Graduate Texts in Mathematics

Peter Petersen

Riemannian Geometry

Springer

Graduate Texts in Mathematics **171**

Springer Science+Business Media, LLC

Graduate Texts in Mathematics

1 TAKEUTI/ZARING. Introduction to Axiomatic Set Theory. 2nd ed.
2 OXTOBY. Measure and Category. 2nd ed.
3 SCHAEFER. Topological Vector Spaces.
4 HILTON/STAMMBACH. A Course in Homological Algebra. 2nd ed.
5 MAC LANE. Categories for the Working Mathematician.
6 HUGHES/PIPER. Projective Planes.
7 SERRE. A Course in Arithmetic.
8 TAKEUTI/ZARING. Axiomatic Set Theory.
9 HUMPHREYS. Introduction to Lie Algebras and Representation Theory.
10 COHEN. A Course in Simple Homotopy Theory.
11 CONWAY. Functions of One Complex Variable I. 2nd ed.
12 BEALS. Advanced Mathematical Analysis.
13 ANDERSON/FULLER. Rings and Categories of Modules. 2nd ed.
14 GOLUBITSKY/GUILLEMIN. Stable Mappings and Their Singularities.
15 BERBERIAN. Lectures in Functional Analysis and Operator Theory.
16 WINTER. The Structure of Fields.
17 ROSENBLATT. Random Processes. 2nd ed.
18 HALMOS. Measure Theory.
19 HALMOS. A Hilbert Space Problem Book. 2nd ed.
20 HUSEMOLLER. Fibre Bundles. 3rd ed.
21 HUMPHREYS. Linear Algebraic Groups.
22 BARNES/MACK. An Algebraic Introduction to Mathematical Logic.
23 GREUB. Linear Algebra. 4th ed.
24 HOLMES. Geometric Functional Analysis and Its Applications.
25 HEWITT/STROMBERG. Real and Abstract Analysis.
26 MANES. Algebraic Theories.
27 KELLEY. General Topology.
28 ZARISKI/SAMUEL. Commutative Algebra. Vol.I.
29 ZARISKI/SAMUEL. Commutative Algebra. Vol.II.
30 JACOBSON. Lectures in Abstract Algebra I. Basic Concepts.
31 JACOBSON. Lectures in Abstract Algebra II. Linear Algebra.
32 JACOBSON. Lectures in Abstract Algebra III. Theory of Fields and Galois Theory.
33 HIRSCH. Differential Topology.
34 SPITZER. Principles of Random Walk. 2nd ed.
35 WERMER. Banach Algebras and Several Complex Variables. 2nd ed.
36 KELLEY/NAMIOKA et al. Linear Topological Spaces.
37 MONK. Mathematical Logic.
38 GRAUERT/FRITZSCHE. Several Complex Variables.
39 ARVESON. An Invitation to C^*-Algebras.
40 KEMENY/SNELL/KNAPP. Denumerable Markov Chains. 2nd ed.
41 APOSTOL. Modular Functions and Dirichlet Series in Number Theory. 2nd ed.
42 SERRE. Linear Representations of Finite Groups.
43 GILLMAN/JERISON. Rings of Continuous Functions.
44 KENDIG. Elementary Algebraic Geometry.
45 LOÈVE. Probability Theory I. 4th ed.
46 LOÈVE. Probability Theory II. 4th ed.
47 MOISE. Geometric Topology in Dimensions 2 and 3.
48 SACHS/WU. General Relativity for Mathematicians.
49 GRUENBERG/WEIR. Linear Geometry. 2nd ed.
50 EDWARDS. Fermat's Last Theorem.
51 KLINGENBERG. A Course in Differential Geometry.
52 HARTSHORNE. Algebraic Geometry.
53 MANIN. A Course in Mathematical Logic.
54 GRAVER/WATKINS. Combinatorics with Emphasis on the Theory of Graphs.
55 BROWN/PEARCY. Introduction to Operator Theory I: Elements of Functional Analysis.
56 MASSEY. Algebraic Topology: An Introduction.
57 CROWELL/FOX. Introduction to Knot Theory.
58 KOBLITZ. p-adic Numbers, p-adic Analysis, and Zeta-Functions. 2nd ed.
59 LANG. Cyclotomic Fields.
60 ARNOLD. Mathematical Methods in Classical Mechanics. 2nd ed.

continued after index

Peter Petersen

Riemannian Geometry

With 60 Illustrations

Springer

Peter Petersen
Department of Mathematics
University of California, Los Angeles
Los Angeles, CA 90095-1555
USA

Mathematics Subject Classification (1991): 53-01, 53C20

Library of Congress Cataloging-in-Publication Data
Petersen, Peter, 1962–
Riemannian geometry/Peter Petersen.
p. cm. – (Graduate texts in mathematics; 171)
Includes bibliographical references (p. -) and index.
ISBN 978-1-4757-6436-9 ISBN 978-1-4757-6434-5 (eBook)
DOI 10.1007/978-1-4757-6434-5
1. Geometry, Riemannian. I. Title. II. Series.
QA649.P386 1997
516.3'73–dc21 97-5786

Printed on acid-free paper.

Production managed by Francine McNeill; manufacturing supervised by Johanna Tschebull.
Photocomposed copy prepared from the author's LaTeX files using Springer's svsing.sty macro.

9 8 7 6 5 4 3 2 1

ISBN 978-1-4757-6436-9 SPIN 10572067

To my wife, Laura

Preface

This book is meant to be an introduction to Riemannian geometry. The reader is assumed to have some knowledge of standard manifold theory, including basic theory of tensors, forms, and Lie groups. At times we shall also assume familiarity with algebraic topology and de Rham cohomology. Specifically, we recommend that the reader is familiar with texts like [14] or [76, vol. 1]. For the readers who have only learned something like the first two chapters of [65], we have an appendix which covers Stokes' theorem, Čech cohomology, and de Rham cohomology. The reader should also have a nodding acquaintance with ordinary differential equations. For this, a text like [59] is more than sufficient. Most of the material usually taught in basic Riemannian geometry, as well as several more advanced topics, is presented in this text. Many of the theorems from Chapters 7 to 11 appear for the first time in textbook form. This is particularly surprising as we have included essentially only the material students of Riemannian geometry must know.

The approach we have taken deviates in some ways from the standard path. First and foremost, we do not discuss variational calculus, which is usually the sine qua non of the subject. Instead, we have taken a more elementary approach that simply uses standard calculus together with some techniques from differential equations.

We emphasize throughout the text the importance of using the correct type of coordinates depending on the theoretical situation at hand. First, we develop our substitute for the second variation formula by using adapted coordinates. These are coordinates naturally associated to a distance function. If, for example, we use the function that measures the distance to a point, then the adapted coordinates are nothing but polar coordinates. Next, we have exponential coordinates, which are of fundamental importance in showing that distance functions are smooth. Then distance coordinates are used first to show that distance-preserving maps are smooth, and then later to give good coordinate systems in which the metric is sufficiently controlled so that one can prove, say, Cheeger's finiteness theorem. Finally, we have harmonic coordinates. These coordinates have some magic properties. One in particular is that in such coordinates the Ricci curvature is essentially the Laplacian of the metric. Our motivation for this treatment has been that examples become a natural and integral part of the text rather than a separate item that much too often is forgotten. Another desirable by-product has been that one actually gets the feeling that gradients, Hessians, Laplacians, curvatures, and many other things are actually computable. Often these concepts are simply abstract notions that are pushed around for fun.

From a more physical viewpoint, the reader will get the idea that we are simply using the Hamilton-Jacobi equations rather than the Euler-Lagrange equations to develop Riemannian geometry (see [4] for an explanation of these matters). It is simply a matter of taste which path one wishes to follow, but surprisingly, the Hamilton-Jacobi approach has never been tried systematically in Riemannian geometry.

The book can be divided into five imaginary parts:

Part I: Tensor geometry, consisting of Chapters 1 to 4.

Part II: Classical geodesic geometry, consisting of Chapters 5 and 6.

Part III: Geometry à la Bochner and Cartan, consisting of Chapters 7 and 8.

Part IV: Comparison geometry, consisting of Chapters 9 to 11.

Appendices: de Rham cohomology, principal bundles, and spinors.

Chapters 1 to 8 give a pretty complete picture of some of the most classical results in Riemannian geometry, while Chapters 9 to 11 explain some of the more recent developments in Riemannian geometry. The individual chapters contain the following material:

Chapter 1: Riemannian manifolds, isometries, immersions, and submersions are defined. Homogeneous spaces and covering maps are also briefly mentioned. We have a discussion on various types of warped products, leading to an elementary account of why the Hopf fibration is also a Riemannian submersion.

