a certain simplifying condition obtains. The condition we had in mind is the orthogonality of $\tilde{t}^a$ and ϕ^a. But, strictly speaking, that is not sufficient to guarantee agreement. We must, in addition, rule out one rather special, singular possibility. We characterize it in the next proposition. (We shall comment on the listed conditions after presenting a proof.)

PROPOSITION 3.2.5. *Suppose that (in addition to satisfying conditions (i) to (vi)), $\tilde{t}^a$ and ϕ^a are orthogonal; i.e., $\tilde{t}^a\phi_a = 0$. Then, for all orbit cylinders $\mathcal{R}$, the following conditions are equivalent.*

(1) $\nabla_a \left(\dfrac{\phi^b\,\phi_b}{\tilde{t}^c\,\tilde{t}_c}\right) = \mathbf{0}$ *on* $\mathcal{R}$.

(2) $\tilde{t}^a + \sqrt{\dfrac{-\tilde{t}^b\tilde{t}_b}{\phi^c\phi_c}}\;\phi^a$ *is a null, geodesic field on* $\mathcal{R}$.

(3) $(\mathcal{R}, k)$ *is non-rotating on the CIR criterion for all* k *(such that* $\tilde{t}^a + k\,\phi^a$ *is timelike on* $\mathcal{R}$*).*

Proof. It follows from our orthogonality assumption that the following conditions all hold on $\mathcal{R}$:

$$\tilde{t}^a\nabla_a\,\phi_b = \mathbf{0}, \tag{3.2.20}$$

$$\phi^a\nabla_a\,\tilde{t}_b = \mathbf{0}, \tag{3.2.21}$$

$$\tilde{t}_{[a}\nabla_b\,\tilde{t}_{c]} = \mathbf{0}, \tag{3.2.22}$$

$$\phi_{[a}\nabla_b\,\phi_{c]} = \mathbf{0}. \tag{3.2.23}$$

The first follows since we have

$$\tilde{t}^a\nabla_a\,\phi_b = -\tilde{t}^a\nabla_b\,\phi_a = -\nabla_b(\phi_a\tilde{t}^a) + \phi^a\nabla_b\,\tilde{t}_a = -\phi^a\nabla_a\,\tilde{t}_b = -\tilde{t}^a\nabla_a\,\phi_b.$$

(Here we use the fact that $\tilde{t}^a$ and ϕ^a are Killing fields for the first and third equalities, as well as the fact that they have a vanishing Lie bracket for the final equality.) That gives us equation (3.2.21) as well. For equation (3.2.22), we use condition (vi) in our original list. We have $\phi_{[a}\,\tilde{t}_b\nabla_c\,\tilde{t}_{d]} = \mathbf{0}$ or, equivalently,

$$\phi_a\,\tilde{t}_{[b}\nabla_c\,\tilde{t}_{d]} - \phi_d\,\tilde{t}_{[a}\nabla_b\,\tilde{t}_{c]} + \phi_c\,\tilde{t}_{[d}\nabla_a\,\tilde{t}_{b]} - \phi_b\,\tilde{t}_{[c}\nabla_d\,\tilde{t}_{a]} = \mathbf{0}.$$

Since contracting ϕ^a on any index in $\tilde{t}_{[a}\nabla_b\,\tilde{t}_{c]}$ yields $\mathbf{0}$, it follows that $(\phi^a\phi_a)\,\tilde{t}_{[b}\nabla_c\,\tilde{t}_{d]} = \mathbf{0}$. Since ϕ^a is spacelike on $\mathcal{R}$, it follows that equation (3.2.22) holds on $\mathcal{R}$ as well. The argument for equation (3.2.23) is very much the same. For that one we start with $\tilde{t}_{[a}\,\phi_b\nabla_c\,\phi_{d]} = \mathbf{0}$.

Let us first check that conditions (1) and (2) are equivalent. Consider the field

$$\eta^a = \tilde{t}^a + \sqrt{\frac{-\tilde{t}^b\tilde{t}_b}{\phi^c\phi_c}}\;\phi^a.$$

It is a null field by our orthogonality assumption. It follows from equations (3.2.20) and (3.2.21) that

$$\eta^a \nabla_a \eta_b = \tilde{t}^a \nabla_a \tilde{t}_b - \frac{(\tilde{t}^b \tilde{t}_b)}{(\phi^c \phi_c)} \phi^a \nabla_a \phi_b.$$

(We know that $\tilde{t}^a \nabla_a \left(\frac{\tilde{t}^b \tilde{t}_b}{\phi^c \phi_c} \right) = \phi^a \nabla_a \left(\frac{\tilde{t}^b \tilde{t}_b}{\phi^c \phi_c} \right) = 0$, even without the orthogonality assumption, just because ϕ^a and $\tilde{t}^a$ are commuting Killing fields.) So (2) holds iff

(3.2.24) $$(\phi^c \phi_c)\, \tilde{t}^a \nabla_a \tilde{t}_b - (\tilde{t}^b \tilde{t}_b)\, \phi^a \nabla_a \phi_b = \mathbf{0}.$$

But $2\, \tilde{t}^a \nabla_a \tilde{t}_b = -\nabla_b\, (\tilde{t}^a \tilde{t}_a)$ and $2\, \phi^a \nabla_a \phi_b = -\nabla_b\, (\phi^a \phi_a)$, since ϕ^a and $\tilde{t}^a$ are Killing fields. So this condition is equivalent to (1).

Now consider condition (3). $(\mathcal{R}, k)$ is non-rotating according to the CIR criterion iff

$$\tilde{t}_{[a} \nabla_b\, \tilde{t}_{c]} + k\, \phi_{[a} \nabla_b\, \tilde{t}_{c]} + k\, \tilde{t}_{[a} \nabla_b\, \phi_{c]} + k^2\, \phi_{[a} \nabla_b\, \phi_{c]} = \mathbf{0}$$

on $\mathcal{R}$. This reduces to

(3.2.25) $$k\, \left(\phi_{[a} \nabla_b\, \tilde{t}_{c]} + \tilde{t}_{[a} \nabla_b\, \phi_{c]} \right) = \mathbf{0}$$

in the case at hand by virtue of equations (3.2.22) and (3.2.23). So (3) holds iff

(3.2.26) $$\phi_{[a} \nabla_b\, \tilde{t}_{c]} + \tilde{t}_{[a} \nabla_b\, \phi_{c]} = \mathbf{0}$$

on $\mathcal{R}$. Now suppose equation (3.2.26) holds at a point. Then, contraction with $\phi^a\, \tilde{t}^b$ yields

$$(\phi^a \phi_a)\, \tilde{t}^b \nabla_b\, \tilde{t}_c + (\tilde{t}^b \tilde{t}_b)\, \phi^a \nabla_c\, \phi_a = \mathbf{0},$$

which is equivalent to equation (3.2.24). So we have the implication (3) $\Longrightarrow$ (2). For the converse, suppose that equation (3.2.24) holds at a point. Contracting equations (3.2.22) and (3.2.23) with $\tilde{t}^a$ and ϕ^a respectively, yields,

(3.2.27) $$(\tilde{t}^n \tilde{t}_n)\, \nabla_b\, \tilde{t}_c = \tilde{t}_b\, \tilde{t}^a \nabla_a\, \tilde{t}_c - \tilde{t}_c\, \tilde{t}^a \nabla_a\, \tilde{t}_b,$$

(3.2.28) $$(\phi^n \phi_n)\, \nabla_b\, \phi_c = \phi_b\, \phi^a \nabla_a\, \phi_c - \phi_c\, \phi^a \nabla_a\, \phi_b.$$

If we substitute for $\phi^a \nabla_a\, \phi_c$ in equation (3.2.28) using equation (3.2.24), it comes out as

(3.2.29) $$(\tilde{t}^n \tilde{t}_n)\, \nabla_b\, \phi_c = \phi_b\, \tilde{t}^a \nabla_a\, \tilde{t}_c - \phi_c\, \tilde{t}^a \nabla_a\, \tilde{t}_b.$$

It now follows from equations (3.2.27) and (3.2.29) that

$$(\tilde{t}^n \tilde{t}_n)\, \phi_{[a}\, \nabla_b\, \tilde{t}_{c]} = -(\tilde{t}^n \tilde{t}_n)\, \tilde{t}_{[a}\, \nabla_b\, \phi_{c]},$$

which gives us equation (3.2.26). So we have the implication (2) $\Longrightarrow$ (3). $\square$

We mention in passing that the conditions listed in the proposition can arise, for example, in Schwarzschild spacetime (Wald [60]). There we have (transferring our notation)

$$(\tilde{t}^c\, \tilde{t}_c) = 1 - \frac{2\,M}{r},$$

$$(\phi^b\, \phi_b) = -r^2,$$

where r is a radial coordinate. A simple calculation shows that

$$\nabla_a \left(\frac{\phi^b\, \phi_b}{\tilde{t}^c\, \tilde{t}_c} \right) = \mathbf{0} \Longleftrightarrow \frac{d}{dr} \left(-\frac{1}{r^2} + \frac{2\,M}{r^3} \right) = \mathbf{0} \Longleftrightarrow r = 3\,M.$$

So the conditions arise only for one special radius.

Notice that condition (1) cannot hold on all rings in an axi-symmetric spacetime if, for example, there are axis points in that spacetime. For if it did hold on all rings, then the function $(\phi^b\, \phi_b)/(\tilde{t}^c\, \tilde{t}_c)$ would be constant on the background manifold M. And since $\phi^a = \mathbf{0}$ at axis points, that constant value would have to be 0. But that is impossible, since ϕ^a is spacelike on non-axis points.

Consider the third condition in the list. It captures the claim that *all* (rigid motion) states of the ring qualify as non-rotating on the CIR criterion. This possibility may seem even more counterintuitive than the one we encountered in the case of a restricted region of Gödel spacetime—the region where $(r_c/2) \leq r < r_c$—where *no* (rigid motion) states of the ring qualified as non-rotating on that criterion. Abramowicz and coworkers [1, 2] has suggested a way of thinking about this situation that may be helpful.

Let us forget about our ring for a moment and consider what would happen if we carried a gyroscope in a straight line at a certain speed (possibly 0). Suppose that at some initial moment the axis of the gyroscope is co-aligned with the direction of motion (figure 3.2.6). Then we would expect it to remain

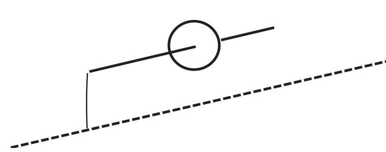

Figure 3.2.6. A gyroscope moving in a “straight line” will not change direction relative to that line.

co-aligned, no matter what the speed of transport. The speed seems irrelevant because the trajectory of the gyroscope involves no change in direction. But in the special case where condition (2) in the proposition obtains (we are now switching back to the case of the ring), there is a sense in which a gyroscope mounted on the ring *is* moving in a "straight line," no matter what the rotational state of the ring—at least if we use light rays as our standard for what constitutes motion in a straight line. For condition (2) asserts that light rays, by themselves, without the intervention of mirrors or lenses or other devices, will follow the ring.

With all this as preparation, we can formulate our proposition about the conditions under which the two criteria for ring non-rotation agree.

PROPOSITION 3.2.6. *Suppose that (in addition to satisfying conditions (i) to (vi) listed above) $\tilde{t}^a$ and ϕ^a are orthogonal. Let $\mathcal{R}$ be an orbit cylinder on which*

$$\nabla_a \left(\frac{\phi^b \, \phi_b}{\tilde{t}^c \, \tilde{t}_c} \right) \neq \mathbf{0}. \tag{3.2.30}$$

Finally, let k be a number for which $\tilde{t}^a + k\,\phi^a$ is timelike on $\mathcal{R}$. Then the following conditions are equivalent.

(1) $(\mathcal{R}, k)$ *is non-rotating according to the ZAM criterion.*
(2) $(\mathcal{R}, k)$ *is non-rotating according to the CIR criterion.*
(3) $k = 0$.

Proof. $(\mathcal{R}, k)$ is non-rotating according to the ZAM criterion iff $0 = (\tilde{t}^a + k\,\phi_a)\,\phi^a = k\,(\phi_a\,\phi^a)$. And ϕ^a is spacelike on $\mathcal{R}$. So the equivalence of (1) and (3) is immediate. (The added assumption about $\mathcal{R}$ is not needed for this equivalence.)

As we saw in the proof of the preceding proposition, $(\mathcal{R}, k)$ is non-rotating according to the CIR criterion iff

$$k\,\left(\phi_{[a}\nabla_b\,\tilde{t}_{c]} + \tilde{t}_{[a}\nabla_b\,\phi_{c]}\right) = \mathbf{0} \tag{3.2.31}$$

on $\mathcal{R}$. (Recall equation (3.2.25).) But we also saw in that proof that equation (3.2.30) is equivalent to

$$\phi_{[a}\nabla_b\,\tilde{t}_{c]} + \tilde{t}_{[a}\nabla_b\,\phi_{c]} \neq \mathbf{0}.$$

So $(\mathcal{R}, k)$ qualifies as non-rotating on the (CIR) criterion iff $k = 0$. $\square$

3.3. A No-Go Theorem about Orbital (Non-)Rotation

We have considered two particular criteria for non-rotation of the ring. Now we switch our attention to a large class of "generalized criteria" of non-rotation. We take any one such criterion (as applied in any one stationary axi-stationary spacetime) to be, simply, a specification, for every striated orbit cylinder $(\mathcal{R}, k)$ in that spacetime, whether it is to count as "non-rotating" or not. We do not insist in advance that the criterion have a natural geometric or quasi-operational formulation. Our plan is to consider three conditions that one might want such a criterion to satisfy—(i) relative rotation condition, (ii) limit condition, and (iii) non-vacuity condition—and then show that, at least in some stationary axi-stationary spacetimes, no generalized criterion of ring non-rotation satisfies all three. The proof of this no-go theorem is entirely elementary when all the definitions are in place. But it may be of some interest to *put* them in place and formulate a result of this type. The idea is to step back from the details of particular proposed criteria of non-rotation and direct attention instead to the conditions they do and do not satisfy.

Let us start with the relative rotation condition. Suppose we have two rings, R_1 and R_2, centered about the same axis of rotational symmetry. (Intuitively, we imagine that the planes of the rings are parallel but not necessarily coincident. See figure 3.3.1.) Suppose further that R_2 is not rotating relative to R_1. Then, one might think, either both rings should qualify as "non-rotating" or neither should. This is the requirement captured in the "relative rotation condition." What it means to say that R_2 is not rotating *relative* to R_1 is not entirely unambiguous. But all we need here is a sufficient condition for relative non-rotation of the rings. And it seems, at least, a plausible sufficient condition for this that, over time, there is no change in the distance between any point on one ring and any point on the other; i.e., the two rings together

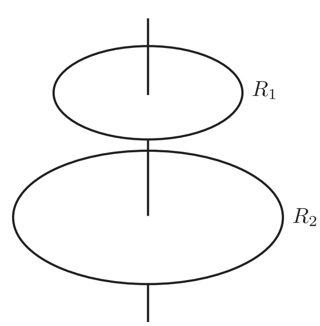

Figure 3.3.1. Two rings centered about the same axis of rotational symmetry.

form a rigid (ganged) system. So we are led to the following first formulation of the condition.

Relative Rotation Condition (intuitive formulation): Given two rings R_1 and R_2, if (i) R_1 is "non-rotating," and if (ii) R_2 is non-rotating relative to R_1 (in the sense that, given any point on R_2 and any point on R_1, the distance between them is constant over time), then R_2 is "non-rotating."

Now let us formulate a more precise version. Let (M, g_{ab}) be a stationary, axi-symmetric spacetime with Killing fields $\tilde{t}^a$ and ϕ^a, and let $(\mathcal{R}_1, k_1)$ and $(\mathcal{R}_2, k_2)$ be two striated orbit cylinders (as determined relative to $\tilde{t}^a$ and ϕ^a). (So, in particular, given how we have defined striated orbit cylinders, $(\tilde{t}^a + k_i\, \phi^a)$ is timelike on $\mathcal{R}_i$ for $i = 1, 2$.) Let γ_i be a striation curve—i.e., an integral curve of $(\tilde{t}^a + k_i\, \phi^a)$—in $\mathcal{R}_i$, for $i = 1, 2$. There are various ways we might try to determine the "distance" between γ_1 and γ_2. For example, we might bounce a light signal back and forth between them and keep track of how much time is needed for the round trip, as measured by a clock following one of the striation curves. But, presumably, no matter what procedure we use, the measured distance will be constant over time if γ_1 and γ_2 are integral curves of a common Killing field. (For, presumably, any reasonable measurement procedure can be characterized in terms of some set of relations and functions that are definable in terms g_{ab}, and all such relations and functions will be preserved under the isometries generated by the common Killing field.) So we seem to have a plausible sufficient condition for the relative non-rotation of $(\mathcal{R}_2, k_2)$ with respect to $(\mathcal{R}_1, k_1)$—namely, that there exists a (single) Killing field κ^a whose restriction to $\mathcal{R}_1$ is proportional to $(\tilde{t}^a + k_1\, \phi^a)$ and whose restriction to $\mathcal{R}_2$ is proportional to $(\tilde{t}^a + k_2\, \phi^a)$. But the latter condition holds immediately, of course, if $k_1 = k_2$.

The upshot of this long-winded discussion is the proposal that it is plausible to regard $(\mathcal{R}_2, k_2)$ as non-rotating relative to $(\mathcal{R}_1, k_1)$ if $k_1 = k_2$. (Again, all we need here is a sufficient condition for relative non-rotation.) So we take the relative rotation condition to be the following.

Relative Rotation Condition (precise formulation): For all k, and all striated orbit cylinders $(\mathcal{R}_1, k)$ and $(\mathcal{R}_2, k)$ sharing that k, if $(\mathcal{R}_1, k)$ qualifies as non-rotating, so does $(\mathcal{R}_2, k)$.

It follows immediately from proposition 3.2.6 that both the ZAM and CIR criteria satisfy the relative rotation condition in any stationary, axi-symmetric spacetime in which the Killing fields $\tilde{t}^a$ and ϕ^a are orthogonal—at least if one restricts attention to rings on which equation (3.2.30) holds. (For in that case, on either criterion, if $(\mathcal{R}_1, k)$ is non-rotating, it follows that $k = 0$; and if $k = 0$, it follows that $(\mathcal{R}_2, k)$ is non-rotating as well.) It also follows immediately from proposition 3.2.4 that neither criterion satisfies the relative rotation condition in Gödel spacetime. (For if, say, $0 < r_1 < r_2 < (r_c/2)$, then it is not the case that $\sqrt{2}/(1 - sh^2 r_1) = \sqrt{2}/(1 - sh^2 r_2)$; and it is not the case that the corresponding expressions that arise with the CIR criterion are equal.)

It is natural to ask whether there is *any* generalized criterion of rotation that satisfies the relative rotation condition in Gödel spacetime. The answer is, trivially, "yes". Indeed, given any stationary, axi-symmetric spacetime, there is a generalized criterion of rotation that satisfies the relative rotation condition in that spacetime. Intuitively, all one has to do is pick one ring in one rotational state arbitrarily, and then take other rings to be non-rotating iff they are non-rotating relative to that one. (Or, in the formal language, one need only pick one striated orbit cylinder $(\mathcal{R}, k)$ arbitrarily, and then take a striated orbit cylinder $(\mathcal{R}', k')$ to be non-rotating iff $k' = k$.)

The point of the no-go theorem that follows is to show that, though there do exist generalized criteria of non-rotation that satisfy the relative rotation condition in any particular stationary, axi-symmetric spacetime, none are fully satisfactory because (at least in some cases) they violate *other* conditions that we would want to see satisfied.

Consider, next, the limit condition. Recall our remarks about the asymptotic agreement of the ZAM and CIR criteria for "infinitely small rings" in Gödel spacetime. We suggested that this agreement should not be surprising because in relativity theory there is an unambiguous notion of non-rotation for a timelike vector field at a point, and we should *expect* any reasonable notion of orbital non-rotation for rings to deliver that notion in the limit. The limit condition simply makes that expectation precise. It asserts that if we have a sequence of orbit cylinders $\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3, \ldots$ that converges to a point p on the axis of rotational symmetry, and if we have a sequence of numbers $k_1, k_2, k_3, \ldots$ such that $(\mathcal{R}_i, k_i)$ qualifies as non-rotating for every i, then the latter sequence has a well-defined limit at p, and that limit is the correct one. What does "correct" mean here? Just as in the Gödel case, the limit value should be that (unique) k for which the field $(\tilde{t}^a + k\,\phi^a)$ is non-rotating at p.

That there *is* a unique k at each axis point satisfying the stated condition (in all stationary, axi-symmetric spacetimes) is confirmed in the following proposition. To avoid interruption, we hold its proof for an appendix.

PROPOSITION 3.3.1. *Let (M, g_{ab}) be a stationary, axi-symmetric spacetime with Killing fields $\tilde{t}^a$ and ϕ^a. Let p be a point at which $\phi^a = \mathbf{0}$. Then there is a unique* *number k such that* $\eta^a = \tilde{t}^a + k\,\phi^a$ *is non-rotating* ($\eta_{[a}\nabla_b\,\eta_{c]} = \mathbf{0}$) *at p. Its value is given by*

$$k_{crit}(p) = -\frac{(\nabla_b\,\tilde{t}_c)(\nabla^b\,\phi^c)}{(\nabla_m\,\phi_n)(\nabla^m\,\phi^n)}. \tag{3.3.1}$$

There is one point concerning our formulation of the limit condition that requires comment. We need to make clear what it means to say that "a sequence of orbit cylinders $\mathcal{R}_1$, $\mathcal{R}_2$, $\mathcal{R}_3$, ... converges to a point on the axis of rotational symmetry." Indeed, that provisional language is somewhat misleading. It must be remembered that the axis set where $\phi^a = \mathbf{0}$ forms a two-dimensional submanifold of our background stationary, axi-symmetric spacetime. (This fact is not brought out by the figures displayed to this point because they suppress one dimension.) So, for example, in Gödel spacetime, the axis set consists of all points with r coordinate 0 but with arbitrary $\tilde{t}$ and $\tilde{z}$ coordinates. What the sequence $\mathcal{R}_1$, $\mathcal{R}_2$, $\mathcal{R}_3$, ... can converge to, strictly speaking, is not a point p in the axis set but rather an integral curve γ of the Killing field $\tilde{t}^a$ that is, itself, fully contained within the two-dimensional axis set. (In the case of Gödel spacetime, these are curves characterized by r value 0, and some fixed value for $\tilde{z}$, but arbitrary values for $\tilde{t}$.) And we can understand convergence here to mean, simply, that given any point p on γ and any open set O containing p, there is an N such that $\mathcal{R}_i$ intersects O for all $i \geq N$.

Finally, note that because these limit curves *are* integral curves of $\tilde{t}^a$ on which $\phi^a = \mathbf{0}$—and so are mapped onto themselves by all isometries generated by $\tilde{t}^a$ and ϕ^a—the number $k_{crit}(p)$ in our proposition must be the same for all points p on them.

With all this by way of preparation, we now formulate the limit condition officially as follows.[25]

> *Limit Condition*: Let γ be an integral curve of $\tilde{t}^a$ on which $\phi^a = \mathbf{0}$. Let $\mathcal{R}_1$, $\mathcal{R}_2$, $\mathcal{R}_3$, ... be a sequence of orbit cylinders that converges to γ. And let $k_1, k_2, k_3, \ldots$ be a sequence of numbers such that $(\mathcal{R}_i, k_i)$ qualifies as non-rotating for every i. Then $\lim\limits_{i\to\infty} k_i = k_{crit}(p)$, where p is any point on γ.

---

25. Our formulation here is slightly different from that in Malament [41] in that we avoid reference to the "center point of the ring." That notion played a role in [41] in the characterization of the CIA criterion of ring non-rotation, but has not been used here.

Though it will play no role in what follows, we claim (without proof) that the ZAM and CIR criteria of non-rotation satisfy this limit condition in *all* stationary, axi-symmetric spacetimes, not just in Gödel spacetime.

The first questions to ask is whether there is any generalized criterion of non-rotation for the ring that satisfies both the relative rotation condition and the limit condition in Gödel spacetime. The answer is certainly "yes" again. In that spacetime, $k_{crit}(p) = \sqrt{2}$ for all points p in the axis set. So it suffices to take the following as our criterion: given any striated orbit cylinder $(\mathcal{R}, k)$, it counts as non-rotating precisely if $k = \sqrt{2}$. It trivially satisfies both the relative rotation and limit conditions.

Moreover, there is a cheap sense in which one can *always* find a generalized criterion of non-rotation that satisfies the two conditions—i.e., in any stationary, axi-symmetric spacetime. It is the degenerate criterion according to which no striated orbit cylinder whatsoever counts as non-rotating. As a matter of simple logic, it vacuously satisfies both conditions. The non-vacuity condition rules out this uninteresting possibility.

Non-Vacuity Condition: There is at least one striated orbit cylinder $(\mathcal{R}, k)$ that qualifies as non-rotating.

We have just seen that there is a criterion of non-rotation that satisfies all three conditions in Gödel spacetime. But Gödel spacetime is rather special within the class of stationary, axi-symmetric spacetimes because it has the Killing field $\tilde{z}^a$ in addition to $\tilde{t}^a$ and ϕ^a. As a result, given any two axis points in Gödel spacetime, there is an isometry that takes the first to the second. So it *must* be the case that the function k_{crit} has the same value at all axis points. But there are stationary, axi-symmetric spacetimes in which it does *not* have the same value at all axis points (we shall give an example in a moment), and in those there *is* no generalized criterion of non-rotation that satisfies all three conditions.

PROPOSITION 3.3.2. *Let (M, g_{ab}) be a stationary, axi-symmetric spacetime. It admits a generalized criterion of ring non-rotation that satisfies the relative rotation, limit, and non-vacuity conditions iff $k_{crit}(p) = k_{crit}(p')$ for all axis points p and p'.*

Proof. (*If*) Suppose there is a number k_{crit} such that $k_{crit}(p) = k_{crit}$ for all axis points p. Then, trivially, there is a criterion of ring non-rotation that satisfies the three conditions, namely the one according to which a striated orbit cylinder $(\mathcal{R}, k)$ counts as non-rotating iff $k = k_{crit}$.

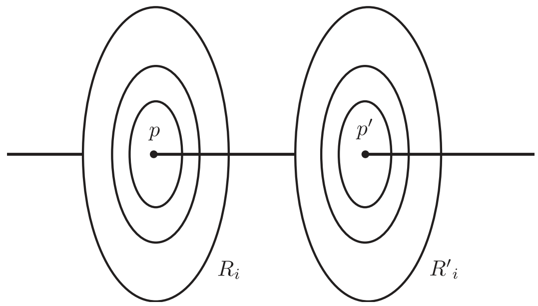

Figure 3.3.2. Two sequences of rings $\{R_i\}$ and $\{R'_i\}$ converging to points p and p', respectively, on the axis of rotational symmetry.

(*Only if*) Suppose there exist axis points p and p' such that $k_{crit}(p) \neq k_{crit}(p')$. Let γ and γ' be the (maximally extended) integral curves of $\tilde{t}^a$ that contain p and p', respectively. Further, let $\mathcal{R}_1$, $\mathcal{R}_2$, $\mathcal{R}_3, \ldots$ and $\mathcal{R}'_1$, $\mathcal{R}'_2$, $\mathcal{R}'_3, \ldots$ be sequences of orbit cylinders that converge to γ and γ', respectively (figure 3.3.2). (Existence is guaranteed. Let $p_1, p_2, p_3, \ldots$ be any sequence of points converging to p and, for all i, let $\mathcal{R}_i$ be the (unique) orbit cylinder the contains p_i. ($\mathcal{R}_i$ is the set of all points of the form $\psi(p)$, where ψ is an isometry generated by $\tilde{t}^a$ and ϕ^a.) Then $\mathcal{R}_2$, $\mathcal{R}_3, \ldots$ converges to γ. And $\mathcal{R}'_1$, $\mathcal{R}'_2$, $\mathcal{R}'_3, \ldots$ can be generated in the same way.) Now assume there is a generalized criterion of ring non-rotation $\mathcal{C}$ that satisfies all three conditions. By the non-vacuity condition, there is a striated orbit cylinder $(\mathcal{R}, k)$ that is non-rotating according to $\mathcal{C}$. For all sufficiently large i, $(\mathcal{R}_i, k)$ and $(\mathcal{R}'_i, k)$ are striated orbit cylinders; i.e., $\tilde{t}^a + k\,\phi^a$ is timelike on $\mathcal{R}_i$ and $\mathcal{R}'_i$. So (because we can always dispose of particular finite initial segments), we may as well assume that $(\mathcal{R}_i, k)$ and $(\mathcal{R}'_i, k)$ are striated orbit cylinders for all i. By the relative rotation condition, then, $(\mathcal{R}_i, k)$ and $(\mathcal{R}'_i, k)$ are non-rotating according to $\mathcal{C}$ for all i. Therefore, by the limit condition applied to $(\mathcal{R}_1, k), (\mathcal{R}_2, k), (\mathcal{R}_3, k), \ldots$ and $(\mathcal{R}'_1, k), (\mathcal{R}'_2, k), (\mathcal{R}'_3, k), \ldots,$ it must be the case that $k_{crit}(p) = k = k_{crit}(p')$, contradicting our initial assumption. So we may conclude that there is no generalized criterion of ring non-rotation $\mathcal{C}$ that satisfies all three conditions. $\square$

For the no-go theorem, we need now only exhibit a stationary, axisymmetric spacetime in which it is *not* the case that $k_{crit}(p) = k_{crit}(p')$ for all axis points p and p'. One example is Kerr spacetime (Wald [60] and O'Neill [47]). We shall say only enough about it to establish this one fact. In Boyer-Lindquist (spherical) coordinates $\tilde{t}, r, \phi, \theta$, the metric is

$$
\begin{aligned}
g_{ab} = {} & \left(1-\frac{2\,M\,r}{\rho^2}\right)(d_a\tilde{t})(d_b\tilde{t}) - \frac{\rho^2}{\Delta}\,(d_a r)(d_b r) - \rho^2\,(d_a\theta)(d_b\theta) \\
& - \left[r^2 + a^2 + \frac{2\,M\,r\,a^2\,sin^2\,\theta}{\rho^2}\right](sin^2\,\theta)\,(d_a\phi)(d_b\phi) \\
& + \frac{4\,M\,r\,a\,sin^2\,\theta}{\rho^2}\,(d_{(a}\tilde{t})(d_{b)}\phi),
\end{aligned}
$$

where

$$
\begin{aligned}
\rho^2 &= r^2 + a^2\,cos^2\,\theta, \\
\Delta &= r^2 - 2\,M\,r + a^2,
\end{aligned}
$$

and M and a are positive constants (O'Neill [47]). The axis set A here consists of all points at which $\sin\theta = 0$, for it is at those points at which the rotational Killing field $\phi^a = (\partial/\partial\,\phi)^a$ vanishes. (So every point in A is uniquely characterized by its $\tilde{t}$ and r coordinates.) It is not the case that $\tilde{t}^a = (\partial/\partial\,\tilde{t})^a$ is timelike and ϕ^a is spacelike at all points in $M^- = (M - A)$. But those conditions do obtain in restricted regions of interest—e.g., in the open set where $r > 2\,M$. If we think of Kerr spacetime as representing the spacetime structure surrounding a rotating black hole, our interest will be in small rings that are positioned close to the axis of rotational symmetry (where $\sin^2\theta$ is small) and far away from the center (where r is large). There we can sidestep all complexities having to do with horizons and singularities. The proposition we need is the following.

PROPOSITION 3.3.3. *Let p be an axis point in Kerr spacetime with coordinates $\tilde{t}$ and $r > 2M$. Then*

$$
k_{crit}(p) = \frac{2\,M\,r\,a}{(r^2+a^2)^2}. \tag{3.3.2}
$$

(So k_{crit} does not assume the same value at all axis points.)

Proof. We can certainly verify equation (3.3.2) directly by computing

$$
\frac{(\nabla_b\,\tilde{t}_c)(\nabla^b\,\phi^c)}{(\nabla_m\,\phi_n)(\nabla^m\,\phi^n)}
$$

at p and then invoking equation (3.3.1). But we can save ourselves a bit of work with an alternate approach that focuses attention on the smooth function $f : M^- \to \mathbb{R}$ defined by

$$
f = -\frac{(\tilde{t}^a\,\phi_a)}{(\phi^n\,\phi_n)} = \frac{2\,M\,r\,a}{(r^2+a^2)\rho^2 + 2\,M\,r\,a^2\,sin^2\,\theta}.
$$

Consider the field $\eta^a = \tilde{t}^a + f\,\phi^a$ on M^-. We claim that it can be expressed in the form

$$\eta^a = \frac{\nabla^a \tilde{t}}{(\nabla_n \tilde{t})(\nabla^n \tilde{t})}. \tag{3.3.3}$$

To see this, let $D = (\tilde{t}^a\,\tilde{t}_a)(\phi^a\,\phi_a) - (\tilde{t}^n\,\phi_n)^2$. Clearly, $D < 0$ on M^- (since ϕ^a is spacelike there). We have

$$\nabla_a\,\tilde{t} = \frac{1}{D}\left[(\phi^n\,\phi_n)\,\tilde{t}_a - (\tilde{t}^n\,\phi_n)\,\phi_a\right].$$

(This follows since both sides yield the same result when contracted with $\tilde{t}^a, \phi^a, r^a$, and θ^a.) Hence

$$(\nabla_n\,\tilde{t})(\nabla^n\,\tilde{t}) = \frac{1}{D}\,(\phi^n\phi_n)$$

and, therefore,

$$\frac{\nabla^a\,\tilde{t}}{(\nabla_n\,\tilde{t})(\nabla^n\,\tilde{t})} = \frac{(\phi^n\,\phi_n)\,\tilde{t}^a - (\tilde{t}^n\,\phi_n)\,\phi^a}{(\phi^n\phi_n)} = \eta^a,$$

as claimed. The right side of equation (3.3.3) has the form $g\,\nabla^a\tilde{t}$. It follows that $\eta_{[a}\nabla_b\,\eta_{c]} = \mathbf{0}$ everywhere on M^-.

Now f and η^a can be smoothly extended to A. At p, the extended function assumes the value

$$f(p) = \frac{2\,M\,r\,a}{(r^2 + a^2)^2}$$

(since, once again, the axis points here are ones where $\sin\theta = 0$). So, at p, the extended vector field satisfies

$$\begin{aligned}\mathbf{0} = \eta_{[a}\nabla_b\,\eta_{c]} &= \tilde{t}_{[a}\nabla_b\,\tilde{t}_{c]} + f(p)\,\tilde{t}_{[a}\nabla_b\,\phi_{c]}\\ &\quad + f(p)\,\phi_{[a}\nabla_b\,\tilde{t}_{c]} + f(p)^2\,\phi_{[a}\nabla_b\,\phi_{c]}.\end{aligned}$$

But we know from proposition 3.3.1 that the final expression on the right can be $\mathbf{0}$ only if $f(p) = k_{crit}(p)$. So we are done. □

Our main result now follows as an immediate corollary.

PROPOSITION 3.3.4. *(No-Go Theorem) There is no criterion of ring non-rotation on Kerr spacetime that satisfies the relative rotation, limit, and non vacuity conditions.*

It is intended to bear this interpretation: *given any (non-vacuous) generalized criterion of ring non-rotation in Kerr spacetime, to the extent that it gives "correct" attributions of non-rotation in the limit for infinitely small rings—the domain where one does have an unambiguous notion of non-rotation—it must violate the relative rotation condition.*

Appendix: The Proof of Proposition 3.3.1

Here we prove proposition 3.3.1. It will be convenient to collect a few facts first that will be used in the proof.

PROPOSITION 3.3.5. *Let (M, g_{ab}) be a stationary, axi-symmetric spacetime with Killing fields $\tilde{t}^a$ and ϕ^a. Let p be an axis point. (So $\phi^a = \mathbf{0}$ at p.) Let ϵ_{abcd} be a volume element defined on some open set O containing p, and let σ^a be the smooth field on O defined by $\sigma^a = \epsilon^{abcd}\,\tilde{t}_b \nabla_c\,\phi_d$. Then at p,*

(1) $\sigma^a \neq \mathbf{0}$

(2) $\nabla_a\,\phi_b = \dfrac{1}{2\,(\tilde{t}^n\tilde{t}_n)}\,\epsilon_{abcd}\,\tilde{t}^c\sigma^d.$

Furthermore, given any smooth field ψ^a (defined on some open set containing p), if $\pounds_\phi\,\psi^a = \mathbf{0}$ at p, then it must be of the form $\psi^a = k_1\,\tilde{t}^a + k_2\,\sigma^a$ at p.

Proof. Note that σ^a is orthogonal to $\tilde{t}^a$ and ϕ^a throughout O. (The first claim follows just because ϵ_{abcd} is anti-symmetric, and the second by clause (vi) in our characterization of stationary, axi-symmetric spacetimes.) Note, as well, that

$$\tilde{t}_{[a}\nabla_b\,\phi_{c]} = \frac{1}{6}\,\epsilon_{abcd}\,\sigma^d \tag{3.3.4}$$

throughout O. (We get this by contracting both sides of $\sigma^d = \epsilon^{dmnp}\,\tilde{t}_m\nabla_n\,\phi_p$ with ϵ_{abcd}.) Now we argue for (1). Suppose that $\sigma^a = \mathbf{0}$ at p. Then, by equation (3.3.4),

$$\mathbf{0} = \tilde{t}^a\,\tilde{t}_{[a}\nabla_b\,\phi_{c]} = \frac{1}{3}\left[(\tilde{t}^a\tilde{t}_a)\nabla_b\,\phi_c + \tilde{t}_c\,\tilde{t}^a\nabla_a\,\phi_b - \tilde{t}_b\,\tilde{t}^a\nabla_a\,\phi_c\right]$$

at p. Now $\tilde{t}^a\nabla_a\,\phi_b = \phi^a\nabla_a\,\tilde{t}_b$ everywhere on O (since the fields $\tilde{t}^a$ and ϕ^a have a vanishing Lie bracket), and $\phi^a = \mathbf{0}$ at p. So the second and third terms on the right vanish there. Thus $\nabla_a\,\phi_b = \mathbf{0}$ at p. But this is impossible. For given any Killing field κ^a on the (connected) manifold M, if κ^a and $\nabla_a\,\kappa_b$ *both* vanish at any one point, then they must vanish everywhere. (See Wald [60, p. 443].) And that is not possible in the present case because ϕ^a is spacelike at all non-axis points (and *some* non-axis points exist). So we have (1). And for (2) we need only contract both sides of equation (3.3.4) with $\tilde{t}^c$, expand the left side, and use much the same argument we have just used to show that two terms in the expansion vanish.

Finally, let ψ^a be a smooth field (defined on some open set containing p) such that $\mathcal{L}_\phi\, \psi^a = \mathbf{0}$ at p. Then $\psi^a \nabla_a \phi_b = \phi^a \nabla_a \psi_b = \mathbf{0}$ at p (since, once again, $\phi^a = \mathbf{0}$ at p). Hence, by (2), $\epsilon_{abcd}\, \psi^a\, \tilde{t}^c \sigma^d = \mathbf{0}$. So the three vectors ψ^a, $\tilde{t}^a$, and σ^a are linearly dependent at p. Since $\tilde{t}^a$ and σ^a are non-zero at p, ψ^a can be expressed as a linear combination of them at p. □

Now for the proof of proposition 3.3.1. The formulation, once again, is as follows.

Let (M, g_{ab}) be a stationary, axi-symmetric spacetime with Killing fields $\tilde{t}^a$ and ϕ^a. Further, let p be a point at which $\phi^a = \mathbf{0}$. Then there is a unique number k such that $\eta^a = \tilde{t}^a + k\,\phi^a$ is non-rotating ($\eta_{[a} \nabla_b\, \eta_{c]} = \mathbf{0}$) at p, and that number is

$$-\frac{(\nabla_b\, \tilde{t}_c)(\nabla^b\, \phi^c)}{(\nabla_m\, \phi_n)(\nabla^m\, \phi^n)}.$$

Proof. For the first claim, what we need to show is that there a unique k such that

$$\tilde{t}_{[a} \nabla_b\, \tilde{t}_{c]} + k\, \tilde{t}_{[a} \nabla_b\, \phi_{c]} = \mathbf{0} \tag{3.3.5}$$

at p. (This is equivalent to $\eta_{[a} \nabla_b\, \eta_{c]} = \mathbf{0}$ at p since $\phi^a = \mathbf{0}$ there.) We know from clause (1) of the preceding proposition and equation (3.3.4) that $\tilde{t}_{[a} \nabla_b\, \phi_{c]} \neq \mathbf{0}$ at p. So uniqueness is immediate. For existence, let ϵ_{abcd} be a volume element defined on some open set containing p, let $\sigma^a = \epsilon^{abcd}\, \tilde{t}_b \nabla_c\, \phi_d$ (as in the preceding proposition), and let $\omega^a = \epsilon^{abcd}\, \tilde{t}_b \nabla_c\, \tilde{t}_d$. The new field ω^a is orthogonal to $\tilde{t}^a$. And it is Lie derived by ϕ^a; i.e., $\mathcal{L}_\phi\, \omega^a = \mathbf{0}$ (since ϕ^a is a Killing field that commutes with $\tilde{t}^a$). So, by the preceding proposition, there is a number k_2 such that $\epsilon^{abcd}\, \tilde{t}_b \nabla_c\, \tilde{t}_d = \omega^a = k_2\, \epsilon^{abcd}\, \tilde{t}_b \nabla_c\, \phi_d$ or, equivalently,

$$\tilde{t}_{[a} \nabla_b\, \tilde{t}_{c]} = k_2\, \tilde{t}_{[a} \nabla_b\, \phi_{c]}$$

at p. Thus equation (3.3.5) holds at p iff $k = -k_2$.

Now we compute k_2. Contracting the preceding line with $\tilde{t}^a\, \nabla^b \phi^c$, and then dividing by $(\tilde{t}^a \tilde{t}_a)$, yields

$$(\nabla_b\, \tilde{t}_c)(\nabla^b \phi^c) = k_2\, (\nabla_b\, \phi_c)(\nabla^b \phi^c)$$

at p. So, to complete the proof, we need only verify that $(\nabla_b \phi_c)(\nabla^b \phi^c) \neq 0$ at p. But this follows from the preceding proposition. By clause (2) we have

$$(\nabla_b \phi_c)(\nabla^b \phi^c) = \frac{1}{2\,(\tilde{t}^n \tilde{t}_n)}\, \epsilon_{bcmn}\, \tilde{t}^m \sigma^n (\nabla^b \phi^c) = -\frac{1}{2\,(\tilde{t}^n \tilde{t}_n)}\, (\sigma_n \sigma^n)$$

at p. And σ^a is spacelike at p, since it is orthogonal to $\tilde{t}^a$ and (by clause (1)) non-zero there. □

4

NEWTONIAN GRAVITATION THEORY

The "geometrized" formulation of Newtonian gravitation theory—also known as "Newton-Cartan theory"—was first introduced by Cartan [5, 6] and Friedrichs [21] and later developed by Dautcourt [10], Dixon [11], Dombrowski and Horneffer [13], Ehlers [15], Havas [28], Künzle [34, 35], Lottermoser [37], Trautman [59], and others. It is significant for several reasons.

First, it shows that several features of relativity theory once thought to be uniquely characteristic of it do not distinguish it from (a suitably reformulated version of) Newtonian gravitation theory. The latter, too, can be cast as a "generally covariant" theory in which (i) gravity emerges as a manifestation of spacetime curvature, and (ii) spacetime structure is dynamical—i.e., participates in the unfolding of physics rather than being a fixed backdrop against which it unfolds.

Second, it clarifies the gauge status of the Newtonian gravitational potential. In the geometrized formulation of Newtonian theory, one works with a single curved derivative operator $\overset{g}{\nabla}$. It can be decomposed (in a sense) into two pieces—a flat derivative operator ∇ and a gravitational potential ϕ—to recover the standard formulation of the theory. But in the absence of special boundary conditions, the decomposition will not be unique. Physically, there is no unique way to divide into "inertial" and "gravitational" components the forces experienced by particles. Neither has any direct physical significance. Only their "sum" does. It is an attractive feature of the geometrized formulation that it trades in two gauge quantities for this sum. (See the discussion at the end of section 4.2.)

Third, the clarification just described also leads to a solution, or dissolution, of an old problem about Newtonian gravitation theory, namely the apparent breakdown of the theory when applied (in cosmology) to a hypothetically infinite, homogeneous mass distribution. (See section 4.4.)

"530-47773_Ch04_2P.tex" — 1/23/2012 — 17:18 — page 248

Fourth, it allows one to make precise, in coordinate-free, geometric language, the standard claim that Newtonian gravitation theory (or, at least, a certain generalized version of it) is the "classical limit" of general relativity. (See Künzle [35], Ehlers [15], and Lottermoser [37].)

4.1. Classical Spacetimes

We begin our discussion by characterizing a new class of geometric models for the spacetime structure of our universe (or subregions thereof) that is broad enough to include the models considered in both the standard and geometrized versions of Newtonian gravitation theory.

We take a *classical spacetime* to be a structure $(M, t_{ab}, h^{ab}, \nabla)$ where (i) M is a smooth, connected, four-dimensional manifold; (ii) t_{ab} is a smooth, symmetric field on M of signature $(1, 0, 0, 0)$; (iii) h^{ab} is a smooth, symmetric field on M of signature $(0, 1, 1, 1)$; (iv) ∇ is a derivative operator on M; and (v) the following two conditions hold:

$$h^{ab}\, t_{bc} = \mathbf{0}. \tag{4.1.1}$$

$$\nabla_a\, t_{bc} = \mathbf{0} \quad \text{and} \quad \nabla_a\, h^{bc} = \mathbf{0}. \tag{4.1.2}$$

We refer to them, respectively, as the "orthogonality" and "compatibility" conditions.

M is interpreted as the manifold of point events (as before). Collectively, the objects t_{ab}, h^{ab}, and ∇ on M represent the spacetime structure presupposed by classical Galilean relativistic dynamics. It will soon emerge how they do so.

We need to explain what we mean by the "signatures" of t_{ab} and h^{ab}, since we are using the term here in a new, somewhat extended sense. The signature condition for t_{ab} is the requirement that, at every point in M, the tangent space there have a basis $\overset{1}{\xi}{}^a, \ldots, \overset{4}{\xi}{}^a$ such that, for all i and j in $\{1, 2, 3, 4\}$, $t_{ab}\, \overset{i}{\xi}{}^a \overset{j}{\xi}{}^b = 0$ if $i \neq j$, and

$$t_{ab}\, \overset{i}{\xi}{}^a \overset{i}{\xi}{}^b = \begin{cases} 1 & \text{if } i = 1 \\ 0 & \text{if } i = 2,\ 3,\ 4. \end{cases}$$

(We shall call this an "orthonormal basis" for t_{ab}, though this does involve a slight extension of ordinary usage.) Hence, given any vectors $\mu^a = \sum_{i=1}^4 \overset{i}{\mu}\, \overset{i}{\xi}{}^a$ and $\nu^a = \sum_{i=1}^4 \overset{i}{\nu}\, \overset{i}{\xi}{}^a$ at the point,

$$t_{ab}\, \mu^a \nu^b = \overset{1}{\mu}\, \overset{1}{\nu} \tag{4.1.3}$$

and

$$t_{ab}\,\mu^a\mu^b = (\overset{1}{\mu})^2 \;\geq\; 0. \tag{4.1.4}$$

Notice that t_{ab} is not a metric as defined in section 1.9, since it does not satisfy the required non-degeneracy condition. (For example, if the vectors $\overset{1}{\xi}{}^a, \ldots, \overset{4}{\xi}{}^a$ are as above at some point, then $t_{ab}\,\overset{2}{\xi}{}^a = \mathbf{0}$ there, even though $\overset{2}{\xi}{}^a \neq \mathbf{0}$.)

The signature condition for h^{ab}, similarly, is the requirement that, at every point, the cotangent space there have a basis $\overset{1}{\sigma}_a, \ldots, \overset{4}{\sigma}_a$ such that, for all i and j in $\{1, 2, 3, 4\}$, $h^{ab}\,\overset{i}{\sigma}_a\overset{j}{\sigma}_b = 0$ if $i \neq j$, and

$$h^{ab}\,\overset{i}{\sigma}_a\overset{i}{\sigma}_b = \begin{cases} 0 & \text{if } i = 1 \\ 1 & \text{if } i = 2,\ 3,\ 4. \end{cases}$$

(We shall extend ordinary usage once again and call this an "orthonormal basis" for h^{ab}.) Hence, given any vectors $\alpha_a = \sum_{i=1}^{4} \overset{i}{\alpha}\,\overset{i}{\sigma}_a$ and $\beta_a = \sum_{i=1}^{4} \overset{i}{\beta}\,\overset{i}{\sigma}_a$ at the point,

$$h^{ab}\alpha_a\beta_b = \overset{2}{\alpha}\,\overset{2}{\beta} + \overset{3}{\alpha}\,\overset{3}{\beta} + \overset{4}{\alpha}\,\overset{4}{\beta} \tag{4.1.5}$$

and

$$h^{ab}\alpha_a\alpha_b = (\overset{2}{\alpha})^2 + (\overset{3}{\alpha})^2 + (\overset{4}{\alpha})^2 \geq 0. \tag{4.1.6}$$

Notice, too, that h^{ab} is not the inverse of a metric (in the sense of section 1.9); i.e., there is no field h_{ab} such that $h_{ab}h^{bc} = \delta^c{}_a$. (Why? If $\overset{1}{\sigma}_a, \ldots, \overset{4}{\sigma}_a$ are as in the preceding paragraph at some point, then $h^{ab}\,\overset{1}{\sigma}_a = \mathbf{0}$. Hence, if there were a tensor h_{ab} at the point such that $h_{ab}\,h^{bc} = \delta^c{}_a$, it would follow that $\mathbf{0} = h_{ab}\,h^{bc}\,\overset{1}{\sigma}_c = \delta^c{}_a\,\overset{1}{\sigma}_c = \overset{1}{\sigma}_a$, contradicting the assumption that $\overset{1}{\sigma}_a, \ldots, \overset{4}{\sigma}_a$ form a basis of the cotangent space there.)

In what follows, let $(M, t_{ab}, h^{ab}, \nabla)$ be a fixed classical spacetime.

Consider, first, t_{ab}. We can think of it as a "temporal metric," even though it is not a metric in the sense of section 1.9. Given any vector ξ^a at a point, we take its "temporal length" to be $(t_{ab}\,\xi^a\,\xi^b)^{\frac{1}{2}}$. (We know from equation (4.1.4) that $(t_{ab}\,\xi^a\,\xi^b)$ must be non-negative.) We further classify ξ^a as either *timelike* or *spacelike*, depending on whether its temporal length is positive or zero. It follows from the signature of t_{ab} that the subspace of spacelike vectors at any point is three-dimensional. (For if $\overset{1}{\xi}{}^a, \ldots, \overset{4}{\xi}{}^a$ is an orthonormal basis for t_{ab} there, $\overset{1}{\xi}{}^a$ is timelike, and the remaining three are spacelike.) Notice too that at any point we can find a co-vector t_a, unique up to sign, such that $t_{ab} = t_a t_b$. (Again, let $\overset{1}{\xi}{}^a, \ldots, \overset{4}{\xi}{}^a$ be an orthonormal basis for t_{ab} at the point. Then $t_a = \pm\, t_{an}\,\overset{1}{\xi}{}^n$ satisfies the stated condition. Conversely, if $t_{ab} = t_a t_b$, then contraction with $\overset{1}{\xi}{}^a \overset{1}{\xi}{}^b$ yields $1 = (t_a \overset{1}{\xi}{}^a)^2$. So $t_a\, \overset{1}{\xi}{}^a = \pm 1$ and, hence, $t_{ab}\, \overset{1}{\xi}{}^b = t_a\, (t_b\, \overset{1}{\xi}{}^b) = \pm\, t_a$.)

So far we have considered only the decomposition $t_{ab} = t_a t_b$ at individual points of M. We say that $(M, t_{ab}, h^{ab}, \nabla)$ is *temporally orientable* if there exists a continuous (globally defined) vector field t_a that satisfies the decomposition condition at every point. (Our assumptions to this point do not guarantee existence.) Any such field t_a (which must, in fact, be smooth since t_{ab} is) will be called a *temporal orientation*. A timelike vector ξ^a qualifies as *future-directed* relative to t_a if $t_a\, \xi^a > 0$; otherwise it is *past-directed*. If a classical spacetime admits one temporal orientation t_a, then it admits two altogether, namely t_a and $-t_a$.

In what follows, we shall restrict attention to classical spacetimes that are temporally orientable and in which a temporal orientation has been selected. (We shall say, for example, "consider the classical spacetime $(M, t_a, h^{ab}, \nabla) \ldots$".) The orthogonality condition and the first compatibility condition can then be formulated directly in terms of t_a:

$$h^{ab}\, t_b = \mathbf{0}, \tag{4.1.7}$$

$$\nabla_a\, t_b = \mathbf{0}. \tag{4.1.8}$$

(These follow easily from the original formulations.)

Clearly, we understand a smooth curve to be *timelike* (respectively *spacelike*) if its tangent vectors are of this character at every point. And a timelike curve is understood to be *future-directed* (respectively *past-directed*) if its tangent vectors are so at every point.

From the compatibility condition, it follows that t_a is closed; i.e. $\nabla_{[a}\, t_{b]} = \mathbf{0}$. So (by proposition 1.8.3), at least locally, it must be exact—i.e., of the form $t_a = \nabla_a t$ for some smooth function t. We call any such function a *time function*. Any two time functions t and t' defined on a (common) open set can differ only by a constant; i.e., there must be a number k such that $t'(p) = t(p) + k$ for all p in the set. Given any time function t, and any smooth, future-directed timelike curve $\gamma : [s_1, s_2] \to M$ with tangent field ξ^a (whose image falls within the domain of t), the temporal length of γ is given by

$$\int_{s_1}^{s_2} (t_a\, \xi^a)\, ds = \int_{s_1}^{s_2} (\xi^a \nabla_a t)\, ds = \int_{s_1}^{s_2} \frac{d(t \circ \gamma)}{ds}\, ds = t(\gamma(s_2)) - t(\gamma(s_1));$$

i.e., it depends only on the endpoints of the curve. This shows that, at least locally, we have a well-defined, path-independent notion of "temporal distance" between points.

Let us say that a hypersurface S in M is *spacelike* if, at all points of S, all vectors tangent to S are spacelike. Notice that the defining condition is equivalent to the requirement that all time functions be constant on S. (A time function t is constant on S iff, given any vector ξ^a tangent to S at some point of S, $\xi^a \nabla_a t = 0$. But $t_a \xi^a = \xi^a \nabla_a t$. So the latter condition holds iff all vectors tangent to S are spacelike.) We can think of spacelike hypersurfaces as (at least local) "simultaneity slices."

If M is simply connected, then there must exist a globally defined time function $t: M \to \mathbb{R}$. In this case, spacetime can be decomposed into a one-parameter family of global ($t =$ constant) simultaneity slices. One can speak of "space" at a given "time." A different choice of (globally defined) time function would result in a different zero-point for the time scale, but would induce the same simultaneity slices and the same temporal distances between points on them.

We are now in a position to formulate interpretive principles corresponding to (C1), (P1), and (P2). (Recall our discussion in sections 2.1 and 2.3.) For all smooth curves $\gamma : I \to M$,

(C1$'$) γ is timelike iff its image $\gamma[I]$ could be the worldline of a point particle.
(P1$'$) γ can be reparametrized so as to be a timelike geodesic (with respect to ∇) iff $\gamma[I]$ could be the worldline of a *free* point particle.[1]
(P2$'$) Clocks record the t_{ab}-length of their worldlines.

Two points should be noted. First, in (C1$'$) and (P1$'$), we make reference to "point particles" without qualification, whereas previously we needed to restrict attention to particles with mass $m > 0$. Here there *are* no zero mass particles to consider, and no null curves whose images might serve as their worldlines. Second, there is an ambiguity as to what we mean by a "free" particle in (P1$'$). In the standard formulation of Newtonian gravitation theory, particles subject to a (non-vanishing) gravitational force do not count as free. But on the geometrized formulation, as in relativity theory, they do.

In what follows, unless indication is given to the contrary, we shall understand a "timelike curve" to be smooth, future-directed, and parametrized by its t_{ab}–length. In this case, its tangent field ξ^a satisfies the normalization condition $t_a \xi^a = 1$. And in this case, if a particle happens to have the image of the curve as its worldline, then we call ξ^a the *four-velocity* field of the particle,

1. We have seen (proposition 2.5.2) that it is possible, in a sense, to recover principle (P1) as a theorem in general relativity. Similarly, one can recover (P1$'$) as a theorem in geometrized Newtonian gravitation theory. Indeed, one can prove a result that is a close counterpart to proposition 2.5.2 (Weatherall [61]).

and call $\xi^n \nabla_n \xi^a$ its *four-acceleration* field. If the particle has mass m, then its four-acceleration field satisfies the equation of motion

$$F^a = m\xi^n \nabla_n \xi^a, \tag{4.1.9}$$

where F^a is a spacelike vector field (on the image of its worldline) that represents the net force acting on the particle. This is our version of Newton's second law of motion. (Recall equation (2.4.13).) Note that the equation makes geometric sense because the four-acceleration field is spacelike. (For, by the first compatibility condition, $t_a \xi^n \nabla_n \xi^a = \xi^n \nabla_n (t_a \xi^a) = \xi^n \nabla_n (1) = 0$.)

Now consider h^{ab}. It serves as a spatial metric, but just how it does so is a bit tricky. In Galilean relativistic mechanics, we have no notion of spatial length for timelike vectors—e.g., four-velocity vectors—since having one is tantamount to a notion of absolute rest. (We can take a particle to be "at rest" if its four-velocity field has spatial length 0 everywhere.) But we *do* have a notion of spatial length for spacelike vectors—e.g., four-acceleration vectors. (We can, for example, use measuring rods to determine distances between simultaneous events.) h^{ab} gives us one without the other.

We cannot take the spatial length of a vector μ^a to be $(h_{ab}\mu^a\mu^b)^{\frac{1}{2}}$ because the latter is not well defined. (As we have seen, there does not *exist* a field h_{ab} satisfying $h^{ab}h_{bc} = \delta^a{}_c$.) But if μ^a is spacelike, we can use h^{ab} to assign a spatial length to it indirectly. Here we need a small result about spacelike vectors.

PROPOSITION 4.1.1. *Let (M, t_a, h^{ab}, ∇) be a classical spacetime. Then the following conditions hold at all points in M.*

(1) *For all σ_b, $h^{ab}\sigma_b = \mathbf{0}$ iff σ_b is a multiple of t_b.*
(2) *For all μ^a, μ^a is spacelike iff there is a σ_b such that $h^{ab}\sigma_b = \mu^a$.*
(3) *For all σ_b and σ'_b, if $h^{ab}\sigma_b = h^{ab}\sigma'_b$, then $h^{ab}\sigma_a\sigma_b = h^{ab}\sigma'_a\sigma'_b$.*

Proof. The "if" halves of (1) and (2) follow immediately from the orthogonality condition (4.1.7). For the "only if" half of (1), let $\overset{1}{\sigma}_a, \ldots, \overset{4}{\sigma}_a$ be an orthonormal basis for h^{ab} in the sense discussed above. (So $h^{ab}\,\overset{i}{\sigma}_a\overset{j}{\sigma}_b = 0$ if $i \neq j$, $h^{ab}\,\overset{1}{\sigma}_a\overset{1}{\sigma}_b = 0$, and $h^{ab}\,\overset{i}{\sigma}_a\overset{i}{\sigma}_b = 1$ for $i = 2, 3, 4$.) We can take $\overset{1}{\sigma}_a$ to be t_a, since the latter satisfies the required conditions. Now consider any vector $\sigma_b = \overset{1}{k}\,\overset{1}{t}_b + \overset{2}{k}\,\overset{2}{\sigma}_b + \overset{3}{k}\,\overset{3}{\sigma}_b + \overset{4}{k}\,\overset{4}{\sigma}_b$, and assume $h^{ab}\sigma_b = \mathbf{0}$. Then, by the orthogonality condition, $\overset{2}{k}\,(h^{ab}\,\overset{2}{\sigma}_b) + \overset{3}{k}\,(h^{ab}\,\overset{3}{\sigma}_b) + \overset{4}{k}\,(h^{ab}\,\overset{4}{\sigma}_b) = \mathbf{0}$. Contraction with $\overset{i}{\sigma}_b$ yields $\overset{i}{k} = 0$ for $i = 2, 3, 4$. So $\sigma_b = \overset{1}{k}\,\overset{1}{t}_b$.

The "only if" half of (2) follows by dimensionality considerations. At any point in M, we can construe h^{ab} as a linear map from the cotangent space

V_b there to the tangent space V^a. Every vector in the image space $h^{ab}[V_b]$ is spacelike (by the "if" half of (2)). Moreover, $h^{ab}[V_b]$ is three-dimensional. (If $\overset{2}{\sigma}_b, \overset{3}{\sigma}_b$, and $\overset{4}{\sigma}_b$ are as above, then the vectors $h^{ab}\overset{2}{\sigma}_b$, $h^{ab}\overset{3}{\sigma}_b$, and $h^{ab}\overset{4}{\sigma}_b$ are linearly independent. For, as we have just seen, if a linear combination $\overset{2}{k}\,(h^{ab}\,\overset{2}{\sigma}_b) + \overset{3}{k}\,(h^{ab}\,\overset{3}{\sigma}_b) + \overset{4}{k}\,(h^{ab}\,\overset{4}{\sigma}_b)$ of the three is $\mathbf{0}$, the three coefficients must all be 0.) So, at every point, $h^{ab}[V_b]$ is a three-dimensional subspace of the vector space of spacelike vectors. But the latter is itself three-dimensional. So every spacelike vector must be in $h^{ab}[V_b]$.

For (3), suppose $h^{ab}\sigma_b = h^{ab}\sigma'_b$. Then, by (1), $(\sigma'_b - \sigma_b) = k\,t_b$ for some k. So $h^{ab}\sigma'_a\,\sigma'_b = h^{ab}(\sigma_a + k\,t_a)(\sigma_b + k\,t_b) = h^{ab}\sigma_a\,\sigma_b$. □

So here is the indirect procedure. If μ^a is spacelike, we take its *spatial length* to be $(h^{ab}\sigma_a\sigma_b)^{\frac{1}{2}}$, where σ_b is a vector such that $h^{ab}\sigma_b = \mu^a$. Clause (2) guarantees existence, and clause (3) guarantees that the choice of σ_b makes no difference.

Proposition 4.1.1 has a number of simple consequences that will be used again and again in what follows. Here is one. Suppose we have a tensor $\gamma_{\ldots a \ldots}$ at a point such that (i) $\gamma_{\ldots a \ldots} h^{ab} = \mathbf{0}$ and (ii) $\gamma_{\ldots a \ldots}\xi^a = \mathbf{0}$ for some timelike vector ξ^a there. Then $\gamma_{\ldots a \ldots} = \mathbf{0}$. (To see this, it suffices to consider any three linearly independent spacelike vectors $\overset{2}{\mu}{}^a, \overset{3}{\mu}{}^a, \overset{4}{\mu}{}^a$ at the point. (Existence is guaranteed by the signature of t_{ab}.) They, together with ξ^a, form a basis for the tangent space there. Since we are given that $\gamma_{\ldots a \ldots}\xi^a = \mathbf{0}$, it suffices to show that $\gamma_{\ldots a \ldots}\overset{i}{\mu}{}^a = \mathbf{0}$ for $i = 2, 3, 4$. But we know from the proposition that, for each $i = 2, 3, 4$, there is a co-vector $\overset{i}{\sigma}_b$ such that $\overset{i}{\mu}{}^a = h^{ab}\overset{i}{\sigma}_b$. So our claim follows from (i).)

This first consequence of proposition 4.1.1 can be generalized. Suppose we have a tensor $\gamma_{\ldots ab \ldots}$ at a point such that, for some timelike vector ξ^a there, (i) $\gamma_{\ldots ab \ldots} h^{am} h^{bn} = \mathbf{0}$, and (ii) $\gamma_{\ldots ab \ldots}\xi^a h^{bn} = \mathbf{0} = \gamma_{\ldots ab \ldots} h^{an}\xi^b$, and (iii) $\gamma_{\ldots ab \ldots}\xi^a\xi^b = \mathbf{0}$. Then $\gamma_{\ldots ab \ldots} = \mathbf{0}$. Other tensors $\gamma_{\ldots a_1 a_2 \ldots a_n \ldots}$ can be handled similarly.

PROPOSITION 4.1.2. *Let (M, t_a, h^{ab}, ∇) be a classical spacetime, and let ξ^a be a smooth, future-directed, unit timelike vector field on M. (So $t_a\xi^a = 1$.) Then there is a (unique) smooth, symmetric field $\hat{h}_{ab}$ on M satisfying the conditions*

$$\hat{h}_{ab}\,\xi^b = \mathbf{0}, \tag{4.1.10}$$

$$\hat{h}_{ab}\,h^{bc} = \delta^c_{\ a} - t_a\,\xi^c. \tag{4.1.11}$$

Proof. It follows by the remark in the preceding paragraph that there can be at most one field $\hat{h}_{ab}$ satisfying the stated conditions. (Given any two candidates, we need only substract one from the other and apply the remark to the resulting difference field.) We can define a symmetric field $\hat{h}_{ab}$ by its specifying its action, at any point, on the unit timelike vector ξ^a and on an arbitrary spacelike vector μ^a. So consider the field $\hat{h}_{ab}$ that annihilates the former and makes the assignment

$$\hat{h}_{ab}\,\mu^b = \sigma_a - t_a(\xi^c \sigma_c)$$

to the latter—where σ_a is *any* vector such that $\mu^a = h^{ab}\sigma_b$. It is easy to check that the choice of σ_a plays no role here. (For suppose that $h^{ab}\overset{1}{\sigma}_b = h^{ab}\overset{2}{\sigma}_b$. Then $\overset{1}{\sigma}_a - t_a(\xi^c\overset{1}{\sigma}_c) = \overset{2}{\sigma}_a - t_a(\xi^c\overset{2}{\sigma}_c)$. The latter follows since we get the same result on both sides if we contract with either ξ^a or h^{ab}.) It now follows, as well, that condition (4.1.11) holds. For by the very way we have defined $\hat{h}_{ab}$, both sides of (4.1.11) yield the same result when contracted with any vector σ_c. □

We call $\hat{h}_{ab}$ the *spatial metric (or spatial projection field) relative to* ξ^a. Our notation is imperfect here because we make no explicit reference to ξ^a. But it will be clear from the context which unit timelike field is intended.

Because h^{ab} is not invertible, we cannot raise *and* lower indices with it. But we can, at least, raise indices, and it is sometimes convenient to do so. So, for example, if $R^a{}_{bcd}$ is the Riemann curvature tensor field associated with ∇, we can understand $R^{ab}{}_{cd}$ to be the field $h^{bn}R^a{}_{ncd}$. Note that

(4.1.12)
$$\hat{h}^a{}_b = \delta^a{}_b - t_b\,\xi^a.$$

(This is simply equation (4.1.11), since $\hat{h}^a{}_b = \hat{h}_{mb}\,h^{ma}$.) It follows immediately from equation (4.1.12) that, given any vector μ^a at a point, we can express it in the form

$$\mu^a = \hat{h}^a{}_b\,\mu^b + (t_b\,\mu^b)\,\xi^a.$$

Here the first term on the right side is spacelike, and the second is proportional to ξ^a. We call $\hat{h}^a{}_b\,\mu^b$ the *spatial projection (or spatial component) of* μ^a *relative to* ξ^a.

We also call $(\hat{h}_{ab}\,\mu^a\,\mu^b)^{\frac{1}{2}}$ the *spatial length of* μ^a *relative to* ξ^a. It is easy to check that this magnitude is just what we would otherwise describe as the spatial length of the spatial component $\hat{h}^a{}_b\,\mu^b$. (According to our prescription, the spatial length of $\hat{h}^a{}_b\,\mu^b$ is given by $(h^{mn}\,\sigma_m\,\sigma_n)^{\frac{1}{2}}$, where σ_m is any vector satisfying $\hat{h}^a{}_b\,\mu^b = h^{am}\,\sigma_m$. But $\hat{h}^a{}_b\,\mu^b = h^{am}\,\hat{h}_{mr}\,\mu^r$. So the spatial length of $\hat{h}^a{}_b\,\mu^b$ is given by

$$\left(h^{mn}\,(\hat{h}_{mr}\,\mu^r)\,(\hat{h}_{ns}\,\mu^s)\right)^{\frac{1}{2}}.$$

But $h^{mn}\,\hat{h}_{mr}\,\hat{h}_{ns} = \hat{h}_{rs}$. So the spatial length of $\hat{h}^a{}_b\,\mu^b$ comes out as $(\hat{h}_{rs}\,\mu^r\,\mu^s)^{\frac{1}{2}}$, as claimed.)

It is important that the compatibility conditions $\nabla_a\,h^{bc} = \mathbf{0}$ and $\nabla_a\,t_b = \mathbf{0}$ (or, equivalently, $\nabla_a\,t_{bc} = \mathbf{0}$) do *not* determine a unique derivative operator. (There is no contradiction here with proposition 1.9.2 since neither t_{ab} nor h^{ab} is an (invertible) metric.) In fact, we have the following characterization result.

PROPOSITION 4.1.3. *Let (M, t_a, h^{ab}, ∇) be a classical spacetime. Let $\nabla' = (\nabla,\ C^a{}_{bc})$ be a second derivative operator on M (i.e., the action of ∇' relative to that of ∇ is given by $C^a{}_{bc}$). Then ∇' is compatible with t_a and h^{ab} iff $C^a{}_{bc}$ is of the form*

$$C^a{}_{bc} = 2\,h^{an}\,t_{(b}\,\kappa_{c)n} \tag{4.1.13}$$

where κ_{ab} is a smooth anti-symmetric field on M.

Proof. Since (M, t_a, h^{ab}, ∇) is a classical spacetime, ∇ is compatible with both t_a and h^{ab}. Hence, by equation (1.7.1), we have

$$\nabla'_a\,t_b = \nabla_a\,t_b + C^r{}_{ab}\,t_r = C^r{}_{ab}\,t_r, \tag{4.1.14}$$

$$\nabla'_a\,h^{bc} = \nabla_a\,h^{bc} - C^b{}_{ar}\,h^{rc} - C^c{}_{ar}\,h^{br} = -C^b{}_{ar}\,h^{rc} - C^c{}_{ar}\,h^{br}. \tag{4.1.15}$$

Assume, first, that $C^a{}_{bc}$ has the indicated form. Then $t_a\,C^a{}_{bc} = \mathbf{0}$ and $C^a{}_{bc}\,h^{cd} = h^{an}\,t_b\,\kappa_{cn}\,h^{cd}$ by the orthogonality condition. It follows immediately that ∇' is compatible with t_a. It also follows that

$$\nabla'_a\,h^{bc} = -t_a\,(h^{bn}\,\kappa_{rn}\,h^{rc} + h^{cn}\,\kappa_{rn}\,h^{br}) = -t_a\,(\kappa^{cb} + \kappa^{bc}).$$

But κ_{ab} is anti-symmetric. So ∇' is compatible with h^{ab} as well.

Conversely, assume ∇' is compatible with t_a and h^{ab}. Then, by equations (4.1.14) and (4.1.15),

$$C^r{}_{ab}\,t_r = \mathbf{0}, \tag{4.1.16}$$

$$C^b{}_{ar}\,h^{rc} + C^c{}_{ar}\,h^{br} = \mathbf{0}. \tag{4.1.17}$$

Now consider the raised index tensor field $C^{abc} = C^a{}_{mn}\,h^{mb}\,h^{nc}$. It is spacelike; i.e., contraction on any index with t_a yields $\mathbf{0}$. Moreover, it satisfies the two conditions

$$C^{abc} = C^{acb}, \tag{4.1.18}$$

$$C^{abc} = -C^{cba}. \tag{4.1.19}$$

(This first follows from the symmetry of $C^a_{\ bc}$ itself, and the second from equation (4.1.17).) By repeated use of these two, we have

$$C^{abc} = -C^{cba} = -C^{cab} = C^{bac} = C^{bca} = -C^{acb} = -C^{abc}.$$

So the field vanishes everywhere:

$$C^{abc} = \mathbf{0}. \tag{4.1.20}$$

Now let ξ^a be a smooth, future-directed, unit timelike field (so $t_a\xi^a = 1$), and let $\hat{h}_{ab}$ be the corresponding spatial projection field. Then we have

$$\begin{aligned}\mathbf{0} &= C^{amn}\,\hat{h}_{mb}\,\hat{h}_{nc} = C^a_{\ rs}\,h^{rm}\,h^{sn}\,\hat{h}_{mb}\,\hat{h}_{nc}\\ &= C^a_{\ rs}(\delta^r_{\ b} - t_b\,\xi^r)(\delta^s_{\ c} - t_c\,\xi^s).\end{aligned}$$

Hence,

$$C^a_{\ bc} = t_c\,C^a_{\ bs}\,\xi^s + t_b\,C^a_{\ rc}\,\xi^r - t_b\,t_c\,C^a_{\ rs}\,\xi^r\,\xi^s. \tag{4.1.21}$$

Now consider

$$\kappa_{cn} = -\hat{h}_{cp}\,\hat{h}_{nq}\,C^{p\ q}_{\ r}\,\xi^r + t_{[c}\,\hat{h}_{n]q}\,C^q_{\ rs}\,\xi^r\,\xi^s. \tag{4.1.22}$$

It is anti-symmetric by equation (4.1.17) and, we claim, it satisfies equation (4.1.13). To verify this, we compute the right side of (4.1.13). We have

$$2\,h^{an}\,t_b\,\kappa_{cn} = \ -2\,(h^{an}\,\hat{h}_{nq})\,t_b\,\hat{h}_{cp}\,C^{p\ q}_{\ r}\,\xi^r + t_b\,t_c\,(h^{an}\,\hat{h}_{nq})\,C^q_{\ rs}\,\xi^r\,\xi^s.$$

Now, by equations (4.1.16) and (4.1.17), $(h^{an}\,\hat{h}_{nq})\,C^{p\ q}_{\ r} = -(\delta^a_{\ q} - t_q\,\xi^a)\,C^{q\ p}_{\ r} = -C^{a\ p}_{\ r}$, and $(h^{an}\,\hat{h}_{nq})\,C^q_{\ rs} = C^a_{\ rs}$. So

$$2\,h^{an}\,t_b\,\kappa_{cn} = 2\,t_b\,\hat{h}_{cp}\,C^{a\ p}_{\ r}\,\xi^r + t_b\,t_c\,C^a_{\ rs}\,\xi^r\,\xi^s.$$

Furthermore, $\hat{h}_{cp}\,C^{a\ p}_{\ r} = \hat{h}_{cp}\,C^a_{\ rs}\,h^{sp} = (\delta^s_{\ c} - t_c\,\xi^s)\,C^a_{\ rs}$. So

$$2\,h^{an}\,t_b\,\kappa_{cn} = 2\,t_b\,C^a_{\ rc}\,\xi^r - t_b\,t_c\,C^a_{\ rs}\,\xi^r\,\xi^s.$$

Hence, by equation (4.1.21),

$$2\,h^{an}\,t_{(b}\,\kappa_{c)n} = 2\,t_{(b}\,C^a_{\ c)r}\,\xi^r - t_b\,t_c\,C^a_{\ rs}\,\xi^r\,\xi^s = C^a_{\ bc}. \qquad \square$$

Now let $R^a{}_{bcd}$ be the curvature tensor associated with ∇. Of course, it satisfies the algebraic conditions listed in proposition 1.8.2:

$$R^a{}_{b(cd)} = \mathbf{0}, \tag{4.1.23}$$

$$R^a{}_{[bcd]} = \mathbf{0}. \tag{4.1.24}$$

The compatibility conditions ($\nabla_a t_b = \mathbf{0}$ and $\nabla_a h^{bc} = \mathbf{0}$) further imply that

$$t_a R^a{}_{bcd} = \mathbf{0}, \tag{4.1.25}$$

$$R^{(ab)}{}_{cd} = \mathbf{0}. \tag{4.1.26}$$

(We have $\mathbf{0} = 2\nabla_{[c}\nabla_{d]}\, t_b = t_a R^a{}_{bcd}$ and $\mathbf{0} = 2\nabla_{[c}\nabla_{d]}\, h^{ab} = -R^a{}_{mcd}\, h^{mb} - R^b{}_{mcd}\, h^{am} = -R^{ab}{}_{cd} - R^{ba}{}_{cd}$.) It follows immediately from the conditions listed so far that if we raise all three indices with h^{ab}, the resulting field R^{abcd} satisfies

$$R^{ab(cd)} = \mathbf{0}, \tag{4.1.27}$$

$$R^{a[bcd]} = \mathbf{0}, \tag{4.1.28}$$

$$R^{(ab)cd} = \mathbf{0}. \tag{4.1.29}$$

These, in turn, jointly imply

$$R^{abcd} = R^{cdab}. \tag{4.1.30}$$

(The argument is the same as in the case where ∇ is determined by a (non-degenerate) metric. Recall our proof of the fourth clause of proposition 1.9.4.)

Now consider the Ricci tensor field $R_{ab} = R^c{}_{abc}$ and the (spatial) scalar curvature field $R = h^{ab} R_{ab}$. We claim that the former is symmetric. To verify this, we consider an arbitrary smooth, future-directed, timelike field ξ^a and use the corresponding projection field $\hat{h}_{ab}$ to lower indices. First, it follows easily from equations (4.1.11), (4.1.25), and (4.1.26) that

$$R^a{}_{acd} = \hat{h}_{ab}\, R^{ab}{}_{cd} = \mathbf{0}, \tag{4.1.31}$$

$$R^{bc} = \hat{h}_{ad}\, R^{abcd}, \tag{4.1.32}$$

$$R = \hat{h}_{ab}\, R^{ab}. \tag{4.1.33}$$

(For example, we have $\hat{h}_{ab}\, R^{ab}{}_{cd} = \hat{h}_{ab}\, h^{br}\, R^a{}_{rcd} = (\delta^r{}_a - t_a\, \xi^r)\, R^a{}_{rcd} = R^a{}_{acd}$. This, with equation (4.1.26), gives us equation (4.1.31).) Hence, by equations (4.1.23) and (4.1.24),

$$R_{ab} - R_{ba} = R^c{}_{abc} - R^c{}_{bac} = R^c{}_{abc} + R^c{}_{bca} = -R^c{}_{cab}. \tag{4.1.34}$$

So, by equation (4.1.31), we have

$$R_{ab} = R_{ba}, \tag{4.1.35}$$

as claimed.

Less straightforward is the following proposition.

PROPOSITION 4.1.4. *Let* $(M, t_a, h^{ab}, \nabla_a)$ *be a classical spacetime. Then the curvature field* $R^a{}_{bcd}$ *associated with* ∇ *satisfies*

$$\begin{aligned} R^{abcd} &= (h^{bc}\, R^{ad} + h^{ad}\, R^{bc} - h^{ac}\, R^{bd} - h^{bd}\, R^{ac}) \\ &\quad + \frac{1}{2}\,(h^{ac}\, h^{bd} - h^{ad}\, h^{bc})\, R. \end{aligned} \tag{4.1.36}$$

Proof. The relation is familiar from the case where we are dealing with a derivative operator determined by an (invertible) metric and the background manifold has dimension 3. It follows from the symmetries (4.1.27)–(4.1.30) and (4.1.35), as well as the crucial fact that all the indices in R^{abcd} are spacelike; i.e., contraction on any of these indices with t_a yields $\mathbf{0}$.

We prove equation (4.1.36) at an arbitrary point p of M by introducing an appropriate basis there and considering the resulting component relations.

Let t_a, $\overset{1}{\sigma}_a$, $\overset{2}{\sigma}_a$, $\overset{3}{\sigma}_a$ be an orthonormal basis for h^{ab} at p in the sense discussed above. (So $h^{ab}\, \overset{i}{\sigma}_a \overset{j}{\sigma}_b = 0$ if $i \neq j$, and $h^{ab}\, \overset{i}{\sigma}_a \overset{i}{\sigma}_b = 1$ for $i = 1, 2, 3$.) Then $h^{ab} = \overset{1}{\sigma}{}^a\, \overset{1}{\sigma}{}^b + \overset{2}{\sigma}{}^a\, \overset{2}{\sigma}{}^b + \overset{3}{\sigma}{}^a\, \overset{3}{\sigma}{}^b$. Further, let ξ^a be a future-directed unit timelike vector at p with corresponding projection tensor $\hat{h}_{ab}$. Now consider the co-vectors $\overset{1}{\alpha}_a$, $\overset{2}{\alpha}_a$, $\overset{3}{\alpha}_a$ at p defined by

$$\overset{i}{\alpha}_a = \hat{h}_{ab}\, h^{bc}\, \overset{i}{\sigma}_c = \overset{i}{\sigma}_a - t_a(\overset{i}{\sigma}_c\, \xi^c).$$

It is easy to check that

(1) $\overset{i}{\alpha}_a\, \xi^a = 0$ for $i = 1, 2, 3$.
(2) $h^{ab}\, \overset{i}{\sigma}_a = h^{ab}\, \overset{i}{\alpha}_a$ for $i = 1, 2, 3$.
(3) t_a, $\overset{1}{\alpha}_a$, $\overset{2}{\alpha}_a$, $\overset{3}{\alpha}_a$ form a co-basis at p.
(4) $\hat{h}_{ab} = \overset{1}{\alpha}_a\, \overset{1}{\alpha}_b + \overset{2}{\alpha}_a\, \overset{2}{\alpha}_b + \overset{3}{\alpha}_a\, \overset{3}{\alpha}_b$.

Since all indices in R^{abcd} and R^{ab} are spacelike, both tensors are determined by their action on the basis vectors $\overset{1}{\alpha}_a$, $\overset{2}{\alpha}_a$, $\overset{3}{\alpha}_a$. Consider the components

$$\overset{ij}{R} = R^{ab}\, \overset{i}{\alpha}_a\, \overset{j}{\alpha}_b,$$

$$\overset{ijkl}{R} = R^{abcd}\, \overset{i}{\alpha}_a\, \overset{j}{\alpha}_b\, \overset{k}{\alpha}_c\, \overset{l}{\alpha}_d$$

where $i, j, k, l \in \{1, 2, 3\}$. Because of the symmetries of R^{abcd} and R^{ab}, each has only six independent (non-zero) components, namely

$$\overset{11}{R} \quad \overset{12}{R} \quad \overset{13}{R} \quad \overset{22}{R} \quad \overset{23}{R} \quad \overset{33}{R}$$

and

$$\overset{1212}{R} \quad \overset{1313}{R} \quad \overset{2323}{R} \quad \overset{1213}{R} \quad \overset{1223}{R} \quad \overset{1323}{R}\,.$$

Now, by equation (4.1.32), $R^{ab} = R^{rabs}\,\hat{h}_{rs} = \sum_{i=1}^{3} R^{rabs}\,\overset{i}{\alpha}_r\,\overset{i}{\alpha}_s$. Hence, for all $j, k \in \{1, 2, 3\}$, $\overset{jk}{R} = \sum_{i=1}^{3} \overset{ijki}{R}$. This gives us

$$\begin{aligned}
\overset{11}{R} &= -\overset{1212}{R} - \overset{1313}{R} & \qquad \overset{12}{R} &= -\overset{1323}{R} \\
\overset{22}{R} &= -\overset{1212}{R} - \overset{2323}{R} & \qquad \overset{13}{R} &= \overset{1223}{R} \\
\overset{33}{R} &= -\overset{1313}{R} - \overset{2323}{R} & \qquad \overset{23}{R} &= -\overset{1213}{R}\,.
\end{aligned}$$

Also, by equation (4.1.33),

$$R = R^{ab}\,\hat{h}_{ab} = \sum_{i=1}^{3} R^{ab}\,\overset{i}{\alpha}_a\,\overset{i}{\alpha}_b \;=\; \overset{11}{R} + \overset{22}{R} + \overset{33}{R}.$$

Using these relations, we can check that the two sides of equation (4.1.36) agree in their action on any quadruple $\overset{i}{\alpha}_a\,\overset{j}{\alpha}_b\,\overset{k}{\alpha}_c\,\overset{l}{\alpha}_d$. As an example, consider $\overset{1}{\alpha}_a\,\overset{2}{\alpha}_b\,\overset{1}{\alpha}_c\,\overset{2}{\alpha}_d$. We have $h^{ab}\overset{1}{\alpha}_a\,\overset{1}{\alpha}_b = h^{ab}\overset{2}{\alpha}_a\,\overset{2}{\alpha}_b = 1$ and $h^{ab}\overset{1}{\alpha}_a\,\overset{2}{\alpha}_b = 0$. So it suffices to confirm that $\overset{1212}{R} = (-\overset{22}{R} - \overset{11}{R}) + \frac{1}{2}R$. But this follows from the entries in our table. □

Next we consider the notion of "spatial flatness." Of course, we say that our background classical spacetime is *flat* at a point if $R^a{}_{bcd} = \mathbf{0}$ there. In parallel, we say that it is *spatially flat* there if $R^{abcd} = \mathbf{0}$. To motivate this definition, we need to say something about "induced derivative operators" on spacelike hypersurfaces. (Recall that a hypersurface is spacelike—in a classical spacetime as well as in a relativistic spacetime— if all smooth curves with images in the hypersurface are spacelike.)

Let S be a spacelike hypersurface, and let ξ^a be an arbitrary smooth, unit, future-directed timelike vector field on S. Let $\hat{h}_{ab}$ be the associated projection field on S. Given any tensor field on S, we say that it is *spacelike relative to* ξ^a if contraction on any of its indices with t_a or ξ^a yields $\mathbf{0}$. We can think of fields spacelike relative to ξ^a as living on the manifold S. (Recall the discussion in section 1.10.) Clearly, h^{ab} and $\hat{h}_{ab}$ both qualify as spacelike relative to ξ^a. So does $\hat{h}^b{}_a = \delta^b{}_a - t_a\,\xi^b$. Notice that $\hat{h}^b{}_a$ preserves all vectors that are spacelike relative to ξ^a; i.e., $\hat{h}^b{}_a\,\mu^a = \mu^b$ and $\hat{h}^b{}_a\,\sigma_b = \sigma_a$, for all μ^a and σ_a spacelike relative to ξ^a. We can thus think of $\hat{h}^b{}_a$ as a "delta (or index substitution) field" for fields on S that are spacelike relative to ξ^a. And we shall, on occasion, write $\hat{\delta}^b{}_a$ rather than $\hat{h}^b{}_a$—just as in the case of a (non-degenerate) metric g_{ab} we often write $\delta^b{}_a$ rather than $g^b{}_a$.

What is most important here is that we can think of $\hat{h}_{ab}$ as a (non-degenerate) metric that lives on S. It is non-degenerate in the relevant sense because it does not annihilate any non-zero vectors that are spacelike relative to ξ^a or, equivalently, because it has an "inverse" h^{ab}; i.e., $\hat{h}_{ab}\,h^{bc} = \hat{\delta}^c{}_a$. (This is just equation 4.1.11.) So there is a unique derivative operator D on S that is compatible with $\hat{h}_{ab}$—i.e., such that $D_a\hat{h}_{bc} = \mathbf{0}$. We can express the action of D in terms of ∇ (as explained in section 1.10). Given any field spacelike relative to ξ^a, the action of D on it is given by first applying ∇ and then projecting all covariant indices with $\hat{h}^b{}_a$. So, for example,

$$D_n\,\alpha^a{}_{bc} = \hat{h}^m{}_n\,\hat{h}^r{}_b\,\hat{h}^s{}_c\,\nabla_m\,\alpha^a{}_{rs}. \tag{4.1.37}$$

The projection insures that the resultant field is spacelike relative to ξ^a. There is no need to project the contravariant indices. Since $\nabla_a t_b = \mathbf{0}$, they remain spacelike even after ∇ is applied. (One can check directly that D satisfies all the defining conditions of a derivative operator on S, and furthermore $D_a\hat{h}_{bc} = \mathbf{0}$ and $D_a h^{bc} = \mathbf{0}$.) We refer to D as the *derivative operator induced on S* relative to ξ^a.

The following proposition serves to motivate our definition of spatial flatness.

PROPOSITION 4.1.5. *(Spatial Flatness Proposition) Let (M, t_a, h^{ab}, ∇) be a classical spacetime. The following conditions are equivalent at every point in M.*

(1) *Space is flat, i.e., $R^{abcd} = \mathbf{0}$.*
(2) $R^{ab} = \mathbf{0}$.
(3) $R_{ab} = t_{(a}\,\varphi_{b)}$ *for some* φ_a.

Furthermore, given any spacelike hypersurface S in M, these conditions hold throughout S iff parallel transport of spacelike vectors within S is, at least locally, path independent.

Proof. Let p be a point in M, and let ξ^a be an arbitrary, future-directed, unit timelike vector at p with corresponding spatial projection tensor $\hat{h}_{ab}$. The equivalence of (1) and (2) follows from equations (4.1.32), (4.1.33), and (4.1.36).

The implication (3) ⇒ (2) is immediate. For the converse, consider the vector

$$\varphi_a = 2\, R_{ab}\xi^b - t_a(R_{mn}\,\xi^m\xi^n).$$

We have $\varphi_a\xi^a = R_{ab}\,\xi^a\xi^b$ and $\varphi_a h^{ar} = 2R_{ab}\,h^{ar}\xi^b$. Therefore, at any point where $R^{ab} = \mathbf{0}$, it must be the case that $R_{ab} = t_{(a}\,\varphi_{b)}$, since both sides agree in their action on $\xi^a\xi^b$, $h^{ar}\xi^b$, and $h^{ar}h^{bs}$. (Recall our remarks following proposition 4.1.1.)

Now let S be a spacelike hypersurface, and let ξ^a be a smooth, unit, future-directed timelike vector field on S. Further, let $\hat{h}_{ab}$ be the associated projection field on S, and let D be the derivative operator induced on S relative to ξ^a (as explained in the preceding paragraphs). Finally, suppose that μ^a and ν^a are spacelike fields on S. Then they automatically qualify as spacelike relative to ξ^a, and by equation (4.1.37) we have $\mu^n D_n\, \nu^a = \mu^n\, \hat{h}^r{}_n\, \nabla_r\, \nu^a = \mu^r\, \nabla_r\, \nu^a$. *It follows that D and ∇ induce the same conditions for parallel transport of spacelike vectors on S.* We know that parallel transport of such vectors on S is, at least locally, path independent iff the Riemann curvature tensor field $\mathcal{R}^a{}_{bcd}$ on S associated with D vanishes. So, for the second half of the proposition, it suffices for us to show that, at all points on S,

$$R^{abcd} = \mathbf{0} \iff \mathcal{R}^a{}_{bcd} = \mathbf{0}. \tag{4.1.38}$$

This just involves a bit of computation. The right-side condition here is equivalent to the requirement that, for all spacelike fields μ^a on S,

$$\mathbf{0} = \mathcal{R}^a{}_{bcd}\,\mu^b = -2\,D_{[c}\,D_{d]}\,\mu^a = -2\,\hat{h}^r{}_c\,\hat{h}^s{}_d\,\nabla_{[r}\,\nabla_{s]}\,\mu^a = \hat{h}^r{}_c\,\hat{h}^s{}_d\,R^a{}_{brs}\,\mu^b.$$

Hence, it is equivalent to the condition

$$\mathbf{0} = \hat{h}^r{}_c\,\hat{h}^s{}_d\,R^a{}_{prs}\,h^{pb} = \hat{h}^r{}_c\,\hat{h}^s{}_d\,R^{ab}{}_{rs}.$$

Contracting this equation with $h^{cm}\,h^{dn}$ yields $R^{abmn} = \mathbf{0}$. Conversely, contracting $R^{abmn} = \mathbf{0}$ with $\hat{h}_{cm}\,\hat{h}_{dn}$ yields $\hat{h}^r{}_c\,\hat{h}^s{}_d\,R^{ab}{}_{rs} = \mathbf{0}$. □

The interest of proposition 4.1.5 will become apparent in the next section when we consider the geometrized formulation of Newtonian gravitation theory. In that formulation, Poisson's equation assumes the form $R_{ab} = 4\,\pi\rho\; t_a t_b$ (where ρ is the mass density function). We see from the proposition that Poisson's equation (in its geometrized formulation) implies the flatness of space. This is striking. It is absolutely fundamental to the idea of geometrized Newtonian theory that *spacetime* is curved (and gravitation is just a manifestation of that curvature). Yet the basic field equation of the theory itself rules out the possibility that *space* is curved.