Chapter 2: Many of the tensor constructions one needs on Riemannian manifolds are developed. First the Riemannian connection is defined, and it is shown how one can use the connection to define the classical notions of Hessian, Laplacian, and divergence on Riemannian manifolds. We proceed to define all of the important curvature concepts and discuss a few simple properties. Aside from these important tensor concepts, we also develop several important formulas that relate curvature and the underlying metric. These formulas are to some extent our replacement for the second variation formula. The chapter ends with a short section where such tensor operations as contractions, type changes, and inner products are briefly discussed.

Chapter 3: First, we set up some general situations where it is possible to compute the curvature tensor. The rest of the chapter is then devoted to carrying out this program in several concrete situations. The curvature tensor of spheres, product spheres, warped products, and doubly warped products is computed. This is used to exhibit some interesting examples that are Ricci flat and scalar flat. In particular, we explain how the Riemannian analogue of the Schwarzschild metric can be constructed. Several different models of hyperbolic spaces are mentioned. Finally, we compute the curvatures of the Berger spheres and use this information as our basis for finding the curvatures of the complex projective plane.

Chapter 4: Here we concentrate on the special case where the Riemannian manifold is a hypersurface in Euclidean space. In this situation, one gets some special relations between the curvatures. We give examples of simple Riemannian manifolds that cannot be represented as hypersurface metrics. Finally, we give a brief introduction to the Gauss-Bonnet theorem and its generalization to higher dimensions.

Chapter 5: The remaining foundational topics for Riemannian manifolds are developed in this chapter. These include parallel translation, geodesics, Riemannian manifolds as metric spaces, exponential maps, geodesic completeness versus metric completeness, and maximal domains on which the exponential map is an embedding.

Chapter 6: Some of the classical results we prove here are: classification of simply connected space forms, the Hadamard-Cartan theorem, Preissmann's theorem, Cartan's center of mass construction in nonpositive curvature and why it shows that the fundamental group of such spaces is torsion free, Bonnet's diameter estimate, and Synge's theorem.

Chapter 7: Many of the classical and more recent results that arise from the Bochner technique are explained. We look at Killing fields and harmonic 1-forms as Bochner did, and finally, discuss some generalizations to harmonic p-forms. For the more advanced audience, we have developed the language of Clifford multiplication for the study of p-forms, as we feel that it is an important way of treating this material. The last section contains some more exotic but also profound situations where the Bochner technique is applied to the curvature tensor. These last two sections can easily be skipped in a more elementary course. The Bochner technique gives many nice bounds on the topology of closed manifolds with nonnegative curvature. In the spirit of comparison geometry, we show how Betti numbers of nonnegatively curved spaces are bounded by the prototypical compact flat manifold: the torus.

The importance of the Bochner technique in Riemannian geometry cannot be sufficiently emphasized. It seems that time and again, when people least expect it, new important developments come out of this simple philosophy.

Chapter 8: Part of the theory of symmetric spaces and holonomy is developed. The standard representations of symmetric spaces as homogeneous spaces and via Lie algebras are explained. We prove Cartan's existence theorem for isometries. We explain how one can compute curvatures in general and make some concrete calculations on several of the Grassmann manifolds including complex projective space. Having done this, we define holonomy for general manifolds, and discuss the de Rham decomposition theorem and several corollaries of it. The above examples are used to give an idea of how one can classify symmetric spaces. Also, we show in the same spirit why symmetric spaces of (non)compact type have (nonpositive) nonnegative curvature operator. Finally, we present a brief overview of how holonomy and symmetric spaces are related with the classification of holonomy groups. This is used in a grand synthesis, with all that has been learned up to this point, to give Gallot and Meyer's classification of compact manifolds with nonnegative curvature operator. A few things from Chapter 9 are used in Chapter 8, namely Myers' theorem and the splitting theorem. However, their use is inessential, and they are there to tie this material together with some of the more geometrical constructions that come later.

Chapter 9: Manifolds with lower Ricci curvature bounds are investigated in further detail. First, we discuss volume comparison and its uses for Cheng's maximal diameter theorem. Then we investigate some interesting relationships between Ricci curvature and fundamental groups. The strong maximum principle for continuous functions is developed. This result is first used in a warm-up exercise to give a simple proof of Cheng's maximal diameter theorem. We then proceed to prove the Cheeger-Gromoll splitting theorem and discuss its consequences for manifolds with nonnegative Ricci curvature.

Chapter 10: Convergence theory is the main focus of this chapter. First, we introduce the weakest form of convergence: Gromov-Hausdorff convergence. This concept is often useful in many contexts as a way of getting a weak form of convergence. The real object is then to figure out what weak convergence implies, given some stronger side conditions. There is a section which breezes through Hölder spaces, Schauder's elliptic estimates, and harmonic coordinates. To facilitate the treatment of the stronger convergence ideas, we have introduced a norm concept for Riemannian manifolds. We hope that these norms will make the subject a little more digestible. The main idea of this chapter is to prove the Cheeger-Gromov convergence theorem, which is called the Convergence Theorem of Riemannian Geometry, and Anderson's generalizations of this theorem to manifolds with bounded Ricci curvature.

Chapter 11: In this chapter we prove some of the more general finiteness theorems that do not fall into the philosophy developed in Chapter 10. Initially, we discuss critical point theory and Toponogov's theorem. These two techniques are used throughout the chapter to prove all of the important theorems. First, we probe the mysteries of sphere theorems. These results, while often unappreciated by a larger audience, have been instrumental in developing most of the new ideas in the subject. Comparison theory, injectivity radius estimates, and Toponogov's theorem were first used in a highly nontrivial way to prove the classical quarter pinched sphere theorem of Rauch, Berger, Toponogov, and Klingenberg. Critical point theory was invented by Grove and Shiohama to prove the diameter sphere theorem. After the sphere theorems, we go through some of the major results of comparison geometry: Gromov's Betti number estimate, the Soul theorem of Cheeger and Gromoll, and the Grove-Petersen homotopy finiteness theorem.

Appendix A: Here, some of the important facts about forms are collected. Stokes' theorem is proved, and we give a very short and streamlined introduction to Čech and de Rham cohomology. The exposition starts with the assumption that we only work with manifolds that can be covered by finitely many charts such that all possible intersections are contractible. This makes it very easy to prove all of the major results, as one can simply use the Poincaré and Meyer-Vietoris lemmas together with induction on the number of charts in the covering.

Appendix B: Here, we develop Cartan formalism for the connection and curvature on a Riemannian manifold. We then develop this in the indexfree work of the frame bundle. Finally, we explain how principal bundles can be used to describe all of this in a very compact and abstract manner.

Appendix C: Using the language of principal bundles developed in the previous appendix, we define spin manifolds, and show why they have some new and interesting bundles that are not tensor bundles. We prove the Lichnerowicz formula for the Dirac Laplacian on spinors. This formula is used in two situations: first, to conclude that the $\hat{A}$-genus vanishes in positive scalar curvature, and secondly, in the positive mass conjecture. In the last section, we also discuss how to square a spinor. The entire treatment is self-contained but does not take the reader into the world of index theory, even though this is where things start to get really interesting. Our intention is simply to give a short and concise account of one of the most important topics in mathematical physics and differential geometry.

At the end of each chapter, we give a list of books and papers that cover and often expand on the material in the chapter. We have whenever possible attempted to refer just to books and survey articles. The reader is then invited to go from those sources back to the original papers. For more recent works, we also give journal references if the corresponding books or surveys do not cover all aspects of the original paper. One particularly exhaustive treatment of Riemannian geometry for the reader who is interested in learning more is [11]. Other valuable texts that expand or complement much of the material covered here are [62], [76], and [79]. There is also a forthcoming historical survey by Berger (see [10]) that complements this text very well.

A first course should definitely contain Chapters 2, 5, and 6 together with whatever one feels is necessary from Chapters 1, 3, and 4. Note that Chapter 4 is really a world unto itself and is not used in a serious way later in the text. A more advanced course could consist of going through either part III or IV as defined earlier. These parts do not depend in a serious way on each other. One can probably not cover the entire book in two semesters, but one can cover parts I, II, and III or alternatively I, II, and IV depending on one's inclination. It should also be noted that, if one does not discuss the section on Killing fields in Chapter 7, then this material can actually be covered without having been through Chapters 5 and 6. Each of the chapters ends with a collection of exercises. These exercises are designed both to reinforce the material covered and to establish some simple results that will be needed later. The reader should at least read and think about all of the exercises, if not actually solve all of them.

There are several people I would like to thank. First and foremost are those students who suffered through my various pedagogical experiments with the teaching of Riemannian geometry. Special thanks go to Marcel Berger, Hao Fang, Chad Sprouse, Semion Shteingold, Marc Troyanov, Gerard Walschap, Nik Weaver, Fred Wilhelm, and Hung-Hsi Wu for their constructive criticism of parts of the book. I would especially like to thank Joseph Borzellino for his very careful reading of this text, and Peter Blomgren for writing the programs that generated Figures 2.1 and 2.2. I would like to thank the New York office of Springer-Verlag for their excellent copy-editing of my manuscript and renderings of my hand-drawn pictures. Their efforts have made the book both more readable and much nicer to look at. Finally, I would like to thank Robert Greene, Karsten Grove, and Gregory Kallo for all the discussions on geometry we have had over the years.