Intermediate between the curvature conditions $R^a{}_{bcd} = \mathbf{0}$ and $R^{abcd} = \mathbf{0}$ is the condition $R^{ab}{}_{cd} = \mathbf{0}$. We shall show later (proposition 4.3.1) that it holds throughout M iff parallel transport of spacelike vectors along arbitrary curves is, at least locally, path independent. (Here we still restrict attention to spacelike vectors (rather than arbitrary vectors), but consider their transport along arbitrary curves in M (not just curves confined to a particular spacelike hypersurface).)

Before continuing with the main line of presentation in this section, we stop briefly to record a fact that will be needed in later sections. We place it here because it concerns the induced derivative operator D that was considered in the preceding proof.

PROPOSITION 4.1.6. *Let (M, t_a, h^{ab}, ∇) be a classical spacetime, and let ϕ^a be smooth spacelike field on M such that $\nabla^{[a}\,\phi^{b]} = \mathbf{0}$. Then, at least locally, there exists a smooth field ϕ such that $\phi^a = \nabla^a \phi$.*

Proof. This is not quite an instance of proposition 1.8.3, but it is close. Let p be any point in M, and let O be any open set containing p that is sufficiently small and well behaved that it has this property: O can be covered by a family $\mathcal{F}$ of spacelike hypersurfaces, each of which is connected and simply connected. Let $\gamma : I \to M$ be any timelike curve whose image contains p and intersects every one of the hypersurfaces in $\mathcal{F}$. Finally, let ξ^a be a smooth, future-directed, unit timelike field on O, and let $\hat{h}_{ab}$ be the associated spatial projection field.

Now consider any hypersurface S in $\mathcal{F}$, and the projected field $\hat{\phi}_a = \hat{h}_{ab}\,\phi^b$ on S. If D is the induced derivative operator on S defined by equation (4.1.37), then on S we have $D_{[a}\,\hat{\phi}_{b]} = \hat{h}_{am}\,\hat{h}_{bn}\,\nabla^{[m}\,\phi^{n]} = \mathbf{0}$. So, by proposition 1.8.3, there is a smooth field ϕ_S on S such that $\hat{\phi}_a = D_a\,\phi_S$. It is determined only up to a constant, but we can pin it down uniquely by requiring, in addition, that it have value 0 at the point where S intersects $\gamma[I]$.

Now let ϕ be the "aggregated" scalar field on O that agrees with ϕ_S on each S in $\mathcal{F}$. We claim without further argument that it is smooth. It satisfies the required condition since, given any spacelike hypersurface S in $\mathcal{F}$, we have $\phi^a = h^{an}\hat{\phi}_n = h^{an} D_n\,\phi_S = h^{an}\,\hat{h}^r{}_n \nabla_r\,\phi = \nabla^a \phi$ on S. □

Now we briefly consider the representation of fluid flow. Our formalism here is related closely to that developed in section 2.8. Let ξ^a be a smooth, unit, future-directed timelike vector field on our background classical spacetime. We think of ξ^a as the four-velocity of a fluid. Let $\hat{h}_{ab}$ be the projection field associated with ξ^a. The *rotation field* ω_{ab} and *expansion field* θ_{ab} associated with ξ^a are defined by

$$\omega_{ab} = \hat{h}^m{}_{[a}\, \hat{h}_{b]n}\, \nabla_m\, \xi^n, \tag{4.1.39}$$

$$\theta_{ab} = \hat{h}^m{}_{(a}\, \hat{h}_{b)n}\, \nabla_m\, \xi^n. \tag{4.1.40}$$

(We can motivate the terminology here much as we did in section 2.8.) It follows that

$$\hat{h}_{bn} \nabla_a\, \xi^n = \omega_{ab} + \theta_{ab} + t_a\, \hat{h}_{bn}\, \xi^m \nabla_m\, \xi^n \tag{4.1.41}$$

and, hence, that

$$\nabla_a\, \xi^b = \omega_a{}^b + \theta_a{}^b + t_a\, \xi^m \nabla_m\, \xi^b \tag{4.1.42}$$

and

$$\nabla^a \xi^b = \omega^{ab} + \theta^{ab}. \tag{4.1.43}$$

As in the relativistic case, we can decompose the expansion field to arrive at the scalar expansion field θ and the shear field σ_{ab}:

$$\theta = \theta_a{}^a = \nabla_a\, \xi^a, \tag{4.1.44}$$

$$\sigma_{ab} = \theta_{ab} - \frac{1}{3}\, \theta\, \hat{h}_{ab}. \tag{4.1.45}$$

(That $\theta_a{}^a = \nabla_a\, \xi^a$ follows from equation (4.1.42) and the anti-symmetry of ω_{ab}.) Clearly, σ_{ab} is "trace-free" since $\sigma_a{}^a = \theta_a{}^a - \frac{1}{3}\, \theta\, \hat{h}_a{}^a = \theta - \frac{1}{3}\theta(\delta_a{}^a - t_a \xi^a) = \theta - \frac{1}{3}\theta(4-1) = 0$. We note for future reference the following equivalences:

$$\omega_{ab} = \mathbf{0} \iff \nabla^{[a}\, \xi^{b]} = \mathbf{0}, \tag{4.1.46}$$

$$\theta_{ab} = \mathbf{0} \iff \nabla^{(a}\, \xi^{b)} = \mathbf{0}. \tag{4.1.47}$$

(In each case, we get the implication from left to right by raising indices with h^{mn}, and the one from right to left by lowering indices with $\hat{h}_{mn}$.) The conditions in the first line capture the claim that ξ^a is *non-rotating* (or *twist-free*).

Finally, we say just a bit about the four-momentum of point particles and the four-momentum density of matter fields. It is instructive to consider the situations in Newtonian and relativistic mechanics side by side. (For a more complete and thorough comparison, see Dixon [12].) Suppose, first, that we have a point particle with mass m and four-velocity field ξ^a. Then, just as in relativity theory, we associate with it a four-momentum field $P^a = m\, \xi^a$ along its worldline. (In the present context we have only particles with positive mass ($m > 0$) to consider.)

Suppose particle O has four-velocity ξ^a at a point, and another particle O' has four-momentum $P^a = m\,\xi'^a$ there. Just as in the relativistic case, we can decompose P^a relative to ξ^a.

Newtonian Mechanics	Relativistic Mechanics
$P^a = \underbrace{(t_b P^b)}_{\text{mass}}\,\xi^a + \underbrace{\hat{h}^a{}_b\, P^b}_{\text{relative 3-momentum}}$	$P^a = \underbrace{(\xi_b P^b)}_{\text{relative energy}}\,\xi^a + \underbrace{h^a{}_b\, P^b}_{\text{relative 3-momentum}}$

But the decomposition works somewhat differently in the two cases. In Newtonian mechanics, we have a component proportional to ξ^a with magnitude $t_b P^b = t_b\,(m\,\xi'^b) = m$, and a spacelike component

$$\hat{h}^a{}_b\, P^b = (\delta^a{}_b - t_b\,\xi^a)\,(m\,\xi'^b) = m\,(\xi'^b - \xi^a),$$

which gives the three-momentum of the particle relative to ξ^a. (The vector $(\xi'^b - \xi^a)$ by itself gives the relative velocity of O' with respect to O.) Thus, in Newtonian mechanics, the four-momentum P^a of a point particle codes its mass and its three-momentum, as determined relative to other background observers. So it is appropriately called the "mass-momentum vector." In relativistic mechanics, in contrast, as we have seen, the component of P^a proportional to ξ^a has magnitude $(\xi_b P^b)$, which gives the energy of the particle as determined relative to ξ^a. And we call P^a the "energy-momentum vector."

In relativistic mechanics, the mass of the particle is given by the length of its four-momentum $(g_{ab} P^a P^b)^{\frac{1}{2}}$. The corresponding statement in Newtonian mechanics is that the mass of the particle is given by the temporal length of its four-momentum $(t_{ab} P^a P^b)^{\frac{1}{2}}$.

Now we switch from point particles to continuous matter fields. Just as in relativity theory, we associate with each matter field $\mathcal{F}$ a smooth, symmetric field T^{ab}. But the interpretation of T^{ab} is different in Newtonian mechanics (parallel to the way that the interpretation of P^a is different), and here we call it the *mass-momentum* field associated with $\mathcal{F}$. In both cases, T^{ab} codes the four-momentum density of $\mathcal{F}$ as determined, at any point, relative to future-directed, unit timelike vectors ξ^a there. But in the Newtonian case, the four-momentum density is the same for all ξ^a. It is given by $T^{ab} t_b$. (In the relativistic case, it is not invariant and is given, instead, by $T^{ab}\xi_b$. Recall section 2.5.)

Newtonian Mechanics	Relativistic Mechanics
$T^{ab} t_b$ is the four-momentum density of $\mathcal{F}$	$T^{ab}\xi_b$ is the four-momentum density of $\mathcal{F}$ as determined relative to ξ^a

The conservation equation carries over intact from relativistic mechanics:

$$\nabla_a T^{ab} = \mathbf{0}. \tag{4.1.48}$$

We can decompose the Newtonian four-momentum density $T^{ab}t_b$ just as we decomposed P^a to determine an invariant mass-density and a relative three-momentum density. The former is given by $\rho = T^{ab}t_a t_b$. We can take it to be a (Newtonian) "mass condition" that $T^{ab}t_a t_b > 0$ whenever $T^{ab} \neq 0$. When the condition is satisfied, we can further define the fields

$$\eta^a = \frac{1}{\rho}\, T^{ab} t_b,$$

$$p^{ab} = T^{ab} - \rho\, \eta^a \eta^b,$$

and arrive at a canonical representation of T^{ab}:

$$T^{ab} = \rho\, \eta^a \eta^b + p^{ab}. \tag{4.1.49}$$

Here η^a is a smooth, future-directed, unit timelike field, and p^{ab} is a smooth, symmetric field that is spacelike in both indices ($t_a p^{ab} = \mathbf{0}$). In the case of a fluid, for example, we can interpret η^a as the four-velocity of the fluid. In terms of this representation, the conservations equation comes out as

$$\mathbf{0} = \nabla_a T^{ab} = \rho\, \eta^a \nabla_a \eta^b + \eta^b\, [\eta^a \nabla_a\, \rho + \rho\, \nabla_a\, \eta^a] + \nabla_a\, p^{ab}. \tag{4.1.50}$$

Contracting with t_b yields the following equivalence:

$$\nabla_a\, T^{ab} = \mathbf{0} \iff \begin{cases} \rho\, \eta^a\, \nabla_a\, \eta^b + \nabla_a\, p^{ab} & = \ \mathbf{0} \\ \eta^a\, \nabla_a\, \rho + \rho\, (\nabla_a\, \eta^a) & = \ 0. \end{cases}$$

The second equation on the right expresses the conservation of mass. (The analysis we gave in the context of relativity theory carries over intact.) The first is an equation of motion. In the case of a perfect fluid, for example, $p^{ab} = p\, h^{ab}$, where p is the (isotropic) pressure of the fluid. In this case, the first equation comes out as Euler's equation:

$$\rho\, \eta^a\, \nabla_a\, \eta^b = -\nabla^b\, p. \tag{4.1.51}$$

For more on the development of Newtonian mechanics within our geometric framework, see, for example, Ellis [17] and Künzle [35].

4.2. Geometrized Newtonian TheoryFirst Version

Now we turn to Newtonian gravitation theory proper. In the standard (non-geometrized) version, one assumes that the background derivative operator ∇

is flat and posits a gravitational potential ϕ. The gravitational force on a point particle with mass m is given by $-m\, h^{ab}\, \nabla_b\, \phi$. (Notice that this is a spacelike vector by the orthogonality condition.) Using our convention for raising indices, we can also express the vector as $-m\, \nabla^a\, \phi$. It follows that if the particle is subject to no forces except gravity, and if it has four-velocity ξ^a, it satisfies the equation of motion

$$-\nabla^a\, \phi = \xi^n\, \nabla_n\, \xi^a. \tag{4.2.1}$$

(Here we have just used $-m\, \nabla^a\, \phi$ for the left side of equation (4.1.9).) It is also assumed that ϕ satisfies Poisson's equation

$$\nabla_a\, \nabla^a\, \phi \;=\; 4\, \pi\, \rho, \tag{4.2.2}$$

where ρ is the Newtonian mass-density function. (The expression on the left side is an abbreviation for $h^{ab}\, \nabla_a\, \nabla_b\, \phi$.)

In the geometrized formulation of the theory, gravitation is no longer conceived of as a fundamental "force" in the world but rather as a manifestation of spacetime curvature, just as in relativity theory. Rather than thinking of point particles as being deflected from their natural straight trajectories in flat spacetime, one thinks of them as traversing geodesics in curved spacetime. So we have a geometry problem. Starting with a classical spacetime (M, t_a, h^{ab}, ∇), with ∇ flat and with field ϕ on M, can we find a new derivative operator $\overset{g}{\nabla}$ on M, also compatible with t_a and h^{ab}, such that a timelike curve satisfies the equation of motion (4.2.1) with respect to the original derivative operator ∇ iff it is a geodesic with respect to $\overset{g}{\nabla}$? The following proposition (essentially due to Trautman [59]) asserts that there is exactly one such $\overset{g}{\nabla}$. It also records several conditions satisfied by the Riemann curvature tensor field $\overset{g}{R}{}^a{}_{bcd}$ associated with $\overset{g}{\nabla}$. We shall consider the geometric significance of these conditions in section 4.3.

PROPOSITION 4.2.1. *(Geometrization Lemma) Let (M, t_a, h^{ab}, ∇) be a classical spacetime with ∇ flat ($R^a{}_{bcd} = \mathbf{0}$). Further, let ϕ and ρ be smooth real valued functions on M satisfying Poisson's equation $\nabla_a\, \nabla^a\, \phi = 4\, \pi\, \rho$. Finally, let $\overset{g}{\nabla} = (\nabla,\ C^a_{\ bc})$, with $C^a_{\ bc} = -t_b t_c \nabla^a \phi$. Then all the following hold.*

- *(G1) $(M, t_a, h^{ab}, \overset{g}{\nabla})$ is a classical spacetime.*
- *(G2) $\overset{g}{\nabla}$ is the unique derivative operator on M such that, for all timelike curves on M with four-velocity field ξ^a,*

$$\xi^n\, \overset{g}{\nabla}_n\, \xi^a = \mathbf{0} \Longleftrightarrow \xi^n\, \nabla_n\, \xi^a = -\nabla^a \phi. \tag{4.2.3}$$

(G3) The curvature field $\overset{g}{R}{}^{a}{}_{bcd}$ *associated with* $\overset{g}{\nabla}$ *satisfies*

$$\overset{g}{R}_{bc} = 4\,\pi\,\rho\,t_b\,t_c, \tag{4.2.4}$$

$$\overset{g}{R}{}^{a}{}_{b}{}^{c}{}_{d} = \overset{g}{R}{}^{c}{}_{d}{}^{a}{}_{b}, \tag{4.2.5}$$

$$\overset{g}{R}{}^{ab}{}_{cd} = \mathbf{0}. \tag{4.2.6}$$

Proof. For (G1), we need to show that $\overset{g}{\nabla}$ is compatible with t_b and h^{ab}. But this follows from proposition 4.1.3, for we can express $C^{a}_{\ bc}$ in the form $C^{a}_{\ bc} = 2\,h^{an}\,t_{(b}\,\kappa_{c)n}$ if we take $\kappa_{cn} = -t_{[c}\,\nabla_{n]}\,\phi$.

For (G2), let $\overset{g}{\nabla} = (\nabla, C^{a}_{\ bc})$ where (temporarily) $C^{a}_{\ bc}$ is an arbitrary smooth symmetric field on M. Let p be an arbitrary point in M, and let ξ^a be the four-velocity field of an arbitrary timelike curve through p. Then, by equation (1.7.1),

$$\xi^n\,\overset{g}{\nabla}_n\,\xi^a = \xi^n\nabla_n\,\xi^a - C^{a}_{\ rn}\,\xi^r\xi^n.$$

It follows that $\overset{g}{\nabla}$ will satisfy (G2) iff $C^{a}_{\ rn}\xi^r\xi^n = -\nabla^a\phi$ or, equivalently,

$$[C^{a}_{\ rn} + (\nabla^a\phi)\,t_r t_n]\,\xi^r\xi^n = \mathbf{0} \tag{4.2.7}$$

for all future-directed unit timelike vectors ξ^a at all points p. But the set of future-directed unit timelike vectors at any p spans the tangent space M_p there. (Why? Let $\overset{1}{\xi}{}^a, \ldots, \overset{4}{\xi}{}^a$ be an orthonormal basis for $t_{ab} = t_a t_b$ in the sense discussed above. [So $t_a\overset{1}{\xi}{}^a = 1$, and $t_a\overset{i}{\xi}{}^a = 0$ for $i = 2, 3, 4$.] Then $\overset{1}{\xi}{}^a$, $(\overset{1}{\xi}{}^a + \overset{2}{\xi}{}^a)$, $(\overset{1}{\xi}{}^a + \overset{3}{\xi}{}^a)$, and $(\overset{1}{\xi}{}^a + \overset{4}{\xi}{}^a)$ are all future-directed unit timelike vectors, and the set is linearly independent.) And the field in brackets in equation (4.2.7) is symmetric in its covariant indices. So, $\overset{g}{\nabla}$ will satisfy (G2) iff $C^{a}_{\ rn} = -(\nabla^a\phi)\,t_r t_n$ everywhere.

Finally, for (G3) we use equation (1.8.2). We have

$$\begin{aligned}\overset{g}{R}{}^{a}{}_{bcd} &= R^{a}{}_{bcd} + 2\,\nabla_{[c}\,C^{a}{}_{d]b} + 2\,C^{n}{}_{b[c}C^{a}{}_{d]n}\\ &= R^{a}{}_{bcd} - 2\,t_b t_{[d}\nabla_{c]}\,\nabla^a\phi = -\,2\,t_b t_{[d}\nabla_{c]}\nabla^a\phi.\end{aligned} \tag{4.2.8}$$

(Here $C^{n}{}_{b[c}C^{a}{}_{d]n} = \mathbf{0}$ by the orthogonality condition, and $\nabla_{[c}\,C^{a}{}_{d]b} = -t_b\,t_{[d}\nabla_{c]}\nabla^a\phi$ by the compatibility condition. For the final equality, we use our assumption that $R^{a}_{\ bcd} = \mathbf{0}$.) Equation (4.2.6) now follows from the orthogonality condition. Equation (4.2.5) follows from that and the fact that $\nabla^{[c}\nabla^{a]}\phi = \mathbf{0}$ for *any* smooth function ϕ. Finally, contraction on a and d yields

$$\overset{g}{R}_{bc} = t_b t_c (\nabla_a \nabla^a \phi). \tag{4.2.9}$$

So equation (4.2.4) follows from our assumption that $\nabla_a \nabla^a \phi = 4\pi\rho$. □

Equation (4.2.4) is the geometrized version of Poisson's equation. In the special case where $\rho = 0$ everywhere, of course, it reduces to $\overset{g}{R}_{bc} = \mathbf{0}$, which we recognize as Einstein's equation in the corresponding special case in which $T_{bc} = \mathbf{0}$. Even in the general case, equation (4.2.4) can be reformulated so as to have almost exactly the same structure as Einstein's equation. Recall our discussion of the mass-momentum field T^{ab} toward the end of section 4.1. We saw there that it encodes ρ via

$$\rho = T^{mn}\, t_{mn}.$$

(We shall temporarily revert to writing t_{ab}, rather than $t_a t_b$, to emphasize the field's relation to a two index Lorentzian metric g_{ab}, but nothing turns on our doing so.) So we can certainly formulate Poisson's equation directly in terms of T^{ab}. Now consider the fields

$$\hat{T}_{bc} = T^{mn}\, t_{mb}\, t_{nc} = \rho\, t_{bc},$$
$$\hat{T} = T^{mn}\, t_{mn} = \rho.$$

(Caution is required here. It must be remembered that we cannot recover T^{bc} from $\hat{T}_{bc}$ by "raising indices" with h^{ab}, since $\hat{T}_{mn}\, h^{mb}\, h^{nc} = \mathbf{0}$.) Using these fields, we can express Poisson's equation in the form

$$\overset{g}{R}_{bc} = 8\pi\,(\hat{T}_{bc} - \frac{1}{2}\, t_{bc}\, \hat{T}), \tag{4.2.10}$$

which is very close indeed to Einstein's equation (2.7.2).

Moreover, if we start with a version of Poisson's equation that incorporates a "cosmological constant"

$$\nabla_a \nabla^a \phi + \Lambda = 4\pi\rho, \tag{4.2.11}$$

then substitution for $\nabla_a \nabla^a \phi$ in equation (4.2.9) yields

$$\overset{g}{R}_{bc} = 4\pi\,\rho\, t_b t_c - \Lambda\, t_b t_c \tag{4.2.12}$$

(but everything else in the proof remains the same). And *this* equation, in turn, can be expressed as

$$\overset{g}{R}_{bc} = 8\pi\,(\hat{T}_{bc} - \frac{1}{2}\, t_{bc}\, \hat{T}) - \Lambda\, t_{bc}, \tag{4.2.13}$$

which matches equation (2.7.4).

So far, we have seen how to pass from a standard to a geometrized formulation of Newtonian theory. It is also possible to work in the opposite direction. In Trautman's [59] version of geometrized Newtonian gravitation theory—one of two we shall consider[2]—one *starts* with a curved derivative operator ∇ satisfying equations (4.2.4), (4.2.5), and (4.2.6), and with the principle that point particles subject to no forces (except "gravity") traverse geodesics with respect to ∇. Equations (4.2.5) and (4.2.6) function as integrability conditions that ensure the possibility of working backwards to recover the standard formulation in terms of a gravitational potential ϕ and flat derivative operator $\overset{f}{\nabla}$. We shall prove this recovery, or de-geometrization, theorem in this section (proposition 4.2.5), and we shall see that, in the absence of special boundary conditions, the pair $(\overset{f}{\nabla}, \phi)$ that one recovers is not unique.

Later, in section 4.5, we shall consider a second, more general version of geometrized Newtonian gravitation theory, developed by Künzle [34, 35] and Ehlers [15], in which one of the two supplemental curvature conditions is dropped.

$$\text{Trautman Version}\ \left\{\begin{array}{l} \left.\begin{array}{lcl} R_{bc} & = & 4\pi\,\rho\, t_b t_c \\ R^a{}_b{}^c{}_d & = & R^c{}_d{}^a{}_b \end{array}\right\}\ \text{Künzle-Ehlers Version} \\ \begin{array}{lcl} R^{ab}{}_{cd} & = & \mathbf{0} \end{array} \end{array}\right.$$

At issue here is whether "Newtonian gravitation theory" is to qualify as a limiting version of relativity theory. The geometrized version of Poisson's equation does, in a natural sense, qualify as a limiting form of Einstein's equation. And the first of Trautman's two supplemental curvature conditions ($R^a{}_b{}^c{}_d = R^c{}_d{}^a{}_b$) holds automatically in relativistic spacetimes. (Recall the fourth clause of proposition 1.9.4.) So it naturally carries over to any limiting version of relativity theory. But the second supplemental curvature condition does not hold in relativistic spacetimes (unless they happen to be flat), and it is therefore not an automatic candidate for inclusion in a limiting version of relativity theory. It is crucially important that the conditions $R^{ab}{}_{cd} = \mathbf{0}$ and $R^a{}_{bcd} = \mathbf{0}$ are *not* equivalent for classical spacetime structures, though they are for relativistic ones.

Starting only from the weaker assumptions of Künzle and Ehlers, one can still prove a recovery theorem of sorts. But the (de-geometrized) gravitation theory one recovers is not Newtonian theory proper, but rather a generalized version of it. In this version, the gravitational force acting on a particle of unit mass is given by a vector field, but it need not be of the form $\nabla^a \phi$. Moreover, the

2. See Bain [4] for a systematic discussion of these and yet other versions.

de-geometrized field equations to which one is led involve a "rotation field" ω_{ab}. We shall eventually prove this recovery theorem for the Künzle-Ehlers version of Newtonian theory (proposition 4.5.2), and also consider special circumstances under which the difference between the two versions of geometrized Newtonian theory collapses. It turns out that the second curvature condition ($R^{ab}{}_{cd} = \mathbf{0}$) is satisfied automatically, for example, in classical spacetimes that are, in a certain weak sense, asymptotically flat (see section 4.5), and also in Newtonian cosmological models that satisfy a natural homogeneity and isotropy condition (see section 4.4).

Before turning to the Trautman Recovery Theorem, we isolate a few needed facts. Let ξ^a be a smooth, future-directed, unit timelike field in a classical spacetime (M, t_a, h^{ab}, ∇). We say that it is *rigid* (or *non-expanding)* if $\mathcal{L}_\xi\, h^{ab} = \mathbf{0}$ or, equivalently, $\nabla^{(a}\xi^{b)} = \mathbf{0}$. (These conditions obtain, we know, iff the expansion field θ_{ab} associated with ξ^a vanishes. Recall equation (4.1.47).) Certain things we have established about Killing fields (which we have defined only in connection with non-degenerate metrics) carry over to rigid fields in classical spacetimes. So, for example, we have the following.

PROPOSITION 4.2.2. *Let (M, t_a, h^{ab}, ∇) be a classical spacetime, and let ξ^a be a smooth, future-directed, unit timelike field that is rigid. Then*

$$\nabla^n\, \nabla^a\, \xi^b = R^{ba}{}_r{}^n\, \xi^r. \tag{4.2.14}$$

Proof. The proof is a just a variant of that used for proposition 1.9.8. Cycling indices, we have

$$\nabla^n\nabla^a\xi^b - \nabla^a\nabla^n\xi^b = -R^b{}_r{}^{na}\xi^r,$$

$$\nabla^b\nabla^n\xi^a - \nabla^n\nabla^b\xi^a = -R^a{}_r{}^{bn}\xi^r,$$

$$\nabla^a\nabla^b\xi^n - \nabla^b\nabla^a\xi^n = -R^n{}_r{}^{ab}\xi^r.$$

Subtracting the third line from the sum of the first two (and using the fact that $\nabla^{(a}\xi^{b)} = \mathbf{0}$) yields

$$2\,\nabla^n\nabla^a\xi^b = (-R^b{}_r{}^{na} - R^a{}_r{}^{bn} + R^n{}_r{}^{ab})\,\xi^r.$$

Finally, we reformulate the expression in parentheses on the right side using the symmetries $R^a{}_{[bcd]} = \mathbf{0}$, $R^{ab}{}_{(cd)} = \mathbf{0}$, and $R^{(ab)}{}_{cd} = \mathbf{0}$:

$$\begin{aligned}
-R^{b}{}_{r}{}^{na} - R^{a}{}_{r}{}^{bn} + R^{n}{}_{r}{}^{ab} &= (R^{ba}{}_{r}{}^{n} + R^{bna}{}_{r}) - R^{a}{}_{r}{}^{bn} + R^{n}{}_{r}{}^{ab} \\
&= R^{ba}{}_{r}{}^{n} + (R^{nb}{}_{r}{}^{a} + R^{n}{}_{r}{}^{ab}) - R^{a}{}_{r}{}^{bn} \\
&= R^{ba}{}_{r}{}^{n} - R^{nab}{}_{r} - R^{a}{}_{r}{}^{bn} \\
&= R^{ba}{}_{r}{}^{n} - (R^{an}{}_{r}{}^{b} + R^{a}{}_{r}{}^{bn}) \\
&= R^{ba}{}_{r}{}^{n} + R^{abn}{}_{r} \;=\; 2\, R^{ba}{}_{r}{}^{n}.
\end{aligned}$$

So we have equation (4.2.14). □

Our proof of the Trautman Recovery Theorem turns on the availability of a unit timelike field η^a that is rigid and twist-free ($\nabla^a \eta^b = \mathbf{0}$). The latter provides a backbone, of sorts, for our construction. The following proposition shows that the condition $R^{ab}{}_{cd} = \mathbf{0}$ insures the existence of such fields (at least locally).

PROPOSITION 4.2.3. *Let (M, t_a, h^{ab}, ∇) be a classical spacetime that is spatially flat ($R^{abcd} = \mathbf{0}$). Let $\gamma : I \to M$ be a smooth, future-directed timelike curve with unit tangent field $\widehat{\eta}^{\,a}$, and let p be any point in $\gamma[I]$. Then there is an open set O containing p, a smooth spacelike field χ^a on O, and a smooth, future-directed, unit timelike field η^a on O such that $\chi^a = \mathbf{0}$ on $\gamma[I]$, $\eta^a = \widehat{\eta}^{\,a}$ on $\gamma[I]$, and*

$$\nabla_a \chi^b = \delta_a{}^b - t_a \eta^b. \tag{4.2.15}$$

Furthermore, (i) if $R^{ab}{}_{cd} = \mathbf{0}$, then $\nabla^a \eta^b = \mathbf{0}$; and (ii) if $R^{a}{}_{bcd} = \mathbf{0}$ and if γ is a geodesic, then $\nabla_a\, \eta^b = \mathbf{0}$.

Proof. First, we claim there exists a smooth spacelike field χ^a on some open set O containing p such that

$$\nabla^a \chi^b = h^{ab} \tag{4.2.16}$$

and $\chi^a = \mathbf{0}$ on $\gamma[I]$. Indeed, as restricted to any one spacelike hypersurface S, χ^a is just the familiar "position vector field" (relative to the point where $\gamma[I]$ intersects S). (Recall proposition 1.7.12. All we need here is that the [three-dimensional, invertible] metric g_{ab} induced on S by h^{ab} is flat and so, at least locally, the pair (S, g_{ab}) is isometric to three-dimensional Euclidean space.) Now let ξ^a be any smooth, future-directed, unit timelike field on O. Consider the field $\eta^b = (-\xi^a \nabla_a \chi^b + \xi^b)$. We claim that it satisfies all the required conditions. First, it satisfies equation (4.2.15). This follows since the two fields $(-\nabla_a \chi^b + \delta_a{}^b)$ and $t_a \eta^b$ yield the same result when contracted with either h^{na} or ξ^a. Next, it is clearly a future-directed, unit timelike field; i.e., $t_b\eta^b = t_b(-\xi^a\nabla_a\chi^b + \xi^b) = 1$. Third, it agrees with $\widehat{\eta}^a$ on $\gamma[I]$. For since χ^a vanishes on $\gamma[I]$, it is certainly constant along the curve; i.e., $\widehat{\eta}^a\nabla_a\chi^b = \mathbf{0}$. So, by equation (4.2.15), we have $\mathbf{0} = \widehat{\eta}^a\nabla_a\chi^b = \widehat{\eta}^a(\delta_a{}^b - t_a\eta^b) = \widehat{\eta}^b - \eta^b$ on $\gamma[I]$.

Now we turn to the curvature conditions. By equation (4.2.15) again,

$$\eta^a\nabla^n\nabla_a\chi^b = \eta^a\nabla^n(\delta_a{}^b - t_a\eta^b) = -\nabla^n\eta^b.$$

Hence,

$$\nabla^n\eta^b = -\eta^a(\nabla_a\nabla^n\chi^b - R^b{}_m{}^n{}_a\chi^m) = \eta^a R^b{}_m{}^n{}_a\chi^m,$$

since, by equation (4.2.16), $\nabla_a\nabla^n\chi^b = \nabla_a h^{nb} = \mathbf{0}$. Since χ^a is spacelike, we can express it in the form $\chi^a = h^{ab}\hat{\chi}_b$. Thus we have

$$\nabla^n\eta^b = R^{bmn}{}_a\,\hat{\chi}_m\eta^a. \tag{4.2.17}$$

So, if $R^{bm}{}_{na} = \mathbf{0}$, it clearly follows that $\nabla^n\eta^b = \mathbf{0}$. This gives us (i).

Now assume that γ is a geodesic and $R^a{}_{bcd} = \mathbf{0}$. Then $\nabla_a\eta^b = \mathbf{0}$ on $\gamma[I]$. (Why? $\eta^a\nabla_a\eta^b = \mathbf{0}$ on $\gamma[I]$ since γ is a geodesic, and $h^{na}\nabla_a\eta^b = \mathbf{0}$ *everywhere* by (i).) We may assume (by moving to a smaller open set O containing p if necessary) that every maximally extended spacelike hypersurface in O intersects $\gamma[I]$. So it will suffice for (ii) to show that $\nabla_a\eta^b$ is constant on spacelike hypersurfaces; i.e., $\nabla^c\nabla_a\eta^b = \mathbf{0}$. But this follows immediately from $R^a{}_{bcd} = \mathbf{0}$ and $\nabla^a\eta^b = \mathbf{0}$, since $\nabla^c\nabla_a\eta^b = \nabla_a\nabla^c\eta^b - R^b{}_m{}^c{}_a\eta^m$. □

Proposition 4.2.3 yields a useful characterization of the relative strength of two curvature conditions. (Here and throughout it should be understood that when we formulate a curvature equation without qualification, as on the left sides of (1) and (2) that follow in proposition 4.2.4, we have in mind the condition that the equation hold at all points.)

PROPOSITION 4.2.4. *Let (M, t_a, h^{ab}, ∇) be a classical spacetime that is spatially flat ($R^{abcd} = \mathbf{0}$). Then the following both hold.*

(1) $R^{ab}{}_{cd} = \mathbf{0}$ *iff there exists, at least locally, a smooth unit timelike field η^a that is rigid and twist-free ($\nabla^a\,\eta^b = \mathbf{0}$).*
(2) $R^a{}_{bcd} = \mathbf{0}$ *iff there exists, at least locally, a smooth unit timelike field η^a that is rigid, twist-free, and acceleration-free ($\nabla_a\,\eta^b = \mathbf{0}$).*

Proof. The "only if" clauses follow from the preceding proposition. The other drections are easy. (1) Assume that for any point p in M, there exists a smooth unit timelike field η^a defined on an open set containing p such that $\nabla^a \eta^b = \mathbf{0}$. We show that $R^{ab}{}_{cd}$ vanishes at p. We have $R^{ab}{}_{cd}\, \eta^c \eta^d = \mathbf{0}$ at p since $R^{ab}{}_{cd}$ is anti-symmetric in the indices c and d. We also have $R^{ab}{}_{cd}\, h^{cr} h^{ds} = \mathbf{0}$ at p (by our assumption of spatial flatness). So to prove that $R^{ab}{}_{cd}$ vanishes at p, it suffices to show that contraction there with $\eta^c\, h^{ds}$ (or $h^{cr}\, \eta^d$) yields $\mathbf{0}$. But this follows since $\nabla^b\, \eta^a = \mathbf{0}$ and hence, by proposition 4.2.2, $R^{ab}{}_{cd}\, \eta^c\, h^{ds} = \nabla^s\, \nabla^b\, \eta^a = \mathbf{0}$.

(2) Next, assume that for any point p in M, there exists a smooth unit timelike field η^a defined on an open set containing p such that $\nabla_a\, \eta^b = \mathbf{0}$. We show that $R^a{}_{bcd}$ vanishes at p. We know from part (1) of the proposition (and the fact that $\nabla_a\, \eta^b = \mathbf{0}$ implies $\nabla^a\, \eta^b = \mathbf{0}$) that $R^{ab}{}_{cd} = \mathbf{0}$ at p. The latter condition implies that $R^a{}_{bcd} = t_b R^a{}_{ncd}\, \eta^n$. (Contracting both sides with either h^{br} or η^b yields the same result.) But since $\nabla_a\, \eta^d = \mathbf{0}$, we also have $R^a{}_{ncd}\, \eta^n = -2\, \nabla_{[c} \nabla_{d]}\, \eta^a = \mathbf{0}$ at p. So we are done. □

Now we turn to our first recovery theorem. Our formulation is purely local in character since we have opted not to impose special global topological constraints on the underlying manifold M. Our proof is a bit different from that in Trautman [59].

PROPOSITION 4.2.5. *(Trautman Recovery Theorem) Let (M, t_a, h^{ab}, ∇) be a classical spacetime that satisfies*

$$R_{bc} = 4\,\pi\,\rho\, t_b t_c, \tag{4.2.18}$$

$$R^a{}_b{}^c{}_d = R^c{}_d{}^a{}_b, \tag{4.2.19}$$

$$R^{ab}{}_{cd} = \mathbf{0} \tag{4.2.20}$$

for some smooth scalar field ρ on M. Then given any point p in M, there is an open set O containing p, a smooth scalar field ϕ on O, and a derivative operator $\overset{f}{\nabla}$ on O such that all the following hold on O.

(R1) $\overset{f}{\nabla}$ is compatible with t_a and h^{ab}.
(R2) $\overset{f}{\nabla}$ is flat.
(R3) For all timelike curves with four-velocity field ξ^a,

$$\xi^n \nabla_n\, \xi^a = \mathbf{0} \iff \xi^n\, \overset{f}{\nabla}_n\, \xi^a = -\overset{f}{\nabla}{}^a \phi. \tag{4.2.21}$$

(R4) $\overset{f}{\nabla}$ satisfies Poisson's equation $\overset{f}{\nabla}_a\, \overset{f}{\nabla}{}^a \phi = 4\,\pi\,\rho$.

The pair $(\overset{f}{\nabla}, \phi)$ *is not unique. A second pair* $(\overset{f}{\nabla}{}', \phi')$ *(defined on the same open set* O*) will satisfy the stated conditions iff*

(U1) $\nabla^a \nabla^b (\phi' - \phi) = \mathbf{0}$, *and*
(U2) $\overset{f}{\nabla}{}' = (\overset{f}{\nabla}, C'^a{}_{bc})$, *where* $C'^a{}_{bc} = t_b\, t_c \nabla^a (\phi' - \phi)$.

Proof. Let p be a point in M. As we have just seen, it follows from $R^{ab}{}_{cd} = \mathbf{0}$ that we can find an open set O containing p, as well as a smooth, future-directed, unit timelike vector field η^a on O that is rigid and twist-free ($\nabla^a \eta^b = \mathbf{0}$). Let ϕ^a be the acceleration field of η^a; i.e., $\phi^a = \eta^n \nabla_n \eta^a$. Then we have

$$\nabla_a \eta^b = t_a \phi^b. \tag{4.2.22}$$

(This follows since contraction of the two sides with both η^a and h^{an} yields the same result.) Further, let $\overset{f}{\nabla}$ be the derivative operator on O defined by $\overset{f}{\nabla} = (\nabla, C^a{}_{bc})$, where $C^a{}_{bc} = t_b\, t_c\, \phi^a$. Clearly, $t_a\, C^a{}_{bc} = \mathbf{0}$, $C^a{}_{bc}\, h^{bn} = \mathbf{0}$, and $C^a{}_{bc}\, h^{cn} = \mathbf{0}$. It follows that

$$\overset{f}{\nabla}_a\, t_b = \nabla_a\, t_b + t_n\, C^n{}_{ab} = \nabla_a\, t_b,$$

$$\overset{f}{\nabla}_a\, h^{bc} = \nabla_a\, h^{bc} - h^{nc}\, C^b{}_{na} - h^{bn}\, C^c{}_{na} = \nabla_a\, h^{bc}.$$

So, since ∇ is compatible with t_a and h^{bc}, $\overset{f}{\nabla}$ is compatible with them as well. So we have (R1). Notice next that $C^b{}_{an}\, \eta^n = t_a\, \phi^b$ and so, by equation (4.2.22),

$$\overset{f}{\nabla}_a\, \eta^b = \nabla_a\, \eta^b - C^b{}_{an}\, \eta^n = t_a\, \phi^b - t_a\, \phi^b = \mathbf{0}. \tag{4.2.23}$$

Thus, η^a is constant with respect to the new derivative operator $\overset{f}{\nabla}$.

Now we consider the curvature field associated with $\overset{f}{\nabla}$. We have $C^n{}_{bc}\, C^a{}_{dn} = \mathbf{0}$ since $\phi^n\, t_n = \mathbf{0}$. So, by equation (1.8.2),

$$\begin{aligned} \overset{f}{R}{}^a{}_{bcd} &= R^a{}_{bcd} + 2\, \nabla_{[c}\, C^a{}_{d]b} + 2\, C^n{}_{b[c}\, C^a{}_{d]n} \\ &= R^a{}_{bcd} + 2\, t_b\, t_{[d}\, \nabla_{c]}\, \phi^a. \end{aligned} \tag{4.2.24}$$

It follows immediately that $\overset{f}{R}{}^{abcd} = R^{abcd} = \mathbf{0}$ (since $R^{ab}{}_{cd} = \mathbf{0}$). So $\overset{f}{\nabla}$ is spatially flat. But now recall the second clause of proposition 4.2.4. We have just verified that there is a smooth, unit timelike field η^a on O that is constant with respect to $\overset{f}{\nabla}$. So (since $\overset{f}{\nabla}$ is spatially flat), the proposition tells us that $\overset{f}{\nabla}$ must be flat outright; i.e., $\overset{f}{R}{}^{a}{}_{bcd} = \mathbf{0}$. So we have (R2). And equation (4.2.24) reduces to

$$R^{a}{}_{bcd} = -2\, t_b\, t_{[d}\, \nabla_{c]}\, \phi^{a}. \tag{4.2.25}$$

Now we extract further information from equation (4.2.25). Raising and contracting indices yields

$$R^{a}{}_{b}{}^{c}{}_{d} = -t_b\, t_d\, \nabla^{c}\, \phi^{a}, \tag{4.2.26}$$

$$R_{bc} = t_b\, t_c\, \nabla_a\, \phi^{a}. \tag{4.2.27}$$

Since we are assuming $R^{a}{}_{b}{}^{c}{}_{d} = R^{c}{}_{d}{}^{a}{}_{b}$, it follows from the first of these assertions that $\nabla^{[c}\, \phi^{a]} = \mathbf{0}$. This implies that (after possibly further restricting O to some smaller open set containing p) there is a smooth scalar field ϕ on O such that $\phi^a = \nabla^a \phi$. (Here we invoke proposition 4.1.6.) And since we are assuming $R_{bc} = 4\pi\, \rho\, t_b\, t_c$, it follows from the second assertion that $\nabla_a \nabla^a \phi = \nabla_a\, \phi^a = 4\pi\rho$. But $C^{a}{}_{an} = t_a\, t_n\, \phi^a = \mathbf{0}$ and, therefore,

$$\overset{f}{\nabla}_a\, \overset{f}{\nabla}{}^{a}\phi = \nabla_a\, \overset{f}{\nabla}{}^{a}\phi - C^{a}{}_{an}\, \overset{f}{\nabla}{}^{n}\phi = \nabla_a \nabla^a \phi. \tag{4.2.28}$$

So $\overset{f}{\nabla}_a\, \overset{f}{\nabla}{}^{a}\phi = 4\pi\rho$. That is, we have (R4).

For (R3), note that for all timelike curves in O with four-velocity field ξ^a,

$$\begin{aligned} \xi^n\, \overset{f}{\nabla}_n\, \xi^a &= \xi^n (\nabla_n \xi^a - C^{a}{}_{nm}\, \xi^m) = \xi^n \nabla_n \xi^a - (t_n\, t_m\, \nabla^a \phi)\, \xi^n\, \xi^m \\ &= \xi^n \nabla_n \xi^a - \nabla^a \phi. \end{aligned}$$

So $\xi^n \nabla_n \xi^a = \mathbf{0}$ iff $\xi^n\, \overset{f}{\nabla}_n\, \xi^a = -\nabla^a \phi$.

Finally, we consider the non-uniqueness of the pair $(\overset{f}{\nabla}, \phi)$. Let $(\overset{f}{\nabla}{}', \phi')$ be a second pair on O. Consider fields $C'^{a}{}_{bc}$ and ψ on O defined by $\overset{f}{\nabla}{}' = (\overset{f}{\nabla}, C'^{a}{}_{bc})$ and $\psi = \phi' - \phi$. We first show that if the new pair satisfies the stated conditions of the proposition, then it must be the case that $\nabla^a \nabla^b \psi = \mathbf{0}$ and $C'^{a}{}_{bc} = t_b\, t_c \nabla^a \psi$.

Assume $(\overset{f}{\nabla}{}', \phi')$ satisfies (R1)–(R4). Then—since $(\overset{f}{\nabla}, \phi)$ and $(\overset{f}{\nabla}{}', \phi')$ *both* satisfy (R3)—we have

$$\xi^n\, \overset{f}{\nabla}_n\, \xi^a + \overset{f}{\nabla}{}^{a}\phi = \mathbf{0} \Longleftrightarrow \xi^n \nabla_n\, \xi^a = \mathbf{0} \Longleftrightarrow \xi^n\, \overset{f}{\nabla}{}'_n\, \xi^a + \overset{f}{\nabla}{}'^{a}\phi' = \mathbf{0}$$

for all timelike curves with four-velocity field ξ^a. But $\overset{f}{\nabla}{}'^a\phi' = \overset{f}{\nabla}{}^a\phi' = \overset{f}{\nabla}{}^a\phi + \overset{f}{\nabla}{}^a\psi$. And $\xi^n\, \overset{f}{\nabla}{}'_n\, \xi^a = \xi^n(\overset{f}{\nabla}_n\, \xi^a - C'^a{}_{nm}\xi^m)$. So it must be the case that, for all future-directed, unit timelike vectors ξ^a at all points in O,

$$C'^a{}_{nm}\xi^m\xi^n = \overset{f}{\nabla}{}^a\psi.$$

And from this it follows that $C'^a{}_{mn} = t_m\, t_n\, \overset{f}{\nabla}{}^a\psi = t_m\, t_n \nabla^a\psi$, as required. (Recall the argument for a corresponding assertion in our proof of the Geometrization Lemma.) Now the curvature fields of $(\overset{f}{\nabla}, \phi)$ and $(\overset{f}{\nabla}{}', \phi')$ are related by

$$\overset{f}{R}{}'^a{}_{bcd} = \overset{f}{R}{}^a{}_{bcd} + 2\, t_b t_{[d}\, \overset{f}{\nabla}_{c]}\, \overset{f}{\nabla}{}^a\psi.$$

(The argument here is exactly the same as given for equation (4.2.24).) Since $\overset{f}{\nabla}$ and $\overset{f}{\nabla}{}'$ are both flat, it follows that $t_{[d}\, \overset{f}{\nabla}_{c]}\overset{f}{\nabla}{}^a\psi = \mathbf{0}$ or, equivalently, $\overset{f}{\nabla}{}^c\, \overset{f}{\nabla}{}^a\psi = \mathbf{0}$. But $\nabla^c\nabla^a\psi = \overset{f}{\nabla}{}^c\, \overset{f}{\nabla}{}^a\psi$. (Indeed, $\overset{f}{\nabla}{}^c$ and ∇^c agree in their action on *all* vector fields λ^a, since $\overset{f}{\nabla}{}^c\lambda^a = \nabla^c\lambda^a - C^{ac}{}_n\lambda^n$ and $C^{ac}{}_n = \mathbf{0}$.) So $\nabla^c\nabla^a\psi = \mathbf{0}$, and we are done with the first direction.

Conversely, assume that $C'^a{}_{bc} = t_b\, t_c \nabla^a\psi$ and $\nabla^a\nabla^b\psi = \mathbf{0}$. The first assumption alone implies that $\overset{f}{\nabla}{}'$ is compatible with t_a and h^{ab}. And by reversing the steps in the preceding paragraphs, we can show that $(\overset{f}{\nabla}{}', \phi')$ satisfies (R2) and (R3). That leaves only (R4). For this, note first that since $C'^a{}_{an} = \mathbf{0}$,

$$\begin{aligned}
\overset{f}{\nabla}{}'_a\, \overset{f}{\nabla}{}'^a\phi' &= \overset{f}{\nabla}_a\, \overset{f}{\nabla}{}'^a\phi' - C'^a{}_{an}\, \overset{f}{\nabla}{}'^n\phi' = \overset{f}{\nabla}_a\, \overset{f}{\nabla}{}^a\phi' \\
&= \overset{f}{\nabla}_a\, \overset{f}{\nabla}{}^a\phi + \overset{f}{\nabla}_a\, \overset{f}{\nabla}{}^a\psi \\
&= 4\pi\,\rho + \overset{f}{\nabla}_a\, \overset{f}{\nabla}{}^a\psi = 4\pi\,\rho + \nabla_a\nabla^a\psi.
\end{aligned}$$

(The penultimate equality holds because $(\overset{f}{\nabla}, \phi)$ satisfies (R3); and the argument for the final equality is exactly the same as the one given for equation (4.2.28).) But $\nabla^a\nabla^b\psi = \mathbf{0}$ and, so, $\nabla_a\nabla^b\psi = t_a\,\xi^n\nabla_n\nabla^b\psi$, where ξ^n is any smooth, future-directed unit timelike field on O. It follows that $\nabla_a\nabla^a\psi = \mathbf{0}$ and, therefore, $\overset{f}{\nabla}{}'_a\, \overset{f}{\nabla}{}'^a\phi' = 4\pi\,\rho$, as required for (R4). □

Just as with the Geometrization Lemma, only a small change is necessary here if we want to work with a cosmological constant. If we replace equation

(4.2.18) with $R_{bc} = 4\pi\,\rho\, t_b t_c - \Lambda\, t_b t_c$, then substitution for R_{bc} in equation (4.2.27) yields $\nabla_a \phi^a + \Lambda = 4\pi\,\rho$. The further argument that ϕ^a is of the form $\nabla^a\phi$ is unaffected. So we are led to equation (4.2.11).

The Trautman Recovery Theorem tells us that if ∇ arises as the geometrization of the pair $(\overset{f}{\nabla}, \phi)$, then, for any field ψ such that $\nabla^a\nabla^b\psi = \mathbf{0}$, it also arises as the geometrization of $(\overset{f}{\nabla}{}', \phi')$ where $\phi' = \phi + \psi$ and $\overset{f}{\nabla}{}' = (\overset{f}{\nabla},\, t_b\, t_c \nabla^a\psi)$.

$$\begin{array}{ccc} (\overset{f}{\nabla}, \phi) & & \\ & \searrow & \\ & & \nabla \\ & \nearrow & \\ (\overset{f}{\nabla}{}', \phi') & & \end{array}$$

We certainly have sufficient freedom here to insure that $\overset{f}{\nabla}{}'$ is, in fact, distinct from $\overset{f}{\nabla}$. We can think of $\nabla^b\psi$ as the "spatial gradient" of ψ. The stated condition on ψ, namely $\nabla^a\nabla^b\psi = \mathbf{0}$, is just the requirement that this spatial gradient be constant on spacelike hypersurfaces. The condition can certainly be satisfied without that gradient vanishing at all points. (Its value can change from one spacelike hypersurface to another.) And if $\nabla^a\psi \neq \mathbf{0}$ at some point p, then $\overset{f}{\nabla}{}'$ cannot be the same operator as $\overset{f}{\nabla}$. Indeed, let ξ^a be the four-velocity field of a timelike curve passing through p. Then at p,

$$\xi^a\, \overset{f}{\nabla}{}'_a \xi^b = \xi^a (\overset{f}{\nabla}_a\, \xi^b - (t_a t_n \nabla^b\psi)\xi^n) = \xi^a\, \overset{f}{\nabla}_a\, \xi^b - \nabla^b\psi \;\neq\; \xi^a\, \overset{f}{\nabla}_a\, \xi^b.$$

We can use the current discussion to capture in precise language the standard claim that gravitational force in (standard) Newtonian theory is a gauge quantity. Consider a point particle with mass m and four-velocity ξ^a that is not accelerating with respect to ∇. According to the de-geometrization $(\overset{f}{\nabla}, \phi)$, the particle has acceleration $\xi^n\, \overset{f}{\nabla}_n\, \xi^a$ and is subject to gravitational force $-m\, \overset{f}{\nabla}{}^a\phi = -m\nabla^a\phi$. (We get this from (R3).) Rather than being subject to no forces at all—the account given by the geometrized formulation of the theory—it is here taken to be subject to two "forces" (inertial and gravitational) that cancel each other. Alternatively, according to the de-geometrization $(\overset{f}{\nabla}{}', \phi')$, it has acceleration $\xi^n\, \overset{f}{\nabla}{}'_n \xi^a = \xi^n\, \overset{f}{\nabla}_n\, \xi^a - \nabla^a\,\psi$ and is subject to gravitational force $-m\, \overset{f}{\nabla}{}'^a\, \phi' = -m\nabla^a\phi - m\nabla^a\psi$. *So the gravitational force on the particle is determined only up to a factor $m\nabla^a\psi$, where $\nabla^a\psi$ is constant on any one spacelike hypersurface but can change over time.*

Of course, if boundary conditions are brought into consideration, we regain the possibility of unique de-geometrization. In particular, if we are dealing with a bounded mass distribution—i.e., if ρ has compact support on every spacelike hypersurface—then it seems appropriate to require that the gravitational field die off as one approaches spatial infinity. But if $\nabla^a\psi$ is constant on spacelike hypersurfaces *and* if it goes to $\mathbf{0}$ at spatial infinity, then it must vanish everywhere.

4.3. Interpreting the Curvature Conditions

In this section, we consider the geometric significance of three curvature conditions that appear in Trautman's formulation of geometrized Newtonian gravitation theory:

$$R_{ab} = 4\pi\rho\, t_a t_b, \tag{4.3.1}$$

$$R^a{}_b{}^c{}_d = R^c{}_d{}^a{}_b, \tag{4.3.2}$$

$$R^{ab}{}_{cd} = \mathbf{0}. \tag{4.3.3}$$

We start with the third. We know already (from proposition 4.2.4) that it holds in a classical spacetime iff the latter is spatially flat ($R^{abcd} = \mathbf{0}$) and, at least locally, admits a unit timelike vector field ξ^a that is rigid and twist-free ($\nabla^a\xi^b = \mathbf{0}$). We also have the following more direct interpretation.

PROPOSITION 4.3.1. *Let (M, t_a, h^{ab}, ∇) be a classical spacetime. Then $R^{ab}{}_{cd} = \mathbf{0}$ throughout M iff parallel transport of spacelike vectors within M is, at least locally, path independent.*

Proof. (*If*) This direction is immediate. Let p be any point in M, and let O be an open set containing p within which parallel transport of spacelike vectors is path independent. We can certainly find three smooth, linearly independent, spacelike fields $\overset{1}{\sigma}{}^a, \overset{2}{\sigma}{}^a, \overset{3}{\sigma}{}^a$ on O that are constant ($\nabla_n \overset{i}{\sigma}{}^a = \mathbf{0}$). (Start with three linearly independent, spacelike vectors at p and parallel transport them, along any curve, to other points in O.) For each one, we have

$$R^a{}_{rcd}\,\overset{i}{\sigma}{}^r = -2\,\nabla_{[c}\,\nabla_{d]}\,\overset{i}{\sigma}{}^a = \mathbf{0}$$

at p. Since $\overset{1}{\sigma}{}^a, \overset{2}{\sigma}{}^a, \overset{3}{\sigma}{}^a$ span the space of spacelike vectors at p, it follows that $R^a{}_{rcd}\,\sigma^r = \mathbf{0}$ for *all* spacelike vectors σ^a there. So $R^a{}_{rcd}\,h^{rb}\,\alpha_b = \mathbf{0}$ for all co-vectors α_b at p; i.e., $R^{ab}{}_{cd} = R^a{}_{rcd}\,h^{rb} = \mathbf{0}$ at p.

(*Only if*) There are various ways to see this. But it is, perhaps, easiest to make use of what we have established and reduce this to a claim about a (different) flat derivative operator. If $R^{ab}{}_{cd} = \mathbf{0}$, then, by proposition 4.2.4, given any point p in M, there is an open set O containing p and a future-directed unit timelike vector field η^a on O such that $\nabla^a \eta^b = \mathbf{0}$. Now recall our proof of the Trautman Recovery Theorem (proposition 4.2.5). Let ϕ^a be the acceleration field of η^a, and let $\overset{f}{\nabla}$ be the derivative operator on O defined by $\overset{f}{\nabla} = (\nabla, C^a{}_{bc})$, where $C^a{}_{bc} = t_b\, t_c\, \phi^a$. We established in our proof of the Recovery Theorem that $\overset{f}{\nabla}$ is flat. (And for *this* part of the proof, we did not need the additional assumptions that appear in our formulation of the theorem, namely $R^a{}_b{}^c{}_d = R^c{}_d{}^a{}_b$ and $R_{bc} = 4\pi\, \rho\, t_b t_c$. We needed only $R^{ab}{}_{cd} = \mathbf{0}$.) So parallel transport of *all* vectors within O relative to $\overset{f}{\nabla}$ is, at least locally, path independent. To complete the proof, it suffices to note that $\overset{f}{\nabla}$ and ∇ agree in their action on *spacelike* vector fields (and so agree in their determinations of parallel transport for such fields on arbitrary curves). This is clear. For let σ^a be a smooth spacelike vector field (defined on some open subset of O). Then

$$\overset{f}{\nabla}_a\, \sigma^b = \nabla_a\, \sigma^b - C^b{}_{an}\sigma^n \; = \; \nabla_a\, \sigma^b,$$

as required, since $C^b{}_{an}\sigma^n = (t_a\, t_n\, \phi^b)\, \sigma^n = \mathbf{0}$. □

The proposition also provides a physical interpretation of the third curvature condition (4.3.3) in terms of the precession, or non-precession, of gyroscopes. Suppose we hold two spinning gyroscopes at a point, side by side, with their axes co-aligned. And suppose we then transport them (without constraint) to another point along different routes. We cannot expect *a priori* that, on arrival, their axes will still be co-aligned. There is no reason why "gyroscope propagation" must be path independent. Indeed, we see from the proposition that it will be path independent (at least locally) iff equation (4.3.3) holds.

Now we consider the geometrized version of Poission's equation $R_{ab} = 4\pi\, \rho\, t_a\, t_b$. The interpretation we offered for Einstein's equation in terms of geodesic deviation has a close counterpart here. Almost everything carries over intact from section 2.7. Let ξ^a be a "geodesic reference frame" defined on some open set in M—i.e., a smooth, future-directed, unit timelike vector field whose associated integral curves are geodesics. Further, let λ^a be a smooth, spacelike vector field along (the image of) one of the integral curves γ satisfying

$\mathcal{L}_{\xi}\lambda^a = \mathbf{0}$. (Once again, we can think of λ^a as a connecting field that joins the image of γ to the image of an "infinitesimally close" neighboring integral curve.) The equation of geodesic deviation

$$\xi^n \nabla_n (\xi^m \nabla_m \lambda^a) = R^a{}_{bcd}\, \xi^b \, \lambda^c \, \xi^d \tag{4.3.4}$$

carries over without alteration, as does the expression we derived for the "average radial acceleration" of ξ^a,

$$ARA = -\frac{1}{3} R_{bd} \xi^b \xi^d. \tag{4.3.5}$$

The latter, in turn, leads to the following proposition (which is proved in almost exactly the same way as proposition 2.7.2).

PROPOSITION 4.3.2. *Let (M, t_a, h^{ab}, ∇) be a classical spacetime, let ρ be a smooth scalar field on M, and let p be a point in M. Then Poisson's equation $R_{ab} = 4\pi\ \rho\ t_a t_b$ holds at p iff for all geodesic reference frames ξ^a (defined on some open set containing p) the average radial acceleration of ξ^a at p is given by $ARA = -\frac{4}{3}\pi\ \rho$.*

We can make the result look even more like proposition 2.7.2 if we use our alternate formulation of Poisson's equation. In that case, the conclusion is this: Poisson's equation $R_{ab} = 8\pi\ (\hat{T}_{ab} - \frac{1}{2}\ t_{ab}\ \hat{T})$ holds at p iff for all geodesic reference frames ξ^a (defined on some open set containing p) the average radial acceleration of ξ^a at p is given by $ARA = -\frac{8\pi}{3}\ \pi\ (\hat{T}_{ab} - \frac{1}{2}\ t_{ab}\ \hat{T})$ $\xi^a \xi^b$.

Finally, we turn to the geometric interpretation of the second condition in our list, $R^a{}_b{}^c{}_d = R^c{}_d{}^a{}_b$. This will require a good deal more work than the others. We show that *it holds in a classical spacetime iff the latter admits, at least locally, a smooth, unit timelike field ξ^a that is geodesic ($\xi^n \nabla_n \xi^a = \mathbf{0}$) and twist-free ($\nabla^{[a}\xi^{b]} = \mathbf{0}$).* This equivalence is proved in Dombrowski and Horneffer [13] and Künzle [34]. Our argument, at least for the "only if" half (proposition 4.3.7), is a bit different from theirs. We begin with the "if" half of the assertion, which is straightforward.

PROPOSITION 4.3.3. *Let (M, t_a, h^{ab}, ∇) be a classical spacetime, and let p be any point in M. Assume there is a smooth, future-directed, unit timelike field ξ^a, defined on some open set containing p, that is geodesic and twist-free. Then $R^a{}_b{}^c{}_d = R^c{}_d{}^a{}_b$ at p.*

Proof. It suffices for us to show that, at p, contracting $(R^{a}{}_{b}{}^{c}{}_{d} - R^{c}{}_{d}{}^{a}{}_{b})$ with (i) $\xi^b \xi^d$, (ii) $h^{br} h^{ds}$, and (iii) $\xi^b h^{ds}$ (or $h^{br} \xi^d$) yields $\mathbf{0}$. The claim in case (ii) comes free, without any assumptions about ξ^a, since $R^{arcs} = R^{csar}$ holds in any classical spacetime. (Recall equation (4.1.30).)

For case (iii), we need only the fact that ξ^a is twist-free. We must show that $R^{a}{}_{b}{}^{cs}\xi^b = R^{csa}{}_{b}\xi^b$. To do so, we recast the right side using symmetries of the curvature field, namely $R^{a}{}_{[bcd]} = \mathbf{0}$, $R^{a}{}_{b(cd)} = \mathbf{0}$, and $R^{(ab)}{}_{cd} = \mathbf{0}$. (The first two hold for any derivative operator. The third follows from the compatibilty of ∇ with h^{ab}. Recall equation (4.1.26). We use the symmetries with some indices in raised position. So, for example, since $R^{a}{}_{bcd} + R^{a}{}_{dbc} + R^{a}{}_{cdb} = \mathbf{0}$, it follows that $R^{a}{}_{b}{}^{cd} + R^{ad}{}_{b}{}^{c} + R^{acd}{}_{b} = \mathbf{0}$.)

$$\begin{aligned}
R^{csa}{}_{b}\,\xi^b &= -R^{sca}{}_{b}\,\xi^b = R^{s}{}_{b}{}^{ca}\,\xi^b + R^{sa}{}_{b}{}^{c}\,\xi^b = R^{s}{}_{b}{}^{ca}\,\xi^b - R^{as}{}_{b}{}^{c}\,\xi^b \\
&= R^{s}{}_{b}{}^{ca}\,\xi^b + (R^{acs}{}_{b} + R^{a}{}_{b}{}^{cs})\,\xi^b \;=\; R^{s}{}_{b}{}^{ca}\,\xi^b + R^{a}{}_{b}{}^{cs}\,\xi^b - R^{cas}{}_{b}\,\xi^b \\
&= R^{s}{}_{b}{}^{ca}\,\xi^b + R^{a}{}_{b}{}^{cs}\,\xi^b + (R^{c}{}_{b}{}^{as} + R^{cs}{}_{b}{}^{a})\,\xi^b.
\end{aligned}$$

Hence,

$$\begin{aligned}
R^{csa}{}_{b}\,\xi^b &= \frac{1}{2}(R^{s}{}_{b}{}^{ca}\,\xi^b + R^{a}{}_{b}{}^{cs}\,\xi^b + R^{c}{}_{b}{}^{as})\,\xi^b \\
&= -(\nabla^{[c}\,\nabla^{a]}\xi^s + \nabla^{[c}\,\nabla^{s]}\xi^a + \nabla^{[a}\,\nabla^{s]}\xi^c).
\end{aligned}$$

If we now expand the final sum and use the fact (for the first time) that $\nabla^a\xi^b = \nabla^b\xi^a$, we arrive at

$$R^{csa}{}_{b}\,\xi^b = -(\nabla^c\,\nabla^s\xi^a - \nabla^s\,\nabla^c\xi^a) = R^{a}{}_{b}{}^{cs}\xi^b.$$

Finally, we consider case (i). Here we need both the fact that ξ^a is twist-free and that it is geodesic. We must show that $R^{a}{}_{b}{}^{c}{}_{d}\xi^b\xi^d = R^{c}{}_{d}{}^{a}{}_{b}\,\xi^b\xi^d$—i.e., that $R^{a}{}_{b}{}^{c}{}_{d}\,\xi^b\xi^d$ is symmetric in a and c. But

$$\begin{aligned}
R^{a}{}_{b}{}^{c}{}_{d}\,\xi^b\xi^d &= -\xi^d\,(\nabla^c\,\nabla_d\,\xi^a - \nabla_d\,\nabla^c\,\xi^a) \\
&= -\nabla^c\,(\xi^d\,\nabla_d\,\xi^a) + (\nabla^c\,\xi^d)(\nabla_d\,\xi^a) + \xi^d\,\nabla_d\,\nabla^c\,\xi^a.
\end{aligned}$$

The first term in the final sum vanishes since ξ^a is geodesic. The third is symmetric in a and c since ξ^a is twist-free. The second is symmetric in a and c for the same reason, since $(\nabla^c\,\xi^d)(\nabla_d\,\xi^a) = (\nabla^d\,\xi^c)(\nabla_d\,\xi^a) = h^{dn}(\nabla_n\,\xi^c)(\nabla_d\,\xi^a) = (\nabla_n\,\xi^c)(\nabla^n\,\xi^a) = (\nabla_n\,\xi^c)(\nabla^a\,\xi^n)$. □

Next we consider a particular class of derivative operators that satisfy the curvature condition (4.3.2) (in addition to being compatible with the background metrics t_a and h^{ab}).

PROPOSITION 4.3.4. *Let (M, t_a, h^{ab}, ∇) be a classical spacetime, and let ξ^a be any smooth, unit, future-directed timelike vector field on M. Then there exists a unique derivative operator $\widetilde{\nabla}$ on M such that (i) $\widetilde{\nabla}$ is compatible with t_a and h^{ab} and (ii) ξ^a is geodesic and twist-free with respect to $\widetilde{\nabla}$.*

When conditions (i) and (ii) obtain, we call $\widetilde{\nabla}$ the *special derivative operator determined by ξ^a*. It follows immediately from the preceding proposition that all special derivative operators (determined by some field) satisfy equation (4.3.2). We shall soon verify (in proposition 4.3.7) that they are the only derivative operators that do so.

Proof. Let $\hat{h}_{ab}$ be the projection field associated with ξ^a, let $\kappa_{ab} = \hat{h}_{n[b}\,\nabla_{a]}\xi^n$, let $C^a{}_{bc} = 2\,t_{(b}\,\kappa_{c)}{}^a$, and, finally, let $\widetilde{\nabla} = (\nabla, C^a{}_{bc})$. Then, by proposition 4.1.3 , $\widetilde{\nabla}$ is compatible with t_a and h^{ab}. Moreover, we claim, ξ^a is geodesic and twist-free with respect to $\widetilde{\nabla}$. To see this, note first that since $\hat{h}_{ab}\,h^{bc} = \delta_a{}^b - t_a\,\xi^b$, we have

$$\kappa_a{}^b = h^{br}\kappa_{ar} = \frac{1}{2}\,h^{br}(\hat{h}_{nr}\nabla_a\,\xi^n - \hat{h}_{na}\nabla_r\,\xi^n) = \frac{1}{2}(\nabla_a\,\xi^b - \hat{h}_{na}\nabla^b\,\xi^n)$$

and, therefore,

$$\kappa^{ab} = \frac{1}{2}\,(\nabla^a\,\xi^b - \nabla^b\,\xi^a) = \nabla^{[a}\,\xi^{b]},$$

$$\kappa_a{}^b\,\xi^a = \frac{1}{2}\,\xi^a\nabla_a\,\xi^b.$$

Now

$$\widetilde{\nabla}_a\,\xi^b = \nabla_a\,\xi^b - C^b{}_{ar}\,\xi^r = \nabla_a\,\xi^b - (t_a\,\kappa_r{}^b + t_r\,\kappa_a{}^b)\xi^r.$$

Hence, since κ_{ab} is anti-symmetric, we have

$$\widetilde{\nabla}^{[a}\,\xi^{b]} = \nabla^{[a}\,\xi^{b]} - \kappa^{ab} = \mathbf{0},$$

$$\xi^a\widetilde{\nabla}_a\,\xi^b = \xi^a\nabla_a\,\xi^b - 2\,\kappa_a{}^b\xi^a = \mathbf{0},$$

as claimed. So we have established existence.

For uniqueness, suppose $\widetilde{\widetilde{\nabla}} = (\widetilde{\nabla}, \widetilde{C}^a_{bc})$ is a second derivative operator on M that satisfies conditions (i) and (ii). We know from proposition 4.1.3 (since both $\widetilde{\nabla}$ and $\widetilde{\widetilde{\nabla}}$ are compatible with t_a and h^{ab}) that there is a smooth, anti-symmetric field κ_{ab} such that $\widetilde{C}^a_{\ bc} = 2\, h^{an}\, t_{(b}\, \kappa_{c)n}$. We show that $\kappa_{cn} = \mathbf{0}$. Now $\kappa_{cn}\xi^c\xi^n = 0$, since κ_{cn} is anti-symmetric. So it will suffice for us to show that $\kappa_{cn}\, \xi^c h^{ns} = \mathbf{0}$ and $\kappa_{cn}\, h^{cr} h^{ns} = \mathbf{0}$. Since ξ^a is geodesic with respect to both $\widetilde{\nabla}$ and $\widetilde{\widetilde{\nabla}}$, we have, first,

$$\mathbf{0} = \xi^c \widetilde{\widetilde{\nabla}}_c\, \xi^a = \xi^c (\widetilde{\nabla}_c\, \xi^a - \widetilde{C}^a_{\ bc}\, \xi^b) = -\widetilde{C}^a_{\ bc}\, \xi^b\, \xi^c = -2\, \kappa_{cn}\, \xi^c\, h^{an}.$$

Next,

$$\widetilde{\widetilde{\nabla}}^r\, \xi^s = h^{rc}\, \widetilde{\widetilde{\nabla}}_c\, \xi^s = h^{rc}\, (\widetilde{\nabla}_c\, \xi^s - \widetilde{C}^s_{\ bc}\, \xi^b) = \widetilde{\nabla}^r \xi^s - \kappa^{rs}.$$

So, since $\widetilde{\widetilde{\nabla}}^{[r}\, \xi^{s]} = \mathbf{0} = \widetilde{\nabla}^{[r}\, \xi^{s]}$ (and since κ_{cn} is anti-symmetric), we also have $h^{rc}\, h^{sn} \kappa_{cn} = \mathbf{0}$. □

Now we extend proposition 4.1.3 and consider the most general form for a connecting field $C^a_{\ bc}$ that links two derivative operators on M that are compatible with t_a and h^{ab} *and also satisfy equation (4.3.2)*.

PROPOSITION 4.3.5. *Let (M, t_a, h^{ab}, ∇) be a classical spacetime such that $R^{a\ c}_{\ b\ d} = R^{c\ a}_{\ d\ b}$. Let $\nabla' = (\nabla,\, C^a_{\ bc})$ be a second derivative operator on M where $C^a_{\ bc} = 2\, h^{an}\, t_{(b}\, \kappa_{c)n}$ and κ_{ab} is a smooth, anti-symmetric field on M. (We know this is the general form for a derivative operator on M that is compatible with t_a and h^{ab}.) Then $R'^{a\ c}_{\ b\ d} = R'^{c\ a}_{\ d\ b}$ iff κ_{ab} is closed; i.e., $\nabla_{[n}\, \kappa_{ab]} = \mathbf{0}$. (Here, of course, $R'^a_{\ bcd}$ is the Riemann curvature field associated with ∇'.)*

Proof. We know (from problem 1.8.1) that

$$R'^a_{\ bcd} = R^a_{\ bcd} + 2\, \nabla_{[c}\, C^a_{\ d]b} + 2\, C^n_{\ b[c}\, C^a_{\ d]n}.$$

In the present case, where $C^a_{\ bc} = 2t_{(b}\, \kappa_{c)}{}^a$, we have $C^n_{\ bc}\, C^a_{\ dn} = t_d\, t_b\, \kappa_c{}^n \kappa_n{}^a + t_d\, t_c\, \kappa_b{}^n \kappa_n{}^a$ and, hence,

$$2\, C^n_{\ b[c}\, C^a_{\ d]n} = 2\, t_b\, t_{[d}\, \kappa_{c]}{}^n \kappa_n{}^a.$$

Similarly, $\nabla_c\, C^a_{\ db} = t_d\, \nabla_c\, \kappa_b{}^a + t_b\, \nabla_c\, \kappa_d{}^a$ and, hence,

$$2\, \nabla_{[c}\, C^a_{\ d]b} = 2\, t_{[d}\, \nabla_{c]}\, \kappa_b{}^a + 2\, t_b\, \nabla_{[c}\, \kappa_{d]}{}^a.$$

If we now raise the index c in all these terms, we arrive at

$$R'^{a\ c}_{\ b\ d} = R^{a\ c}_{\ b\ d} + (t_d\, \nabla^c\, \kappa_b{}^a + t_b\, \nabla^c\, \kappa_d{}^a - t_b\, \nabla_d\, \kappa^{ca}) + t_b\, t_d\, \kappa^{cn}\, \kappa_n{}^a$$

and, therefore, also

$$R'^{c\ a}_{\ d\ b} = R^{c\ a}_{\ d\ b} + (t_b\, \nabla^a\, \kappa_d{}^c + t_d\, \nabla^a\, \kappa_b{}^c - t_d\, \nabla_b\, \kappa^{ac}) + t_d\, t_b\, \kappa^{an}\, \kappa_n{}^c.$$

We are assuming that $R^a{}_b{}^c{}_d = R^c{}_d{}^a{}_b$. And, by the anti-symmetry of κ_{ab}, $\kappa^{cn}\kappa_n{}^a = \kappa^{an}\kappa_n{}^c$. So we see that $R'^a{}_b{}^c{}_d = R'^c{}_d{}^a{}_b$ iff the respective middle terms (in parentheses) in the two lines are equal—i.e., iff

$$t_b\,(-\nabla^a \kappa_d{}^c + \nabla^c \kappa_d{}^a - \nabla_d \kappa^{ca}) = t_d\,(-\nabla^c \kappa_b{}^a + \nabla^a \kappa_b{}^c - \nabla_b \kappa^{ac}). \tag{4.3.6}$$

In turn, this equation holds iff

$$-\nabla^a \kappa_d{}^c + \nabla^c \kappa_d{}^a - \nabla_d \kappa^{ca} = \mathbf{0}. \tag{4.3.7}$$

(Why? If equation (4.3.7) holds, then both sides of equation (4.3.6) vanish. Conversely, assume equation (4.3.6) holds, let $\psi^{ac}{}_d = (-\nabla^a \kappa_d{}^c + \nabla^c \kappa_d{}^a - \nabla_d \kappa^{ca})$, and let ξ^a be any unit timelike vector field. Contracting both sides of equation (4.3.6) with $\xi^b h^{dr}$ yields $h^{dr}\psi^{ac}{}_d = \mathbf{0}$. Contracting both sides with $\xi^b\xi^d$ yields $\xi^d\psi^{ac}{}_d = \mathbf{0}$. So it must be the case that $\psi^{ac}{}_d = \mathbf{0}$.) We can express equation (4.3.7) in the form

$$h^{ar}\,h^{cs}\,\nabla_{[r}\,\kappa_{sd]} = \mathbf{0}. \tag{4.3.8}$$

But this condition is equivalent to

$$\nabla_{[r}\,\kappa_{sd]} = \mathbf{0}. \tag{4.3.9}$$

For if equation (4.3.8) holds, then, by the anti-symmetry of $\nabla_{[r}\,\kappa_{sd]}$, contraction with $\xi^r\xi^s\xi^d$, $\xi^r\xi^s h^{dn}$, $\xi^r h^{sc} h^{dn}$, and $h^{ra}h^{sc}h^{dn}$ all yield $\mathbf{0}$. Thus, as claimed, $R'^a{}_b{}^c{}_d = R'^c{}_d{}^a{}_b$ iff κ_{ab} is closed. □

Now we make precise a sense in which condition $R^a{}_b{}^c{}_d = R^c{}_d{}^a{}_b$ rules out the possibility of "spontaneous rotation."

PROPOSITION 4.3.6. *Let (M, t_a, h^{ab}, ∇) be a classical spacetime such that $R^a{}_b{}^c{}_d = R^c{}_d{}^a{}_b$. Let ξ^a be a smooth, future-directed, unit timelike field on M that is geodesic (with respect to ∇). Then its associated rotation and expansion fields satisfy*

$$\xi^n\nabla_n\,\omega^{ab} = 2\,\omega^{n[a}\,\theta_n{}^{b]}. \tag{4.3.10}$$

Hence, given any integral curve $\gamma : I \to M$ of ξ^a, if ξ^a is twist-free ($\omega^{ab} = \mathbf{0}$) at one point on $\gamma[I]$, it is twist-free at all points on it. (Or, more colloquially, if it is twist-free at one time, it is twist-free at all times.)

Proof. We know that $\nabla^{[a}\,\xi^{b]} = \omega^{ab}$. (This follows immediately from equation (4.1.43).) Hence,

$$\begin{aligned} 2\,\xi^n\nabla_n\,\omega^{ab} &= \xi^n h^{am}\nabla_n\nabla_m\,\xi^b - \xi^n h^{bm}\,\nabla_n\nabla_m\,\xi^a \\ &= \xi^n h^{am}\left(\nabla_m\nabla_n\,\xi^b - R^b_{\ snm}\xi^s\right) - \xi^n h^{bm}\left(\nabla_m\nabla_n\,\xi^a - R^a_{\ snm}\xi^s\right). \end{aligned}$$

Since ξ^a is geodesic,

$$\xi^n h^{am} \nabla_m \nabla_n \xi^b = h^{am} \left[\nabla_m (\xi^n \nabla_n \xi^b) - (\nabla_m \xi^n)(\nabla_n \xi^b) \right] = -(\nabla^a \xi^n)(\nabla_n \xi^b)$$

and, similarly, $-\xi^n h^{bm} \nabla_m \nabla_n \xi^a = (\nabla^b \xi^n)(\nabla_n \xi^a)$. Furthermore, since $R^b{}_{s(nm)} = \mathbf{0}$,

$$\begin{aligned} -\xi^n h^{am} R^b{}_{snm} \xi^s + \xi^n h^{bm} R^a{}_{snm} \xi^s &= \left(R^b{}_s{}^a{}_n - R^a{}_s{}^b{}_n \right) \xi^n \xi^s \\ &= \left(R^b{}_s{}^a{}_n - R^a{}_n{}^b{}_s \right) \xi^n \xi^s = \mathbf{0}. \end{aligned}$$

So,

$$\begin{aligned} 2\, \xi^n \nabla_n \, \omega^{ab} &= -(\nabla^a \xi^n)(\nabla_n \xi^b) + (\nabla^b \xi^n)(\nabla_n \xi^a) \\ &= -[2\, \nabla^{[a} \xi^{n]} + \nabla^n \xi^a](\nabla_n \xi^b) + [2\, \nabla^{[b} \xi^{n]} + \nabla^n \xi^b](\nabla_n \xi^a) \\ &= -2\, (\nabla^{[a} \xi^{n]})(\nabla_n \xi^b) + 2\, (\nabla^{[b} \xi^{n]})(\nabla_n \xi^a) \\ &= -2\, \omega^{an} (\theta_n{}^b + \omega_n{}^b) + 2\, \omega^{bn} (\theta_n{}^a + \omega_n{}^a) \\ &= -2\, \omega^{an} \theta_n{}^b + 2\, \omega^{bn} \theta_n{}^a = 4\, \omega^{n[a}\, \theta_n{}^{b]}. \end{aligned}$$

Now let $\gamma : I \to M$ be an integral curve of ξ^a, and suppose $\omega_{ab} = \mathbf{0}$ at some point $\gamma(s_0)$. It follows from the basic uniqueness theorem for systems of first-order ordinary differential equations that equation (4.3.10) will be satisfied at all points on $\gamma[I]$ iff $\omega_{ab} = \mathbf{0}$ vanishes everywhere on that set. (To see this in detail, let $\overset{1}{\sigma}_a, \ldots, \overset{4}{\sigma}_a$ be a basis for the co-tangent space at some point on $\gamma[I]$ that is orthonormal with respect to h^{ab} (in our extended sense of "orthonormal"). We can extend the vectors (by parallel transport) to fields $\overset{i}{\sigma}_a$ on $\gamma[I]$—we use the same notation—that satisfy $\xi^n \nabla_n \overset{i}{\sigma}_a = \mathbf{0}$. Since ∇ is compatible with h^{ab}, the generated fields will be orthonormal everywhere. Now consider the scalar (coefficient) fields $\overset{ij}{\omega} = \omega^{ab}\, \overset{i}{\sigma}_a\, \overset{j}{\sigma}_b$. Equation (4.3.10) can then be expressed as a system of first-order differential equations,

$$\frac{d\, \overset{ij}{\omega}}{dt} = f_{ij}(\overset{11}{\omega}, \overset{12}{\omega}, \ldots, \overset{44}{\omega}),$$

to which the uniqueness theorem is applicable.) □

We have claimed that condition (4.3.2) holds iff, at least locally (in a neighborhood of every point), there exists a unit timelike vector field that is geodesic and twist-free. We have proved the "if" half of the claim in proposition 4.3.3. Now, finally, we turn to the converse.

PROPOSITION 4.3.7. *Let (M, t_a, h^{ab}, ∇) be a classical spacetime such that $R^a{}_b{}^c{}_d = R^c{}_d{}^a{}_b$. Then, given any point p in M, there is a smooth, future-directed, unit timelike vector field, defined on some open set containing p, that is geodesic and twist-free (with respect to ∇).*

Proof. Let p be given. Our proof will proceed in two steps and make reference to three smooth, future-directed, unit timelike fields: ξ^a, ξ'^a, and ξ''^a. (They will be defined on open sets O, O', and O'', respectively, where $p \in O'' \subseteq O' \subseteq O$.) ξ^a will be an arbitrary field. ξ'^a will be twist-free. ξ''^a will be geodesic *and* twist-free. (It is the existence of the third that we need to establish.)

(Step 1) Let ξ^a be a smooth, future-directed, unit timelike field defined on some open set O containing p. By proposition 4.3.4, there is a derivative operator $\widetilde{\nabla}$ on O such that $\widetilde{\nabla}$ is compatible with t_a and h^{ab}, and such that ξ^a is geodesic and twist-free with respect to $\widetilde{\nabla}$. Let $C^a_{\ bc}$ be the connecting field (on O) such that $\nabla = (\widetilde{\nabla}, C^a_{\ bc})$. Now, by proposition 4.3.3, $\widetilde{R}^a{}_b{}^c{}_d = \widetilde{R}^c{}_d{}^a{}_b$. So, since both ∇ and $\widetilde{\nabla}$ satisfy equation (4.3.2), it follows by proposition 4.3.5 that there is a smooth, closed, anti-symmetric field κ_{ab} on O such that $C^a_{\ bc} = 2\, h^{an}\, t_{(b}\, \kappa_{c)n}$. Since κ_{ab} is closed, we know by proposition 1.8.3 that it is, at least locally, exact. So there is an open subset O' of O containing p, and a smooth field κ_a on O' such that $\kappa_{ab} = \widetilde{\nabla}_{[a}\, \kappa_{b]}$. Now consider the field $\xi'^a = \xi^a + \kappa^a$ on O'. It is a smooth, future-directed, unit timelike field. (It is of unit timelike length since $t_a \kappa^a = t_a h^{ab} \kappa_b = \mathbf{0}$.) We claim that it is twist-free with respect to ∇. We have

$$\nabla_n\, \xi'^a = \nabla_n\, (\xi^a + \kappa^a) \ = \widetilde{\nabla}_n\, (\xi^a + \kappa^a) - C^a_{\ mn}(\xi^m + \kappa^m).$$

But,

$$C^a_{\ mn}(\xi^m + \kappa^m) = (t_m\, \kappa_n{}^a + t_n\, \kappa_m{}^a)(\xi^m + \kappa^m) = \kappa_n{}^a + t_n\, \kappa_m{}^a(\xi^m + \kappa^m).$$

Hence,

$$\nabla^n\, \xi'^a = \widetilde{\nabla}^n \xi^a + \widetilde{\nabla}^n \kappa^a - \kappa^{na}$$

and, therefore (since ξ^a is twist-free with respect to $\widetilde{\nabla}$, and $\kappa_{ab} = \widetilde{\nabla}_{[a}\, \kappa_{b]}$),

$$\nabla^{[n}\, \xi'^{a]} = \widetilde{\nabla}^{[n}\, \xi^{a]} + \widetilde{\nabla}^{[n} \kappa^{a]} - \kappa^{na} = \mathbf{0}.$$

(Step 2) So far, we established the existence of a field ξ'^a, defined on some open set O' containing p, that is twist-free with respect to ∇. Now let S be a spacelike hypersurface within O' that contains p. Then we can find a smooth, future-directed, unit timelike vector field ξ''^a, defined on some open subset O'' of O' containing p, that is geodesic with respect to ∇ and agrees with ξ'^a on S. (We first restrict ξ'^a to S, and then use each vector in this restricted field $\xi'^a{}_{|S}$

to generate a geodesic. This gives us a congruence of curves. We take ξ''^a to be its tangent field.) Now, since ξ'^a is twist-free on S, so is ξ''^a. (The difference field $(\xi''^a - \xi'^a)$ vanishes on S. So, at any point of S, its directional derivative in any spacelike direction vanishes as well; i.e., $h^{an}\, \nabla_n (\xi''^b - \xi'^b) = \mathbf{0}$. Hence, on S, $\nabla^{[a}\, \xi''^{b]} = \nabla^{[a}\, \xi'^{b]} = \mathbf{0}$.) But now, since the geodesic field ξ''^a is twist-free on S, it follows from proposition 4.3.6 that is it everywhere twist-free. So we are done. □

4.4. A Solution to an Old Problem about Newtonian Cosmology

The geometrized formulation of Newtonian theory provides a satisfying solution to an old problem about Newtonian cosmology. We present it in this section.[3]

At issue is whether Newtonian gravitation theory provides a sensible prescription for what the gravitational field should be like in a hypothetically infinite, homogeneous universe. Let us first think about this in terms of a traditional, non-geometrized, three-dimensional formulation of the theory. Let $(\mathbb{R}^3, g_{ab})$ be three-dimensional Euclidean space. We take it to represent physical space at a given time. Further, let ρ and ϕ be two smooth functions on $\mathbb{R}^3$ that, respectively, give the mass density and the gravitational potential at different points of space.[4] We assume that they satisfy Poisson's equation $\nabla_a \nabla^a \phi = 4\,\pi\,\rho$. (Here ∇ is the derivative operator on $\mathbb{R}^3$ compatible with g_{ab}.)

Suppose that we are dealing with a homogeneous distribution of matter; i.e., suppose that ρ is constant. Then, presumably, the gravitational field associated with this matter distribution should be homogeneous as well. (Why should it be different here from the way it is there?) The gravitational force felt by a particle of unit mass at any point is given by $-\nabla^a \phi$. So, it would seem, the natural way to capture the homogeneity condition on the gravitational field is to require that the field $\nabla^a \phi$ be constant—i.e., require that $\nabla_b \nabla^a \phi = \mathbf{0}$. But now we have a problem. If $\nabla_b \nabla^a \phi = \mathbf{0}$, and if Poisson's equation is satisfied, then $4\,\pi\,\rho = \nabla_a \nabla^a \phi = 0$. So we *cannot* satisfy the homogeneity condition except in the degenerate case where the mass density ρ is everywhere 0. Here is another version of the problem. It directs attention to a particular class of solutions to

3. For further discussion of the problem and its history, see Norton [43, 44, 45] and Malament [40].

4. Caution: we have previously understood ρ and ϕ to be objects defined on a four-dimensional spacetime manifold, and shall soon do so again. But now, temporarily, we take them to be defined on a three-dimensional manifold (representing space a given time) instead.

Poisson's equation $\nabla_a \nabla^a \phi = 4\,\pi\,\rho$ that *do* exist in the case where ρ is constant ($\nabla_a\,\rho = \mathbf{0}$). Let o be any point in $\mathbb{R}^3$, and let χ^a be the position field determined relative to o. So $\nabla_a \chi^b = \delta_a{}^b$, and $\chi^a = \mathbf{0}$ at o. Let us say that a smooth field ϕ on $\mathbb{R}^3$ is a *canonical solution centered at* o if

$$\nabla^a \phi = \frac{4}{3}\,\pi\,\rho\,\chi^a \tag{4.4.1}$$

i.e., if $\nabla^a \phi$ is a spherically symmetric, outward-directed, radial vector field, centered at o, whose assignment to any point p has length $\frac{4}{3}\pi\,\rho\,r$, where r is the Euclidean distance between o and p.

Note that if this condition holds, then (since $\nabla_a\,\rho = \mathbf{0}$),

$$\nabla_a\,\nabla^a \phi = \frac{4}{3}\,\pi\,\rho\,(\nabla_a \chi^a) = 4\,\pi\,\rho.$$

So canonical solutions centered at o (if they exist) *are* solutions. And they certainly *do* exist; e.g.,

$$\phi = \frac{2}{3}\,\pi\,\rho\,(\chi_n \chi^n).$$

Not all solutions to Poisson's equation (in the present case where ρ is constant) are canonical solutions centered at some point or other. (If ϕ is a solution, then so is $(\phi + \psi)$, where ψ is *any* smooth field that satisfies $\nabla_a \nabla^a \psi = \mathbf{0}$.) But canonical solutions are the only solutions that satisfy a certain natural constraint, and for this reason they are the only ones that are usually considered in discussions of Newtonian cosmology. The constraint arises if we consider not just the distribution of cosmic matter at a given time, but also its motion under the influence of that potential. It turns out that if we require that the motion be isotropic in a certain natural sense, then all solutions are ruled out *except* those that are canonical for some center point o. (We shall, in effect, prove this. See proposition 4.4.3.) In any case, our problem re-emerges when we direct our attention to the class of canonical solutions. The gravitational field associated with any one of them is a radial field that vanishes at a unique center point. Why, one wants to ask, should there be any such distinguished point in a homogeneous universe? And why should it be one point rather than another; i.e., why should any one canonical solution be a better choice for the gravitational field in a homogeneous universe than another?

That is the problem. A solution, or dissolution, can be found in the recognition that the gravitational field (in standard formulations of Newtonian theory) is a kind of "gauge field"—i.e., a field that is, in general, systematically underdetermined by all experimental evidence. Despite appearances, canonical solutions centered at different points really *are* empirically equivalent. No experimental test could ever distinguish one from another (or distinguish the center point of any one of them). Canonical solutions centered at different points should be viewed as but alternative mathematical representations of the same underlying state of gravitational affairs—a state that *is* perfectly homogenous in the appropriate sense.

One can certainly argue for these claims directly, without reference to geometrized formulations of Newtonian theory.[5] (See, for example, Heckmann and Schücking [31] and Norton [44].) But some insight is achieved if we *do* think about this old problem in Newtonian cosmology using the ideas developed in section 4.2. *We can develop an account of Friedmann-like cosmological models within geometrized Newtonian gravitation theory, and then recover the class of canonical solutions (centered at different points) as but alternative "degeometrizations" of the initial curved derivative operator*—exactly as described at the end of that section. The choice between different canonical solutions emerges as a choice between different ways to decompose into "gravitational" and "inertial" components the net force experienced by a point particle. Nothing more.

Before proceeding, we give an alternative characterization of the class of canonical solutions—at least in the case of interest where $\rho > 0$—that will be convenient later.

PROPOSITION 4.4.1. *Let $(\mathbb{R}^3, g_{ab})$ be three-dimensional Euclidean space, and let ρ be a constant field on $\mathbb{R}^3$ with $\rho > 0$. Then for all smooth fields ϕ on $\mathbb{R}^3$, the following conditions are equivalent.*

(1) *ϕ is a canonical solution (to Poisson's equation $\nabla^a \nabla_a \phi = 4\pi\rho$) centered at some point in $\mathbb{R}^3$.*
(2) $\nabla^a \nabla^b \phi = \frac{4}{3}\pi\rho\, g^{ab}$.