Los Angeles, California Peter Petersen

Contents

1

Riemannian Metrics

In this chapter we shall introduce the category (i.e., sets and maps) that we wish to work with. Without discussing any theory we shall present many examples of Riemannian manifolds and Riemannian maps. All of these examples will form the foundation for future investigations into constructions of Riemannian manifolds with various interesting properties.

The abstract definition of a Riemannian manifold used today dates back only to the 1930s. It was not really until Whitney's work in 1936 that mathematicians obtained a clear understanding of what manifolds were, other than as submanifolds of Euclidean space. Riemann himself defined Riemannian metrics only on domains in Euclidean space. Before Riemann, Gauss and others really understood only 2-dimensional geometry. The invention of Riemannian geometry is quite curious. The story goes that Gauss was on Riemann's defense committee for his Habilitation (super doctorate). In those days, the candidate was asked to submit three topics in advance, with the implicit understanding that the committee would ask to hear about the first topic (the actual thesis was on Fourier series and the Riemann integral.) Riemann's third topic was "On the hypotheses which lie at the foundations of geometry." Clearly he was hoping that the committee would select from the first two topics, which were on material he had already worked on. Gauss, however, always being in an inquisitive mood, decided he wanted to hear whether Riemann had anything to say about the subject on which he, Gauss, was the reigning expert. So, much to Riemann's dismay he had to go home and invent Riemannian geometry to satisfy Gauss's curiosity. No doubt Gauss was suitably impressed, a very rare occurrence for him indeed.

From Riemann's work it appears that he worked with changing metrics mostly by multiplying them by a function (conformal change). With this technique he was able to construct all three constant-curvature geometries in one fell swoop for the first time ever. Soon after Riemann's discoveries it was realized that in polar coordinates one can change the metric in a different way, now referred to as a warped product. This also yields in a unified way all constant curvature geometries. Of course, Gauss already knew about polar coordinate representations on surfaces, and rotationally symmetric metrics were studied even earlier. But these examples are much simpler than the higher-dimensional analogues. Throughout this book we shall emphasize the importance of these special warped products and polar coordinates. It is not far to go from warped products to doubly warped products, which will also be defined in this chapter, but they don't seem to have attracted much attention until Schwarzschild discovered a vacuum space-time that wasn't flat. Since then, doubly warped products have been at the heart of many examples and counterexamples in Riemannian geometry.

Another important way of finding Riemannian metrics is by using left-invariant metrics on Lie groups. This leads us to, among other things, the Hopf fibration and Berger spheres. Both of these are of fundamental importance and are at the core of a large number of examples in Riemannian geometry. These will also be defined here and studied throughout the book.

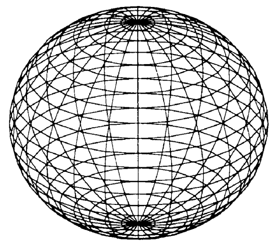
1.1 Riemannian Manifolds and Maps

A *Riemannian manifold* (M, g) consists of a (C^∞) manifold M and a Euclidean inner product g_p on all of the tangent spaces T_pM of M. We shall assume that g_p varies smoothly. This means that for any two smooth vector fields X, Y, the inner product $g_p(X, Y)$ should be a smooth function of p. The subscript p will be suppressed throughout the book. At several places we shall also need M to be connected, and thus we make the assumption throughout the book that we work only with connected manifolds.

All inner product spaces of the same dimension are isometric; therefore all tangent spaces T_pM on a Riemannian manifold (M, g) are isometric to n-dimensional Euclidean space $\mathbb{R}^n$ endowed with its canonical inner product. Hence, all Riemannian manifolds have the same infinitesimal structure not only as manifolds but also as manifolds with a Riemannian metric.

Example 1.1 By far the most important Riemannian manifold is Euclidean space $(\mathbb{R}^n, \mathrm{can})$. The canonical Riemannian structure "can" is defined by identifying the tangent bundle $T\mathbb{R}^n \simeq \mathbb{R}^n \times \mathbb{R}^n$ via the map $(x, v) \to$ [equivalence class of curves through x represented by $s \mapsto x + s \cdot v$]. Thus the standard inner product on $\mathbb{R}^n$ induces a Riemannian structure on $\mathbb{R}^n$.

A *Riemannian isometry* between Riemannian manifolds (M, g) and (N, h) is a diffeomorphism $\varphi : M \to N$ such that $\varphi^* h = g$, i.e., $h(D\varphi(v), D\varphi(w)) = g(v, w)$ for all tangent vectors $v, w \in T_pM$ and all $p \in M$. Clearly, φ^{-1} is a Riemannian isometry as well.



Unit sphere

FIGURE 1.1.

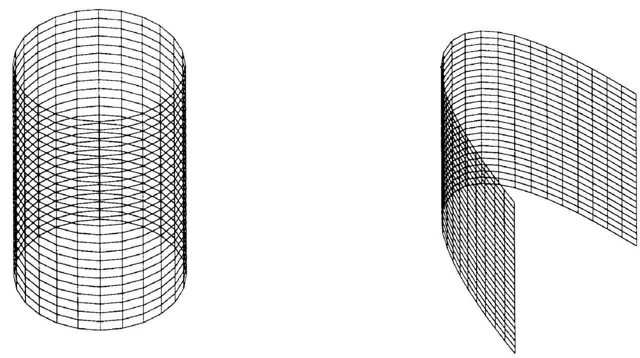
Example 1.2 Whenever we have a finite-dimensional vector space E with an inner product, we can construct a Riemannian manifold by declaring that $g((x,v),\ (x,w)) = v \cdot w$, where $(x,v) \to [s \to x + s \cdot v]$ is the usual trivialization of TE. If we have two such Riemannian manifolds (E,g) and (F,h) of the same dimension, then they are isometric. Recall that both spaces admit orthonormal bases $(e_1,\ldots,e_n)$ and $(f_1,\ldots,f_n)$ with respect to their respective inner products. The Riemannian isometry $\varphi : E \to F$ is defined as $\varphi(\Sigma\, \alpha^i e_i) = \Sigma\, \alpha^i f_i$. (You should check that this is an isometry.) Thus $(\mathbb{R}^n, \text{can})$ is not only the only n-dimensional inner product space, but also the only Riemannian manifold of this simple type.

Suppose that we have an immersion (or embedding) $\varphi : M \to N$, and that (N,h) is a Riemannian manifold. We can then construct a Riemannian metric on M by pulling back h to $g = \varphi^* h$ on M, in other words, $g(v,w) = h\,(D\varphi\,(v),\ D\varphi\,(w))$. Notice that this gives an inner product because $D\varphi\,(v)$ is never zero unless $v = 0$.

A *Riemannian immersion* (or *Riemannian embedding*) is thus an immersion (or embedding) $\varphi : M \to N$ such that $g = \varphi^* h$. Riemannian immersions are also called *isometric immersions.*

Example 1.3 We now come to the second most important example. Define $S^n(r) = \{x \in \mathbb{R}^{n+1} : |x| = r\}$. This is the Euclidean sphere of radius r. The metric induced from the embedding $S^n(r) \hookrightarrow \mathbb{R}^{n+1}$ is the canonical metric on $S^n(r)$. The unit sphere, or standard sphere, is $S^n = S^n(1) \subset \mathbb{R}^{n+1}$ with the induced metric. In Figure 1.1 is a picture of the unit sphere in $\mathbb{R}^3$ shown with latitudes and longitudes.

Example 1.4 If $k < n$ there are, of course, several linear isometric immersions $(\mathbb{R}^k, \text{can}) \to (\mathbb{R}^n, \text{can})$. Those are, however, not the only ones. Any curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ with unit speed, i.e., $|\dot{\gamma}(t)| = 1$ for all $t \in \mathbb{R}$, is an example of an isometric immersion. If the curve has no self-intersections then it will in fact become an embedding. One could, for example, take $t \to (\cos t,\ \sin t)$ as an immersion,



Cylinder $(\mathbb{R}^2, \text{can}) \subset (\mathbb{R}^3, \text{can})$

FIGURE 1.2.

while $t \to \left(\log\left(t + \sqrt{1+t^2}\right), \sqrt{1+t^2}\right)$ gives an embedding. A map of the form: $\varphi : \mathbb{R}^k \to \mathbb{R}^{k+1}$ where $\varphi(x^1, \ldots, x^k) = (\gamma(x^1), x^2, \ldots, x^k)$ (where γ fills up the first two entries) will then give an isometric immersion (or embedding) that is not linear. This is counterintuitive in the beginning, but serves to illustrate the difference between a Riemannian immersion and a distance-preserving map. In Figure 1.2 there are two pictures, one of the cylinder, the other of the isometric embedding of $\mathbb{R}^2$ into $\mathbb{R}^3$ just described.