Proof. One direction is immediate. If ϕ is a canonical solution centered at point o (and if χ^a is the position field relative to o),

$$\nabla^a \nabla^b \phi = \nabla^a \left(\frac{4}{3}\pi\rho\,\chi^b\right) = \frac{4}{3}\pi\rho\,(\nabla^a \chi^b) = \frac{4}{3}\pi\rho\, g^{ab}.$$

Conversely, suppose ϕ satisfies condition (2). Let ϕ' be a canonical solution centered at some point o', let χ'^a be the position field relative to o', and let κ^b be the difference field

5. The important point is that if ϕ and ϕ' are canonical solutions, based at o and o', respectively, the difference field $(\nabla^a\phi - \nabla^a\phi')$ is constant, and constant gravitational fields are undetectable. Only field differences can be detected. The difference field is constant since

$$\nabla_a(\nabla^b\phi - \nabla^b\phi') = \nabla_a\left(\frac{4}{3}\pi\rho\,\chi^b - \frac{4}{3}\pi\rho\,\chi'^b\right) = \frac{4}{3}\pi\rho\,(\delta_a{}^b - \delta_a{}^b) = \mathbf{0}.$$

$$\kappa^b = \nabla^b \phi - \nabla^b \phi' = \nabla^b \phi - \frac{4}{3}\pi\,\rho\,\chi'^b.$$

Then κ^b is constant ($\nabla^a \kappa^b = \mathbf{0}$) and

$$\nabla^b \phi \;=\; \frac{4}{3}\pi\,\rho\,\left(\chi'^b + (\frac{3}{4\,\pi\,\rho})\,\kappa^b\right).$$

Now let o be the (unique) point where the vector field on the right side vanishes. (We can think of o as the point one gets if one displaces o' by the vector $-(3/4\,\pi\,\rho)\,\kappa^b$. This makes sense since we can identify vectors at different points in three-dimensional Euclidean space.) Then $(\chi'^b + (3/4\,\pi\,\rho)\,\kappa^b)$ is just what we would otherwise describe as the position field χ^b relative to o. (Note that when we apply ∇_a to the field, we get ${\delta_a}^b$.) So ϕ qualifies as a canonical solution centered at o. □

Note that the proposition fails if $\rho = 0$. In that case, the implication (1) ⇒ (2) still holds, but not the converse. For then all canonical solutions have vanishing gradient ($\nabla^a \phi = (4/3)\,\pi\,\rho\,\chi^a = \mathbf{0}$), whereas condition (2) requires only that $\nabla^a \phi$ be constant.

Condition (2) in the proposition naturally lifts to the context of classical spacetimes where it becomes

$$\nabla^a\,\nabla^b\,\phi = \frac{4}{3}\pi\,\rho\,h^{ab}. \tag{4.4.2}$$

(That is why it will be convenient later.) The latter holds iff the restriction of ϕ to any spacelike hypersurface S (together with the restrictions of ∇ and h^{ab} to S) satisfies (2).

Let us now shift back to the framework of geometrized Newtonian gravitation theory. Our first task is to introduce a class of cosmological models that correspond to the Friedmann spacetimes we considered in section 2.11. We could proceed just as we did there—i.e., start with a condition of spatial homogeneity and isotropy (relative to some smooth, future-directed, unit timelike field ξ^a) and derive the consequences of that assumption. We could show again that ξ^a is necessarily geodesic, twist-free, and shear-free; that any vector field definable in terms of the basic elements of structure t_a, h^{ab}, ∇, and ξ^a is necessarily proportional to ξ^a; and so forth. Instead, we proceed directly to an explicit characterization.

Let us first take a *(classical) cosmological model* to be a a structure of the form $(M, t_a, h^{ab}, \nabla, \xi^a, \rho)$, where (M, t_a, h^{ab}, ∇) is a classical spacetime; ξ^a is a smooth, future-directed unit timelike field on M; and ρ is a smooth field on M. We take ξ^a be the four-velocity of a cosmic fluid that fills all of spacetime, and take ρ to be the mass-density of the fluid. Next, let us say that $(M, t_a, h^{ab}, \nabla, \xi^a, \rho)$ is *Friedmann-like* if the following conditions are satisfied.

(1) ξ^a is geodesic, twist-free, and shear-free; i.e.,

$$\nabla_a \xi^b = \frac{1}{3}(\delta_a{}^b - t_a \xi^b)\,\theta. \tag{4.4.3}$$

(Here $\theta = \nabla_a \xi^a$ is the scalar expansion field associated with ξ^a. Note that equation (4.4.3) follows from equations (4.1.42), (4.1.45), and (4.1.12). In more detail, since $\omega_{ab} = \sigma_{ab} = \mathbf{0}$ and $\xi^n \nabla_n \xi^a = \mathbf{0}$, we have

$$\nabla_a \xi^b = \theta_a{}^b = \theta_{an} h^{nb} = \left(\frac{1}{3}\theta\,\hat{h}_{an}\right) h^{nb} = \frac{1}{3}(\delta_a{}^b - t_a \xi^b)\,\theta.)$$

(2) $\nabla^a \rho = \mathbf{0}$; i.e., ρ is constant on all spacelike hypersurfaces.
(3) Poisson's equation $R_{ab} = 4\pi\rho\, t_{ab}$ holds.

Note that we have not included Trautman's two supplemental integrability conditions ($R^a{}_b{}^c{}_d = R^c{}_d{}^a{}_b$ and $R^{ab}{}_{cd} = \mathbf{0}$) in the list. We have not done so because, as we now show, they follow from the other assumptions. *So in this special case—the case of Friedmann-like cosmological models — the difference between our two formulations of geometrized Newtonian theory collapses.* (In section 4.5, we shall consider another case where it collapses.)

PROPOSITION 4.4.2. *Let $(M, t_a, h^{ab}, \nabla, \xi^a, \rho)$ be a Friedmann-like cosmological model. Then the following conditions hold.*

(1) $R^a{}_b{}^c{}_d = R^c{}_d{}^a{}_b$ *and* $R^{ab}{}_{cd} = \mathbf{0}$.
(2) $\xi^n \nabla_n \theta = -4\pi\rho - \frac{1}{3}\theta^2$.

Proof. (1) The first condition, $R^a{}_b{}^c{}_d = R^c{}_d{}^a{}_b$, follows immediately from proposition 4.3.3. (We need only that ξ^a be geodesic and twist-free for this much.) For the second condition, $R^{ab}{}_{cd} = \mathbf{0}$, it will suffice to establish the existence, at least locally, of a smooth, future-directed, unit timelike field η^a on M that is rigid and twist-free ($\nabla^a \eta^b = \mathbf{0}$). For then we can invoke proposition 4.2.4.

Let p be any point in M. All Friedmann-like cosmological models are spatially flat (by proposition 4.1.5). So there must be an open set O containing p and a smooth spacelike field χ^a on O such that $\nabla^a \chi^b = h^{ab}$. (Recall the very beginning of our proof of proposition 4.2.3.) Now consider the field

$$\eta^a = \xi^a - \frac{1}{3}\theta\,\chi^a$$

on O. It is certainly a smooth, future-directed, unit timelike field. We claim that it is rigid and twist-free, as required. To see this, note first that, by equation (4.4.3),

$$\nabla^a \eta^b = \nabla^a \xi^b - \frac{1}{3}\,\theta\,\nabla^a \chi^b - \frac{1}{3}(\nabla^a \theta)\,\chi^b = \left(\frac{1}{3}\,\theta\,h^{ab} - \frac{1}{3}\,\theta\,h^{ab}\right) - \frac{1}{3}\,(\nabla^a \theta)\,\chi^b.$$

Furthermore, $\nabla^a \theta = \mathbf{0}$. This follows, since by equation (4.4.3) and Poisson's equation,

$$\begin{aligned}\nabla^a \theta &= h^{an}\,\nabla_n \nabla_m\,\xi^m = -h^{an}\,R^m{}_{rnm}\,\xi^r + h^{an}\,\nabla_m \nabla_n\,\xi^m \\ &= -h^{an}\,R_{rn}\,\xi^r + \nabla_m(\nabla^a\,\xi^m) = -h^{an}\,(4\pi\rho t_{rn})\,\xi^r + \nabla_m\left(\frac{1}{3}h^{am}\,\theta\right) = \frac{1}{3}\nabla^a \theta.\end{aligned}$$

So $\nabla^a \eta^b = \mathbf{0}$, as claimed.

(2) Here we start as we did in our derivation of Raychaudhuri's equation (2.8.17):

$$\begin{aligned}\xi^a \nabla_a\,\theta = \xi^a \nabla_a \nabla_b\,\xi^b &= -\xi^a R^b{}_{cab}\,\xi^c + \xi^a \nabla_b \nabla_a\,\xi^b \\ &= -\,R_{ca}\,\xi^c \xi^a + \nabla_b(\xi^a \nabla_a\,\xi^b) - (\nabla_b\,\xi^a)(\nabla_a\,\xi^b).\end{aligned}$$

But now, by equation (4.4.3), $\xi^a \nabla_a\,\xi^b = \mathbf{0}$ and $(\nabla_b\,\xi^a)(\nabla_a\,\xi^b) = \frac{1}{3}\,\theta^2$. And by Poisson's equation (the third condition in our characterization of Friedmann-like cosmological models), $R_{ca}\,\xi^c \xi^a = 4\,\pi\,\rho$. So we are done. □

Note that condition (2) in the proposition—the equation that governs the rate of change of θ in Friedmann-like cosmological models—agrees with equation (2.11.9) in the case where $p = 0$. This makes sense. Though in general relativity the "gravitational field" generated by a blob of perfect fluid depends on its internal pressure as well as on its mass density, only the latter plays a role in Newtonian gravitation theory.

Now we make precise our claim about the recovery of canonical solutions. Condition (4.4.4) in the following proposition is the condition we motivated using proposition 4.4.1. At least if $\rho \neq 0$, we can understand it to capture the claim that the restriction of ϕ to any spacelike hypersurface is a canonical solution to Poisson's equation. (If $\rho = 0$, it asserts instead that $\nabla^a \phi$ is constant on spacelike hypersurfaces.)

PROPOSITION 4.4.3. *Let $(M, t_a, h^{ab}, \nabla, \xi^a, \rho)$ be a Friedmann-like cosmological model, and let ϕ be a smooth field on some open set in M. If ϕ arises as part of a de-geometrization $(\overset{f}{\nabla}, \phi)$ of ∇ (on that open set), then*

$$\nabla^a \nabla^b \phi = \frac{4}{3} \pi \rho h^{ab}. \tag{4.4.4}$$

Conversely, if ϕ satisfies equation (4.4.4), then, at least locally, there is a derivative operator $\overset{f}{\nabla}$ such that $(\overset{f}{\nabla}, \phi)$ is a de-geometrization of ∇. (Once again, to say that $(\overset{f}{\nabla}, \phi)$ is a de-geometrization of ∇ is to say that it satisfies conditions (R1)–(R4) in the Trautman Recovery Theorem.)

Proof. We begin the proof of showing that, given any point p in M, there is an open set O containing p and *some* de-geometrization $(\overset{f}{\nabla}{}^*, \phi^*)$ of ∇ on O such that

$$\nabla^a \nabla^b \phi^* = \frac{4}{3} \pi \rho h^{ab}. \tag{4.4.5}$$

This will require a bit of work. But once we have established this much, our principal claims will follow easily.

We begin just as we did in our proof of the Trautman Recovery Theorem. (Note that all the assumptions needed for the theorem hold. In particular, the supplemental integrability conditions $R^a{}_b{}^c{}_d = R^c{}_d{}^a{}_b$ and $R^{ab}{}_{cd} = \mathbf{0}$ hold. We know this from proposition 4.4.2.) Let p be any point in M. Then we can find an open set O containing p, and a smooth, future-directed, unit timelike field η^a on O that is rigid and twist-free. Now consider the derivative operator $\overset{f}{\nabla}{}^*$ on O defined by $\overset{f}{\nabla}{}^* = (\nabla, C^a{}_{bc})$, where $C^a{}_{bc} = t_a t_b \phi^a$ and $\phi^a = \xi^n \nabla_n \xi^a$. As we know from our proof of the Trautman Recovery Theorem, we can (after possibly restricting O to some smaller open set containing p) find a smooth scalar field ϕ^* on O such that $\phi^a = \nabla^a \phi^*$ and such that $(\overset{f}{\nabla}{}^*, \phi^*)$ qualifies as a de-geometrization of ∇ on O. We claim that ϕ^* satisfies equation (4.4.5).

To see this, consider the field ξ^a. (It gives the four-velocity of matter in our Friedmann-like cosmological model.) It is a geodesic field with respect to ∇. So, by condition (R3) in the Trautman Recovery Theorem,

$$\xi^n \overset{f}{\nabla}{}^*_n \xi^a = - \overset{f}{\nabla}{}^{*a} \phi^*.$$

Hence,

$$\begin{aligned} \overset{f}{\nabla}{}^{*a} \overset{f}{\nabla}{}^{*b} \phi^* &= - \overset{f}{\nabla}{}^{*a} (\xi^n \overset{f}{\nabla}{}^*_n \xi^b) = -(\overset{f}{\nabla}{}^{*a} \xi^n)(\overset{f}{\nabla}{}^*_n \xi^b) \\ &\quad - \xi^n \overset{f}{\nabla}{}^{*a} \overset{f}{\nabla}{}^*_n \xi^b \\ &= -(\overset{f}{\nabla}{}^{*a} \xi^n)(\overset{f}{\nabla}{}^*_n \xi^b) - \xi^n \overset{f}{\nabla}{}^*_n \overset{f}{\nabla}{}^{*a} \xi^b. \end{aligned} \tag{4.4.6}$$

(We use the fact that $\overset{f}{\nabla}{}^{*}$ is flat for the final equality.) Next, we derive an expression for $\overset{f}{\nabla}{}^{*}{}_{n}\, \xi^{b}$. We have, by equation (4.4.3),

$$\begin{aligned}\overset{f}{\nabla}{}^{*}{}_{n}\, \xi^{b} &= \nabla_{n}\, \xi^{b} - C^{b}{}_{nm}\, \xi^{m} = \nabla_{n}\, \xi^{b} - (t_{n}\, t_{m}\, \phi^{b})\xi^{m} = \nabla_{n}\, \xi^{b} - t_{n}\, \phi^{b} \\ &= \frac{1}{3}(\delta_{n}{}^{b} - t_{n}\, \xi^{b})\, \theta - t_{n}\, \phi^{b}.\end{aligned}$$

It follows that

$$\overset{f}{\nabla}{}^{*a}\, \xi^{b} = \frac{1}{3}\, h^{ab}\, \theta.$$

Substituting these expression for $\overset{f}{\nabla}{}^{*}{}_{n}\, \xi^{b}$ and $\overset{f}{\nabla}{}^{*a}\, \xi^{b}$ in equation (4.4.6) yields

$$\overset{f}{\nabla}{}^{*a}\, \overset{f}{\nabla}{}^{*b}\, \phi^{*} = -\frac{1}{9}\, h^{ab}\, \theta^{2} - \frac{1}{3}\, h^{ab}\, \xi^{n}\, \overset{f}{\nabla}{}^{*}{}_{n}\, \theta. \tag{4.4.7}$$

Now by the second clause of proposition 4.4.2,

$$\xi^{n}\, \overset{f}{\nabla}{}^{*}{}_{n}\, \theta = \xi^{n} \nabla_{n}\, \theta = -\frac{1}{3}\, \theta^{2} - 4\, \pi\, \rho.$$

So, after substituting this expression for $\xi^{n}\, \overset{f}{\nabla}{}^{*}{}_{n}\, \theta$ in equation (4.4.7), we have

$$\overset{f}{\nabla}{}^{*a}\, \overset{f}{\nabla}{}^{*b}\, \phi^{*} = \frac{4}{3}\, \pi\, \rho\, h^{ab}.$$

But

$$\begin{aligned}\overset{f}{\nabla}{}^{*a}\, \overset{f}{\nabla}{}^{*b}\, \phi^{*} = \overset{f}{\nabla}{}^{*a}\, \nabla^{b}\, \phi^{*} &= h^{am}\, (\nabla_{m}\, \nabla^{b}\, \phi^{*} - C^{b}{}_{mn}\, \nabla^{n}\, \phi^{*}) \\ &= h^{am}\, [\nabla_{m}\, \nabla^{b} \phi^{*} - (t_{m}\, t_{n}\, \phi^{b})\, \nabla^{n}\, \phi^{*}] = \nabla^{a}\, \nabla^{b} \phi^{*}.\end{aligned}$$

So

$$\nabla^{a}\, \nabla^{b}\, \phi^{*} = \frac{4}{3}\, \pi\, \rho\, h^{ab},$$

as claimed. This completes the first part of the proof.

Now let ϕ be a smooth field on some open subset U of M. Let p be any point in U. We know from what we have just proved that we can find an open subset O of U containing p and a de-geometrization $(\overset{f}{\nabla}{}^{*}, \phi^{*})$ of ∇ on O such that ϕ^{*} satisfies equation (4.4.5). Suppose first that ϕ arises as part of a de-geometrization $(\overset{f}{\nabla}, \phi)$ of ∇ on U. Then we have *two* de-geometrizations of ∇

on O, namely $(\overset{f}{\nabla}{}^{*}, \phi^{*})$ and $(\overset{f}{\nabla}, \phi)$. By the final part of the Trautman Recovery Theorem governing the non-uniqueness of de-geometrizations, it follows that

$$\mathbf{0} = \nabla^{a}\,\nabla^{b}\,(\phi - \phi^{*}) = \nabla^{a}\,\nabla^{b}\,\phi - \frac{4}{3}\,\pi\,\rho\,h^{ab}.$$

(Here the roles of $(\overset{f}{\nabla}{}', \phi')$ and $(\overset{f}{\nabla}, \phi)$ in that theorem are played, respectively, by $(\overset{f}{\nabla}, \phi)$ and $(\overset{f}{\nabla}{}^{*}, \phi^{*})$.) So ϕ satisfies equation (4.4.4) throughout the open set O containing p. But p was chosen arbitrarily. So ϕ satisfies equation (4.4.4) everywhere in U.

Conversely, suppose ϕ satisfies equation (4.4.4). Then, by equation (4.4.5) again, we have

$$\nabla^{a}\,\nabla^{b}\,(\phi - \phi^{*}) = \frac{4}{3}\,\pi\,\rho\,h^{ab} - \frac{4}{3}\,\pi\,\rho\,h^{ab} = \mathbf{0}$$

on O. Hence, by the final part of the Trautman theorem again, if we set $\overset{f}{\nabla} = (\overset{f}{\nabla}{}^{*},\ t_b\,t_c\,\nabla^{a}(\phi - \phi^{*}))$, then $(\overset{f}{\nabla},\ \phi)$ qualifies as a de-geometrization of ∇ on O. □

Let us think about what we would experience if we resided in a Friedmann-like Newtonian universe of the sort we have been considering. Suppose we were at rest in the cosmic fluid—i.e., moving along an integral curve of the background four-velocity field ξ^{a}. Then we would experience no net force and would observe all other mass points in the fluid moving uniformly away from, or toward, us. If we were inclined to describe the situation in terms of traditional, non-geometrized Newtonian theory, we would say (adopting, implicitly, a particular de-geometrization) that the the gravitational field is centered where we are and vanishes there. (That is why we experience no net force.) But we would offer a different account for why our colleagues co-moving with *other* cosmic mass points experience no net force. From our point of view (i.e., according to *our* de-geometrization), they *do* experience a non-zero gravitational force. But it is perfectly balanced by a corresponding "inertial" force. And it is for *this* reason that they experience no net force. (Of course, those colleagues have their own story to tell with the roles reversed. They take themselves to be the ones residing where the gravitational field vanishes.)

4.5. Geometrized Newtonian TheorySecond Version

In this section, we prove a recovery or de-geometrization theorem for the Künzle-Ehlers version of geometrized Newtonian gravitation theory. It is the counterpart to the recovery theorem we proved for the Trautman version (proposition 4.2.5) and actually subsumes that earlier result as a special case.

We also consider a second set of special circumstances in which the difference between our two versions of the theory collapses. We saw in section 4.4 that Trautman's second integrability condition $R^{ab}{}_{cd} = \mathbf{0}$ holds automatically in Friedmann-like cosmological models. Here we show that it holds automatically if we restrict attention to classical spacetimes that are, in a certain weak sense, asymptotically flat.

We start with a lemma. Our proof of the Trautman Recovery Theorem turned on the availability of a rigid, twist-free field η^a. Existence was guaranteed by the second integrability condition (proposition 4.2.4). Now we have to work with less. We cannot count on the existence of rigid, twist-free fields. But, as we now show, we *can* still count on the existence of fields that are, at least, rigid. And this will suffice. To prove the new recovery theorem, we need only rerun the argument for the old one using a field η^a that is merely rigid. The computations are a bit more complicated, but no new ideas are involved. (We could have proved this version of the theorem first and then recovered the Trautman version simply by considering what happens when η^a is also twist-free. But there is some advantage to taking on complications one at a time.)

PROPOSITION 4.5.1. *Let (M, t_a, h^{ab}, ∇) be a classical spacetime that is spatially flat ($R^{abcd} = \mathbf{0}$). Then, given any point p in M, there exist an open set O containing p and a smooth, future-directed, unit timelike field η^a on O that is rigid ($\nabla^{(a}\eta^{b)} = \mathbf{0}$).*

Proof. Let p be any point in M, and let $\gamma : I \to M$ be a smooth, future-directed, timelike curve—with four-velocity field $\widehat{\eta}^a$—that passes through p. We claim first that we can find three smooth, linearly independent, spacelike fields $\overset{1}{\sigma}{}^a, \overset{2}{\sigma}{}^a, \overset{3}{\sigma}{}^a$ on some open set O containing p with these properties (for all i):

(i) $h^{ab} = \overset{1}{\sigma}{}^a\,\overset{1}{\sigma}{}^b + \overset{2}{\sigma}{}^a\,\overset{2}{\sigma}{}^b + \overset{3}{\sigma}{}^a\,\overset{3}{\sigma}{}^b$.
(ii) $\nabla^a\,\overset{i}{\sigma}{}^b = \mathbf{0}$.
(iii) $\widehat{\eta}^n\nabla_n\,\overset{i}{\sigma}{}^a = \mathbf{0}$ on $\gamma[I]$.

We can generate the fields as follows. First we find three linearly independent, spacelike vectors at p that satisfy condition (i)—just as we did in the proof of proposition 4.1.4. Then we extend the vectors to an open set containing p in two stages. First, we extend them by parallel transport along γ. (So condition (iii) is satisfied.) Then we extend them "outward" from $\gamma[I]$ by parallel transport along spacelike curves. The latter operation works this way. Let S be a spacelike hypersurface that intersects the image of γ at the point q. Then, because of spatial flatness, parallel transport of spacelike vectors *within* S is, at least locally, path independent. (Recall proposition 4.1.5.) So we can unambiguously extend the triple $\overset{1}{\sigma}{}^a, \overset{2}{\sigma}{}^a, \overset{3}{\sigma}{}^a$ at q by parallel transport to points on S sufficiently close to q. The fields generated by this construction are "constant in spacelike directions"; i.e., $\lambda^n \nabla_n \overset{i}{\sigma}{}^a = \mathbf{0}$ for all spacelike vectors λ^a. The latter condition is equivalent to (ii). Finally, we claim, condition (i) holds everywhere. Consider the difference field $(h^{ab} - (\overset{1}{\sigma}{}^a \overset{1}{\sigma}{}^b + \overset{2}{\sigma}{}^a \overset{2}{\sigma}{}^b + \overset{3}{\sigma}{}^a \overset{3}{\sigma}{}^b))$. It vanishes at p. Hence, by (iii), it vanishes along $\gamma[I]$. And therefore, by (ii), it vanishes on spacelike hypersurfaces that intersect $\gamma[I]$. So it vanishes everywhere. Thus, as claimed, we can find three smooth, spacelike fields $\overset{1}{\sigma}{}^a, \overset{2}{\sigma}{}^a, \overset{3}{\sigma}{}^a$ on some open set O containing p that satisfy the three listed conditions. And the fields must certainly be linearly independent throughout O—because we started with three linearly independent vectors at p, and linear independence is preserved under parallel transport.

It follows from (ii), of course, that $\nabla^{[a}\, \overset{i}{\sigma}{}^{b]} = \mathbf{0}$ for all i. So, restricting O to a smaller open set containing p if necessary, we can find smooth scalar fields $\overset{i}{x}$ on O such that $\overset{i}{\sigma}{}^a = \nabla^a \overset{i}{x} = h^{ab} \nabla_b \overset{i}{x}$. (Here we invoke proposition 4.1.6.) We can pin them down uniquely by requiring that they assume the value 0 at points on $\gamma[I]$. This guarantees that

(4.5.1)
$$\widehat{\eta}^n \nabla_n \overset{i}{x} = 0$$

on $\gamma[I] \cap O$ for all i. And, by condition (i),

(4.5.2)
$$h^{ab} (\nabla_a \overset{i}{x}) (\nabla_b \overset{j}{x}) = \delta_{ij}$$

holds everywhere for all i and j. (Why? Contracting (i) with $(\nabla_b \overset{j}{x})$ yields

$$\overset{j}{\sigma}{}^a = h^{ab} (\nabla_b \overset{j}{x}) = \sum_{i=1}^{3} \overset{i}{\sigma}{}^a (\overset{i}{\sigma}{}^b \nabla_b \overset{j}{x}).$$

But the vectors $\overset{1}{\sigma}{}^a, \overset{2}{\sigma}{}^a, \overset{3}{\sigma}{}^a$ are linearly independent at every point. So $\overset{i}{\sigma}{}^b \nabla_b \overset{j}{x} = \delta_{ij}$ and, therefore, $h^{ab}(\nabla_a \overset{i}{x})(\nabla_b \overset{j}{x}) = \overset{i}{\sigma}{}^b \nabla_b \overset{j}{x} = \delta_{ij}$.)

Now we extend the tangent field $\widehat{\eta}^a$ to a smooth field η^a on O by requiring that $t_a \eta^a = 1$ and $\eta^n \nabla_n \overset{i}{x} = 0$ hold *everywhere* for all i. (The fields $t_a, (\nabla_a \overset{1}{x}), (\nabla_a \overset{2}{x}), (\nabla_a \overset{3}{x})$ form a co-basis at every point, and so a vector field is uniquely determined by its contractions with them.) We claim that the resultant field η^a is rigid; i.e., $\mathcal{L}_\eta h^{ab} = \mathbf{0}$. We have $\mathcal{L}_\eta \overset{i}{x} = \eta^n \nabla_n \overset{i}{x} = 0$ for all i. And $\mathcal{L}_\eta(\nabla_a \varphi) = \nabla_a(\mathcal{L}_\eta \varphi)$ for *all* smooth scalar fields φ. (This is easily checked using proposition 1.7.4.) So $\pounds_\eta(\nabla_a \overset{i}{x}) = \nabla_a(\pounds_\eta \overset{i}{x}) = \mathbf{0}$ for all i. Hence, by equation (4.5.2),

$$0 = \pounds_\eta(h^{ab}\,(\nabla_a \overset{i}{x})\,(\nabla_b \overset{j}{x})) = (\nabla_a \overset{i}{x})\,(\nabla_b \overset{j}{x})\,\pounds_\eta\,h^{ab}$$

for all i and j. But we also have $\pounds_\eta\,t_a = \eta^n \nabla_n\,t_a + t_n \nabla_a\,\eta^n = \mathbf{0}$ and, therefore, $t_a\,\pounds_\eta h^{ab} = \pounds_\eta(t_a\,h^{ab}) = \mathbf{0}$. Thus, contracting $\pounds_\eta h^{ab}$ with any of the basis elements t_a, $(\nabla_a \overset{1}{x})$, $(\nabla_a \overset{2}{x})$, $(\nabla_a \overset{3}{x})$ yields $\mathbf{0}$. So $\pounds_\eta h^{ab} = \mathbf{0}$, as claimed. □

Now we turn to the recovery theorem.

PROPOSITION 4.5.2. *(Künzle-Ehlers Recovery Theorem) Let (M, t_a, h^{ab}, ∇) be a classical spacetime that satisfies*

$$R_{bc} = 4\,\pi\,\rho\,t_{bc}, \tag{4.5.3}$$

$$R^a{}_b{}^c{}_d = R^c{}_d{}^a{}_b, \tag{4.5.4}$$

for some smooth scalar field ρ on M. Let η^a be a smooth, future-directed, unit timelike vector field on some open subset O of M that is rigid. (Existence of such fields, at least locally, is guaranteed by the preceding proposition and proposition 4.1.5.) Let $\hat{h}_{ab}$ be the projection field associated with η^a, and let ϕ^a and ω_{ab} be the associated acceleration and rotation fields:

$$\phi^a = \eta^n \nabla_n\,\eta^a,$$
$$\omega_{ab} = \hat{h}^m{}_{[a}\,\hat{h}_{b]n}\,\nabla_m\,\eta^n.$$

Then there exists a unique derivative operator $\overset{f}{\nabla}$ on O such that all the following hold on O.

(RR1) $\overset{f}{\nabla}$ is compatible with t_a and h^{ab}.
(RR2) η^a constant with respect to $\overset{f}{\nabla}$ (i.e., $\overset{f}{\nabla}_a\,\eta^b = \mathbf{0}$).
(RR3) $\overset{f}{\nabla}$ is flat.
(RR4) For all timelike curves with four-velocity field ξ^a,

$$\xi^n \nabla_n \xi^a = \mathbf{0} \iff \xi^n\,\overset{f}{\nabla}_n\,\xi^a = -\phi^a - 2\,\omega_n{}^a \xi^n.$$

(RR5) ϕ^a and ω_{ab} satisfy the "field equations":

(4.5.5) $$\overset{f}{\nabla}{}^{[a}\omega^{bc]} = \mathbf{0},$$

(4.5.6) $$\overset{f}{\nabla}_a \omega^{ab} = \mathbf{0},$$

(4.5.7) $$\overset{f}{\nabla}{}^{[a}\phi^{b]} = \eta^n \overset{f}{\nabla}_n \omega^{ab},$$

(4.5.8) $$\overset{f}{\nabla}_a \phi^a = 4\pi\rho - \omega_{ab}\,\omega^{ab}.$$

Note that, as promised, the Trautman Recovery Theorem emerges as a corollary. If we add the supplemental condition ($R^{ab}{}_{cd} = \mathbf{0}$), then, by proposition 4.2.4 again, we can find timelike fields locally that are rigid *and* twist-free. But if $\omega_{ab} = \mathbf{0}$, it follows from equation (4.5.7) (and proposition 4.1.6) that ϕ^a must, at least locally, be of the form $\phi^a = \nabla^a\phi$ for some smooth scalar field ϕ. And in *this* case ($\omega_{ab} = \mathbf{0}$ and $\phi^a = \nabla^a\phi$), we fully recover the conclusions of the Trautman Recovery Theorem.

The de-geometrization presented here is relativized to a rigid unit timelike vector field η^a. Given *that* field, there is a unique derivative operator satisfying the listed conditions (relative to it). But it will be clear from the proof that, in general, different choices for η^a lead to different derivative operators—i.e., lead to different de-geometrizations. Indeed, one has, here, much the same non-uniqueness that we encountered in the Trautman Recovery Theorem.

Proof. The argument here is similar in structure to the one we gave for the Trautman Recovery Theorem, and many individual steps carry over intact or with only minimal change. We just have to remember that whereas previously we had the condition $\nabla^a\eta^b = \mathbf{0}$ to work with, we now have only $\nabla^{(a}\eta^{b)} = \mathbf{0}$.

Consider the fields

(4.5.9) $$\kappa_{ab} = \hat{h}_{n[b}\,\nabla_{a]}\eta^n,$$

(4.5.10) $$C^a{}_{bc} = 2\,t_{(b}\,\kappa_{c)}{}^a$$

on O. It is easy to check that they satisfy the following conditions.

(4.5.11) $$2\,\kappa_a{}^b = \nabla_a\eta^b - \hat{h}_{na}\nabla^b\eta^n,$$

(4.5.12) $$\kappa^{ab} = \nabla^a\eta^b \;=\; \omega^{ab},$$

(4.5.13) $$2\,\kappa_a{}^b\eta^a = \eta^a\nabla_a\,\eta^b \;=\; \phi^b,$$

$$2\,\kappa_a{}^b = 2\,\nabla_a\,\eta^b - t_a\,\phi^b, \tag{4.5.14}$$

$$C^a{}_{ac} = \mathbf{0}. \tag{4.5.15}$$

We get the second from the fact that η^a is rigid, and so $\nabla^a\eta^b = \nabla^{[a}\eta^{b]} = \omega^{ab}$. The fourth follows from the second and third. (Note that contracting both sides with either h^{ar} or η^a yields the same result.) The fifth follows from the anti-symmetry of κ_{ab}.

Next consider the derivative operator $\overset{f}{\nabla} = (\nabla, C^a{}_{bc})$ on O. We claim that it satisfies all the listed conditions. (RR1) follows immediately from proposition 4.1.3. For (RR2), note that, by equations (4.5.13) and (4.5.14),

$$\begin{aligned}\overset{f}{\nabla}_a\,\eta^b &= \nabla_a\,\eta^b - C^b{}_{an}\,\eta^n = \nabla_a\,\eta^b - (t_a\,\kappa_n{}^b + t_n\,\kappa_a{}^b)\,\eta^n\\ &= \nabla_a\,\eta^b - \frac{1}{2}\,t_a\,\phi^b - \left(\nabla_a\,\eta^b - \frac{1}{2}\,t_a\,\phi^b\right) = \mathbf{0}.\end{aligned}$$

Thus, as required for (RR2), η^a is constant with respect to the new derivative operator $\overset{f}{\nabla}$.

Now we turn to the Riemann curvature field associated with $\overset{f}{\nabla}$. We have, by equation (1.8.2),

$$\begin{aligned}\overset{f}{R}{}^a{}_{bcd} &= R^a{}_{bcd} + 2\,\nabla_{[c}\,C^a{}_{d]b} + 2\,C^n{}_{b[c}\,C^a{}_{d]n}\\ &= R^a{}_{bcd} + 2\,t_{[d}\,\nabla_{c]}\,\kappa_b{}^a + 2\,t_b\nabla_{[c}\,\kappa_{d]}{}^a + 2\,t_b\,t_{[d}\,\kappa_{c]}{}^n\,\kappa_n{}^a.\end{aligned} \tag{4.5.16}$$

It follows immediately that $\overset{f}{R}{}^{abcd} = R^{abcd}$. But $R^{abcd} = \mathbf{0}$. (By proposition 4.1.5, this is a consequence of the geometrized version of Poisson's equation (4.5.3).) So $\overset{f}{\nabla}$ is spatially flat. Now recall the second clause of proposition 4.2.4. We have just verified that there is smooth unit timelike field η^a on O that is constant with respect to $\overset{f}{\nabla}$. So (since $\overset{f}{\nabla}$ is spatially flat), the proposition tells us that $\overset{f}{\nabla}$ must be flat outright; i.e., $\overset{f}{R}{}^a{}_{bcd} = \mathbf{0}$. So we have (RR3). And equation (4.5.16) reduces to

$$R^a{}_{bcd} = -2\,t_{[d}\,\nabla_{c]}\,\kappa_b{}^a - 2\,t_b\nabla_{[c}\,\kappa_{d]}{}^a - 2\,t_b\,t_{[d}\,\kappa_{c]}{}^n\,\kappa_n{}^a. \tag{4.5.17}$$

For (RR4), note first that for all timelike curves on O with four-velocity field ξ^a,

$$\xi^n \overset{f}{\nabla}_n \xi^a = \xi^n(\nabla_n\xi^a - C^a{}_{nm}\,\xi^m) \;=\; \xi^n\nabla_n\xi^a - (t_m\,\kappa_n{}^a + t_n\,\kappa_m{}^a)\xi^m\xi^n$$
$$= \xi^n\nabla_n\xi^a - 2\,\kappa_n{}^a\xi^n.$$

But, by equations (4.5.14) and (4.1.42) and the fact that η^a is rigid ($\theta_n{}^a = \mathbf{0}$),

$$2\,\kappa_n{}^a = \;2\,\nabla_n\,\eta^a \;-\; t_n\,\phi^a$$
$$= 2\,(\omega_n{}^a + t_n\,\phi^a) \;-\; t_n\,\phi^a = \;2\,\omega_n{}^a + t_n\,\phi^a. \tag{4.5.18}$$

So $\xi^n\nabla_n\xi^a = \mathbf{0}$ iff $\xi^n \overset{f}{\nabla}_n \xi^a = -\phi^a - 2\,\omega_n{}^a\,\xi^n$. Thus we have (RR4).

Notice also that there can be at most one derivative operator $\overset{f}{\nabla}$ on O satisfying condition (RR4), so we get our uniqueness claim. For suppose that $\overset{f}{\nabla}{}' = (\overset{f}{\nabla}, C'^a{}_{bc})$ satisfies it as well. Then, for all timelike geodesics on O (with respect to ∇) with four-velocity field ξ^a, we have

$$\xi^n \overset{f}{\nabla}_n \xi^a = -\phi^a - 2\,\omega_n{}^a\xi^n = \xi^n \overset{f}{\nabla}{}'_n\xi^a = \xi^n(\overset{f}{\nabla}_n\,\xi^a - C'^a{}_{nm}\xi^m).$$

So, $C'^a{}_{nm}\,\xi^m\,\xi^n = \mathbf{0}$ holds at every point. But every future-directed unit timelike vector ξ^a at a point in O is the tangent vector of some geodesic (with respect to ∇) through the point, and the collection of future-directed unit timelike vectors at a point spans the tangent space there. So it follows that $C'^a{}_{nm} = \mathbf{0}$ at every point in O.

Now, finally, we turn to (RR5). The four conditions we must verify all follow from equation (4.5.17). Contracting a with d yields

$$4\,\pi\,\rho\,t_b\,t_c = R_{bc} = t_c\,\nabla_a\,\kappa_b{}^a + t_b\,\nabla_a\,\kappa_c{}^a + t_b\,t_c\,\kappa_a{}^n\kappa_n{}^a. \tag{4.5.19}$$

And raising c yields

$$R^a{}_b{}^c{}_d = -t_d\,\nabla^c\,\kappa_b{}^a \;-\; t_b\nabla^c\,\kappa_d{}^a + t_b\nabla_d\,\kappa^{ca} - t_b\,t_d\,\kappa^{cn}\kappa_n{}^a. \tag{4.5.20}$$

Let us now contract (4.5.19) with $\eta^b\,h^{cr}$. This, together with equation (4.5.12), gives us $\mathbf{0} = \nabla_a\,\kappa^{ra} = \nabla_a\,\omega^{ra}$. It follows that

$$\overset{f}{\nabla}_a\,\omega^{ab} = \nabla_a\,\omega^{ab} - \omega^{nb}C^a{}_{an} - \omega^{an}C^b{}_{an} = \mathbf{0}. \tag{4.5.21}$$

(Here $C^a{}_{an} = \mathbf{0}$ by (4.5.15), and $\omega^{an}C^b{}_{an} = \mathbf{0}$ because of the respective antisymmetry and symmetry of ω^{an} and $C^b{}_{an}$.) So we have the second in our list of four (RR5) conditions. Next, let us contract equation (4.5.19) with $\eta^b\eta^c$. Then, using equations (4.5.13), (4.5.14), and (4.5.18), we get

$$\begin{aligned}
4\,\pi\,\rho &= 2\,\eta^b\,\nabla_a\,\kappa_b{}^a + \kappa_a{}^n\kappa_n{}^a = 2\,[\nabla_a(\kappa_b{}^a\,\eta^b) - (\nabla_a\,\eta^b)\kappa_b{}^a] + \kappa_a{}^n\kappa_n{}^a \\
&= \nabla_a\,\phi^a - 2\,(\kappa_a{}^b + \frac{1}{2}t_a\,\phi^b)\kappa_b{}^a + \kappa_a{}^n\kappa_n{}^a \\
&= \nabla_a\,\phi^a - \kappa_a{}^b\,\kappa_b{}^a = \nabla_a\,\phi^a - \omega_a{}^b\,\omega_b{}^a.
\end{aligned}$$

So, by equation (4.5.15) again,

$$\overset{f}{\nabla}_a\,\phi^a \;=\; \nabla_a\,\phi^a - C^a{}_{an}\,\phi^n \;=\; \nabla_a\,\phi^a \;=\; 4\,\pi\,\rho - \omega_{ab}\,\omega^{ab}. \tag{4.5.22}$$

Thus we have the fourth condition in the (RR5) list. That leaves the first and the third.

Now, for the first time, we use the fact that ∇ satisfies the first supplemental curvature condition (4.5.4). Since $\overset{f}{\nabla}$ satisfies it as well—as it clearly does since $\overset{f}{R}{}^a{}_{bcd} = \mathbf{0}$—we know from proposition 4.3.5 that κ_{ab} must be closed; i.e, $\overset{f}{\nabla}_{[a}\,\kappa_{bc]} = \mathbf{0}$. So, by equation (4.5.12), $\overset{f}{\nabla}{}^{[a}\omega^{bc]} = \overset{f}{\nabla}{}^{[a}\kappa^{bc]} = \mathbf{0}$. That is the first condition in the list. Finally, contracting equation (4.5.20) with $\eta^b\eta^d$ and using equations (4.5.13), (4.5.12), and (4.5.18) yields

$$\begin{aligned}
R^a{}_b{}^c{}_d\,\eta^b\eta^d &= \;-2\,\eta^b\,\nabla^c\,\kappa_b{}^a + \eta^d\,\nabla_d\,\kappa^{ca} - \kappa^{cn}\kappa_n{}^a \\
&= \;-2\left(\nabla^c(\kappa_b{}^a\,\eta^b) - (\nabla^c\,\eta^b)\kappa_b{}^a\right) + \eta^d\,\nabla_d\,\kappa^{ca} - \kappa^{cn}\kappa_n{}^a \\
&= \;-\nabla^c\,\phi^a + \eta^d\,\nabla_d\,\kappa^{ca} + \kappa^{cb}\kappa_b{}^a = -\nabla^c\,\phi^a + \eta^d\,\nabla_d\,\omega^{ca} + \omega^{cb}\omega_b{}^a.
\end{aligned}$$

So, since $R^a{}_b{}^c{}_d = R^c{}_d{}^a{}_b$ (and since ω^{cn} is anti-symmetric), $\nabla^{[a}\,\phi^{c]} - \eta^d\,\nabla_d\,\omega^{ac} = \mathbf{0}$. But, as one can easily check (with a computation much like ones we have seen before), $\overset{f}{\nabla}{}^a\,\phi^c = \nabla^a\,\phi^c$ and $\eta^d\,\overset{f}{\nabla}_d\,\omega^{ac} = \eta^d\nabla_d\,\omega^{ac}$. This gives us the third condition in the (RR5) list, and we are done. □

We have claimed that the difference between the two versions of geometrized Newtonian gravitation theory collapses if one restricts attention to classical spacetimes that are, in a certain weak sense, "asymptotically flat." (In that case, the second supplemental curvature condition, $R^{ab}{}_{cd} = \mathbf{0}$, follows from the other assumptions.) Now we make the claim precise. Toward that goal, we first prove a result of Ehlers's [15].

PROPOSITION 4.5.3. *Let (M, t_a, h^{ab}, ∇) be a classical spacetime that is spatially flat ($R^{abcd} = \mathbf{0}$). Then there is a smooth scalar field Ψ on M such that*

$$R^a{}_{bcd}\, R^b{}_a{}^c{}_e = \Psi\, t_d\, t_e. \tag{4.5.23}$$

Moreover,

$$R^{ab}{}_{cd} = \mathbf{0} \Longleftrightarrow \Psi = 0. \tag{4.5.24}$$

Proof. Let p be any point in M; let η^a be a smooth, future-directed, rigid, unit timelike field defined on some open set containing p (existence is guaranteed, once again, by proposition 4.5.1); and let ω_{ab} be the rotation field determined by η^a. Further, let κ_{ab} and $\overset{f}{\nabla}$ be defined (relative to η^a) as in the preceding proof. Then, by equations (4.5.17), (4.5.20), and (4.5.18),

$$\begin{aligned} R^a{}_{bcd}\, R^b{}_a{}^c{}_e = {} & \big(-2\, t_{[d}\, \nabla_{c]}\, \kappa_b{}^a - 2\, t_b \nabla_{[c}\, \kappa_{d]}{}^a - 2\, t_b\, t_{[d}\, \kappa_{c]}{}^n\, \kappa_n{}^a\big) \\ & \big(-t_e\, \nabla^c\, \kappa_a{}^b - t_a \nabla^c\, \kappa_e{}^b + t_a \nabla_e\, \kappa^{cb} - t_a\, t_e\, \kappa^{cn} \kappa_n{}^b\big) \\ = {} & \big(-t_d\, \nabla_c\, \kappa_b{}^a\big)\big(-t_e\, \nabla^c\, \kappa_a{}^b\big) = t_d\, t_e\, (\nabla_c\, \omega_b{}^a)(\nabla^c\, \omega_a{}^b). \end{aligned}$$

So we need only take $\Psi = -(\nabla_c\, \omega_{ab})(\nabla^c\, \omega^{ab})$ at p. Now, by equation (4.5.17) again, we also have $R^{ab}{}_{cd} = -2\, t_{[d}\, \nabla_{c]}\, \kappa^{ba} = -2\, t_{[d}\, \nabla_{c]}\, \omega^{ba}$. Hence (since contracting $R^{ab}{}_{cd}$ with either $\eta^c \eta^d$ or $h^{cr} h^{ds}$ yields $\mathbf{0}$),

$$R^{ab}{}_{cd} = \mathbf{0} \Longleftrightarrow \nabla^c \omega^{ba} = \mathbf{0}. \tag{4.5.25}$$

So the assertion that remains for us to prove, namely equation (4.5.24), is equivalent to

$$\nabla^c \omega^{ba} = \mathbf{0} \Longleftrightarrow (\nabla_c\, \omega_{ab})(\nabla^c \omega^{ab}) = \mathbf{0}. \tag{4.5.26}$$

One direction is trivial, of course. And the other (right to left) follows just from the fact that the indices in $\nabla^c \omega^{ba}$ are spacelike (and the metric induced by h^{ab} on the space of spacelike vectors at any point is positive definite). For future reference, we give the argument in detail. Let $\overset{1}{\sigma}{}^a, \overset{2}{\sigma}{}^a, \overset{3}{\sigma}{}^a$ be three linearly independent, smooth spacelike fields on some open set containing p such that (i) $h^{ab} = \sum_{i=1}^3 \overset{i}{\sigma}{}^a\, \overset{i}{\sigma}{}^b$ and (ii) $\nabla^a\, \overset{i}{\sigma}{}^b = \mathbf{0}$. (Existence is guaranteed by our assumption of spatial flatness. Recall the proof of proposition 4.5.1.) Let $\overset{1}{\lambda}_a, \overset{2}{\lambda}_a, \overset{3}{\lambda}_a$ be three smooth fields such that $\overset{i}{\sigma}{}^a = h^{ab} \overset{i}{\lambda}_b$ (or, equivalently, $\overset{i}{\sigma}{}^a \overset{j}{\lambda}_a = \delta_{ij}$) for all i and j. Now, for all i, j, and k, let $\overset{ijk}{\omega}$ be the scalar field defined by

$$\overset{ijk}{\omega} = \overset{k}{\sigma}{}^c\, \overset{i}{\sigma}{}^a\, \overset{j}{\sigma}{}^b\, (\nabla_c\, \omega_{ab}) = \overset{k}{\lambda}_c\, \overset{i}{\lambda}_a\, \overset{j}{\lambda}_b\, (\nabla^c \omega^{ab}).$$

Then

$$\nabla^c\, \omega^{ab} = \sum_{i,j,k=1}^{3} \overset{ijk}{\omega}\, \overset{k}{\sigma}{}^c\, \overset{i}{\sigma}{}^a\, \overset{j}{\sigma}{}^b,$$

and, hence,

$$(\nabla_c\, \omega_{ab})(\nabla^c\, \omega^{ab}) = \sum_{i,j,k=1}^{3} \overset{ijk}{\omega}\, (\overset{k}{\sigma}{}^c\, \overset{i}{\sigma}{}^a\, \overset{j}{\sigma}{}^b \nabla_c\, \omega_{ab}) = \sum_{i,j,k=1}^{3} (\overset{ijk}{\omega})^2. \tag{4.5.27}$$

So, clearly, $(\nabla_c\, \omega_{ab})(\nabla^c\, \omega^{ab})$ can vanish only if $\overset{ijk}{\omega} = 0$ for all i, j, and k; i.e., only if $\nabla^c \omega^{ab} = \mathbf{0}$. □

Now we can formulate our notion of asymptotic flatness. It is intended to capture the intuitive claim that "$R^{ab}{}_{cd}$ goes to $\mathbf{0}$ at spatial infinity." (We could certainly impose a restriction on the limiting behavior of $R^a{}_{bcd}$ but, in fact, it suffices for our purposes to work with a weaker condition that is formulated in terms of $R^{ab}{}_{cd}$.) With equivalence (4.5.24) in mind, we shall use the condition $\Psi \to 0$ as a surrogate for the condition $R^{ab}{}_{cd} \to \mathbf{0}$.

We first have to insure that there is an asymptotic regime in which spacetime curvature *can* go (or fail to go) to zero. We do so by restricting attention to classical spacetimes that can be foliated by a family of spacelike hypersurfaces that are simply connected and geodesically complete. Each of these hypersurfaces (together with the metric induced on it by h^{ab}) is then, in effect, a copy of ordinary three-dimensional Euclidean space. Given a classical spacetime (M, t_a, h^{ab}, ∇) satisfying this condition, we say officially that *$R^{ab}{}_{cd}$ goes to $\mathbf{0}$ at spatial infinity* if, for all spacelike geodesics $\gamma : \mathbb{R} \to M$, $\Psi(\gamma(s)) \to 0$ as $s \to \infty$.

Now we can formulate the collapse result (due to Künzle [35] and Ehlers [15]).

PROPOSITION 4.5.4. *Let (M, t_a, h^{ab}, ∇) be a classical spacetime that is spatially flat. Suppose the following conditions hold.*

(1) *For all p in M, there is a spacelike hypersurface containing p that is simply connected and geodesically complete.*
(2) *$R^{ab}{}_{cd}$ goes to $\mathbf{0}$ at spatial infinity (in the sense discussed above).*

Then $R^{ab}{}_{cd} = \mathbf{0}$ (everywhere).

Proof. Arguing as in the proof of proposition 4.5.1, but now using assumption (1), we can show that there exist three smooth, linearly independent, *globally defined* spacelike fields $\overset{1}{\sigma}{}^a, \overset{2}{\sigma}{}^a, \overset{3}{\sigma}{}^a$ satisfying (i) $h^{ab} = \overset{1}{\sigma}{}^a\, \overset{1}{\sigma}{}^b + \overset{2}{\sigma}{}^a\, \overset{2}{\sigma}{}^b + \overset{3}{\sigma}{}^a\, \overset{3}{\sigma}{}^b$

and (ii) $\nabla^a \overset{i}{\sigma}{}^b = \mathbf{0}$, and there exists a smooth *globally defined* future-directed, unit timelike field ξ^a that is rigid ($\nabla^{(a}\,\xi^{b)} = \mathbf{0}$). Let ω_{ab} be the rotation field associated with the latter, and let $\overset{f}{\nabla}$ be its associated flat derivative operator (as constructed in the proof of proposition 4.5.2). Finally, let the scalar component fields $\overset{ijk}{\omega}$ be defined by

$$\overset{ijk}{\omega} = \overset{k}{\sigma}{}^c\,\overset{i}{\sigma}{}^a\,\overset{j}{\sigma}{}^b(\nabla_c\,\omega_{ab})$$

as in the preceding proof. We are assuming that $\Psi = (\nabla_c\,\omega_{ab})(\nabla^c\,\omega^{ab})$ goes to 0 as one approaches spatial infinity. But, by equation (4.5.27), $(\nabla_c\,\omega_{ab})(\nabla^c\,\omega^{ab}) = \sum_{i,j,k=1}^{3}(\overset{ijk}{\omega})^2$. Hence, for all i, j, and k,

(a) $\overset{ijk}{\omega} \to 0$ at spatial infinity.

We claim now that the fields $\overset{ijk}{\omega}$ are all harmonic; i.e.,

(b) $\overset{f}{\nabla}_n\,\overset{f}{\nabla}{}^n\,\overset{ijk}{\omega} = 0$.

(We could equally well take the claim to be $\nabla_n\,\nabla^n\,\overset{ijk}{\omega} = 0$, but it is more convenient to work with the flat derivative operator $\overset{f}{\nabla}$.) Once we show this, we will be done. Because it will then follow by the "minimum principle" that the fields $\overset{ijk}{\omega}$ all vanish.[6] That, in turn, will imply that $\Psi = \sum_{i,j,k=1}^{3}(\overset{ijk}{\omega})^2 = \mathbf{0}$ and, hence, by equation (4.5.24), that $R^{ab}{}_{cd} = \mathbf{0}$.

As in the preceding proof, let $\overset{1}{\lambda}_a, \overset{2}{\lambda}_a, \overset{3}{\lambda}_a$ be three smooth fields such that $\overset{i}{\sigma}{}^a = h^{ab}\overset{i}{\lambda}_b$. Now $\overset{f}{\nabla}{}^a$ and ∇^a agree in their action on contravariant fields that are spacelike in all indices. In particular, for all i,

$$\overset{f}{\nabla}{}^a\overset{i}{\sigma}{}^b = \nabla^a\overset{i}{\sigma}{}^b,$$

$$\overset{f}{\nabla}{}^a\omega^{bc} = \nabla^a\omega^{bc}.$$

6. See, e.g., Flanders [20], p. 85. The principle asserts that a harmonic function defined on a compact set in three-dimensional Euclidean space assumes its minimum value on its boundary. It follows—consider a nested sequence of closed balls with radii going to infinity—that if a harmonic function defined on all of three-dimensional Euclidean space goes to 0 asymptotically along any (or even just one) geodesic, then it must be 0 everywhere.

We here apply the principle to the fields $\overset{ijk}{\omega}$ or, rather, the restrictions of those fields to individual spacelike hypersurfaces that are simply connected and geodesically complete. Note that condition (b) can be construed as a constraint on the restricted fields. If D is the (three-dimensional) derivative operator induced on a spaceike hypersurface by $\overset{f}{\nabla}$—which is the same as the one induced by ∇—then it follows from (b) that $D_n\,D^n\,\overset{ijk}{\omega} = 0$.

(Here $C^a{}_{bc}$ has the form $C^a{}_{bc} = 2\, t_{(b}\, \kappa_{c)}{}^a$. So $C^{ab}{}_c = t_c \kappa^{ba}$ and, hence, $\overset{f}{\nabla}{}^a \overset{i}{\sigma}{}^b = \nabla^a \overset{i}{\sigma}{}^b - C^{ba}{}_m \overset{i}{\sigma}{}^m = \nabla^a \overset{i}{\sigma}{}^b - t_m \kappa^{ab} \overset{i}{\sigma}{}^m = \nabla^a \overset{i}{\sigma}{}^b$. The other case is handled similarly.) It follows that $\overset{f}{\nabla}{}^a \overset{i}{\sigma}{}^b = \mathbf{0}$, and $\overset{ijk}{\omega} = \overset{k}{\lambda}_c \overset{i}{\lambda}_a \overset{j}{\lambda}_b (\overset{f}{\nabla}{}^c \omega^{ab}) = \overset{k}{\sigma}{}^c \overset{i}{\sigma}{}^a \overset{j}{\sigma}{}^b (\overset{f}{\nabla}_c\, \omega_{ab})$. Hence,

$$\begin{aligned}
\overset{f}{\nabla}_n \overset{f}{\nabla}{}^n \overset{ijk}{\omega} &= \overset{f}{\nabla}_n \overset{f}{\nabla}{}^n \,(\overset{k}{\sigma}{}^c \overset{i}{\sigma}{}^a \overset{j}{\sigma}{}^b \overset{f}{\nabla}_c\, \omega_{ab}) \\
&= \overset{f}{\nabla}_n \,(\overset{k}{\sigma}{}^c \overset{i}{\sigma}{}^a \overset{j}{\sigma}{}^b \overset{f}{\nabla}{}^n \overset{f}{\nabla}_c\, \omega_{ab}) \;=\; \overset{f}{\nabla}{}^n \,(\overset{k}{\sigma}{}^c \overset{i}{\sigma}{}^a \overset{j}{\sigma}{}^b \overset{f}{\nabla}_n \overset{f}{\nabla}_c\, \omega_{ab}) \\
&= \overset{k}{\sigma}{}^c \overset{i}{\sigma}{}^a \overset{j}{\sigma}{}^b (\overset{f}{\nabla}{}^n \overset{f}{\nabla}_n \overset{f}{\nabla}_c\, \omega_{ab}) = \overset{k}{\lambda}_c \overset{i}{\lambda}_a \overset{j}{\lambda}_b (\overset{f}{\nabla}_n \overset{f}{\nabla}{}^n \overset{f}{\nabla}{}^c\, \omega^{ab})
\end{aligned}$$

for all i, j, and k. So, to complete the proof, it suffices for us to show

(c) $\overset{f}{\nabla}_n \overset{f}{\nabla}{}^n \overset{f}{\nabla}{}^c\, \omega^{ab} = \mathbf{0}$.

And *this* condition follows easily from the fact that $\overset{f}{\nabla}{}^{[a} \omega^{bc]} = \mathbf{0}$ and $\overset{f}{\nabla}_a\, \omega^{ab} = \mathbf{0}$. (Recall equations (4.5.5) and (4.5.6) in the formulation of proposition 4.5.2.) Since $\overset{f}{\nabla}$ is flat, we can switch derivative operator position and, therefore,

$$\begin{aligned}
\overset{f}{\nabla}_n \overset{f}{\nabla}{}^n \overset{f}{\nabla}{}^c \omega^{ab} &= \overset{f}{\nabla}{}^c \overset{f}{\nabla}_n \overset{f}{\nabla}{}^n \omega^{ab} = \overset{f}{\nabla}{}^c \overset{f}{\nabla}_n (- \overset{f}{\nabla}{}^b \omega^{na} - \overset{f}{\nabla}{}^a \omega^{bn}) \\
&= - \overset{f}{\nabla}{}^c \overset{f}{\nabla}{}^b \,(\overset{f}{\nabla}_n\, \omega^{na}) - \overset{f}{\nabla}{}^c \overset{f}{\nabla}{}^a \,(\overset{f}{\nabla}_n\, \omega^{bn}) = \mathbf{0}.
\end{aligned}$$

So we are done.

□

Solutions to Problems

PROBLEM 1.1.1. *Let* $(M, \mathcal{C})$ *be an n-manifold, let* (U, φ) *be an n-chart in* $\mathcal{C}$*, let* $\widehat{O}$ *be an open subset of* $\varphi[U]$*, and let* O *be its pre-image* $\varphi^{-1}[\widehat{O}]$*. Show that* $(O, \varphi|_O)$ *is also an n-chart in* $\mathcal{C}$.

Let φ' be the restricted map $\varphi|_O$. (We write it this way just to simplify our notation.) Clearly, $\varphi'[O]$ is open, since $\varphi'[O] = \varphi[O] = \widehat{O}$. And φ' is one-to-one (since it is a restriction of φ). So (O, φ') qualifies as an n-chart on M. To show that it belongs to $\mathcal{C}$, we must verify that it is compatible with every n-chart in $\mathcal{C}$.

Let (V, ψ) be one such. We may assume that $U \cap V$ is non-empty, since otherwise the charts are automatically compatible. Since φ' is a restriction of φ, and O is a subset of U (and φ is one-to-one), we have

$$\varphi'[O \cap V] = \varphi[O \cap V] = \varphi[O \cap (U \cap V)] = \varphi[O] \cap \varphi[U \cap V].$$

But $\varphi[O]$ is open (since it is equal to $\widehat{O}$), and $\varphi[U \cap V]$ is open (since the charts (U, φ) and (V, ψ) are compatible). So $\varphi'[O \cap V]$ is open. Furthermore, $\psi[O \cap V]$ is open since it is the pre-image of the open set $\varphi[O \cap V]$ under the smooth (hence continuous) map

$$\varphi \circ \psi^{-1} \colon \psi[U \cap V] \to \varphi[U \cap V].$$

(That the map is smooth follows, again, by the compatibility of the charts (U, φ) and (V, ψ).) Finally, the maps

$$\varphi' \circ \psi^{-1} \colon \psi[O \cap V] \to \varphi'[O \cap V],$$

$$\psi \circ \varphi'^{-1} \colon \varphi'[O \cap V] \to \psi[O \cap V]$$

are smooth since they are the restrictions to open sets, respectively, of the smooth maps

$$\varphi \circ \psi^{-1} : \psi[U \cap V] \to \varphi[U \cap V],$$

$$\psi \circ \varphi^{-1} : \varphi[U \cap V] \to \psi[U \cap V].$$

PROBLEM 1.1.2. *Let $(M, \mathcal{C})$ be an n-manifold, let (U, φ) be an n-chart in $\mathcal{C}$, and let O be an open set in M such that $U \cap O \neq \emptyset$. Show that $\left(U \cap O, \varphi|_{U \cap O}\right)$ is also an n-chart in $\mathcal{C}$.*

We claim, first, that $\varphi[U \cap O]$ is open. To see this, let $\varphi(p)$ be any point in $\varphi[U \cap O]$. Since O is open, there exists an n-chart (V, ψ) in $\mathcal{C}$ where $p \in V \subseteq O$. Since (V, ψ) and (U, φ) are compatible, $\varphi[U \cap V]$ qualifies as an open subset of $\varphi[U \cap O]$ containing $\varphi(p)$. So $\varphi[U \cap O]$ is open, as claimed. It now follows by the result of problem 1.1.1 (taking $\widehat{O} = \varphi[U \cap O]$) that the pair $\left(U \cap O, \varphi|_{U \cap O}\right)$ is an n-chart in $\mathcal{C}$.

PROBLEM 1.1.3. *Let $(M, \mathcal{C})$ be an n-manifold and let $\mathcal{T}$ be the set of open subsets of M. (i) Show that $\mathcal{T}$ is a topology on M; i.e., it contains the empty set and the set M, and is closed under finite intersections and arbitrary unions. (ii) Show that $\mathcal{T}$ is the coarsest topology on M with respect to which $\varphi : U \to \mathbb{R}^n$ is continuous for all n—charts (U, φ) in $\mathcal{C}$.*

(i) The empty set qualifies, vacuously, as open, and M qualifies as open since $(M, \mathcal{C})$ satisfies condition (M2). So we need only show that $\mathcal{T}$ is closed under finite intersections and arbitrary unions. For the first claim, it suffices to show that if O_1 and O_2 are both open, then their intersection $O_1 \cap O_2$ is as well. (The claim will then follow by induction.) So assume that O_1 and O_2 are open, and let p be a point in $O_1 \cap O_2$. (If the intersection is empty, it is automatically open.) Since O_2 is open, there is an n-chart (U, φ) in $\mathcal{C}$ such that $p \in U \subseteq O_2$. Then, by the result in problem 1.1.2, the pair $\left(U \cap O_1, \varphi_{|U \cap O_1}\right)$ is an n-chart in $\mathcal{C}$. Thus, given an arbitrary point p in $O_1 \cap O_2$, there is an n-chart in C (namely, $(U \cap O_1, \varphi_{|U \cap O_1})$) whose domain contains p and is a subset of $O_1 \cap O_2$. It follows that $O_1 \cap O_2$ is open, as claimed. Finally, let S be a set of open sets, and let p be a point in its union $\cup S$. (Again, if the union is empty, it is automatically open.) Let O be a set in S such that $p \in O$. Since O is open, there is an n-chart (U, ϕ) in $\mathcal{C}$ such that $p \in U \subseteq O \subseteq (\cup S)$. So, given our arbitrary point in $\cup S$, there is an n-chart in $\mathcal{C}$ (namely (U, φ)) whose domain contains p and is a subset of $\cup S$. It follows that $\cup S$ is open.

(ii) First, we claim that given any n-chart (U, φ) in $\mathcal{C}$, $\varphi \colon U \to \mathbb{R}^n$ is continuous with respect to $\mathcal{T}$. Let (U, φ) be one such. We need to show that, given any open subset $\widehat{O}$ of $\varphi[U]$, its pre-image $\varphi^{-1}[\widehat{O}]$ is open. But by the result in problem 1.1.1, we know that there is an n-chart in $\mathcal{C}$ whose domain is $\varphi^{-1}[\widehat{O}]$. And the domain of an n-chart in $\mathcal{C}$ is certainly open. So our claim follows easily. Next, assume that $\mathcal{T}'$ is a topology on M with respect to which $\varphi : U \to \mathbb{R}^n$ is continuous for all n-charts (U, φ) in $\mathcal{C}$. We show that $\mathcal{T} \subseteq \mathcal{T}'$. Let O be a set in $\mathcal{T}$, and let p be a point in O. (If O is empty, then it certainly belongs to $\mathcal{T}'$ since the latter is a topology on M.) Since O is open, there is an n-chart (U, φ) in $\mathcal{C}$ such that $p \in U \subseteq O$. By assumption, φ is continuous with respect to $\mathcal{T}'$. And $\varphi[U]$ is an open set in $\mathbb{R}^n$ (by the definition of an n-chart). So its pre-image U must belong to $\mathcal{T}'$. Thus given any point p in O, there is a $\mathcal{T}'$-open set (namely, U) that contains p and is a subset of O. It follows that O itself is open with respect to $\mathcal{T}'$. Thus, as claimed, every set O that belongs to $\mathcal{T}$ belongs to $\mathcal{T}'$ as well.

PROBLEM 1.1.4. *Let $(M, \mathcal{C})$ be an n-manifold. Show that a map $\alpha : M \to \mathbb{R}$ is smooth according to our first definition of "smoothness" (which applies only to real-valued maps on manifolds) iff it is smooth according to our second definition (which applies to maps between arbitrary manifolds).*

α is smooth in the first sense iff for all n-charts (U, φ) in $\mathcal{C}$, the map $\alpha \circ \varphi^{-1} : \varphi[U] \to \mathbb{R}$ is smooth. It is smooth in the second sense iff for all smooth maps $\beta : \mathbb{R} \to \mathbb{R}$, the composed map $\beta \circ \alpha : M \to \mathbb{R}$ is smooth in the first sense (i.e., $(\beta \circ \alpha) \circ \varphi^{-1} : \varphi[U] \to \mathbb{R}$ is smooth for all n-charts (U, φ) in $\mathcal{C}$). To see that the second sense implies the first, we need only consider the special case where β is the identity map on $\mathbb{R}$. For the converse, suppose that α is smooth in the first sense, let $\beta : \mathbb{R} \to \mathbb{R}$ be any smooth map on $\mathbb{R}$, and let (U, φ) be any n-chart in $\mathcal{C}$. Then $(\beta \circ \alpha) \circ \varphi^{-1} : \varphi[U] \to \mathbb{R}$ is smooth since $(\beta \circ \alpha) \circ \varphi^{-1} = \beta \circ (\alpha \circ \varphi^{-1})$; i.e., it is composition of smooth maps $\alpha \circ \varphi^{-1} : \varphi[U] \to \mathbb{R}$ and $\beta : \mathbb{R} \to \mathbb{R}$.

In what follows, let $(M, \mathcal{C})$ be an n-manifold, let p be a point in M, and let $\mathcal{C}(p)$ be the set of charts in $\mathcal{C}$ whose domains contain p.

PROBLEM 1.2.1. *Let ξ be a non-zero vector at p, and let $(k^1, \ldots, k^n)$ be a non-zero element of $\mathbb{R}^n$. Show there exists an n-chart in $\mathcal{C}(p)$ with respect to which ξ has components $(k^1, \ldots, k^n)$.*

Let (U_1, φ_1) be an n-chart in $\mathcal{C}(p)$, and let $(\xi^1, \ldots, \xi^n)$ be the components of ξ with respect to (U_1, φ_1). These components cannot all be 0, since ξ is not the zero vector. So there is an isomorphism L of (the vector space) $\mathbb{R}^n$ onto itself that takes $(\xi^1, \ldots, \xi^n)$ to $(k^1, \ldots, k^n)$. Let its associated matrix have elements $\{a_{ij}\}$. Then, for all $i = 1, \ldots, n$, $k^i = \sum_{j=1}^{n} a_{ij}\, \xi^j$.

Now consider a new n-chart (U_2, φ_2) in $\mathcal{C}(p)$ where $U_2 = U_1$ and $\varphi_2 = L \circ \varphi_1$: $U_2 \to \mathbb{R}^n$. (That it *is* an n-chart and *does* belong to $\mathcal{C}$ must be checked. But these claims follow easily from the fact that L, now construed as a map from the manifold $\mathbb{R}^n$ to itself, is a diffeomorphism.) We claim that the components of ξ with respect to (U_2, φ_2) are $(k^1, \ldots, k^n)$. To see this, we invoke proposition 1.2.5. As in the notes, for all $i = 1, \ldots, n$, let $x'^i : \varphi_1[U_1 \cap U_2] \to \mathbb{R}$ be the coordinate map defined by $x'^i = x^i \circ \varphi_2 \circ \varphi_1^{-1}$. Since $\varphi_2 = L \circ \varphi_1$, we have

$$x^i \circ \varphi_2 = \sum_{j=1}^{n} a_{ij}\, (x^j \circ \varphi_1)$$

and, therefore,

$$x'^i = x^i \circ \varphi_2 \circ \varphi_1^{-1} = \sum_{j=1}^{n} a_{ij}\, x^j.$$

It now follows by proposition 1.2.5 that the components of ξ with respect to (U_2, φ_2) are

$$\xi'^i = \sum_{j=1}^{n} \xi^j\, \frac{\partial x'^i}{\partial x^j}(\varphi_1(p)) = \sum_{j=1}^{n} \xi^j\, a_{ij} = k^i,$$

for all i.

PROBLEM 1.3.1. *Let ξ be the vector field $x^1 \dfrac{\partial}{\partial x^1} - x^2 \dfrac{\partial}{\partial x^2}$ on $\mathbb{R}^2$. Show that the maximal integral curve of ξ with initial value $p = (p^1, p^2)$ is the map $\gamma : \mathbb{R} \to \mathbb{R}^2$ with $\gamma(s) = (p^1\, e^s, p^2\, e^{-s})$.*

γ has initial value (p^1, p^2). It is an integral curve of the given vector field since, for all $s \in \mathbb{R}$, and all $f \in \mathcal{S}(\gamma(s))$, by the chain rule,

$$\begin{aligned}
\vec{\gamma}_{\gamma(s)}(f) &= \frac{d}{ds}(f \circ \gamma)(s) = \frac{d}{ds}\big(f\big(p^1\, e^s, p^2\, e^{-s}\big)\big) \\
&= \frac{\partial f}{\partial x^1}(\gamma(s))\, (p^1\, e^s) + \frac{\partial f}{\partial x^2}(\gamma(s))\, (-p^2\, e^{-s}) \\
&= \frac{\partial f}{\partial x^1}(\gamma(s))\, x^1(\gamma(s)) - \frac{\partial f}{\partial x^2}(\gamma(s))\, x^2(\gamma(s)) \\
&= \left[x^1 \frac{\partial}{\partial x^1} - x^2 \frac{\partial}{\partial x^2}\right]_{|\gamma(s)} (f).
\end{aligned}$$

Finally, it is maximal because its domain is $\mathbb{R}$.

PROBLEM 1.3.2. *Let ξ be a smooth vector field on M, let p be a point in M, and let s_0 be any real number (not necessarily 0). Show that there is an integral curve $\gamma: I \to M$ of ξ with $\gamma(s_0) = p$ that is maximal in the sense that given any integral curve $\gamma': I' \to M$ of ξ, if $\gamma'(s_0) = p$, then $I' \subseteq I$ and $\gamma'(s) = \gamma(s)$ for all s in I'.*

Given an interval J, let us understand $J + a$ to be the translation of J by the number a. Let $\sigma: J \to M$ be the maximal integral curve of ξ with initial value p. (Existence is guaranteed by proposition 1.3.1.) Let I be the shifted interval $J + s_0$, and let $\gamma: I \to M$ be the curve defined by $\gamma(s) = \sigma(s - s_0)$. Then γ is an integral curve of ξ by the first clause of proposition 1.3.2, and $\gamma(s_0) = \sigma(0) = p$. We claim that γ satisfies the stated maximality condition.

To see this, suppose $\gamma': I' \to M$ is an integral curve of ξ, and $\gamma'(s_0) = p$. Let $J' = I' - s_0$ and let $\sigma': J' \to M$ be defined by $\sigma'(s) = \gamma'(s + s_0)$. Then σ' is an integral curve of ξ (by the first clause of proposition 1.3.2 again) with initial value 0 (since $\sigma'(0) = \gamma'(s_0) = p$). So, by the maximality of σ, $J' \subseteq J$ and $\sigma'(s) = \sigma(s)$ for all s in J'. It follows immediately that $I' = J' + s_0 \subseteq J + s_0 = I$ and $\gamma'(s) = \sigma'(s - s_0) = \sigma(s - s_0) = \gamma(s)$ for all s in I'.

PROBLEM 1.3.3. *(Integral curves that go nowhere) Let ξ be a smooth vector field on M, and let $\gamma: I \to M$ be an integral curve of ξ. Suppose that ξ vanishes (i.e., assigns the zero vector) at some point $p \in \gamma[I]$. Then the following both hold.*

(1) $\gamma(s) = p$ for all s in I (i.e., γ is a constant curve).
(2) The reparametrized curve $\gamma' = \gamma \circ \alpha: I' \to M$ is an integral curve of ξ for all diffeomorphisms $\alpha: I' \to I$.

(1) Suppose $s_0 \in I$ and $\gamma(s_0) = p$. It follows from problem 1.3.2 that there is a unique maximal integral curve of ξ whose value at s_0 is p. The only possibility is the constant curve $\hat{\gamma}: \mathbb{R} \to M$ that assigns p to all s. ($\hat{\gamma}$ is an integral curve of ξ since, for all $f \in \mathcal{S}(p)$, $f \circ \hat{\gamma}$ is constant and, so,

$$\vec{\hat{\gamma}}_{|\hat{\gamma}(s)}(f) = \frac{d}{ds}(f \circ \hat{\gamma})(s) = 0 = \xi_{|\hat{\gamma}(s)}(f)$$

for all s. It is maximal since its domain is $\mathbb{R}$.) Hence, by maximality, $\gamma(s) = \hat{\gamma}(s) = p$ for all s in I.

(2) Let $\alpha: I' \to I$ be a diffeomorphism and let γ' be the composed map $\gamma' = \gamma \circ \alpha: I' \to M$. We know equation from equation (1.3.3) that γ' is an integral curve of ξ iff

$$\xi\big(\gamma(\alpha(s))\big)\,\frac{d\alpha}{ds}(s) \;=\; \xi\big(\gamma(\alpha(s))\big)$$

for all s in I'. But $\gamma(\alpha(s)) = p$ for all s in I' (by the first part of the problem) and, therefore, $\xi\big(\gamma(\alpha(s))\big) = \xi(p) = 0$ for all s in I'. So the required equation holds for all s in I'. (Both sides are 0.)

PROBLEM 1.3.4. *(Integral curves cannot cross) Let $\gamma : I \to M$ and $\gamma' : I' \to M$ be integral curves of ξ that are maximal (in the sense of problem 1.3.2) and satisfy $\gamma(s_0) = \gamma'(s'_0)$. Then the two curves agree up to a parameter shift: $\gamma(s) = \gamma'\big(s + (s'_0 - s_0)\big)$ for all s in I.*

Let $I'' = I' - (s'_0 - s_0)$, and let $\gamma'' : I'' \to M$ be the curve defined by

$$\gamma''(s) = \gamma'\big(s + (s'_0 - s_0)\big).$$

It is an integral curve of ξ by proposition 1.3.2, and $\gamma''(s_0) = \gamma'(s'_0) = \gamma(s_0)$. So by the maximality of γ, $I'' \subseteq I$ and $\gamma''(s) = \gamma(s)$ for all s in I''; i.e., $\gamma(s) = \gamma'(s + (s'_0 - s_0))$ for all s in I''. It remains to verify only that $I'' = I$. Since $I'' \subseteq I$, it follows that $I' \subseteq I + (s'_0 - s_0)$. If we rerun the argument with the roles of I, γ, and s_0 interchanged with those of I', γ', and s'_0, we arrive at the symmetric conclusion that $I \subseteq I' + (s_0 - s'_0)$. Putting the two set inclusions together, we arrive at $I \subseteq I' + (s_0 - s'_0) \subseteq I + (s'_0 - s_0) + (s_0 - s'_0) = I$. So $I'' = I' + (s_0 - s'_0) = I$, as claimed.

PROBLEM 1.3.5. *Let ξ be a smooth vector field on M that is complete. Let p be a point in M. Show that the restriction of ξ to the punctured set $M - \{p\}$ is complete (as a field on $M - \{p\}$) iff ξ vanishes at p.*

Let ξ' be the restriction of ξ to $M - \{p\}$. Suppose first that ξ vanishes at p. Then, as we know from problem 1.3.3, every integral curve of ξ that passes through p is necessarily a degenerate constant curve that sits at p. It follows, we claim, that ξ' is complete. For let q be any point in M distinct from p. Since ξ is complete (as a field on M), there is an integral curve $\gamma : \mathbb{R} \to M$ of ξ with initial value q. The image of γ is fully contained in $M - \{p\}$ (since otherwise γ would be an integral curve of ξ passing through p that does not sit at p). So γ qualifies as an integral curve of ξ'. Since the domain of γ is $\mathbb{R}$ (and since q was chosen arbitrarily), we see that ξ' is complete, as claimed.

Conversely, suppose ξ does not vanish at p. Since ξ is complete (as a field on M), there is an integral curve $\gamma : \mathbb{R} \to M$ of ξ with initial value p. γ cannot be a constant curve that sits at p. (Otherwise, we would have $\vec{\gamma}_p = \mathbf{0}$ and, hence, $\xi(p) = \mathbf{0}$.) So the set $D = \{s \in \mathbb{R} : \gamma(s) \neq p\}$ is non-empty. It is a disjoint union of open intervals. (If 0 is the only number s in $\mathbb{R}$ such that $\gamma(s) = p$, then D

will be the union of $(-\infty, 0)$ and $(0, \infty)$. Other possibilities arise because γ may pass through p more than once.) Let I' be any one of these intervals, let $\gamma' \colon I' \to M$ be the restriction of γ to I', and let q be any point in $\gamma'[I']$. Then γ' qualifies as a maximal integral curve of ξ' in $M - \{p\}$ that passes through q. By shifting initial values, we can generate a maximal integral curve γ'' of ξ' in $M - \{p\}$ that has initial value q. But the domain of γ'' is not $\mathbb{R}$ (since the pre-shifted domain I' of γ' is not $\mathbb{R}$). So we may conclude that ξ' is not complete.

PROBLEM 1.4.1. *Show that lemma 1.4.1 can also be derived as a corollary to the following fact about square matrices: if M is an $(r \times r)$ matrix $(r \geq 1)$ and M^2 is the zero matrix, then the trace of M is* 0.

Assume the left-side condition $\sum_{k=1}^{r} \overset{k}{\varphi}{}^{a} \overset{k}{\psi}{}_{c} = \mathbf{0}$ holds, and let M be the $r \times r$ matrix with entries $M_{ij} = \overset{i}{\varphi}{}^{a} \overset{j}{\psi}{}_{a}$. Then M^2 is the zero matrix since

$$(M^2)_{ij} = \sum_{k=1}^{r} M_{ik} M_{kj} = \sum_{k=1}^{r} (\overset{i}{\varphi}{}^{a} \overset{k}{\psi}{}_{a})(\overset{k}{\varphi}{}^{b} \overset{j}{\psi}{}_{b}) = (\overset{i}{\varphi}{}^{a} \overset{j}{\psi}{}_{b}) \sum_{k=1}^{r} (\overset{k}{\varphi}{}^{b} \overset{k}{\psi}{}_{a}) = 0.$$

So, by the stated fact, $0 = tr(M) = \sum_{k=1}^{r} M_{kk} = \sum_{k=1}^{r} \overset{k}{\varphi}{}^{a} \overset{k}{\psi}{}_{a}$.

PROBLEM 1.6.1. *Show that for all smooth vector fields ξ^a on M, $\mathcal{L}_\xi \, \delta^b_a = \mathbf{0}$.*

For all smooth vector fields λ^a on M, we have

$$\lambda^a \mathcal{L}_\xi \, \delta^b_a = \mathcal{L}_\xi \, (\delta^b_a \, \lambda^a) - \delta^b_a \, \mathcal{L}_\xi \, \lambda^a = \mathcal{L}_\xi \, \lambda^b - \mathcal{L}_\xi \, \lambda^b = \mathbf{0}.$$

(The first equality follows from the Leibniz rule, and the second from the fact that δ^b_a functions as an index substitution operator.) Since this holds for all smooth fields λ^a (at all points in M), we may conclude that $\mathcal{L}_\xi \, \delta^b_a = \mathbf{0}$.

Here is a second argument. By the Leibniz rule, and the fact that δ^b_a functions as an index substitution operator, we have

$$\mathcal{L}_\xi \, \delta^b_a = \mathcal{L}_\xi \, (\delta^b_c \, \delta^c_a) = \delta^b_c \, \mathcal{L}_\xi \, \delta^c_a + \delta^c_a \, \mathcal{L}_\xi \, \delta^b_c = \mathcal{L}_\xi \, \delta^b_a + \mathcal{L}_\xi \, \delta^b_a.$$

It follows immediately that $\mathcal{L}_\xi \, \delta^b_a = \mathbf{0}$.

PROBLEM 1.6.2. *Let ξ^a and η^a be smooth vector fields on M, and let the latter be non-vanishing. Show that if $\mathcal{L}_\xi (\eta^a \eta^b) = \mathbf{0}$, then $\mathcal{L}_\xi \, \eta^a = \mathbf{0}$.*

Assume that $\mathcal{L}_\xi(\eta^a\eta^b) = \mathbf{0}$, and let p be any point in M. Since η^a is non-vanishing, there exists a smooth field λ_a on M such that the scalar field $\eta^a\lambda_a$ is non-zero at p. At all points we have

$$0 = \lambda_a\lambda_b\mathcal{L}_\xi\,(\eta^a\eta^b) = \lambda_a\lambda_b(\eta^a\mathcal{L}_\xi\,\eta^b + \eta^b\mathcal{L}_\xi\,\eta^a) = 2(\lambda_a\,\eta^a)\lambda_b\mathcal{L}_\xi\,\eta^b.$$

Hence, $\lambda_b\mathcal{L}_\xi\,\eta^b = 0$ at p. But we also have

$$\mathbf{0} = \lambda_b\mathcal{L}_\xi\,(\eta^a\eta^b) = \lambda_b(\eta^a\mathcal{L}_\xi\,\eta^b + \eta^b\mathcal{L}_\xi\,\eta^a) = \eta^a\lambda_b\mathcal{L}_\xi\,\eta^b + (\lambda_b\,\eta^b)\mathcal{L}_\xi\,\eta^a$$

at all points. So $(\lambda_b\,\eta^b)\mathcal{L}_\xi\,\eta^a = \mathbf{0}$ at p and, therefore, $\mathcal{L}_\xi\,\eta^a = \mathbf{0}$ at p. Since p is an arbitrary point in M, we are done.

PROBLEM 1.6.3. *Show that the set of smooth contravariant vector fields on M forms a Lie algebra under the bracket operation; i.e., show that for all smooth vector fields ξ, η, λ on M,*

$$[\xi,\eta] = -[\eta,\xi] \quad \text{and} \quad \big[\lambda,[\xi,\eta]\big] + \big[\eta,[\lambda,\xi]\big] + \big[\xi,[\eta,\lambda]\big] = \mathbf{0}.$$

The anti-symmetry of the bracket operation is immediate. We can establish the second condition with a straightforward computation. Let ξ, η, λ be smooth contravariant vector fields on M, and let α be a smooth scalar field on M. Then

$$\begin{aligned}\big[\lambda,[\xi,\eta]\big](\alpha) &= \lambda\,\big([\xi,\eta](\alpha)\big) - [\xi,\eta](\lambda(\alpha))\\ &= \big[\lambda(\xi(\eta(\alpha))) - \lambda(\eta(\xi(\alpha)))\big] - \big[\xi(\eta(\lambda(\alpha))) - \eta(\xi(\lambda(\alpha)))\big].\end{aligned}$$

Similarly,

$$\begin{aligned}\big[\eta,[\lambda,\xi]\big](\alpha) &= \big[\eta(\lambda(\xi(\alpha))) - \eta(\xi(\lambda(\alpha)))\big] - \big[\lambda(\xi(\eta(\alpha))) - \xi(\lambda(\eta(\alpha)))\big]\\ \big[\xi,[\eta,\lambda]\big](\alpha) &= \big[\xi(\eta(\lambda(\alpha))) - \xi(\lambda(\eta(\alpha)))\big] - \big[\eta(\lambda(\xi(\alpha))) - \lambda(\eta(\xi(\alpha)))\big].\end{aligned}$$

When we add the three lines, we get $\mathbf{0}$ on the right side because each term has a mate with the opposite sign. Since this holds for all smooth scalar fields α on M, we have our second claim.

PROBLEM 1.6.4. *Show that for all smooth vector fields ξ^a, η^a on M, and all smooth scalar fields α on M,*

$$\mathcal{L}_{(\alpha\,\xi)}\,\eta^a = \alpha\big(\mathcal{L}_\xi\,\eta^a\big) - \big(\mathcal{L}_\eta\,\alpha\big)\xi^a.$$

Given any smooth scalar field β on M, we have

$$\begin{aligned}(\pounds_{(\alpha\,\xi)}\,\eta^a)(\beta) &= (\alpha\,\xi)(\eta(\beta)) - \eta\,((\alpha\,\xi)(\beta))\\ &= \alpha\,(\xi(\eta(\beta))) + [\alpha\,\eta(\xi(\beta)) - \eta(\alpha)\,\xi(\beta)]\\ &= \alpha\,[\xi(\eta(\beta)) - \eta(\xi(\beta))] - \eta(\alpha)\,\xi(\beta)\\ &= \alpha\,(\pounds_\xi\,\eta^a)(\beta) - (\pounds_\eta\,\alpha)\,\xi(\beta)\\ &= [\alpha\,(\pounds_\xi\,\eta^a) - (\pounds_\eta\,\alpha)\,\xi^a](\beta).\end{aligned}$$

Since this is true for all smooth scalar fields β, it follows that $\pounds_{(\alpha\,\xi)}\,\eta^a = \alpha(\pounds_\xi\,\eta^a) - (\pounds_\eta\,\alpha)\xi^a$.

PROBLEM 1.6.5. *One might be tempted to take a smooth tensor field to be "constant" if its Lie derivatives with respect to all smooth vector fields are zero. But this idea does not work. Any contravariant vector field that is constant in this sense would have to vanish everywhere. Prove this.*

Let η^a be a smooth vector field on M. Assume that $\pounds_\xi\,\eta^a = \mathbf{0}$ for all smooth vector fields ξ^a on M. Then, given any smooth scalar field α on M, it follows from the preceding problem that

$$\mathbf{0} = \pounds_{(\alpha\,\xi)}\,\eta^a = \alpha(\pounds_\xi\,\eta^a) - (\pounds_\eta\,\alpha)\xi^a = -(\pounds_\eta\,\alpha)\xi^a.$$

Since this is true for all smooth vector fields ξ^a on M, $\pounds_\eta\,\alpha = \mathbf{0}$. Equivalently, $\eta(\alpha) = \mathbf{0}$. But this is true for all smooth scalar fields α on M. So $\eta^a = \mathbf{0}$.

PROBLEM 1.6.6. *Show that for all smooth vector fields ξ^a, η^a on M, and all smooth tensor fields $\alpha^{a\ldots b}_{c\ldots d}$ on M,*

$$(\pounds_\xi\,\pounds_\eta - \pounds_\eta\,\pounds_\xi)\alpha^{a_1\ldots a_r}_{b_1\ldots b_s} = \pounds_\theta\,\alpha^{a_1\ldots a_r}_{b_1\ldots b_s},$$

where θ^a is the field $\pounds_\xi\,\eta^a$.

Consider first the case of a smooth scalar field α on M. The assertion follows since

$$(\pounds_\xi\,\pounds_\eta - \pounds_\eta\,\pounds_\xi)(\alpha) = \xi(\eta(\alpha)) - \eta(\xi(\alpha)) = (\pounds_\xi\,\eta)(\alpha) = \pounds_\theta\,\alpha.$$

Next consider the case of a smooth vector field α^a on M. Given any smooth scalar field β on M, we have

$$\begin{aligned}[(\pounds_\xi\,\pounds_\eta - \pounds_\eta\,\pounds_\xi)\alpha^a](\beta) &= (\pounds_\xi\,\pounds_\eta\,\alpha^a)(\beta) - (\pounds_\eta\,\pounds_\xi\,\alpha^a)(\beta)\\ &= [\xi,\,[\eta,\,\alpha]](\beta) - [\eta,\,[\xi,\,\alpha]](\beta)\\ &= -[\alpha,\,[\xi,\,\eta]](\beta)\\ &= [[\xi,\,\eta],\,\alpha](\beta) = (\pounds_\theta\,\alpha^a)(\beta).\end{aligned}$$

(Note that the third and fourth equalities follow from the assertions in problem 1.6.3.) Since this is true for all smooth scalar fields β, $(\mathcal{L}_\xi\, \mathcal{L}_\eta - \mathcal{L}_\eta\, \mathcal{L}_\xi)\, \alpha^a = \mathcal{L}_\theta\, \alpha^a$. The other cases now follow in standard computational sequence. To compute $(\mathcal{L}_\xi\, \mathcal{L}_\eta - \mathcal{L}_\eta\, \mathcal{L}_\xi)\, \alpha_b$, we consider an arbitrary smooth field λ^b and make use of our previous derived expressions for $(\mathcal{L}_\xi\, \mathcal{L}_\eta - \mathcal{L}_\eta\, \mathcal{L}_\xi)\,(\alpha_b\, \lambda^b)$ and $(\mathcal{L}_\xi\, \mathcal{L}_\eta - \mathcal{L}_\eta\, \mathcal{L}_\xi)\, \lambda^b$. And so forth.

PROBLEM 1.7.1. *Let ∇ be a derivative operator on a manifold. Show that $\nabla_n\, \delta^b_a = \boldsymbol{0}$.*

We can use much the same argument here as used for problem 1.6.1. By the Leibniz rule, and the fact that δ^b_a functions as an index substitution operator,

$$\nabla_n\, \delta^b_a = \nabla_n\, (\delta^b_c\, \delta^c_a) = \delta^b_c\, \nabla_n\, \delta^c_a + \delta^c_a\, \nabla_n\, \delta^b_c = 2\, \nabla_n\, \delta^b_a.$$

So $\nabla_n\, \delta^b_a = \mathbf{0}$.