There is of course also the concept of a *Riemannian submersion* $\varphi : (M, g) \to (N, h)$. This is a submersion $\varphi : M \to N$ such that for each $p \in M$, $D\varphi : \ker^{\perp}(D\varphi) \to T_{\varphi(p)}N$ is a linear isometry. In other words, if $v, w \in T_pM$ are perpendicular to the kernel of $D\varphi : T_pM \to T_{\varphi(p)}N$, then $g(v, w) = h(D\varphi(v), D\varphi(w))$.

Example 1.5 Orthogonal projections $(\mathbb{R}^n, \text{can}) \to (\mathbb{R}^k, \text{can})$ where $k < n$ are examples of Riemannian submersions.

Example 1.6 A much less trivial example is the *Hopf fibration* $S^3(1) \to S^2(\frac{1}{2})$. This map can be written as $(z, w) \to zw^{-1}$ if we think of $S^3(1) \subset \mathbb{C}^2$ and $S^2(\frac{1}{2})$ as being $\hat{\mathbb{C}}$ with the right sort of description of the metric. Later we will examine this example more closely.

1.2 Groups and Riemannian Manifolds

We shall look into groups of Riemannian isometries on Riemannian manifolds and see how this can be useful in constructing new Riemannian manifolds.

1.2.1 Isometry Groups

For a Riemannian manifold (M, g) let $\mathrm{Iso}(M) = \mathrm{Iso}(M, g)$ denote the group of Riemannian isometries $\varphi : (M, g) \to (M, g)$ and $\mathrm{Iso}_p(M, g)$ the *isotropy (sub)group* at p, i.e., those $\varphi \in \mathrm{Iso}(M, g)$ with $\varphi(p) = p$. A Riemannian manifold is said to be *homogeneous* if its isometry group acts transitively, i.e., for each pair of points $p, q \in M$ there is $\varphi \in \mathrm{Iso}\,(M, g)$ such that $\varphi\,(p) = q$.

Example 2.1 $\mathrm{Iso}(\mathbb{R}^n, \mathrm{can}) = \mathbb{R}^n \rtimes O(n) = \{\varphi : \mathbb{R}^n \to \mathbb{R}^n : \varphi(x) = v + Ox,\ v \in \mathbb{R}^n$ and $O \in O(n)\}$. (Here $H \rtimes G$ is the semidirect product, with G acting on H in some way.) The translational part v and rotational part O are uniquely determined. It is clear that these maps indeed are isometries. To see the converse first observe that $\psi\,(x) = \varphi(x) - \varphi(0)$ is also a Riemannian isometry. Using that it is a Riemannian isometry, we observe that at $x = 0$ we can find $\left(O_i^j\right) \in O\,(n)$ such that

$$D\psi\,(\partial_i) = O_i^j \partial_j.$$

Thus, we have two isometries on Euclidean space, both of which preserve the origin and have the same differential there. It is then not hard to see that they must be equal, by using that they must both map unit speed lines through the origin to unit speed lines through the origin.

The isotropy group Iso_p is apparently always isomorphic to $O(n)$, so we see that $\mathbb{R}^n = \mathrm{Iso}/\mathrm{Iso}_p$ for any $p \in \mathbb{R}^n$. This is in fact always true for homogeneous spaces.

Example 2.2 $\mathrm{Iso}(S^n(r), \mathrm{can}) = O(n+1) = \mathrm{Iso}_0(\mathbb{R}^{n+1}, \mathrm{can})$. It is again clear that $O(n+1) \subset \mathrm{Iso}(S^n(r), \mathrm{can})$. Conversely, if $\varphi \in \mathrm{Iso}(S^n(r), \mathrm{can})$ extend it to $\tilde{\varphi} : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ by $\tilde{\varphi}(x) = |x| \cdot r^{-1} \cdot \varphi\left(x \cdot |x|^{-1} \cdot r\right)$ and $\tilde{\varphi}(0) = 0$. Then check that $\tilde{\varphi} \in \mathrm{Iso}_0(\mathbb{R}^{n+1}, \mathrm{can}) = O(n+1)$. This time the isotropy groups are isomorphic to $O(n)$, that is, those elements of $O(n+1)$ fixing a 1-dimensional linear subspace of $\mathbb{R}^{n+1}$. In particular, $O(n+1)/O(n) = S^n$.

1.2.2 Lie Groups

More generally, consider a Lie group G. The tangent space $TG \simeq G \times T_eG$ by using left (or right) translations on G. Therefore, any inner product on T_eG induces a *left-invariant* Riemannian metric on G i.e., left translations are Riemannian isometries. It is obviously also true that any Riemannian metric on G for which all left translations are Riemannian isometries is of this form. In contrast to $\mathbb{R}^n$, not all of these Riemannian metrics are isometric if the identity component of G is not $\mathbb{R}^n$. Lie groups therefore do not come with any canonical metrics.

If H is a closed subgroup of G, then we know that G/H is a manifold. If we endow G with one of the left-invariant metrics, then H acts by isometries (on the left) and one sees that there is a unique Riemannian metric on G/H making the projection $G \to G/H$ into a Riemannian submersion. If in addition the metric is also right invariant then G acts by isometries on G/H (on the right) thus making

G/H into a homogeneous space. It is, in fact, not too hard to prove that $\mathrm{Iso}(M, g)$ is always a Lie group. Thus, all homogeneous spaces look like G/H.

Example 2.3 Consider $S^{2n-1}(1) \subset \mathbb{C}^n$. $S^1 = \{\lambda \in \mathbb{C} : |\lambda| = 1\}$ acts by complex scalar multiplication on both S^{2n-1} and $\mathbb{C}^n$; furthermore this action is by isometries. We know that the quotient $S^{2n-1}/S^1 = \mathbb{C}P^{n-1}$, and since the action of S^1 is by isometries, we induce a metric on $\mathbb{C}P^{n-1}$ such that $S^{2n-1} \to \mathbb{C}P^{n-1}$ is a Riemannian submersion. This metric is called the Fubini-Study metric. When $n = 2$, this turns into the Hopf fibration $S^3(1) \to \mathbb{C}P^1 = S^2(\frac{1}{2})$.

Example 2.4 One of the most important nontrivial Lie groups is $SU(2)$, which is defined as

$$\begin{aligned} SU(2) &= \left\{A \in M_{2\times 2}(\mathbb{C}) : \det A = 1,\ A^* = A^{-1}\right\} \\ &= \left\{\begin{pmatrix} z & -w \\ \bar{w} & \bar{z} \end{pmatrix} : |z|^2 + |w|^2 = 1\right\} = S^3(1). \end{aligned}$$

The Lie algebra $\mathfrak{su}(2)$ of $SU(2)$ is

$$\mathfrak{su}(2) = \left\{\begin{pmatrix} i\alpha & \beta + i\gamma \\ -\beta + i\gamma & -i\alpha \end{pmatrix} : \alpha, \beta, \gamma \in \mathbb{R}\right\}$$

and is spanned by $X_1 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$, $X_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, $X_3 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$. We can think of these matrices as left-invariant vector fields on $SU(2)$. If we declare them to be orthonormal, then we get a left-invariant metric on $SU(2)$, which as we shall later see is $S^3(1)$. If instead we declare the vectors merely to be orthogonal, X_1 to have length ε, and the other two to be unit vectors we get a very important 1-parameter family of metrics g_ε on $SU(2) = S^3$. These distorted spheres are called Berger spheres. Note that scalar multiplication on $S^3 \subset \mathbb{C}^2$ corresponds to multiplication on the left by the matrices $\begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}$ on $SU(2)$. Thus X_1 is exactly tangent to the orbits of the Hopf circle action. The Berger spheres are therefore obtained from the canonical metric by multiplying the metric on the Hopf fiber by ε.

1.2.3 Covering Maps

Groups occur in other ways in geometry, namely, as deck transformations or covering groups. Suppose that $\varphi : M \to N$ is a covering map. Then φ is, in particular, an immersion and a submersion as well. Thus, any Riemannian metric on N induces a Riemannian metric on M, making φ into an isometric immersion, also called a Riemannian covering. Since $\dim M = \dim N$, φ must, in fact, be a local isometry, i.e., for every $p \in M$ there is a neighborhood $U \ni p$ in M such that $\varphi|U : U \to \varphi(U)$ is a Riemannian isometry. Notice that the pullback metric on M has considerable symmetry. For if $q \in V \subset N$ is evenly covered by $\{U_p\}_{p\in\varphi^{-1}(q)}$, then all the sets V and U_p are isometric to each other. In fact, if φ is a normal covering, i.e., there is a group Γ of deck transformations acting on M such that: $M/\Gamma = N$ and $\varphi(gx) = \varphi(x)$ for $g \in \Gamma$, then Γ acts by isometries on the pullback metric. This can be used in the opposite direction. Namely, if $N = M/\Gamma$ and M is a Riemannian manifold, where Γ acts by isometries, then there is a unique Riemannian metric on N such that the quotient map is an isometric immersion.

Example 2.5 If we fix a basis v_1, v_2 for $\mathbb{R}^2$, then $\mathbb{Z}^2$ acts by isometries by $(n, m) \to (x \to x + nv_1 + mv_2)$. The orbit of the origin looks like a lattice. The quotient is a torus T^2 with some metric on it. Note that T^2 is itself an Abelian Lie group and that these metrics are invariant with respect to the Lie group multiplication. However, these metrics are not all isometric to each other.

By adding a reflection to the action by $\mathbb{Z}^2$ we get an action by $\mathbb{Z}^2 \rtimes \mathbb{Z}_2$, and the quotient is the Klein bottle with various Riemannian metrics. One can also use orientation-reversing involutions on T^2 to get these Klein bottles.

Example 2.6 The involution $-\mathrm{id}$ on $S^n(1) \subset \mathbb{R}^{n+1}$ is an isometry and induces a Riemannian covering $S^n \to \mathbb{R}P^n$.

1.3 Local Representations of Metrics

1.3.1 Einstein Summation Convention

We shall often use the index and summation convention that Einstein introduced. Given a vector space V, such as the tangent space of a manifold, we shall always use subscripts for vectors in V. Thus a basis of V is denoted by $v_1, \ldots, v_n$. Given a vector $v \in V$ we can then write it as a linear combination of these basis vectors as follows:

$$v = \sum_i \alpha^i v_i = \alpha^i v_i.$$

Here we use superscripts on the coefficients and then automatically sum over indices that are repeated as both sub- and superscripts. If we define a dual basis v^i for the dual space $V^* = \hom(V, \mathbb{R})$ as follows:

$$v^i\left(v_j\right) = \delta^i_j,$$

then the coefficients can also be computed via

$$\alpha^i = v^i(v).$$

It is therefore convenient to use superscripts for dual bases in V^*. The matrix representation $\left(\alpha_i^j\right)$ of a linear map $L : V \to V$ is usually found by solving

$$L(v_i) = \alpha_i^j v_j.$$

In other words

$$\alpha_i^j = v^j(L(v_i)).$$

Another convenient convention is that subscripts should correspond to rows, while superscripts correspond to columns. Thus, the components of a vector v are arranged in a column, as is standard. But we can then also think of $(L(v_i))$ and (v_i) as row vectors. With this in mind, the matrix representation of a linear map can also be found as the matrix that satisfies

$$(L(v_i)) = (v_j)\left(\alpha_i^j\right).$$

When the objects under consideration are defined on manifolds, the conventions carry over as follows. Cartesian coordinates on $\mathbb{R}^n$ and coordinates on a manifold have superscripts (x^i), as they are the coefficients of the vector corresponding to this point. Coordinate vector fields therefore look like

$$\partial_i = \frac{\partial}{\partial x^i},$$

and consequently they have subscripts. This is natural, as they form a basis for the tangent space. The dual 1-forms

$$dx^i$$

satisfy

$$dx^j(\partial_i) = \delta_i^j$$

and therefore form the natural dual basis for the cotangent space.

Einstein notation is not only useful when one doesn't want to write summation symbols, it also shows when certain coordinate- (or basis-) dependent definitions are invariant under change of coordinates. Examples occur throughout the book. For now, let us just consider a very simple situation, namely, the velocity field of a curve $c : I \to \mathbb{R}^n$. In coordinates, the curve is written

$$\begin{aligned} c(t) &= \left(c^i(t)\right) \\ &= c^i(t)\, e_i, \end{aligned}$$

if e_i is the standard basis for $\mathbb{R}^n$. The velocity field is now defined as the vector

$$\dot{c}(t) = \left(\dot{c}^i(t)\right).$$

Using the coordinate vector fields this can also be written as

$$\dot{c}(t) = \dot{c}^i(t)\,\partial_i.$$

In a coordinate system on a general manifold we could then try to use this as our definition for the velocity field of a curve. But then we must show that indeed it gives the same answer in different coordinates. This is simply because the chain rule tells us that

$$\dot{c}^i(t) = dx^i(\dot{c}(t)),$$

and then observing that, we have simply used the above definition for finding the components of a vector in a given basis.

Generally speaking, we shall, when it is convenient, use Einstein notation. When giving coordinate-dependent definitions we shall be careful that they are given in a form where they obviously conform to this philosophy and therefore can easily be seen to be invariantly defined.

1.3.2 Coordinate Representations

On a manifold M we can multiply 1-forms to get bilinear forms: $\theta_1 \cdot \theta_2(v, w) = \theta_1(v) \cdot \theta_2(w)$. Given coordinates $x(p) = (x^1, \ldots, x^n)$ on an open set U of M, we can thus construct bilinear forms $dx^i \cdot dx^j$. If in addition M has a Riemannian metric g, then we can write

$$g = g(\partial_i, \partial_j) dx^i \cdot dx^j$$

because

$$\begin{aligned} g(v, w) &= g(dx^i(v)\partial_i,\ dx^j(w)\partial_j) \\ &= g(\partial_i, \partial_j) dx^i(v) \cdot dx^j(w). \end{aligned}$$

The functions $g(\partial_i, \partial_j)$ are denoted by g_{ij}. This gives us a representation of g in local coordinates as a positive definite symmetric matrix with entries parametrized over U. Initially one might think that this gives us a way of concretely describing Riemannian metrics. That, however, is a mere illusion. Just think about how many manifolds you know with a good covering of coordinate charts together with corresponding transition functions. On the other hand, coordinate representations are often a good theoretical tool for doing abstract calculations rather than concrete ones.

Example 3.1 The canonical metric on $\mathbb{R}^n$ in the identity chart is $g = \delta_{ij} dx^i dx^j = \sum_{i=1}^n (dx^i)^2$.

Example 3.2 On $\mathbb{R}^2 - \{\text{half line}\}$ we also have polar coordinates (r, θ). In these coordinates the canonical metric looks like $g = dr^2 + r^2 d\theta^2$. In other words,

$$g_{rr} = 1,\ g_{r\theta} = g_{\theta r} = 0,\ g_{\theta\theta} = r^2.$$

Recall that $x^1 = r\cos\theta,\ x^2 = r\sin\theta$. Thus,

$$\begin{aligned} dx^1 &= \cos\theta dr - r\sin\theta d\theta, \\ dx^2 &= \sin\theta dr + r\cos\theta d\theta, \end{aligned}$$

which gives

$$g = (dx^1)^2 + (dx^2)^2$$

$$\begin{aligned}
&= (\cos\theta dr - r\sin\theta d\theta)^2 + (\sin\theta dr + r\cos\theta d\theta)^2 \\
&= (\cos^2\theta + \sin^2\theta)dr^2 + 2(r\cos\theta\sin\theta - r\cos\theta\sin\theta)dr d\theta \\
&\quad + (r^2\sin^2\theta)d\theta^2 + (r^2\cos^2\theta)d\theta^2 \\
&= dr^2 + r^2 d\theta^2.
\end{aligned}$$

1.3.3 Frame Representations

A different but similar way of representing the metric is by choosing a *frame* $X_1, \ldots X_n$ on an open set U of M, i.e., n linearly independent vector fields on U, where $n = \dim M$. If $\sigma^1, \ldots, \sigma^n$ is the coframe, i.e., the 1-forms such that $\sigma^i(X_j) = \delta^i_j$, then the metric can be written as

$$g = g_{ij}\sigma^i\sigma^j,$$

where $g_{ij} = g(X_i, X_j)$.

Example 3.3 Any left-invariant metric on a Lie group G can be written as $(\sigma^1)^2 + \cdots + (\sigma^n)^2$ for a coframing dual to left-invariant vector fields $X_1, \ldots, X_n$, which form an orthonormal basis for T_eG. If instead we just begin with a framing of left-invariant vector fields $X_1, \ldots, X_n$ and dual coframing $\sigma^1, \ldots, \sigma^n$, then any left-invariant metric g depends only on its value on T_eG and can therefore be written $g = g_{ij}\sigma^i\sigma^j$, where g_{ij} is a positive definite symmetric matrix with real-valued entries. The Berger sphere can, for example, be written $g_\varepsilon = \varepsilon^2(\sigma^1)^2 + (\sigma^2)^2 + (\sigma^3)^2$, where $\sigma^i(X_j) = \delta^i_j$.