PROBLEM 1.7.2. *Let ∇ and ∇' be derivative operators on a manifold, and let $\alpha_{a_1 \ldots a_n}$ be a smooth n-form on it. Show that*

$$\nabla_{[b}\, \alpha_{a_1 \ldots a_n]} = \nabla'_{[b}\, \alpha_{a_1 \ldots a_n]}.$$

There is a smooth, symmetric field C^a_{bc} on the manifold such that $\nabla' = (\nabla,\ C^a_{bc})$. For any smooth n-form $\alpha_{a_1 \ldots a_n}$ on M, we have

$$\nabla'_b\ \alpha_{a_1 \ldots a_n} = \nabla_b\, \alpha_{a_1 \ldots a_n} + \alpha_{r\, a_2 \ldots a_n}\, C^r_{b\, a_1} + \cdots + \alpha_{a_1 \ldots a_{n-1}\, r}\, C^r_{b\, a_n}.$$

So, anti-symmetrizing,

$$\nabla'_{[b}\, \alpha_{a_1 \ldots a_n]} = \nabla_{[b}\, \alpha_{a_1 \ldots a_n]} + \alpha_{r\, [a_2 \ldots a_n}\, C^r_{b\, a_1]} + \cdots + \alpha_{[a_1 \ldots a_{n-1}\, |r|}\, C^r_{b\, a_n]}.$$

Since $C^a_{[b\,c]} = \mathbf{0}$, all terms involving $C^a_{b\,c}$ in the sum on the right-hand side are $\mathbf{0}$. (Notice, for example, that $\alpha_{r\,[a_2 \ldots a_n}\, C^r_{b\, a_1]} = \alpha_{r\,[a_2 \ldots a_n}\, C^r_{[b\, a_1]]} = \mathbf{0}$.) It follows that

$$\nabla_{[b}\, \alpha_{a_1 \ldots a_n]} = \nabla'_{[b}\, \alpha_{a_1 \ldots a_n]}.$$

PROBLEM 1.7.3. *Let ∇ be the coordinate derivative operator canonically associated with $(U,\ \varphi)$ on the n-manifold M. Let u^i be the coordinate maps on U determined by the chart. Further, let ∇' be another derivative operator on U. We know (from proposition 1.7.3) that there is a smooth field C^a_{bc} on U such that $\nabla' = (\nabla, C^a_{bc})$.*

Show that if

$$C^a_{bc} \;=\; \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} \overset{ijk}{C}\left(\frac{\partial}{\partial u^i}\right)^{a} (d_b u^j)\,(d_c u^k),$$

then a smooth vector field $\xi^a = \sum\limits_{i=1}^{n} \overset{i}{\xi}\left(\dfrac{\partial}{\partial u^i}\right)^{a}$ *on* U *is constant with respect to* ∇' *(i.e.,* $\nabla'_a \xi^b = \mathbf{0}$*) iff*

$$\frac{\partial \overset{i}{\xi}}{\partial u^j} = \sum_{k=1}^{n} \overset{ijk}{C}\,\overset{k}{\xi}$$

for all i *and* j.

We have

$$\begin{aligned}
\nabla'_b \xi^a &= \nabla_b \xi^a - C^a_{bc}\,\xi^c \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial \overset{i}{\xi}}{\partial u^j}\left(\frac{\partial}{\partial u^i}\right)^{a}(d_b u^j) \\
&\quad - \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} \overset{ijk}{C}\,\overset{l}{\xi}\left(\frac{\partial}{\partial u^i}\right)^{a}(d_b u^j)(d_c u^k)\left(\frac{\partial}{\partial u^l}\right)^{c}.
\end{aligned}$$

But $(d_c u^k)\left(\dfrac{\partial}{\partial u^l}\right)^{c} = \delta_{kl}$. So, continuing,

$$\begin{aligned}
\nabla'_b\, \xi^a &= \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial \overset{i}{\xi}}{\partial u^j}\left(\frac{\partial}{\partial u^i}\right)^{a}(d_b u^j) - \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} \overset{ijk}{C}\,\overset{k}{\xi}\left(\frac{\partial}{\partial u^i}\right)^{a}(d_b u^j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\left[\frac{\partial \overset{i}{\xi}}{\partial u^j} - \sum_{k=1}^{n} \overset{ijk}{C}\,\overset{k}{\xi}\right]\left(\frac{\partial}{\partial u^i}\right)^{a}(d_b u^j).
\end{aligned}$$

Thus $\nabla'_b\, \xi^a = \mathbf{0}$ iff every coefficient (in brackets) in the sum on the right side is 0—i.e., iff

$$\frac{\partial \overset{i}{\xi}}{\partial u^j} - \sum_{k=1}^{n} \overset{ijk}{C}\,\overset{k}{\xi} \;=\; 0$$

for all i and j in $\{1, \ldots, n\}$.

PROBLEM 1.8.1. *Let* ∇ *and* ∇' *be derivative operators on a manifold with* $\nabla'_m = (\nabla_m, C^a_{bc})$*, and let their respective curvature fields be* R^a_{bcd} *and* R'^a_{bcd}*. Show that*

$$R'^a_{bcd} = R^a_{bcd} + 2\nabla_{[c}\, C^a_{d]b} + 2C^n_{b[c}\, C^a_{d]n}.$$

Given any smooth field α_b,

$$\begin{aligned}\nabla'_c\, \nabla'_d\, \alpha_b &= \nabla'_c\, (\nabla_d\, \alpha_b + \alpha_a\, C^a_{db})\\ &= \nabla_c\, (\nabla_d\, \alpha_b + \alpha_a\, C^a_{db}) + (\nabla_p\, \alpha_b + \alpha_a\, C^a_{pb}) C^p_{cd} + (\nabla_d\, \alpha_p + \alpha_a\, C^a_{dp}) C^p_{cb}.\end{aligned}$$

Expanding the first term, anti-symmetrizing on the indices c and d, and using the fact that $C^p_{[cd]} = 0$, we arrive at

$$\begin{aligned}\frac{1}{2}\, {R'}^a{}_{bcd}\, \alpha_a = \frac{1}{2}\, R^a_{bcd}\, \alpha_a &+ (\nabla_{[c}\, \alpha_{|a|})\, C^a_{d]b} + \alpha_a\, \nabla_{[c}\, C^a_{d]b}\\ &+ (\nabla_{[d}\, \alpha_{|p|})\, C^p_{c]b} + \alpha_a\, C^p_{b[c}\, C^a_{d]p}.\end{aligned}$$

The second and fourth terms on the right-hand side differ only in their respective indices of contraction and the order in which the indices c and d occur. So their sum is 0. Hence,

$$\frac{1}{2}\, {R'}^a{}_{bcd}\, \alpha_a = \frac{1}{2}\, R^a_{bcd}\, \alpha_a + \alpha_a\, \nabla_{[c}\, C^a_{d]b} + \alpha_a\, C^p_{b[c}\, C^a_{d]p}.$$

But this holds for all smooth fields α_a. So our conclusion follows.

PROBLEM 1.8.2. *Show that the exterior derivative operator d on any manifold satisfies $d^2 = \mathbf{0}$; i.e., $d_n(d_m\, \alpha_{b_1\ldots b_p}) = \mathbf{0}$ for all smooth p-forms $\alpha_{b_1\ldots b_p}$.*

Here we use the fact that $\lambda_{[a\ldots[b\ldots c]\ldots d]} = \lambda_{[a\ldots b\ldots c\ldots d]}$ for all tensor fields $\lambda_{a\ldots b\ldots c\ldots d}$. It follows from this that

$$\begin{aligned}d_n(d_m\, \alpha_{b_1\ldots b_p}) = \nabla_{[n}\, \nabla_{[m}\, \alpha_{b_1\ldots b_p]]} &= \nabla_{[n}\, \nabla_m\, \alpha_{b_1\ldots b_p]} = \nabla_{[[n}\, \nabla_{m]}\, \alpha_{b_1\ldots b_p]}\\ &= \frac{1}{2}\big[\alpha_{r\,[b_2\ldots b_p}\, R^r_{b_1\, n\, m]} + \cdots + \alpha_{[b_1\ldots b_{p-1}|r|}\, R^r_{b_p\, n\, m]}\big]\\ &= \frac{1}{2}\big[\alpha_{r\,[b_2\ldots b_p}\, R^r_{[b_1\, n\, m]]} + \cdots + \alpha_{[b_1\ldots b_{p-1}|r|}\, R^r_{[b_p\, n\, m]]}\big].\end{aligned}$$

Since $R^a_{[bcd]} = \mathbf{0}$, each of the terms in the final sum is $\mathbf{0}$. So we are done.

PROBLEM 1.8.3. *Show that given any smooth field ξ^a, and any derivative operator ∇ on a manifold, $\mathcal{L}_\xi$ commutes with ∇ (in its action on any tensor field) iff $\nabla_a \nabla_b\, \xi^m = R^m_{bna}\, \xi^n$.*

Let $K^m_{ab} = R^m_{bna}\, \xi^n - \nabla_a \nabla_b\, \xi^m$. We claim that for all smooth fields $\alpha^{a_1\ldots a_r}_{b_1\ldots b_s}$,

$$\begin{aligned}(\mathcal{L}_\xi\, \nabla_n - \nabla_n\, \mathcal{L}_\xi)\, \alpha^{a_1\ldots a_r}_{b_1\ldots b_s} = \alpha^{a_1\ldots a_r}_{m\, b_2\ldots b_s}\, K^m_{n\, b_1} &+ \cdots + \alpha^{a_1\ldots a_r}_{b_1\ldots b_{s-1}\, m}\, K^m_{n\, b_s}\\ &- \alpha^{m\, a_2\ldots a_r}_{b_1\ldots b_s}\, K^{a_1}_{n\, m} - \cdots - \alpha^{a_1\ldots a_{r-1}\, m}_{b_1\ldots b_s}\, K^{a_r}_{n\, m}.\end{aligned}$$

Consider first the case of a scalar field α. By proposition 1.6.4 (and the fact that $\nabla_{[m}\nabla_{n]}\alpha = \mathbf{0}$),

$$(\mathcal{L}_\xi \nabla_n - \nabla_n \mathcal{L}_\xi)\alpha = (\xi^m \nabla_m \nabla_n \alpha + (\nabla_m \alpha)\nabla_n \xi^m) - \nabla_n(\xi^m \nabla_m \alpha)$$

$$= \xi^m \nabla_m \nabla_n \alpha + (\nabla_m \alpha)(\nabla_n \xi^m) - (\nabla_n \xi^m)\nabla_m \alpha - \xi^m \nabla_n \nabla_m \alpha = \mathbf{0}.$$

Similarly, in the case of a smooth vector field α^a, we have

$$\begin{aligned}
&(\mathcal{L}_\xi \nabla_n - \nabla_n \mathcal{L}_\xi)\alpha^a \\
&\quad = [\xi^m \nabla_m \nabla_n \alpha^a + (\nabla_m \alpha^a)(\nabla_n \xi^m) - (\nabla_n \alpha^m)(\nabla_m \xi^a)] \\
&\qquad - \nabla_n(\xi^m \nabla_m \alpha^a - \alpha^m \nabla_m \xi^a) \\
&\quad = [\xi^m \nabla_m \nabla_n \alpha^a + (\nabla_m \alpha^a)(\nabla_n \xi^m) - (\nabla_n \alpha^m)(\nabla_m \xi^a)] \\
&\qquad - [\xi^m \nabla_n \nabla_m \alpha^a + (\nabla_n \xi^m)(\nabla_m \alpha^a) - \alpha^m \nabla_n \nabla_m \xi^a - (\nabla_n \alpha^m)(\nabla_m \xi^a)] \\
&\quad = 2\xi^m \nabla_{[m}\nabla_{n]}\alpha^a + \alpha^m \nabla_n \nabla_m \xi^a \\
&\quad = -\xi^m R^a_{\ pmn}\alpha^p + \alpha^p \nabla_n \nabla_p \xi^a = -\alpha^p K^a_{\ np}.
\end{aligned}$$

The other cases now follow with a standard march through the indices. To compute $(\mathcal{L}_\xi \nabla_n - \nabla_n \mathcal{L}_\xi)\alpha_b$, for example, we consider an arbitrary smooth field λ^b and make use of our derived expressions for $(\mathcal{L}_\xi \nabla_n - \nabla_n \mathcal{L}_\xi)(\alpha_b \lambda^b)$ and $(\mathcal{L}_\xi \nabla_n - \nabla_n \mathcal{L}_\xi)\lambda^b$. And so forth.

Now if $K^m_{\ ab} = \mathbf{0}$, it follows immediately from our equation that $\mathcal{L}_\xi$ commutes with ∇ in its action on any smooth tensor field. Conversely, if the commutation condition holds, then $\alpha_m K^m_{\ nb} = \mathbf{0}$ for all smooth fields α_b. So $K^m_{\ nb} = \mathbf{0}$.

PROBLEM 1.8.4. *Show that given any smooth field ξ^a on a manifold, the operators $\mathcal{L}_\xi$ and d_n commute in their action on all smooth p-forms.*

Given any smooth p-form $\alpha_{b_1\ldots b_p}$, we have, by the preceding problem,

$$\begin{aligned}
(\mathcal{L}_\xi d_n - d_n \mathcal{L}_\xi)\alpha_{b_1\ldots b_p} &= \mathcal{L}_\xi \nabla_{[n}\alpha_{b_1\ldots b_p]} - \nabla_{[n}\mathcal{L}_\xi \alpha_{b_1\ldots b_p]} \\
&= \alpha_{m[b_2\ldots b_p} K^m_{\ n b_1]} + \cdots + \alpha_{[b_1\ldots b_{p-1}|m|} K^m_{\ n b_p]} \\
&= \alpha_{m[b_2\ldots b_p} K^m_{\ [n b_1]]} + \cdots + \alpha_{[b_1\ldots b_{p-1}|m|} K^m_{\ [n b_p]]}.
\end{aligned}$$

Each of the terms in the final sum is $\mathbf{0}$, since $K^m_{\ [rs]} = \mathbf{0}$. (This follows, since by the symmetries of the Riemann tensor field,

$$\begin{aligned}
2K^m_{\ [rs]} &= 2(R^m_{\ [s|n|r]}\xi^n - \nabla_{[r}\nabla_{s]}\xi^m) = R^m_{\ snr}\xi^n - R^m_{\ rns}\xi^n + R^m_{\ nrs}\xi^n \\
&= R^m_{\ snr}\xi^n + R^m_{\ rsn}\xi^n + R^m_{\ nrs}\xi^n = 3R^m_{\ [snr]}\xi^n = \mathbf{0}.)
\end{aligned}$$

So $(\mathcal{L}_\xi d_n - d_n \mathcal{L}_\xi)\alpha_{b_1\ldots b_p} = \mathbf{0}$.

PROBLEM 1.9.1. *Let ∇ be a derivative operator on a manifold that is compatible with the metric g_{ab}. Use the Bianchi identity to show that*

$$\nabla_a\big(R^{ab} - \frac{1}{2}\, g^{ab} R\big) = \mathbf{0}.$$

By the Bianchi identity, and various symmetries of the Riemann tensor field, we have

$$\mathbf{0} = \nabla_m R_{abcd} + \nabla_d R_{abmc} + \nabla_c R_{abdm} = \nabla_m R_{abcd} - \nabla_d R_{bamc} - \nabla_c R_{abmd}.$$

If we raise indices a and b, and then perform (a, d) and (b, c) contraction, we arrive at

$$\mathbf{0} = \nabla_m\, R - \nabla_a\, R^a_{\ m} - \nabla_b\, R^b_{\ m}.$$

Contracting with g^{mc} (and changing indices of contraction) yields

$$\mathbf{0} = \nabla_m (g^{mc} R) - 2\,\nabla_a R^{ac} = \nabla_a (g^{ac} R - 2\, R^{ac}).$$

So, $\nabla_a\big(R^{ac} - \frac{1}{2}\, g^{ac} R\big) = \mathbf{0}$.

PROBLEM 1.9.2. *Let ξ^a be a smooth vector field on M. Show that*

$$\pounds_\xi\, g^{ab} = \boldsymbol{0} \iff \pounds_\xi\, g_{ab} = \boldsymbol{0}.$$

We know that $\pounds_\xi\, \delta^a_{\ c} = \mathbf{0}$ (Problem 1.6.1). Hence

$$\mathbf{0} = \pounds_\xi\, \delta^a_{\ c} = \pounds_\xi\, (g^{ab}\, g_{bc}) = g^{ab}\, \pounds_\xi\, g_{bc} + g_{bc}\, \pounds_\xi\, g^{ab}.$$

Assume that $\pounds_\xi\, g_{ab} = \mathbf{0}$. Then $g_{bc}\, \pounds_\xi\, g^{ab} = \mathbf{0}$ and, therefore,

$$\mathbf{0} = g^{cd}\, g_{bc}\, \pounds_\xi\, g^{ab} = \delta_b^{\ d}\, \pounds_\xi\, g^{ab} = \pounds_\xi\, g^{ad}.$$

This gives us the implication from left to right The converse is handled similarly.

PROBLEM 1.9.3. *Show that Killing fields on M with respect to g_{ab} are affine collineations with respect to ∇.*

Let ξ^a be a Killing field. By proposition 1.9.8 (and various symmetries of the Riemann curvature tensor),

$$\nabla_a \nabla_b\, \xi_m = -R^n_{\ abm}\, \xi_n = -R_{nabm}\, \xi^n = -R_{bmna}\, \xi^n = R_{mbna}\, \xi^n.$$

So $\nabla_a \nabla_b\, \xi^m = R^m_{\ bna}\, \xi^n$. It now follows immediately from problem 1.8.3 that ξ^a is an affine collineation with respect to ∇.

PROBLEM 1.9.4. *Show that if ξ^a is a Killing field on M with respect to g_{ab}, then the Lie derivative operator $\mathcal{L}_\xi$ annihilates the fields R_{abcd}, R_{ab}, and R determined by g_{ab}.*

Given any smooth vector field η^a, we have

$$\mathcal{L}_\xi\,(R^a{}_{bcd}\,\eta_a) = \mathcal{L}_\xi\,(2\,\nabla_{[c}\,\nabla_{d]}\,\eta_b) = 2\,\nabla_{[c}\,\nabla_{d]}\,(\mathcal{L}_\xi\,\eta_b) = R^a{}_{bcd}\,\mathcal{L}_\xi\,\eta_a.$$

(The second equality follows from the preceding problem. Since ξ^a is a Killing field, it is an affine collineation with respect to ∇; i.e., $\mathcal{L}_\xi$ commutes with ∇.) But by the Leibniz rule, we also have

$$\mathcal{L}_\xi\,(R^a{}_{bcd}\,\eta_a) = R^a{}_{bcd}\,\mathcal{L}_\xi\,\eta_a + \eta_a\,\mathcal{L}_\xi\,R^a{}_{bcd}.$$

Comparing these two expressions, we see that $\eta_a\,\mathcal{L}_\xi\,R^a{}_{bcd} = \mathbf{0}$. But this is true for all smooth fields η_a. So $\mathcal{L}_\xi\,R^a{}_{bcd} = \mathbf{0}$. Hence, since $\mathcal{L}_\xi\,\delta^n{}_m = \mathbf{0}$,

$$\mathcal{L}_\xi\,R_{ab} = \mathcal{L}_\xi\,(\delta^n{}_m\,R^m{}_{abn}) = \delta^n{}_m\,\mathcal{L}_\xi\,R^m{}_{abn} = \mathbf{0}.$$

Since ξ^a is a Killing field, $\mathcal{L}_\xi\,g^{mn} = \mathbf{0}$. (See problem 1.9.2.) So it follows that

$$\mathcal{L}_\xi\,R = \mathcal{L}_\xi\,(g^{ab}\,R_{ab}) = g^{ab}\,\mathcal{L}_\xi\,R_{ab} = \mathbf{0}.$$

PROBLEM 1.9.5. *Show that if ξ^a and η^a are Killing fields on M with respect to g_{ab}, and k is a real number, then $(\xi^a + \eta^a)$, $(k\xi^a)$, and the commutator $[\xi, \eta]^a = \mathcal{L}_\xi\,\eta^a$ are all Killing fields with respect to g_{ab} as well.*

$\lambda^a = (\xi^a + \eta^a)$ is a Killing field since $\nabla_{(a}\,\lambda_{b)} = \nabla_{(a}\,\xi_{b)} + \nabla_{(a}\,\eta_{b)} = \mathbf{0}$.

Similarly, $\chi^a = (k\xi^a)$ is a Killing field since $\nabla_{(a}\,\chi_{b)} = k\,\nabla_{(a}\,\xi_{b)} = \mathbf{0}$.

Finally, $\theta^a = \mathcal{L}_\xi\,\eta^a$ is a Killing field since, by problem 1.6.6, $\mathcal{L}_\theta\,g_{ab} = \mathcal{L}_\xi\,\mathcal{L}_\eta\,g_{ab} - \mathcal{L}_\eta\,\mathcal{L}_\xi\,g_{ab} = \mathbf{0}$.

PROBLEM 1.9.6. *Let η^a be a Killing field on M with respect to g_{ab}. (i) Let γ be a geodesic with tangent field ξ^a. Show that the function $E = \xi^a\,\eta_a$ is constant on γ. (ii) Let T^{ab} be a smooth tensor field that is symmetric and divergence-free (i.e., $\nabla_a\,T^{ab} = \mathbf{0}$), and let J^a be the field $T^{ab}\,\eta_b$. Show that $\nabla_a\,J^a = \mathbf{0}$.*

Let η^a, γ, ξ^a, and E be as stated. Then we have

$$\xi^n\,\nabla_n\,E = \xi^n\,\nabla_n\,(\xi^a\,\eta_a) = \xi^n\,\xi^a\,\nabla_n\,\eta_a + \eta_a\,\xi^n\,\nabla_n\,\xi^a.$$

Since ξ^a is a Killing field, $\nabla_n\,\eta_a$ is anti-symmetric. So $\xi^n\,\xi^a\,\nabla_n\,\eta_a = \mathbf{0}$. And since ξ^a is the tangent field of a geodesic, $\xi^n\,\nabla_n\,\xi^a = \mathbf{0}$. So, $\xi^n\,\nabla_n\,E = \mathbf{0}$. This gives us (1). The computation for (2) is much the same:

$$\nabla_a J^a = \nabla_a (T^{ab} \eta_b) = T^{ab} \nabla_a \eta_b + \eta_b \nabla_a T^{ab}.$$

The second term on the right side vanishes since $\nabla_a T^{ab} = \mathbf{0}$. The first vanishes since T^{ab} is symmetric (and hence $T^{ab} \nabla_a \eta_b = T^{ab} \nabla_{(a} \eta_{b)} = \mathbf{0}$). So $\nabla_a J^a$ is $\mathbf{0}$.

PROBLEM 1.9.7. *Show that if η^a is a conformal Killing field on M, and M has dimension n, then*

$$\nabla_{(a} \eta_{b)} = \frac{1}{n} (\nabla_c \eta^c) g_{ab}.$$

Assume $\mathcal{L}_\eta (\Omega^2 g_{ab}) = \mathbf{0}$. Then, by proposition 1.7.4 (and the fact that $\nabla_m g_{ab} = \mathbf{0}$),

$$\begin{aligned} \mathbf{0} &= \Omega^2 \mathcal{L}_\eta g_{ab} + g_{ab} \mathcal{L}_\eta \Omega^2 = \Omega^2 [\eta^m \nabla_m g_{ab} + \nabla_a \eta_b + \nabla_b \eta_a] + g_{ab} \mathcal{L}_\eta \Omega^2 \\ &= \Omega^2 [\nabla_a \eta_b + \nabla_b \eta_a] + g_{ab} \mathcal{L}_\eta \Omega^2. \end{aligned}$$

If we raise the index b and then contract, we obtain

$$\mathbf{0} = 2\,\Omega^2 (\nabla_a \eta^a) + n\, \mathcal{L}_\eta \Omega^2.$$

Our two equations jointly yield $\nabla_{(a} \eta_{b)} = \dfrac{1}{n} g_{ab} (\nabla_c \eta^c)$.

PROBLEM 1.10.1. *Let S be a k-dimensional imbedded submanifold of the n-dimensional manifold M, and let p be a point in S.*

(1) Show that the space of co-vectors $\eta_a \in (M_p)_a$ normal to S has dimension $(n-k)$.
(2) Show that a vector $\xi^a \in (M_p)^a$ is tangent to S iff $\eta_a \xi^a = 0$ for all co-vectors $\eta_a \in (M_p)_a$ that are normal to S.

(1) The subspace of vectors in $(M_p)^a$ tangent to S has dimension k. Let $\{\overset{1}{\xi}{}^a, \overset{2}{\xi}{}^a, \ldots, \overset{k}{\xi}{}^a\}$ be any set of k linearly independent vectors from that subspace. We can extend it to a basis for $(M_p)^a$ by adding $(n-k)$ more (appropriately chosen) vectors $\overset{k+1}{\xi}{}^a, \ldots, \overset{n}{\xi}{}^a$. Now let $\{\overset{1}{\alpha}_a, \ldots, \overset{n}{\alpha}_a\}$ be the dual basis. So $\overset{i}{\alpha}_a \overset{j}{\xi}{}^a = \delta_{ij}$. We claim that the subspace of co-vectors at p normal to S is spanned by $\{\overset{k+1}{\alpha}_a, \ldots, \overset{n}{\alpha}_a\}$. To see this, consider any co-vector $\alpha_a = \overset{1}{\alpha} \overset{1}{\alpha}_a + \cdots + \overset{n}{\alpha} \overset{n}{\alpha}_a$ at p. It is normal to S iff $\overset{i}{\alpha} = \alpha_a \overset{i}{\xi}{}^a = 0$ for all $i = 1, \ldots, k$ (because every vector at p tangent to S is a linear combination of $\overset{1}{\xi}{}^a, \overset{2}{\xi}{}^a, \ldots, \overset{k}{\xi}{}^a$). Thus α is normal iff

it is in the linear span of $\{\overset{k+1}{\alpha}_a, \ldots, \overset{n}{\alpha}_a\}$. So the latter is a basis for the subspace of co-vectors at p normal to S—and therefore that subspace has dimension $(n-k)$.

(2) The argument is much the same. We continue to work with the basis and dual basis described in (1). Consider any vector $\xi^a = \overset{1}{\xi}\,\overset{1}{\xi}{}^a + \cdots + \overset{n}{\xi}\,\overset{n}{\xi}{}^a$ at p. It is killed by every covariant vector at p normal to S iff it is killed by all the vectors $\overset{k+1}{\alpha}_a, \ldots, \overset{n}{\alpha}_a$. And the latter condition holds iff $\overset{i}{\xi} = \overset{i}{\alpha}_a \xi^a = 0$ for all $i = k+1, \ldots, n$. So ξ^a is killed by every covariant vector at p normal to S iff it is a linear combination of $\overset{1}{\xi}{}^a, \overset{2}{\xi}{}^a, \ldots, \overset{k}{\xi}{}^a$—i.e., iff it is tangent to S.

PROBLEM 1.10.2. *Let S be a k-dimensional imbedded submanifold of the n-dimensional manifold M, and let g_{ab} be a metric on M. Show that S is a metric submanifold (relative to g_{ab}) iff, for all p in S, the pull-back tensor $(id_p)^*(g_{ab})$ is non-degenerate; i.e., there is no non-zero vector $\tilde{\xi}^a \in (S_p)^a$ such that $\big((id_p)^*(g_{ab})\big)\,\tilde{\xi}^a = \mathbf{0}$.*

Let p be any point in S. The pull-back tensor $(id_p)^*(g_{ab})$ is degenerate there iff there is a $\tilde{\xi}^a \in (S_p)^a$ such that, for all $\tilde{\eta}^a \in (S_p)^a$, $\big((id_p)^*(g_{ab})\big)\,\tilde{\xi}^a\,\tilde{\eta}^b = g_{ab}\,\big((id_p)_*(\tilde{\xi}^a)\big)\,\big((id_p)_*(\tilde{\eta}^b)\big) = 0$. Since a vector in $(M_p)^a$ is tangent to S precisely if it is of the form $(id_p)_*(\tilde{\eta^a})$ for some $\tilde{\eta}^a \in (S_p)^a$, we see that $(id_p)^*(g_{ab})$ is degenerate at p iff there is a $\tilde{\xi}^a \in (S_p)^a$ such that $g_{ab}\,\big((id_p)_*(\tilde{\xi}^b)\big)$ is normal to S; i.e., there is a vector in $(M_p)^a$ tangent to S that is also normal to S.

PROBLEM 1.10.3. *Prove the following generalization of clause (2) in proposition 1.10.3. For all M-tensor fields $\alpha^{\cdots a \cdots}$ on S, the following conditions both hold.*

(1) $\alpha^{\cdots a \cdots}$ is tangent to S in the index $a \iff h^a{}_b\,\alpha^{\cdots b \cdots} = \alpha^{\cdots a \cdots} \iff k^a{}_b\,\alpha^{\cdots b \cdots} = \mathbf{0}$.

(2) $\alpha^{\cdots a \cdots}$ is normal to S in the index $a \iff k^a{}_b\,\alpha^{\cdots b \cdots} = \alpha^{\cdots a \cdots} \iff h^a{}_b\,\alpha^{\cdots b \cdots} = \mathbf{0}$.

We work with a representative case. (The proof is exactly the same no matter how many indices are involved.) Consider the M-field α^{amn} on S. Suppose first that $h^a{}_b\,\alpha^{bmn} = \alpha^{amn}$. Then α^{amn} is certainly tangent to S in a since $h^a{}_b$ is. Conversely, suppose α^{amn} is tangent to S in a. Then, we claim, $h^a{}_b\,\alpha^{bmn}$ and α^{amn} have the same action on any co-vector η_a (at any point of S) that is either tangent to, or normal to, S. In the first case, $h^a{}_b\,\alpha^{bmn}\,\eta_a = \alpha^{amn}\,\eta_a$, since $h^a{}_b\,\eta_a = \eta_b$, In the second case, $h^a{}_b\,\alpha^{bmn}\,\eta_a = 0 = \alpha^{amn}\,\eta_a$, because both sides are tangent to S in a. This gives us the first equivalence in (1). The second is immediate since $k^a_{\ b}\,\alpha^{bmn} = (g^a_{\ b} - h^a_{\ b})\,\alpha^{bmn} = \alpha^{amn} - h^a_{\ b}\,\alpha^{bmn}$. The equivalences in (2) are handled similarly.

PROBLEM 1.10.4. *Prove that* $h^m_{\ a}\, k^n_{\ b}\, k^p_{\ c}\, \nabla_m h_{np} = \mathbf{0}$.

We have

$$h^m_{\ a}\, k^n_{\ b}\, k^p_{\ c}\, \nabla_m h_{np} = h^m_{\ a}\, k^n_{\ b}\, [\nabla_m (h_{np}\, k^p_{\ c}) - h_{np} \nabla_m k^p_{\ c}].$$

But $h_{np}\, k^p_{\ c} = \mathbf{0} = k^n_{\ b} h_{np}$ (by the third clause of proposition 1.10.3). So both terms on the right are $\mathbf{0}$.

PROBLEM 1.10.5. *Derive the second Gauss-Codazzi equation:*

$$h^m_{\ [a}\, h^n_{\ b]}\, h^p_{\ c}\, k^r_{\ d} \nabla_m \pi_{npr} = \frac{1}{2}\, h^m_{\ a}\, h^n_{\ b}\, h^p_{\ c}\, k^r_{\ d}\, R_{mnpr}.$$

We have

(1) $h^m_{\ a}\, h^n_{\ b}\, h^p_{\ c}\, \nabla_m h_{np} = \mathbf{0}$.
(2) $h^m_{\ a}\, k^n_{\ b}\, k^p_{\ c}\, \nabla_m h_{np} = \mathbf{0}$.
(3) $\pi_{abc} = h^m_{\ a}\, h^n_{\ b}\, k^p_{\ c}\, \nabla_m h_{np} = h^m_{\ a}\, h^n_{\ b}\, (g^p_{\ c} - h^p_{\ c})\, \nabla_m h_{np} = h^m_{\ a}\, h^n_{\ b}\, \nabla_m h_{nc}$.

(The first is equation (1.10.3); the second was proved in the preceding problem; and the third follows from the first.) Hence

$$h^m_{\ [a}\, h^n_{\ b]}\, h^p_{\ c}\, k^r_{\ d} \nabla_m \pi_{npr} = h^m_{\ [a}\, h^n_{\ b]}\, h^p_{\ c}\, k^r_{\ d} \nabla_m\, (h^q_{\ n}\, h^s_{\ p}\, \nabla_q h_{sr}) = A + B + C,$$

where

$$A = h^m_{\ [a}\, h^n_{\ b]}\, h^p_{\ c}\, k^r_{\ d} (\nabla_m h^q_{\ n})\, h^s_{\ p}\, (\nabla_q h_{sr}),$$

$$B = h^m_{\ [a}\, h^n_{\ b]}\, h^p_{\ c}\, k^r_{\ d} h^q_{\ n}\, (\nabla_m h^s_{\ p})\, (\nabla_q h_{sr}),$$

$$C = h^m_{\ [a}\, h^n_{\ b]}\, h^p_{\ c}\, k^r_{\ d} h^q_{\ n}\, h^s_{\ p}\, (\nabla_m \nabla_q h_{sr}).$$

By (3) and lemma 1.10.6, $h^m_{\ [a}\, h^n_{\ b]}\, (\nabla_m h^q_{\ n}) = \pi_{[ab]}{}^q = \mathbf{0}$. So $A = \mathbf{0}$. Furthermore, since $h^n_{\ b}\, h^q_{\ n} = h^q_{\ b}$, we have, by (3),

$$\begin{aligned} B &= h^m_{\ [a}\, h^q_{\ b]}\, h^p_{\ c}\, k^r_{\ d}\, (\nabla_m h^s_{\ p})\, (\nabla_q h_{sr}) = h^q_{\ [b}\, h^m_{\ a]}\, h^p_{\ c}\, (\nabla_m h^s_{\ p})\, k^r_{\ d}\, (\nabla_q h_{sr}) \\ &= h^q_{\ [b}\, \pi_{a]c}{}^s\, k^r_{\ d}\, (\nabla_q h_{sr}). \end{aligned}$$

Now π_{acs} is tangent to S in the index a and normal to it in s. So $\pi_{ac}{}^s = h^u_{\ a}\, k^s_{\ v}\, \pi_{uc}{}^v$ and therefore, continuing the computation,

$$B = h^q_{\ [b}\, \pi_{a]c}{}^s\, k^r_{\ d}\, (\nabla_q h_{sr}) = h^q_{\ [b}\, h^u_{\ a]}\, k^s_{\ v}\, k^r_{\ d}\, (\nabla_q h_{sr}) \pi_{uc}{}^v = \mathbf{0}.$$

(The final equality follows from (2).) Finally, since $h^n_{\ b}\, h^q_{\ n} = h^q_{\ b}$,

$$C = h^m_{\ [a}\, h^q_{\ b]}\, h^p_{\ c}\, k^r_{\ d}\, h^s_{\ p}\, (\nabla_m \nabla_q\, h_{sr}) = h^m_{\ a}\, h^q_{\ b}\, h^p_{\ c}\, k^r_{\ d}\, h^s_{\ p}\, (\nabla_{[m} \nabla_{q]}\, h_{sr})$$

$$= \frac{1}{2}\, h^m_{\ a}\, h^q_{\ b}\, h^p_{\ c}\, k^r_{\ d}\, h^s_{\ p}\, (h_{ur}\, R^u_{\ smq} + h_{su}\, R^u_{\ rmq}).$$

Now $k^r_{\ d}\, h_{ur} = \mathbf{0}$ (by proposition 1.10.3) and $h^s_{\ p}\, h_{su} = h_{pu}$. So, continuing,

$$h^m_{\ [a}\, h^n_{\ b]}\, h^p_{\ c}\, k^r_{\ d} \nabla_m\, \pi_{npr} = C \ = \ \frac{1}{2}\, h^m_{\ a}\, h^q_{\ b}\, h^p_{\ c}\, k^r_{\ d}\, h_{pu}\, R^u_{\ rmq}$$

$$= \frac{1}{2}\, h^m_{\ a}\, h^q_{\ b}\, h^p_{\ c}\, k^r_{\ d}\, (g_{pu} - k_{pu})\, R^u_{\ rmq}$$

$$= \frac{1}{2}\, h^m_{\ a}\, h^q_{\ b}\, h^p_{\ c}\, k^r_{\ d}\, R_{prmq}.$$

(The final equality follows from the fact that, once again, $h^p_{\ c}\, k_{pu} = \mathbf{0}$.) But $R_{prmq} = R_{mqpr}$. So we are done.

PROBLEM 1.11.1. *One learns in the study of ordinary vector analysis that, for all vectors ξ, η, θ, λ at a point, the following identities hold.*

(1) $(\xi \times \eta) \cdot (\theta \times \lambda) = (\xi \cdot \theta)(\eta \cdot \lambda) - (\xi \cdot \lambda)(\eta \cdot \theta)$.
(2) $(\xi \times (\eta \times \theta)) + (\theta \times (\xi \times \eta)) + (\eta \times (\theta \times \xi)) = \mathbf{0}$.

Reformulate these assertions in our notation and prove them.

The two come out as follows.

(1′) $(\epsilon^{abc} \xi_b \eta_c)(\epsilon_{amn} \theta^m \lambda^n) = (\xi^b \theta_b)(\eta^c \lambda_c) - (\xi^b \lambda_b)(\eta^c \theta_c)$.
(2′) $\epsilon^{abc} \xi_b (\epsilon_{cmn} \eta^m \theta^n) + \epsilon^{abc} \theta_b (\epsilon_{cmn} \xi^m \eta^n) + \epsilon^{abc} \eta_b (\epsilon_{cmn} \theta^m \xi^n) = \mathbf{0}$.

They follow easily from equation (1.11.6)—in the case where $n = 3$ and $n^- = 0$. First, we have

$$(\epsilon^{abc} \xi_b \eta_c)(\epsilon_{amn} \theta^m \lambda^n) = (\epsilon^{abc} \epsilon_{amn})\, \xi_b\, \eta_c\, \theta^m\, \lambda^n = 2\, \delta^{[b}_{\ m}\, \delta^{c]}_{\ n}\, \xi_b\, \eta_c\, \theta^m\, \lambda^n$$

$$= 2\, \xi_b\, \eta_c\, \theta^{[b}\, \lambda^{c]} = (\xi^b \theta_b)(\eta^c \lambda_c) - (\xi^b \lambda_b)(\eta^c \theta_c).$$

And for the second, we have

$$\epsilon^{abc} \xi_b (\epsilon_{cmn} \eta^m \theta^n) + \epsilon^{abc} \theta_b (\epsilon_{cmn} \xi^m \eta^n) \ + \ \epsilon^{abc} \eta_b (\epsilon_{cmn} \theta^m \xi^n)$$

$$= (\epsilon^{cab} \epsilon_{cmn})\, \xi_b\, \eta^m\, \theta^n + (\epsilon^{cab} \epsilon_{cmn})\, \theta_b\, \xi^m\, \eta^n + (\epsilon^{cab} \epsilon_{cmn})\, \eta_b\, \theta^m\, \xi^n$$

$$= 2\, \delta^{[a}_{\ m}\, \delta^{b]}_{\ n}\, \xi_b\, \eta^m\, \theta^n + 2\, \delta^{[a}_{\ m}\, \delta^{b]}_{\ n}\, \theta_b\, \xi^m\, \eta^n + 2\, \delta^{[a}_{\ m}\, \delta^{b]}_{\ n}\, \eta_b\, \theta^m\, \xi^n$$

$$= 2\, \xi_b\, \eta^{[a}\, \theta^{b]} + 2\, \theta_b\, \xi^{[a}\, \eta^{b]} + 2\, \eta_b\, \theta^{[a}\, \xi^{b]}$$

$$= [(\xi_b \theta^b)\, \eta^a - (\xi_b \eta^b)\, \theta^a] + [(\theta_b \eta^b)\, \xi^a - (\theta_b \xi^b)\, \eta^a] + [(\eta_b \xi^b)\, \theta^a$$
$$- (\eta_b \theta^b)\, \xi^a]$$
$$= \mathbf{0}.$$

PROBLEM 1.11.2. *Do the same for the following assertion:*

$$div(\xi \times \eta) = \eta \cdot curl(\xi) - \xi \cdot curl(\eta).$$

We have

$$\begin{aligned} \nabla_a (\epsilon^{abc}\, \xi_b\, \eta_c) &= \xi_b\, \epsilon^{abc}\, \nabla_a \eta_c + \eta_c\, \epsilon^{abc}\, \nabla_a \xi_b \\ &= \eta_c\, \epsilon^{cab}\, \nabla_a \xi_b - \xi_b\, \epsilon^{bac}\, \nabla_a \eta_c. \end{aligned}$$

PROBLEM 1.11.3. *We have seen that every Killing field ξ^a in n-dimensional Euclidean space ($n \geq 1$) can be expressed uniquely in the form*

$$\xi_b = \chi^a F_{ab} + k_b,$$

where F_{ab} and k_b are constant, F_{ab} is anti-symmetric, and χ^a is the position field relative to some point p. Consider the special case where $n = 3$. Let ϵ_{abc} be a volume element. Show that (in this special case) there is a unique constant field W^a such that $F_{ab} = \epsilon_{abc} W^c$.

Let $W^a = \frac{1}{2}\epsilon^{abc} F_{bc}$. Then

$$\epsilon_{abc}\, W^c = \epsilon_{abc} \left(\frac{1}{2}\, \epsilon^{cmn} F_{mn} \right) = \frac{1}{2}\, (\epsilon^{cmn}\, \epsilon_{cab})\, F_{mn} = \delta^{[m}{}_a\, \delta^{n]}{}_b\, F_{mn} = F_{[ab]} = F_{ab}.$$

(The final equality follows from the fact that F_{ab} is anti-symmetric.) W^a is constant, since

$$\nabla_b\, W^a = \frac{1}{2}\, \nabla_b\, (\epsilon^{amn} F_{mn}),$$

and both ϵ^{amn} and F_{mn} are constant. Finally, W^a is the unique field satisfying the given constraint, for if we also have $F_{ab} = \epsilon_{abc}\, \widehat{W}^c$, then $\epsilon_{abc}(\widehat{W}^c - W^c) = \mathbf{0}$, and so

$$\mathbf{0} = \epsilon^{abn} \epsilon_{abc}(\widehat{W}^c - W^c) = 2\, \delta^n_{\ c}(\widehat{W}^c - W^c) = 2\, (\widehat{W}^n - W^n).$$

PROBLEM 2.1.1. *Consider our characterization of timelike vectors in terms of null vectors in the proof of proposition 2.1.1. Why does it fail if $n = 2$?*

If $n = 2$, the stated condition holds for spacelike as well as timelike vectors. Indeed, in that dimension, given *any* two non-zero null vectors α^a and γ^a

that are not proportional to one another, every spacelike (as well as every timelike) vector η^a can be expressed in the form $\eta^a = k\alpha^a + l\gamma^a$, where $k \neq 0$ and $l \neq 0$. So, of course, if we take β^a to be the null vector $l\gamma^a$, then we have $\eta^a = k\alpha^a + \beta^a$.

PROBLEM 2.1.2. *(i) Show that it is possible to characterize timelike vectors in terms of causal vectors. (ii) Show that it is possible to characterize timelike vectors in terms of spacelike vectors.*

The following equivalences hold for all $n \geq 2$.

A vector η^a at p is timelike iff for all causal vectors α^a at p, there is an $\epsilon > 0$ such that, for all k, if $|k| < \epsilon$, then $\eta^a + k\alpha^a$ is causal.

A vector η^a at p is timelike iff for all spacelike vectors α^a at p, there is an $\epsilon > 0$ such that, for all k, if $|k| < \epsilon$, then $\eta^a + k\alpha^a$ is not spacelike.

PROBLEM 2.1.3. *Does proposition 2.1.3 still hold if condition (1) is left intact but (2) is replaced by*

$(2')$ $\alpha^{a_1 \ldots a_r}{}_{b_1 \ldots b_s}\, \xi^{b_1} \ldots \xi^{b_s} = \mathbf{0}$ *for all spacelike vectors ξ^a at the point?*

And what if it is replaced by

$(2'')$ $\alpha^{a_1 \ldots a_r}{}_{b_1 \ldots b_s}\, \xi^{b_1} \ldots \xi^{b_s} = \mathbf{0}$ *for all null vectors ξ^a at the point?*

Condition $(2'')$ is certainly not sufficient. For example, if g_{ab} is a spacetime metric and p is a point in the underlying manifold, then $g_{ab}\,\xi^a\xi^b = \mathbf{0}$ for all null vectors ξ^a at p, but $g_{ab} \neq \mathbf{0}$. On the other hand, condition $(2')$ *is* sufficient, and the proof is almost the same as for the original version of proposition 2.1.3. Only one change is needed. Before we used the fact that if ξ^a is a timelike vector at some point, and η^a is an arbitrary vector there, then there is an $\epsilon > 0$ such that, for all x, if $|x| < \epsilon$, then $(\xi^a + x\eta^a)$ is timelike. Now we use the corresponding assertion with both occurences of "timelike" changed to "spacelike."

PROBLEM 2.2.1. *Let p be a point in M. Show that there is no two-dimensional subspace of M_p all of whose elements are causal (timelike or null).*

Assume there are non-zero, linearly independent vectors α^a and β^a at p such that, for all k and l, the vector $(k\,\alpha^a + l\,\beta^a)$ is causal. We derive a contradiction.

There are two cases to consider. Either (i) one of the two is timelike, or (ii) both are null. Assume first that one of the two, say α^a, is timelike. If we set

$$k = -\frac{(\alpha_m \beta^m)}{(\alpha_n \alpha^n)} \qquad l = 1,$$

then $\alpha^n(k\,\alpha_n + l\,\beta_n) = 0$; i.e., $(k\,\alpha^a + l\,\beta^a)$ is orthogonal to the timelike vector α^a. Since $(k\,\alpha^a + l\,\beta^a)$ is causal, it follows from the first clause of proposition 2.2.1 that $(k\,\alpha^a + l\,\beta^a) = \mathbf{0}$. This contradicts our assumption that α^a and β^a are linearly independent. Assume next that α^a and β^a are both null. Then, for all k and l,

$$0 \le (k\,\alpha^n + l\,\beta^n)(k\,\alpha_n + l\,\beta_n) = 2\,k\,l\,(\alpha^n \beta_n),$$

since $(k\,\alpha^a + l\,\beta^a)$ is causal. But this can hold for all k and l only if $(\alpha^n \beta_n) = 0$. Hence, by the second clause of proposition 2.2.1, α^a and β^a must be proportional to one another. Once again, this contradicts our assumption that they are linearly independent.

PROBLEM 2.2.2. *Let g'_{ab} be a second metric on M (not necessarily of Lorentz signature). Show that the following conditions are equivalent.*

(1) For all p in M, g_{ab} and g'_{ab} agree on which vectors at p are orthogonal.
(2) g'_{ab} is conformally equivalent to either g_{ab} or $-g_{ab}$.

The implication (2) $\Rightarrow$ (1) is immediate. For the other direction, assume (1) holds. It follows from (1) that g_{ab} and g'_{ab} agree as to which vectors are null—i.e., orthogonal to themselves. So it will suffice to show that g'_{ab} has signature $(1, 3)$ or $(3, 1)$. For then we can invoke proposition 2.1.1 and conclude that g'_{ab} is conformally equivalent to g_{ab} (in the first case) or to $-g_{ab}$ (in the second case).

Let p be any point in M, and let $\overset{1}{\xi}{}^a, \ldots, \overset{4}{\xi}{}^a$ be an orthonormal basis at p with respect to g_{ab}. Consider the vector $(\overset{1}{\xi}{}^a + \overset{2}{\xi}{}^a)$. It is null with respect to g_{ab}. So it must be null with respect to g'_{ab}. Furthermore, since $\overset{1}{\xi}{}^a$ and $\overset{2}{\xi}{}^a$ are orthogonal with respect to g_{ab}, they must be orthogonal with respect to g'_{ab}. So we have

$$0 = g'_{ab}\,(\overset{1}{\xi}{}^a + \overset{2}{\xi}{}^a)(\overset{1}{\xi}{}^b + \overset{2}{\xi}{}^b) = g'_{ab}\,\overset{1}{\xi}{}^a\,\overset{1}{\xi}{}^b + g'_{ab}\,\overset{2}{\xi}{}^a\,\overset{2}{\xi}{}^b.$$

Similarly, we have

$$0 = g'_{ab}\,\overset{1}{\xi}{}^a\,\overset{1}{\xi}{}^b + g'_{ab}\,\overset{3}{\xi}{}^a\,\overset{3}{\xi}{}^b$$
$$0 = g'_{ab}\,\overset{1}{\xi}{}^a\,\overset{1}{\xi}{}^b + g'_{ab}\,\overset{4}{\xi}{}^a\,\overset{4}{\xi}{}^b.$$

Now let $X_i = g'_{ab} \overset{i}{\xi}{}^a \overset{i}{\xi}{}^b$, for $i = 1, \ldots, 4$. The X_i are non-zero since the vectors $\overset{i}{\xi}{}^a$ are non-null with respect to g_{ab}. So there are only two possibilities. Either $X_1 > 0$ and $X_2, X_3, X_4 < 0$, or $X_1 < 0$ and $X_2, X_3, X_4 > 0$. In the first case, g'_{ab} has signature $(1, 3)$; in the second, it has signature $(3, 1)$. (In either case, we need only normalize the vectors $\overset{1}{\xi}{}^a, \ldots, \overset{4}{\xi}{}^a$ to arrive at an orthonormal basis at p of the appropriate type for g'_{ab}.)

PROBLEM 2.2.3. *Prove the second clause of proposition 2.2.3.*

Let μ^a and ν^a be co-oriented, non-zero causal vectors at a point p. Then either $(\mu^n \nu_n) > 0$, or both vectors are null and $\mu^a = k\,\nu^a$ for some $k > 0$. In the latter case, $\|\mu^a + \nu^a\| = \|\mu^a\| = \|\nu^a\| = 0$, and the assertion follows trivially. So we may assume $(\mu^n \nu_n) > 0$. Hence, by the first clause of proposition 2.2.3, $(\mu^n \nu_n) \geq \|\mu^a\|\,\|\nu^a\|$. Therefore,

$$\big(\|\mu^a\| + \|\nu^a\|\big)^2 = \|\mu^a\|^2 + 2\,\|\mu^a\|\,\|\nu^a\| + \|\nu^a\|^2 \;\leq\; (\mu^n \mu_n) + 2\,(\mu^n \nu_n) + (\nu^n \nu_n)$$
$$= \big(\mu^n + \nu^n\big)\,(\mu_n + \nu_n) = \|\mu^a + \nu^a\|^2.$$

(For the final equality we need the fact μ^a and ν^a are co-oriented. Otherwise, $(\mu^a + \nu^a)$ need not be causal.) Equality holds here iff $(\mu^n \nu_n) = \|\mu^a\|\,\|\nu^a\|$. But by the first half of the proposition, again, this is the case iff μ^a and ν^a are proportional.

PROBLEM 2.5.1. *Give examples for each of the following possibilities.*

(1) A smooth symmetric field T_{ab} that does not satisfy the WEC
(2) A smooth symmetric field T_{ab} that satisfies the WEC but not the DEC
(3) A smooth symmetric field T_{ab} that satisfies the DEC but not the SDEC

(1) $T_{ab} = -g_{ab}$. (2) $T_{ab} = \sigma_a\,\sigma_b$, where σ^a is a smooth spacelike field. (3) $T_{ab} = \lambda_a\,\lambda_b$, where λ^a is a smooth, non-zero null field.

PROBLEM 2.5.2. *Show that the DEC holds iff given any two co-oriented timelike vectors ξ^a and η^a at a point, $T_{ab}\,\xi^a\,\eta^b \geq 0$.*

Suppose first that the DEC holds, and let ξ^a be a timelike vector at some point. Then $T_{ab}\,\xi^a\,\xi^b \geq 0$ and $T^a{}_b\,\xi^b$ is a causal vector. We claim that $T_{ab}\,\xi^a\,\eta^b \geq 0$ for all timelike η^a at the point that are co-oriented with ξ^a. We may assume that $T_{ab}\,\xi^a \neq \mathbf{0}$, since otherwise the claim is trivial. And in this case it follows that $T_{ab}\,\xi^a\,\xi^b > 0$ (since otherwise $T_{ab}\,\xi^a$ is a non-zero causal vector that is orthogonal to the timelike vector ξ^b, and this is impossible by proposition 2.2.1). So $T^a{}_b\,\xi^b$ is a non-zero causal vector that is co-oriented with ξ^a. Now let η^a be any timelike vector at the point that is co-oriented with ξ^a. It must be co-oriented with $T^a{}_b\,\xi^b$ as well (since co-orientation is an equivalence relation). So $T_{ab}\,\xi^a\,\eta^b > 0$.

For the converse, suppose that given any two co-oriented timelike vectors ξ^a and η^a at a point, $T_{ab}\,\xi^a\,\eta^b \geq 0$. Let ξ^a be a timelike vector at some point. It follows immediately (taking $\eta^a = \xi^a$) that $T_{ab}\,\xi^a\,\xi^b \geq 0$. So what we have to show is that $T^a{}_b\,\xi^b$ is a causal vector. Suppose to the contrary that it is spacelike. Then we can find a timelike vector $\mathring{\eta}^a$ at the point, co-oriented with ξ^a, that is orthogonal to $T^a{}_b\,\xi^b$. But since $\mathring{\eta}^a$ is timelike, $(\mathring{\eta}^a + k\,T^a{}_b\,\xi^b)$ is also timelike and co-oriented with ξ^a for all sufficiently small $k > 0$. Hence, by our initial assumption,

$$0 \leq T_{ab}\,\xi^a(\mathring{\eta}^b + k\,T^b{}_n\,\xi^n) = k\,(T_{ab}\,\xi^a)\,(T^b{}_n\,\xi^n)$$

for all sufficiently small $k > 0$. But this is impossible since $(T_{ab}\,\xi^a)(T^b{}_n\,\xi^n) < 0$. So, as claimed, $T^a{}_b\,\xi^b$ is causal.

PROBLEM 2.5.3. *Consider a perfect fluid with four-velocity η^a, energy density ρ, and pressure p. (i) Show that it satisfies the DEC iff $|p| \leq \rho$. (ii) Show that it satisfies the SDEC iff it satisfies the DEC.*

(i) Suppose $T_{ab} = \rho\,\eta_a\,\eta_b - p\,(g_{ab} - \eta_a\,\eta_b)$. Then T_{ab} satisfies the DEC condition at a point iff for all unit timelike vectors ξ^a at that point, $T_{ab}\,\xi^a\,\xi^b \geq 0$ and $T^a{}_b\,\xi^b$ is causal. Now for all such vectors,

$$T_{ab}\,\xi^a\,\xi^b = (\rho + p)(\eta^a \xi_a)^2 - p,$$

$$(T^a{}_b\,\xi^b)(T_{ac}\,\xi^c) = (\rho^2 - p^2)(\eta^a \xi_a)^2 + p^2.$$

So the DEC holds iff both right-side expressions are non-negative for all choices of ξ^a.

Assume first that $|p| \leq \rho$, and let ξ^a be a unit timelike vector at the point in question. Then, by the wrong-way Schwarz inequality (proposition 2.2.3), $(\eta^a \xi_a)^2 \geq \|\eta^a\|^2\,\|\xi^a\|^2 = 1$. Hence,

$$(\rho + p)(\eta^a \xi_a)^2 - p \geq (\rho + p) - p = \rho \geq 0,$$

$$(\rho^2 - p^2)(\eta^a \xi_a)^2 + p^2 \geq (\rho^2 - p^2) + p^2 = \rho^2 \;\geq\; 0.$$

So the DEC holds at the point. Conversely, suppose that $T_{ab}\,\xi^a\,\xi^b \geq 0$ and $T^a{}_b\,\xi^b$ is causal for all unit timelike vectors ξ^a at the point. Then, in particular,

$T_{ab}\, \eta^a\, \eta^b \geq 0$ and, therefore, $0 \leq (\rho + p)(\eta^a \eta_a)^2 - p = (\rho + p) - p = \rho$ there. Next we use the fact that there is no upper bound to the value of $(\eta^a \xi_a)^2$ as ξ^a ranges over unit timelike vectors at the point. It cannot possibly be the case that $(\rho^2 - p^2)(\eta^a \xi_a)^2 + p^2 \;\geq\; 0$ for all such vectors unless $(\rho^2 - p^2) \geq 0$. So we have $\rho \geq 0$ and $(\rho^2 - p^2) \geq 0$. These two together are jointly equivalent to $|p| \leq \rho$, as required.

(ii) The SDEC implies the DEC (always, not just for perfect fluids). Suppose that at some point $T_{ab} = \rho\, \eta_a\, \eta_b - p\,(g_{ab} - \eta_a\, \eta_b)$ satisfies the DEC but not the SDEC. Then there is a timelike vector ξ^a at the point such that $T^a{}_b\, \xi^b$ is null even though $T_{ab} \neq \mathbf{0}$ there. We claim this is impossible. If $T^a{}_b\, \xi^b$ is null, then, $(\rho^2 - p^2)(\eta^a \xi_a)^2 + p^2 = 0$. But $|p| \leq \rho$, since we are assuming that the DEC holds, and $\eta^a \xi_a \neq 0$ (since no two timelike vectors are orthogonal). So this equation can hold only if $\rho = p = 0$, and this contradicts our assumption that $T_{ab} \neq \mathbf{0}$ at the point in question.

PROBLEM 2.6.1. *Show that Maxwell's equations in the source free case ($J^a = \mathbf{0}$) are conformally invariant.*

Let $g'_{ab} = \Omega^2\, g_{ab}$ be a second metric on the underlying manifold M, whose dimension n we leave open. Let its associated derivative operator be ∇'. It will suffice for us to show that

$$\nabla'_a\,(g'^{am}\, g'^{bn}\, F_{mn}) = \frac{1}{\Omega^4}\,(\nabla_a F^{ab}) + \frac{(n-4)}{\Omega^5}\, F^{ab}\, \nabla_a\, \Omega.$$

We know from proposition 1.9.5 that $\nabla' = (\nabla, C^a{}_{bc})$, where

$$C^a_{\ bc} = -\frac{1}{2\,\Omega^2}\big[\delta^a_{\ b}\, \nabla_c\, \Omega^2 + \delta^a_{\ c}\, \nabla_b\, \Omega^2 - g_{bc}\, g^{ar}\, \nabla_r\, \Omega^2\big].$$

We have

$$\begin{aligned}
\nabla'_a\,(g'^{am}\, g'^{bn}\, F_{mn}) &= g'^{am}\, g'^{bn}\, \nabla'_a\, F_{mn} = \Omega^{-4}\, g^{am}\, g^{bn}\, \nabla'_a\, F_{mn} \\
&= \Omega^{-4}\, g^{am}\, g^{bn}\, \big[\nabla_a\, F_{mn} + C^r{}_{am}\, F_{rn} + C^r{}_{an}\, F_{mr}\big].
\end{aligned}$$

Now

$$\begin{aligned}
\Omega^{-4}\, g^{am}\, g^{bn}\, C^r{}_{am}\, F_{rn} &= -\frac{1}{2\,\Omega^6}\, g^{am}\, g^{bn}\, [\delta^r{}_a\, \nabla_m\, \Omega^2 + \delta^r{}_m\, \nabla_a\, \Omega^2 - g_{am}\, g^{rs}\, \nabla_s\, \Omega^2]\, F_{rn} \\
&= -\frac{1}{2\,\Omega^6}\, [F^{mb} \nabla_m\, \Omega^2 + F^{ab} \nabla_a\, \Omega^2 - n\, F^{sb} \nabla_s\, \Omega^2] \\
&= \frac{(n-2)}{\Omega^5}\, F^{ab}\, \nabla_a\, \Omega
\end{aligned}$$

and (since F_{ab} is anti-symmetric and, therefore, $g^{rm}\,F_{mr}=0$),

$$\begin{aligned}\Omega^{-4}\,g^{am}\,g^{bn}\,C^r{}_{an}\,F_{mr} &= -\frac{1}{2\,\Omega^6}\,g^{am}\,g^{bn}\,[\delta^r{}_a\,\nabla_n\,\Omega^2+\delta^r{}_n\,\nabla_a\,\Omega^2-g_{an}\,g^{rs}\,\nabla_s\,\Omega^2]\,F_{mr}\\ &= -\frac{1}{2\,\Omega^6}\,[g^{rm}\,F_{mr}\,g^{bn}\,\nabla_n\,\Omega^2+F^{ab}\nabla_a\,\Omega^2-F^{bs}\nabla_s\,\Omega^2]\\ &= -\frac{2}{\Omega^5}\,F^{ab}\,\nabla_a\,\Omega.\end{aligned}$$

So, as needed, we have

$$\nabla'_a\,(g'^{am}\,g'^{bn}\,F_{mn})=\frac{1}{\Omega^4}\,(\nabla_a F^{ab})+\frac{(n-4)}{\Omega^5}\,F^{ab}\,\nabla_a\,\Omega.$$

PROBLEM 2.6.2. *Prove equation (2.6.19).*

We have

$$\epsilon^{abcd}\,F_{ab}F_{cd}=\epsilon^{abcd}\left[2\,E_{[a}\,\xi_{b]}+\epsilon_{abrs}\,\xi^r\,B^s\right]\left[2\,E_{[c}\,\xi_{d]}+\epsilon_{cdmn}\,\xi^m\,B^n\right].$$

When we expand the right side, we get four terms. Two of them vanish because of the anti-symmetry of ϵ^{abcd}:

$$\begin{aligned}\epsilon^{abcd}\,E_{[a}\,\xi_{b]}\,E_{[c}\,\xi_{d]} &= \epsilon^{abcd}\,E_a\,\xi_b\,E_c\,\xi_d=0,\\ \epsilon^{abcd}\,\epsilon_{abrs}\,\xi^r\,B^s\,\epsilon_{cdmn}\,\xi^m\,B^n &= -4\,\delta^c{}_{[r}\,\delta^d{}_{s]}\,\xi^r\,B^s\,\epsilon_{cdmn}\,\xi^m\,B^n\\ &= -4\,\xi^c\,B^d\,\epsilon_{cdmn}\,\xi^m\,B^n=0.\end{aligned}$$

One of the other terms yields

$$\begin{aligned}2\,\epsilon^{abcd}\,E_{[a}\,\xi_{b]}\,\epsilon_{cdmn}\,\xi^m\,B^n &= 2\,\epsilon^{cdab}\,\epsilon_{cdmn}\,E_a\,\xi_b\,\xi^m\,B^n\\ &= -8\,\delta^a{}_{[m}\,\delta^b{}_{n]}\,E_a\,\xi_b\,\xi^m\,B^n=4\,E^a\,B_a,\end{aligned}$$

since $\xi^a\,E_a=\xi^a\,B_a=0$. The other yields $4\,E^a\,B_a$ as well. (The computation is almost exactly the same.) So we have

$$\epsilon^{abcd}\,F_{ab}F_{cd}=8\,E^a\,B_a.$$

PROBLEM 2.6.3. *Prove equation (2.6.21).*

By equation (2.6.17), we have

$$\begin{aligned}(T_{ab}\,\xi^b)(T^{ac}\,\xi_c) &= \left[\frac{1}{2}(-E^n\,E_n-B^n\,B_n)\,\xi_a-\epsilon_{ars}\,E^r\,B^s\right]\\ &\qquad\left[\frac{1}{2}(-E^m\,E_m-B^m\,B_m)\,\xi^a-\epsilon^{apq}\,E_p\,B_q\right].\end{aligned}$$

The two "cross-terms" on the right vanish because $\xi_a \epsilon^{apq} = \xi_a \epsilon^{apqn}\, \xi_n = \mathbf{0}$. So

$$(T_{ab}\, \xi^b)(T^{ac}\, \xi_c) = \frac{1}{4}(E^n E_n + B^n B_n)^2 + \epsilon_{ars}\, E^r B^s\, \epsilon^{apq}\, E_p B_q.$$

But

$$\begin{aligned}\epsilon_{ars}\, E^r B^s\, \epsilon^{apq}\, E_p B_q &= \epsilon_{arsn}\, \xi^n\, \epsilon^{apqm}\, \xi_m\, E^r B^s E_p B_q = -6\, E^{[p}\, B^q\, \xi^{m]}\, E_p\, B_q\, \xi_m \\ &= -\left[(E^p E_p)(B^q B_q) - (E^p B_p)^2\right].\end{aligned}$$

So we may conclude, as required, that

$$(T_{ab}\, \xi^b)(T^{ac}\, \xi_c) = \frac{1}{4}(E^n E_n - B^n B_n)^2 + (E^n B_n)^2.$$

PROBLEM 2.6.4. *Prove the following equivalence.*

$$\nabla_a\, F^{ab} = J^b \Longleftrightarrow \begin{cases} D_b\, E^b & = \quad \mu \\ \epsilon^{abc}\, D_b\, B_c & = \quad \xi^b\, \nabla_b\, E^a + j^a. \end{cases}$$

Clearly, $(\nabla_a\, F^{ab} - J^b)$ vanishes iff its projections tangent to, and orthogonal to, ξ^b both vanish; i.e.,

$$\nabla_a\, F^{ab} = J^b \Longleftrightarrow \begin{cases} \xi_b\, (\nabla_a\, F^{ab} - J^b) & = \quad \mathbf{0} \\ h^c{}_b\, (\nabla_a\, F^{ab} - J^b) & = \quad \mathbf{0}. \end{cases}$$

We shall work on the right-side equations separately. Since ξ^a (and, therefore, h_{ab}) are constant, and since E^a is orthogonal to ξ^a,

$$\begin{aligned}\xi_b\, (\nabla_a\, F^{ab} - J^b) &= \nabla_a\, (F^{ab} \xi_b) - (J^b \xi_b) = \nabla_a\, E^a - \mu = \nabla_a\, (h^a{}_n\, h^n{}_m\, E^m) - \mu \\ &= h^a{}_b\, h^b{}_m\, \nabla_a\, E^m - \mu = D_b\, E^b - \mu.\end{aligned}$$

This gives us the first equivalence. The second is handled similarly using equations (2.6.12) and (2.6.13). We have

$$h^c{}_b\, (\nabla_a\, F^{ab} - J^b) = \nabla_a\, (F^{ab} h^c{}_b) - (\, J^b h^c{}_b) = \nabla_a\, (F^{ab} h^c{}_b) - j^c$$

and

$$\begin{aligned}\nabla_a\, (F^{ab} h^c{}_b) &= \nabla_a \left[(2\, E^{[a}\, \xi^{b]} + \epsilon^{abrs}\, \xi_r\, B_s)\, h^c{}_b\right] = -\xi^a \nabla_a\, E^c + \epsilon^{abrs}\, \xi_r\, h^c{}_b\, \nabla_a\, B_s \\ &= -\xi^a \nabla_a\, E^c + \epsilon^{cas}\, \nabla_a\, B_s = -\xi^a \nabla_a\, E^c + (\epsilon^{cmn}\, h^a{}_m\, h^s{}_n)\, \nabla_a\, B_s \\ &= -\xi^a \nabla_a\, E^c + \epsilon^{cmn}\, D_m\, B_n.\end{aligned}$$

So

$$h^c{}_b\, (\nabla_a\, F^{ab} - J^b) = \mathbf{0} \Longleftrightarrow \epsilon^{cmn}\, D_m\, B_n = \xi^a \nabla_a\, E^c + j^c.$$

PROBLEM 2.7.1. *Show that in the general case ($n \geq 3$), inversion of equation (2.7.3) leads to*

$$R_{ab} = 8\pi \left(T_{ab} - \frac{1}{(n-2)}\, T\, g_{ab} \right) - \frac{2}{(n-2)}\, \Lambda\, g_{ab}.$$

Contraction of

$$R_{ab} - \frac{1}{2}\, R g_{ab} - \Lambda\, g_{ab} = 8\pi\, T_{ab}$$

yields

$$R - \frac{1}{2}\, R n - \Lambda\, n = 8\pi\, T,$$

or, equivalently,

$$\frac{(2-n)}{2}\, R = 8\pi\, T + n\,\Lambda.$$

So, substitution for R in the first equation yields

$$\begin{aligned} R_{ab} &= 8\pi\, T_{ab} + \frac{1}{2}\, R g_{ab} + \Lambda\, g_{ab} \\ &= 8\pi\, T_{ab} + \frac{1}{(2-n)}\, (8\pi\, T + n\,\Lambda)\, g_{ab} + \Lambda\, g_{ab} \\ &= 8\pi \left(T_{ab} - \frac{1}{(n-2)}\, T\, g_{ab} \right) - \frac{2}{(n-2)}\, \Lambda\, g_{ab}, \end{aligned}$$

as required.

PROBLEM 2.7.2. *Give examples of the following.*

(1) A smooth symmetric field T_{ab} that satisfies the SDEC (and so also the WEC and DEC) but not the SEC

(2) A smooth symmetric field T_{ab} that satisfies the SEC, but not the WEC (and so not the DEC or SDEC, either)

For (1), take $T_{ab} = g_{ab}$. It satisfies the SDEC. But in this case, $T_{ab} - \frac{1}{2} g_{ab} T = -g_{ab}$, and so it does not satisfy the SEC.

For (2), take $T_{ab} = -g_{ab}$. It does not satisfy the WEC. But in this case, $T_{ab} - \frac{1}{2} g_{ab} T = g_{ab}$, so it *does* satisfy the SEC.

PROBLEM 2.7.3. *Consider a perfect fluid with four-velocity η^a, energy density ρ, and pressure p. Show that it satisfies the strong energy condition iff $(\rho + p) \geq 0$ and $(\rho + 3p) \geq 0$.*

If

$$T_{ab} = \rho\, \eta_a\, \eta_b - p\, (g_{ab} - \eta_a\, \eta_b),$$

then $T = (\rho - 3\,p)$, and

$$T_{ab} - \frac{1}{2}\, g_{ab}\, T = (\rho + p)\, \eta_a\, \eta_b + \frac{(p - \rho)}{2}\, g_{ab}.$$

It follows that T_{ab} satisfies the SEC iff, given any unit timelike vector ξ^a at any point,

$$(\rho + p)(\eta_a \xi^a)^2 + \frac{(p - \rho)}{2} \geq 0.$$

Now $(\eta_a \xi^a)^2 \geq 1$ by the Schwarz inequality. So if $(\rho + p) \geq 0$ and $(\rho + 3p) \geq 0$, then

$$(\rho + p)(\eta_a \xi^a)^2 + \frac{(p - \rho)}{2} \geq (\rho + p) + \frac{(p - \rho)}{2} = \frac{(\rho + 3p)}{2} \geq 0,$$

and the inequality is satisfied. Conversely, suppose it is satisfied for all unit timelike vectors ξ^a at some point. Then, in particular, it is satisfied for $\xi^a = \eta^a$, which yields $(\rho + 3p) \geq 0$. And since $(\eta_a \xi^a)^2$ can assume arbitrarily large values as ξ^a ranges over all unit timelike vectors at a point, it must be the case that $(\rho + p) \geq 0$.

PROBLEM 2.8.1. *Prove equation (2.8.8).*

It follows from the definition (2.8.6) of the twist vector that

$$\begin{aligned} \epsilon_{abcd}\, \xi^c \omega^d &= \frac{1}{2}\, \epsilon_{abcd}\, \xi^c \epsilon^{dmnr}\, \xi_m\, \omega_{nr} = 3\, \delta^m{}_{[a}\, \delta^n{}_b\, \delta^r{}_{c]}\, \xi^c\, \xi_m\, \omega_{nr} \\ &= 3\, \xi^c\, \xi_{[a}\, \omega_{bc]} = \omega_{ab}. \end{aligned}$$

For the final equality, we use the fact that ξ^a is orthogonal to ω_{ab} in both indices (and ω_{ab} is anti-symmetric).

PROBLEM 2.8.2. *Show that, at any point,* $\omega^a = \mathbf{0}$ *iff* $\xi_{[a} \nabla_b\, \xi_{c]} = \mathbf{0}$.

We know from equation (2.8.7) and the anti-symmetry of ϵ^{abcd} that

$$\omega^a \;=\; \frac{1}{2}\, \epsilon^{abcd}\, \xi_b\, \nabla_c\, \xi_d = \frac{1}{2}\, \epsilon^{abcd}\, \xi_{[b}\, \nabla_c\, \xi_{d]}.$$

So the "if" half of the equivalence follows immediately. For the other direction, assume that $\omega^a = \mathbf{0}$ holds at some point. Then at that point we have

$$\begin{aligned} \mathbf{0} = \epsilon_{amnr}\, \omega^a &= \frac{1}{2}\, \epsilon_{amnr}\, \epsilon^{abcd}\, \xi_{[b}\, \nabla_c\, \xi_{d]} = -3\, \delta^b{}_{[m}\, \delta^c{}_n\, \delta^d{}_{r]}\, \xi_{[b}\, \nabla_c\, \xi_{d]} \\ &= -3\, \xi_{[m}\, \nabla_n\, \xi_{r]}. \end{aligned}$$

PROBLEM 2.8.3. *Complete the calculation in equation (2.8.10).*

We have to compute

$$\frac{\omega_b{}^a\,\eta^b\,\omega_{ca}\,\eta^c}{\rho^n\,\rho_n}.$$

We work separately with the numerator and denominator. It follows, first, from equation (2.8.8) that

$$\begin{aligned}
\omega_b{}^a\,\eta^b\,\omega_{ca}\,\eta^c &= (\epsilon_b{}^a{}_{mn}\,\xi^m\,\omega^n)\,\eta^b\,(\epsilon_{cars}\,\xi^r\,\omega^s)\,\eta^c \\
&= \epsilon^{abmn}\,\eta_b\,\xi_m\,\omega_n\,\epsilon_{acrs}\,\eta^c\,\xi^r\,\omega^s \\
&= -6\,\delta^b{}_{[c}\,\delta^m{}_r\,\delta^n{}_{s]}\,\eta_b\,\xi_m\,\omega_n\,\eta^c\,\xi^r\,\omega^s \\
&= -6\;\eta_{[c}\,\xi_r\,\omega_{s]}\,\eta^c\,\xi^r\,\omega^s \\
&= -\left[(\eta_c\eta^c)(\omega_s\omega^s) - (\eta_s\,\omega^s)^2\right].
\end{aligned}$$

(For the final equality, we use the fact that we are doing the computation at the "initial point" where η^a is orthogonal to ξ^a.) And, since

$$\rho^n = \eta^n - \frac{\eta^b\,\omega_b}{\omega^m\,\omega_m}\,\omega^n,$$

we have

$$\begin{aligned}
\rho^n\rho_n &= \left[\eta^n - \frac{\eta^b\,\omega_b}{\omega^m\,\omega_m}\,\omega^n\right]\left[\eta_n - \frac{\eta^c\,\omega_c}{\omega^r\,\omega_r}\,\omega_n\right] \\
&= (\eta^n\,\eta_n) - 2\,\frac{(\eta^c\,\omega_c)^2}{\omega^r\,\omega_r} + \frac{(\eta^c\,\omega_c)^2}{\omega^r\,\omega_r} \\
&= \frac{1}{(\omega^r\,\omega_r)}\,\left[(\eta^n\eta_n)(\omega^s\,\omega_s) - (\eta^c\,\omega_c)^2\right].
\end{aligned}$$

So,

$$\frac{\omega_b{}^a\,\eta^b\,\omega_{ca}\,\eta^c}{\rho^n\,\rho_n} = -\omega^r\omega_r,$$

as required.

PROBLEM 2.9.1. *Let κ^a be a timelike Killing field that is locally hypersurface orthogonal ($\kappa_{[a}\nabla_b\,\kappa_{c]} = \mathbf{0}$). Further, let κ be the length of κ^a. (So $\kappa^2 = \kappa^n\kappa_n$.) Show that*

$$\kappa^2\,\nabla_a\,\kappa_b = -\kappa_{[a}\nabla_{b]}\,\kappa^2.$$

This follows with a simple direct computation:

$$\begin{aligned}\mathbf{0} &= 3\,\kappa^c\,\kappa_{[a}\nabla_b\,\kappa_{c]} = \kappa_a\,\kappa^c\nabla_b\,\kappa_c + (\kappa^c\,\kappa_c)\,\nabla_a\,\kappa_b + \kappa_b\,\kappa^c\nabla_c\,\kappa_a\\ &= \kappa_a\,\kappa^c\nabla_b\,\kappa_c + (\kappa^c\,\kappa_c)\,\nabla_a\,\kappa_b - \kappa_b\,\kappa^c\nabla_a\,\kappa_c\\ &= \frac{1}{2}\,\kappa_a\,\nabla_b\,\kappa^2 + \kappa^2\,\nabla_a\,\kappa_b - \frac{1}{2}\,\kappa_b\,\nabla_a\,\kappa^2 = \kappa^2\,\nabla_a\,\kappa_b + \kappa_{[a}\nabla_{b]}\,\kappa^2.\end{aligned}$$

PROBLEM 2.9.2. *Consider a non-trivial boost Killing field $\kappa_b = 2\chi^a E_{[a}\,\xi_{b]}$ on Minkowksi spacetime (as determined relative to some point p and some constant unit timelike field ξ^a). "Non-trivial" here means that $E^a \neq \mathbf{0}$. Let η^a be a constant field on Minkowski spacetime. Show that $\pounds_\kappa\,\eta^a = \mathbf{0}$ iff η^a is orthogonal to both to ξ^a and E^a.*

Since η^a is constant,

$$\begin{aligned}\pounds_\kappa\,\eta^a &= \kappa^n\nabla_n\,\eta^a - \eta^n\nabla_n\,\kappa^a = -\eta^n\nabla_n\,\kappa^a = -2\,\eta^n\nabla_n\,(\chi_m\,E^{[m}\,\xi^{a]})\\ &= -2\,E^{[m}\,\xi^{a]}\,\eta^n\nabla_n\,\chi_m = -2\,E^{[m}\,\xi^{a]}\,\eta^n\,g_{nm} = (\xi^m\,\eta_m)E^a - (E^m\,\eta_m)\xi^a.\end{aligned}$$

Since ξ^a and E^a are linearly independent, we see that $\pounds_\kappa\,\eta^a = \mathbf{0}$ iff $(\xi^m\,\eta_m) = \mathbf{0} = (E^m\,\eta_m)$.

PROBLEM 2.9.3. *This time, consider a non-trivial rotational Killing field $\kappa_b = \chi^a\,\epsilon_{abcd}\,\xi^c\,B^d$ on Minkowski spacetime (with $B^a \neq \mathbf{0}$). Again, let η^a be a constant field on Minkowski spacetime. Show that $\pounds_\kappa\,\eta^a = \mathbf{0}$ iff η^a is a linear combination of ξ^a and B^a.*

The argument is very much the same as with the preceding problem. If η^a is constant,

$$\begin{aligned}\pounds_\kappa\,\eta^a &= -\eta^n\nabla_n\,\kappa^a = -\eta^n\nabla_n\,(\chi_m\,\epsilon^{ma}{}_{cd}\,\xi^c\,B^d)\\ &= -\epsilon^{ma}{}_{cd}\,\xi^c\,B^d\,\eta^n\nabla_n\,\chi_m = -\,\epsilon^{ma}{}_{cd}\,\xi^c\,B^d\,\eta^n\,g_{nm} = \epsilon^a{}_{mcd}\,\eta^m\,\xi^c\,B^d.\end{aligned}$$

Thus $\pounds_\kappa\,\eta^a = \mathbf{0}$ iff η^a has no component orthogonal to both ξ^a and B^a.

PROBLEM 2.9.4. *Let κ^a be a Killing field; let $\gamma : I \to M$ be a smooth, future-directed, timelike curve, with unit tangent field ξ^a; and let $J = (P^a\kappa_a)$, where $P^a = m\,\xi^a$. Finally, let $\alpha^a = \xi^n\nabla_n\xi^a$ and $\alpha = (-\alpha^n\,\alpha_n)^{\frac{1}{2}}$. Show that*

$$|\xi^n\nabla_n\,J| \leq \alpha\,\sqrt{J^2 - m^2\,(\kappa^n\kappa_n)}.$$

We have seen that

$$\xi^n \nabla_n J = m\, \kappa_a\, \xi^n \nabla_n\, \xi^a = m\, \kappa_a\, \alpha^a.$$

Now consider the projected spatial metric $h_{ab} = g_{ab} - \xi_a \xi_b$. It is negative definite. So by the Schwarz inequality (as applied to $-h_{ab}$) and the fact that $\xi^a \alpha_a = \mathbf{0}$,

$$\begin{aligned} |\xi^n \nabla_n J| &= |m\, \kappa_a\, \alpha^a| = |m\, h_{ab}\, \alpha^a\, \kappa^b| \leq (-\,h_{ab}\, \alpha^a\, \alpha^b)^{\frac{1}{2}}\, (-\,m^2\, h_{ab}\, \kappa^a\, \kappa^b)^{\frac{1}{2}} \\ &= \alpha\, [J^2 - m^2\, (\kappa^n \kappa_n)]^{\frac{1}{2}}. \end{aligned}$$

PROBLEM 2.11.1. *Confirm that the three stated solutions do, in fact, satisfy equation (2.11.18).*

We consider just the $k = -1$ case. The others are handled similarly. We have to show that the solution (in parametric form),

$$a(x) = \frac{C}{2}\,(\cosh x - 1)$$

$$t(x) = \frac{C}{2}\,(\sinh x - x)$$

does, in fact, satisfy equation (2.11.18) for all $x \in (0, \infty)$. Note that (dt/dx) is strictly positive in this interval. So by the inverse function theorem, (dx/dt) is everywhere well defined and equal to $(dt/dx)^{-1}$. Thus, we have

$$\dot{a} = \frac{da}{dt} = \frac{da}{dx}\left(\frac{dt}{dx}\right)^{-1} = \frac{\sinh x}{\cosh x - 1}.$$

Therefore

$$\dot{a}^2 - \frac{C}{a} - 1 = \left(\frac{\sinh x}{\cosh x - 1}\right)^2 - \frac{2}{(\cosh x - 1)} - 1 = 0.$$

PROBLEM 2.11.2. *Consider a second equation of state, namely that in which $\rho = 3\,p$. Show that in this case there is a number C' such that*

$$\dot{a}^2\, a^2 + k\, a^2 = \frac{8\pi}{3}\rho\, a^4 = C'.$$

If we multiply the right side of equation (2.11.16) by 3, and equate it with the right side of equation (2.11.15), we arrive at

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2} = -2\,\frac{\ddot{a}}{a} - \left(\frac{\dot{a}}{a}\right)^2 - \frac{k}{a^2}$$

or, equivalently,

$$\ddot{a}\,a + \dot{a}^2 + k = 0.$$

It follows (by integration) that $\dot{a}^2\,a^2 + k\,a^2 = C'$, for some number C'. It then further follows from equation (2.11.15) that $C' = (8\,\pi/3)\rho \mathrm{a}^4$.

Bibliography

[1] M. Abramowicz and J. Lasota. On traveling around without feeling it and uncurving curves. *American Journal of Physics*, 54:936–939, 1986.

[2] M. Abramowicz, P. Nurowski, and N. Wex. Optical reference geometry for stationary and axially symmetric spacetimes. *Classical and Quantum Gravity*, 12:1467–1472, 1995.

[3] A. Ashtekar and A. Magnon. The Sagnac effect in general relativity. *Journal of Mathematical Physics*, 16:341–344, 1975.

[4] J. Bain. Theories of Newtonian gravity and empirical distinguishability. *Studies in the History and Philosophy of Modern Physics*, 35:345–376, 2004.

[5] E. Cartan. Sur les variétés a connexion affine et la théorie de la relativité généralisée. *Annales Scientifiques de l'Ecole Normale Supérieure*, 40:325–412, 1923.

[6] E. Cartan. Sur les variétés a connexion affine et la théorie de la relativité généralisée. *Annales Scientifiques de l'Ecole Normale Supérieure*, 41:1–25, 1924.

[7] S. Chakrabarti, R. Geroch, and Can bin Liang. Timelike curves of limited acceleration in general relativity. *Journal of Mathematical Physics*, 24:597–598, 1983.

[8] Y. Choquet-Bruhat and R. Geroch. Global aspects of the Cauchy problem in general relativity. *Communications in Mathematical Physics*, 14:329–335, 1969.

[9] R. Coleman and H. Korté. Hermann Weyl: Mathematician, physicist, philosopher. In E. Scholz, editor, *Hermann Weyl's Raum-Zeit-Materie and a General Introduction to His Scientific Work*. Birkhäuser Verlag, 2001.

[10] G. Dautcourt. Die Newtonische Gravitationstheorie als strenger Grenzfall der allgemeinen Relativitätstheorie. *Acta Physica Polonica*, 65:637–646, 1964.

[11] W. G. Dixon. On the uniqueness of the Newtonian theory as a geometric theory of gravitation. *Communications in Mathematical Physics*, 45:167–182, 1975.

[12] W. G. Dixon. *Special Relativity*. Cambridge University Press, 1978.

[13] H. Dombrowski and K. Horneffer. Die Differentialgeometrie des Galileischen Relativitätsprinzips. *Mathematische Zeitscrift*, 86:291–311, 1964.

[14] J. Earman. Lambda: The constant that refuses to die. *Archives for History of Exact Sciences*, 55:189–220, 2001.

[15] J. Ehlers. Über den Newtonschen Grenzwert der Einsteinschen Gravitationstheorie. In J. Nitsch, J. Pfarr, and E. W. Stachow, editors, *Grundlagen Probleme der Modernen Physik*. Wissenschaftsverlag, 1981.

[16] J. Ehlers and R. Geroch. Equation of motion of small bodies in relativity. *Annals of Physics*, 309:232–236, 2004.

[17] G. F. R. Ellis. Relativistic cosmology. In R. K. Sachs, editor, *General Relativity and Cosmology*. Academic Press, 1971.

[18] G. F. R. Ellis. Contributions of K. Gödel to relativity and cosmology. In P. Hayek, editor, *Gödel '96: Logical Foundations of Mathematics, Computer Science and Physics—Kurt Gödel's Legacy (Springer Lecture Notes in Logic 6)*. Springer Verlag, 1996.

[19] D. Farnsworth and R. Kerr. Homogeneous dust-filled cosmological solutions. *Journal of Mathematical Physics*, 7:1625–1632, 1966.

[20] H. Flanders. *Differential Forms*. Academic Press, 1963.

[21] K. Friedrichs. Eine Invariante Formulierung des Newtonschen Gravitationsgesetzes und der Grenzüberganges vom Einsteinschen zum Newtonschen Gesetz. *Mathematische Annalen*, 98:566–575, 1927.

[22] R. Geroch. Differential geometry. Unpublished lecture notes, University of Chicago, http://home.uchicago.edu/~geroch/Links_to_Notes.html.

[23] R. Geroch. Spacetime structure from a global viewpoint. In R. K. Sachs, editor, *General Relativity and Cosmology*. Academic Press, 1971.

[24] R. Geroch and P. S. Jang. Motion of a body in general relativity. *Journal of Mathematical Physics*, 16:65–67, 1975.

[25] K. Gödel. An example of a new type of cosmological solutions [sic] of Einstein's field equations of gravitation. *Reviews of Modern Physics*, 21:447–450, 1949.

[26] K. Gödel. Rotating universes in general relativity theory. In *Proceedings of the International Congress of Mathematicians: Cambridge, Massachusetts, U.S.A. August 30–September 6, 1950*, vol. 1. American Mathematical Society, 1952.

[27] K. Gödel. Lecture on rotating universes. In S. Feferman, J. W. Dawson Jr., W. Goldfarb, C. Parsons, and R. M. Solovay, editors, *Kurt Gödel, Collected Papers*, vol. 3. Oxford University Press, 1995.

[28] P. Havas. Four-dimensional formulations of Newtonian mechanics and their relation to the special and general theory of relativity. *Reviews of Modern Physics*, 36:938–965, 1964.

[29] S. Hawking, A. R. King, and P. J. McCarthy. A new topology for curved space-time which incorporates the causal, differential, and conformal structure. *Journal of Mathematical Physics*, 17:174–181, 1976.

[30] S. W. Hawking and G. F. R. Ellis. *The Large Scale Structure of Spacetime*. Cambridge University Press, 1972.

[31] O. Heckmann and E. Schücking. Bemerkungen zur Newtonschen Kosmologie I. *Zeitschrift für Astrophysik*, 38:95–109, 1955.

[32] I. Herstein. *Topics in Algebra*. Blaisdell Publishing, 1964.

[33] E. H. Kronheimer and R. Penrose. On the structure of causal spaces. *Proceedings of the Cambridge Philosophical Society*, 63:481–501, 1967.

[34] H. Künzle. Galilei and Lorentz structures on space-time: Comparison of the corresponding geometry and physics. *Annales Institute Henri Poincaré*, 17:337–362, 1972.

[35] H. Künzle. Covariant Newtonian limts of Lorentz space-times. *General Relativity and Gravitation*, 7:445–457, 1976.

[36] S. Lang. *Linear Algebra*. Springer Verlag, 1987.

[37] M. Lottermoser. Über den Newtonschen Grenzwert der Allgemeinen Relativitätstheorie und die relativistische Erweiterung Newtonscher Anfangsdaten. Unpublished dissertation, 1988.

[38] D. Malament. The class of continuous timelike curves determines the topology of spacetime. *Journal of Mathematical Physics*, 18:1399–1404, 1977.

[39] D. Malament. Introductory essay to a previously unpublished lecture by Kurt Gödel on "rotating universes." In S. Feferman, J. W. Dawson Jr., W. Goldfarb, C. Parsons, and R. M. Solovay, editors, *Kurt Gödel, Collected Papers*, vol. 3. Oxford University Press, 1995.

[40] D. Malament. Is Newtonian cosmology really inconsistent? *Philosophy of Science*, 62: 489–510, 1995.

[41] D. Malament. A no-go theorem about rotation in relativity theory. In D. Malament, editor, *Reading Natural Philosophy, A Festschrift for Howard Stein*. Open Court Press, 2003.

[42] D. Malament. Classical general relativity. In J. Butterfied and J. Earman, editors, *Philosophy of Physics*. Elsevier, 2006.

[43] J. Norton. A paradox in Newtonian gravitation theory. In M. Forbes, D. Hull, and K. Okruhlik, editors, *PSA 1992: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, vol. 2. Philosophy of Science Association, 1993.

[44] J. Norton. The force of Newtonian cosmology: Acceleration is relative. *Philosophy of Science*, 62:511–522, 1995.

[45] J. Norton. The cosmological woes of Newtonian gravitation theory. In H. Goenner, J. Renn, J. Ritter, and T. Sauer, editors, *The Expanding Worlds of General Relativity*. Birkhäuser, 1999.

[46] B. O'Neill. *Semi-Riemannian Geometry with Applications to Relativity*. Academic Press, 1983.

[47] B. O'Neill. *The Geometry of Kerr Black Holes*. A. K. Peters, 1995.

[48] I. Ozsváth. New homogeneous solutions of Einstein's field equations with incoherent matter obtained by a spinor technique. *Journal of Mathematical Physics*, 6:590–610, 1965.

[49] D. Page. How big is the universe today? *General Relativity and Gravitation*, 15:181–185, 1983.

[50] D. Page. Maximal acceleration is non-rotating. *Classical and Quantum Gravitation*, 15:1669–1719, 1998.

[51] R. Penrose and W. Rindler. *Spinors and Spacetime*, vol. 1: *Two-Spinor Calculus and Relativistic Fields*. Cambridge University Press, 1984.

[52] W. Reynolds. Hyperbolic geometry on a hyperboloid. *American Mathematical Monthly*, 100:442–455, 1993.

[53] W. Rindler. Public and private space curvature in Robertson-Walker universes. *General Relativity and Gravitation*, 13:457–461, 1981.

[54] W. Rindler. *Relativity: Special, General, and Cosmological*. Oxford University Press, 2nd edition, 2001.

[55] R. D. Sorkin. A specimen of theory construction from quantum gravity. In J. Leplin, editor, *The Creation of Ideas in Physics*. Kluwer, 1995.

[56] R. D. Sorkin. Causal sets: Discrete gravity. In A. Gomberoff and D. Marolf, editors, *Lectures on Quantum Gravity (Proceedings of the Valdivia Summer School)*. Plenum, 2005.

[57] M. Spivak. *A Comprehensive Introduction to Differential Geometry*. Publish or Perish Press, 2nd edition, 1979.

[58] H. Stein. On the paradoxical time-structures of Gödel. *Philosophy of Science*, 37:589–601, 1970.

[59] A. Trautman. Foundations and current problems of general relativity. In S. Deser and K. W. Ford, editors, *Lectures on General Relativity*. Prentice-Hall, 1965.

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

[61] J. Weatherall. The motion of a body in Newtonian theories. arXiv:1010.0379v1. *Journal of Mathematical Physics*, 52, 2011, 16 pp.

[62] H. Weyl. Zur Infinitesimalgeometrie: Einordnung der projecktiven und konformen Auffassung. *Nachrichten der Königlichen Gesellschaft der Wissenschaften zu Göttingen; Mathematisch-Physikalische Klasse*, 1921, 99–112, 1921.

[63] H. Weyl. *Philosophy of Mathematics and Natural Science*. Princeton University Press, 2nd edition, 1950.

[64] J. Wolf. *Spaces of Constant Curvature*. Publish or Perish Press, 5th edition, 1984.