Example 3.4 A *surface of revolution* consists of a curve $\gamma(t) = (x(t), y(t), 0) : I \to \mathbb{R}^3$, where $I \subset \mathbb{R}$ is open and $y(t) > 0$ for all t. By rotating this curve around the x-axis, we get a surface that can be represented as $(t, \theta) \to f(t, \theta) = (x(t), y(t)\cos\theta, y(t)\sin\theta)$. This is a cylindrical coordinate representation, and we have a natural frame $\partial_t, \partial_\theta$ on all of the surface with dual coframe $dt, d\theta$. We wish to write down the induced metric $dx^2 + dy^2 + dz^2$ from $\mathbb{R}^3$ in this frame. Observe that

$$\begin{aligned}
dx &= \dot{x}dt, \\
dy &= \dot{y}\cos(\theta)dt - y\sin(\theta)d\theta, \\
dz &= \dot{y}\sin(\theta)dt + y\cos(\theta)d\theta,
\end{aligned}$$

so

$$\begin{aligned}
dx^2 + dy^2 + dz^2 &= (\dot{x}dt)^2 + (\dot{y}\cos(\theta)dt - y\sin(\theta)d\theta)^2 \\
&\quad + (\dot{y}\sin(\theta)dt + y\cos(\theta)d\theta)^2 \\
&= \left(\dot{x}^2 + \dot{y}^2\right)dt^2 + y^2 d\theta^2.
\end{aligned}$$

Thus

$$g = (\dot{x}^2 + \dot{y}^2)dt^2 + y^2 d\theta^2.$$

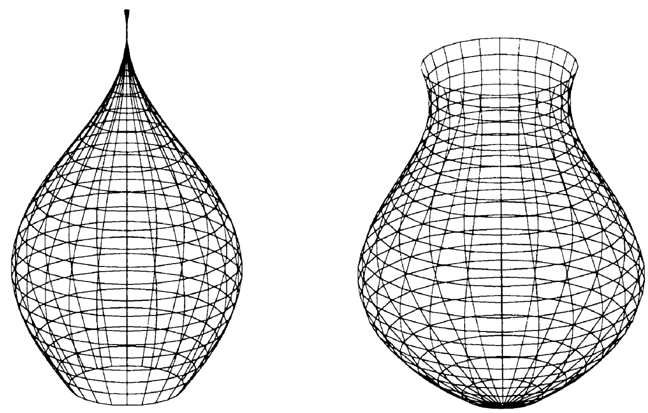


FIGURE 1.3.

If, therefore, the curve is parametrized by arc length, we have the simple formula:

$$g = dt^2 + y^2 d\theta^2,$$

which is reminiscent of our polar coordinate description of $\mathbb{R}^2$. In Figure 1.3 there are two pictures of surfaces of revolution. The first shows that when $y = 0$ the metric looks pinched and therefore destroys the manifold. In the second, y starts out being zero, but this time the metric appears smooth, as y has vertical tangent to begin with.

Example 3.5 On $I \times S^1$ we also have the frame $\partial_t, \partial_\theta$ with coframe $dt, d\theta$. Metrics of the form

$$g = \eta^2(t)dt^2 + \varphi^2(t)d\theta^2$$

are called *rotationally symmetric* since η and φ do not depend on θ. We can, by change of coordinates on I, generally assume that $\eta = 1$. Note that not all rotationally symmetric metrics come from surfaces of revolution. For if $dt^2 + y^2 d\theta^2$ is a surface of revolution, then $\dot{x}^2 + \dot{y}^2 = 1$. Whence $|\dot{y}| \leq 1$.

Example 3.6 $S^2(r) \subset \mathbb{R}^3$ is a surface of revolution. Just revolve $t \to (r\cos(tr^{-1}), r\sin(tr^{-1}), 0)$ around the x-axis. The metric looks like

$$dt^2 + r^2 \sin^2\left(\frac{t}{r}\right) d\theta^2.$$

Note that $r\sin(tr^{-1}) \to t$ as $r \to \infty$, so very large spheres look like Euclidean space. By changing r to ir, we arrive at some interesting rotationally symmetric metrics: $dt^2 + r^2 \sinh^2(tr^{-1})d\theta^2$, which are not surfaces of revolution. If we let $\mathrm{sn}_k(t)$ denote the unique solution to

$$\begin{aligned} \ddot{x}(t) + k \cdot x(t) &= 0, \\ x(0) &= 0, \\ \dot{x}(0) &= 1, \end{aligned}$$

then we have a 1-parameter family $dt^2 + \mathrm{sn}_k^2(t)d\theta^2$ of rotationally symmetric metrics. (The notation sn_k will be used throughout the text; it should not be confused with Jacobi's elliptic function $\mathrm{sn}(k, u)$.) When $k = 0$, this is $\mathbb{R}^2$; when $k > 0$, we get $S^2(1/\sqrt{k})$; and when $k < 0$, we arrive at the *hyperbolic* (from sinh) metrics from above.

1.3.4 Polar Versus Cartesian Coordinates

In these rotationally symmetric examples, we haven't discussed what happens when $\varphi(t) = 0$. In the revolution case, the curve clearly needs to have a vertical tangent in order to look smooth. To be specific, assume that we have $dt^2 + \varphi^2(t)d\theta^2$, $\varphi : [0, b) \to [0, \infty)$, where $\varphi(0) = 0$ and $\varphi(t) > 0$ for $t > 0$. All other situations can be translated or reflected into this position. We assume that φ is smooth, so we can rewrite it as $\varphi(t) = t\psi(t)$ for some smooth $\psi(t) > 0$ for $t > 0$. Now introduce "Cartesian coordinates"

$$\begin{aligned} x &= t\cos\theta, \\ y &= t\sin\theta \end{aligned}$$

near $t = 0$. Then $t^2 = x^2 + y^2$ and

$$\begin{aligned} \begin{pmatrix} dt \\ d\theta \end{pmatrix} &= \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -t^{-1}\sin(\theta) & t^{-1}\cos(\theta) \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix} \\ &= \begin{pmatrix} t^{-1}x & t^{-1}y \\ -t^{-2}y & t^{-2}x \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix}. \end{aligned}$$

Thus,

$$\begin{aligned} dt^2 + \varphi^2(t)d\theta^2 &= dt^2 + t^2\psi^2(t)d\theta^2 \\ &= t^{-2}(xdx + ydy)^2 + (t^2)\psi^2(t)t^{-4}(-ydx + xdy)^2 \\ &= t^{-2}x^2dx^2 + t^{-2}2xydxdy \\ &\quad + t^{-2}y^2dy^2 + t^{-2}\psi^2(t)(xdy - ydx)^2 \\ &= t^{-2}(x^2 + \psi^2(t)y^2)dx^2 \\ &\quad + t^{-2}(2xy - 2xy\psi^2(t))dxdy + t^{-2}(\psi^2(t)x^2 + y^2)dy^2, \end{aligned}$$

whence

$$g_{xx} = \frac{(x^2 + \psi^2(t)y^2)}{x^2 + y^2} = 1 + \frac{\psi^2(t) - 1}{t^2} \cdot y^2,$$

$$g_{xy} = g_{yx} = \frac{1 - \psi^2(t)}{t^2} \cdot xy,$$
$$g_{yy} = \frac{(\psi^2(t)x^2 + y^2)}{x^2 + y^2} = 1 + \frac{\psi^2(t) - 1}{t^2} \cdot x^2,$$

and we need to check for smoothness of the functions at $(x, y) = 0$ (or $t = 0$). For this we must obviously check that the function

$$\frac{\psi^2(t) - 1}{t^2}$$

is smooth at $t = 0$. First, it is clearly necessary that $\psi(0) = 1$; this is the vertical tangent condition. Second, some calculus calculations show that we must further assume that all odd derivatives $\dot{\psi}(0) = \psi^{(3)}(0) = \cdots = 0$. If we translate back to φ, we get that the metric is smooth at $t = 0$ iff $\varphi^{(\text{even})}(0) = 0$ and $\dot{\varphi}(0) = 1$.

These conditions are all satisfied by the metrics $dt^2 + \text{sn}_k^2(t)d\theta^2$, where $t \in [0, \infty)$ when $k \leq 0$ and $t \in [0, \frac{\pi}{\sqrt{k}}]$ for $k > 0$.

1.4 Doubly Warped Products

1.4.1 Doubly Warped Products in General

We can more generally consider metrics on $I \times S^{n-1}$ of the type $dt^2 + \varphi^2(t)ds_{n-1}^2$, where ds_{n-1}^2 is the canonical metric on $S^{n-1}(1) \subset \mathbb{R}^n$. Even more general are metrics of the type: $dt^2 + \varphi^2(t)ds_p^2 + \psi^2(t)ds_q^2$ on $I \times S^p \times S^q$. The first type are again called *rotationally symmetric*, while those of the second type are a special type of *doubly warped product*. As for smoothness, when $\varphi(t) = 0$ we can easily check that the situation for rotationally symmetric metrics is identical to what happened in the previous section. For the doubly warped product observe that nondegeneracy of the metric implies that φ and ψ cannot both be zero at the same time. However, we have the following lemmas:

Lemma 4.1 *If $\varphi : (0, b) \to (0, \infty)$ is smooth and $\varphi(0) = 0$, then we get a smooth metric at $t = 0$ iff*

$$\varphi^{(even)}(0) = 0,$$
$$\dot{\varphi}(0) = 1,$$

and

$$\psi(0) > 0,$$
$$\psi^{(odd)}(0) = 0.$$

The topology near $t = 0$ in this case is $\mathbb{R}^{p+1} \times S^q$.

Lemma 4.2 *If $\varphi : (0, b) \to (0, \infty)$ is smooth and $\varphi(b) = 0$, then we get a smooth metric at $t = b$ iff*

$$\varphi^{(even)}(b) = 0,$$
$$\dot{\varphi}(b) = -1,$$

and

$$\psi(b) > 0,$$
$$\psi^{(odd)}(b) = 0.$$

The topology near $t = b$ in this case is again $\mathbb{R}^{p+1} \times S^q$.

Depending on what happens with φ and ψ as t increases, we can get three different types of topologies.

- $\varphi, \psi : [0, \infty) \to [0, \infty)$ are both positive on all of $(0, \infty)$. Then we have a smooth metric on $\mathbb{R}^{p+1} \times S^q$ if φ, ψ satisfy Lemma 4.1.
- $\varphi, \psi : [0, b] \to [0, \infty)$ are both positive on $(0, b)$ and satisfy Lemma 4.1 and 4.2. Then we get a smooth metric on $S^{p+1} \times S^q$.
- $\varphi, \psi : [0, b] \to [0, \infty)$ as in the second type but the roles of ψ and φ are interchanged at $t = b$. Then we get a smooth metric on S^{p+q+1}!!

1.4.2 Spheres as Warped Products

First let us show how the standard sphere can be written as a rotationally symmetric metric in all dimensions. The metrics $dr^2 + \mathrm{sn}_k^2(r)ds_{n-1}^2$ are analogous to the surfaces from the last section. So when $k = 0$ we get $(\mathbb{R}^n, \mathrm{can})$, and when $k = 1$ we get $(S^n(1), \mathrm{can})$. To see the last statement observe that we have a map

$$f : (0, \pi) \times \mathbb{R}^n \to \mathbb{R} \times \mathbb{R}^n,$$
$$f(r, z) = (t, x) = (\cos(r), \sin(r) \cdot z),$$

which reduces to a map

$$g : (0, \pi) \times S^{n-1} \to \mathbb{R} \times \mathbb{R}^n,$$
$$g(r, z) = (\cos(r), \sin(r) \cdot z).$$

Thus, g really maps into the unit sphere in $\mathbb{R}^{n+1}$. To see that g is a Riemannian isometry we just compute the canonical metric on $\mathbb{R} \times \mathbb{R}^n$ using the coordinates $(\cos(r), \sin(r) \cdot z)$:

$$\mathrm{can} = dt^2 + \left(dx^1\right)^2 + \cdots + \left(dx^n\right)^2$$

$$
\begin{aligned}
&= (d\cos(r))^2 + \left(d\left(\sin(r)\,z^1\right)\right)^2 + \cdots + \left(d\left(\sin(r)\,z^n\right)\right)^2 \\
&= \sin^2(r)\,dr^2 + 2\sin(r)\cos(r)\left(z^1dz^1 + \cdots + z^ndz^n\right) \\
&\quad + \cos^2(r)\,dr^2\left(\left(z^1\right)^2 + \cdots + \left(z^n\right)^2\right) + \sin^2(r)\left(\left(dz^1\right)^2 + \cdots + \left(dz^n\right)^2\right) \\
&= \sin^2(r)\,dr^2 + \cos^2(r)\,dr^2 + \sin^2(r)\left(\left(dz^1\right)^2 + \cdots + \left(dz^n\right)^2\right) \\
&= dr^2 + \sin^2(r)\left(\left(dz^1\right)^2 + \cdots + \left(dz^n\right)^2\right).
\end{aligned}
$$

Now observe that $\left(dz^1\right)^2 + \cdots + (dz^n)^2$ restricted to S^{n-1} is exactly the canonical metric ds_{n-1}^2 and also that $\left(z^1\right)^2 + \cdots + (z^n)^2 = 1$ implies $2\left(z^1dz^1 + \cdots + z^ndz^n\right) = 0$. Thus the claim follows.

The metrics $dt^2+\sin^2(t)ds_p^2+\cos^2(t)ds_q^2$, $t \in [0, \frac{\pi}{2}]$, are also $(S^{p+q+1}(1), \text{can})$. Namely, we have $S^p \subset \mathbb{R}^{p+1}$ and $S^q \subset \mathbb{R}^{q+1}$, so we can map

$$
\begin{aligned}
\left(0, \frac{\pi}{2}\right) \times S^p \times S^q &\to \mathbb{R}^{p+1}\times\mathbb{R}^{q+1}, \\
(t, x, y) &\to (x \cdot \sin(t), y \cdot \cos(t)),
\end{aligned}
$$

where $x \in \mathbb{R}^{p+1}$, $y\in \mathbb{R}^{q+1}$ have $|x| = |y| = 1$. These embeddings clearly map into the unit sphere. The computations that the map is a Riemannian isometry are similar to the above calculations.

1.4.3 The Hopf Fibration

With all this in mind, let us revisit the Hopf fibration $S^3(1) \to S^2\left(\frac{1}{2}\right)$ and show that it is a Riemannian submersion between the spaces indicated. On $S^3(1)$, write the metric as

$$
dt^2 + \sin^2(t)d\theta_1^2 + \cos^2(t)d\theta_2^2,\ t \in \left[0, \frac{\pi}{2}\right],
$$

and use complex coordinates

$$
(t, e^{i\theta_1}, e^{i\theta_2}) \to (\sin(t)e^{i\theta_1}, \cos(t)e^{i\theta_2})
$$

to describe the isometric embedding

$$
\left(0, \frac{\pi}{2}\right) \times S^1 \times S^1 \hookrightarrow S^3(1) \subset \mathbb{C}^2.
$$

Since the Hopf fibers come from complex scalar multiplication, we see that they are of the form $\theta \to (t, e^{i(\theta_1+\theta)}, e^{i(\theta_2+\theta)})$. On $S^2\left(\frac{1}{2}\right)$ use the metric

$$
dr^2 + \frac{\sin^2(2r)}{4}d\theta^2,\ r \in \left[0, \frac{\pi}{2}\right],
$$

with coordinates

$$
(r, e^{i\theta}) \to \left(\frac{1}{2}\cos(2r), \frac{1}{2}\sin(2r)e^{i\theta}\right).
$$

The Hopf fibration in these coordinates, therefore, looks like $(t, e^{i\theta_1}, e^{i\theta_2}) \to (t, e^{i(\theta_1-\theta_2)})$. Now, on $S^3(1)$ we have an orthogonal framing

$$\left\{\partial_{\theta_1}+\partial_{\theta_2},\partial_t,\frac{\cos^2(t)\partial_{\theta_1}-\sin^2(t)\partial_{\theta_2}}{\cos(t)\sin(t)}\right\},$$

where the first vector is tangent to the Hopf fiber and the two other vectors have unit length. On $S^2\left(\frac{1}{2}\right)$

$$\left\{\partial_r,\frac{2}{\sin(2r)}\partial_\theta\right\}$$

is an orthonormal frame. The Hopf map clearly maps

$$\partial_t \to \partial_r,$$

$$\frac{\cos^2(t)\partial_{\theta_1}-\sin^2(t)\partial_{\theta_2}}{\cos(t)\sin(t)} \to \frac{\cos^2(r)\partial_\theta+\sin^2(r)\partial_\theta}{\cos(r)\sin(r)} = \frac{2}{\sin(2r)}\cdot\partial_\theta,$$

thus showing that it is an isometry on vectors perpendicular to the fiber.

Notice also that the map

$$(t, e^{i\theta_1}, e^{i\theta_2}) \to (\cos(t)e^{i\theta_1}, \sin(t)e^{i\theta_2}) \to \begin{pmatrix} \cos(t)e^{i\theta_1} & -\sin(t)e^{i\theta_2} \\ \sin(t)e^{-i\theta_2} & \cos(t)e^{-i\theta_1} \end{pmatrix}$$

gives us the promised isometry from $S^3(1)$ to $SU(2)$, where $SU(2)$ has the left-invariant metric described earlier.

The map $(t, e^{i\theta_1}, e^{i\theta_2}) \to (t, e^{i(\theta_1-\theta_2)})$ from $I\times S^1\times S^1$ to $I\times S^1$ is actually always a Riemannian submersion when the domain is endowed with the doubly warped product metric

$$dt^2+\varphi^2(t)d\theta_1^2+\psi^2(t)d\theta_2^2$$

and the target has the rotationally symmetric metric

$$dr^2+\frac{(\varphi(t)\cdot\psi(t))^2}{\varphi^2(t)+\psi^2(t)}d\theta^2.$$

This submersion can be generalized to higher dimensions as follows: On $I\times S^{2n+1}\times S^1$ consider the doubly warped product metric $dt^2+\varphi^2(t)ds_{2n+1}^2+\psi^2(t)d\theta^2$. The unit circle acts by complex scalar multiplication on both S^{2n+1} and S^1 and consequently induces a free isometric action on this space (if $\lambda\in S^1$ and $(z,w)\in S^{2n+1}\times S^1$, then $\lambda\cdot(z,w)=(\lambda z,\lambda w)$.) The quotient map $I\times S^{2n+1}\times S^1\to I\times\left(\left(S^{2n+1}\times S^1\right)/S^1\right)$ can be made into a Riemannian submersion by choosing the right metric on the quotient space. To find the metric, we split the canonical metric $ds_{2n+1}^2=h+g$, where h corresponds to the metric along the Hopf fiber and g is the orthogonal component. In other words, if $pr : T_pS^{2n+1}\to T_pS^{2n+1}$ is the orthogonal projection (with respect to ds_{2n+1}^2) whose image is the distribution generated by the Hopf action, then

$h(v,w) = ds^2_{2n+1}(pr(v), pr(w))$ and $g(v,w) = ds^2_{2n+1}(v - pr(v), w - pr(w))$. We can then define

$$dt^2 + \varphi^2(t) ds^2_{2n+1} + \psi^2(t) d\theta^2 = dt^2 + \varphi^2(t) g + \varphi^2(t) h + \psi^2(t) d\theta^2.$$

Now notice that $\left(S^{2n+1} \times S^1\right)/S^1 = S^{2n+1}$ and that the S^1 only collapses the Hopf fiber while leaving the orthogonal component to the Hopf fiber unchanged. In analogy with the above example, we therefore get that the metric on $I \times S^{2n+1}$ can be written

$$dt^2 + \varphi^2(t) g + \frac{(\varphi(t) \cdot \psi(t))^2}{\varphi^2(t) + \psi^2(t)} h.$$

In the case where $n = 0$ we recapture the previous case, as g doesn't appear. When $n = 1$, the decomposition: $ds_3^2 = h + g$ can also be written $ds_3^2 = (\sigma^1)^2 + (\sigma^2)^2 + (\sigma^3)^2$, where $(\sigma^1)^2 = h$, $(\sigma^2)^2 + (\sigma^3)^2 = g$, and $\{\sigma^1, \sigma^2, \sigma^3\}$ is the coframing coming from the identification $S^3 \simeq SU(2)$. The Riemannian submersion in this case can therefore be written

$$\begin{gathered}\left(I \times S^3 \times S^1,\ dt^2 + \varphi^2(t)\left[(\sigma^1)^2 + (\sigma^2)^2 + (\sigma^3)^2\right] + \psi^2(t) d\theta^2\right) \\ \downarrow \\ \left(I \times S^3,\ dt^2 + \varphi^2(t)\left[(\sigma^2)^2 + (\sigma^3)^2\right] + \tfrac{(\varphi(t)\cdot\psi(t))^2}{\varphi^2(t)+\psi^2(t)}(\sigma^1)^2\right).\end{gathered}$$

If we let $\varphi = \sin(t)$, $\psi = \cos(t)$, and $t \in I = \left[0, \frac{\pi}{2}\right]$, then we get the generalized Hopf fibration $S^{2n+3} \to \mathbb{C}P^{n+1}$ defined by $\left(0, \frac{\pi}{2}\right) \times \left(S^{2n+1} \times S^1\right) \to \left(0, \frac{\pi}{2}\right) \times \left(\left(S^{2n+1} \times S^1\right)/S^1\right)$ as a Riemannian submersion, and the Fubini-Study metric on $\mathbb{C}P^{n+1}$ can be represented as $dt^2 + \sin^2(t)(g + \cos^2(t) h)$.

1.5 Exercises

1. On product manifolds $M \times N$ one has special product metrics $g = g_1 + g_2$, where g_1, g_2 are metrics on M, N respectively. Show that $(\mathbb{R}^n, \mathrm{can}) = \left(\mathbb{R}, dt^2\right) \times \cdots \times \left(\mathbb{R}, dt^2\right)$. Show that the flat square torus $T^2 = \mathbb{R}^2/\mathbb{Z}^2 = \left(S^1, \left(\frac{1}{2\pi}\right)^2 d\theta^2\right) \times \left(S^1, \left(\frac{1}{2\pi}\right)^2 d\theta^2\right)$. Show that $\varphi(\theta_1, \theta_2) = \frac{1}{\sqrt{2}}(\cos\theta_1, \sin\theta_1, \cos\theta_2, \sin\theta_2)$ is a Riemannian embedding: $T^2 \to \mathbb{R}^4$.

2. Suppose we have an isometric group action G on (M, g) such that the quotient space M/G is a manifold and the quotient map a submersion. Show that there is a unique Riemannian metric on the quotient making the quotient map a Riemannian submersion.

3. Construct paper models of the nonsmooth Riemannian manifolds $\left(\mathbb{R}^2, dt^2 + a^2 d\theta^2\right)$. If $a = 1$, this is of course the Euclidean plane, and when $a < 1$, they look like cones.

4. Suppose φ and ψ are positive on $(0,\infty)$ and consider the Riemannian submersion

$$\left((0,\infty)\times S^3\times S^1,\ dt^2+\varphi^2(t)\left[(\sigma^1)^2+(\sigma^2)^2+(\sigma^3)^2\right]+\psi^2(t)d\theta^2\right)$$
$$\downarrow$$
$$\left((0,\infty)\times S^3,\ dt^2+\varphi^2(t)[(\sigma^2)^2+(\sigma^3)^2]+\frac{(\varphi(t)\cdot\psi(t))^2}{\varphi^2(t)+\psi^2(t)}(\sigma^1)^2\right).$$

Define $f=\varphi$ and $h=(\varphi(t)\cdot\psi(t))^2/\varphi^2(t)+\psi^2(t)$ and assume that

$$f(0)>0,$$
$$f^{(\text{odd})}(0)=0,$$

and

$$h(0)=0,$$
$$h'(0)=k,$$
$$h^{(\text{even})}(0)=0,$$

where k is a positive integer. Show that the above construction yields a smooth metric on the vector bundle over S^2 with Euler number $\pm k$. Hint: Away from the zero section this vector bundle is $(0,\infty)\times S^3/\mathbb{Z}_k$, where $S^3/\mathbb{Z}_k$ is the quotient of S^3 by the cyclic group of order k acting on the Hopf fiber. You should use the submersion description and then realize this vector bundle as a submersion of $S^3\times\mathbb{R}^2$. When $k=2$, this becomes the tangent bundle to S^2. When $k=1$, it looks like $\mathbb{C}P^2-\{\text{point}\}$.

5. Show that any compact Lie group G admits a bi-invariant metric. Show that the inner automorphism $i_h : g\to h^{-1}gh$ is a Riemannian isometry. Conclude that the adjoint action

$$\mathfrak{g}\to\mathfrak{g},$$
$$\mathrm{ad}_U(X) = [X,U]$$

is skew-symmetric, i.e.,

$$g([X,U],Y)=-g(X,[Y,U]).$$

Hint: use that $\mathrm{Ad}_h=Di_h$ is an isometry and that it satisfies

$$\mathrm{Ad}_{\exp U}=\exp(\mathrm{ad}_U).$$

Here the exponential map on the left-hand side, $\exp U : \mathfrak{g}\to G$, is the Lie group exponential map, defined by the property that $t\to\exp(tY)$ is the unique homomorphism $\mathbb{R}\to G$ whose differential at $e\in G$ is $Y\in T_eG=\mathfrak{g}$. The exponential map on the right-hand side is the usual exponential of a linear map $L : V\to V$ on a finite-dimensional vector space defined by the power series

$$\exp(L)=\sum_{i=0}^{\infty}\frac{L^i}{i!}.$$

2

Curvature

With the comforting feeling that there is indeed a variety of Riemannian manifolds out there, we shall now delve into the theory. Initially, we shall confine ourselves to infinitesimal considerations. The most important and often also least understood object of Riemannian geometry is that of the Riemannian connection. From this concept it will be possible to define curvature and more familiar items like gradients and Hessians of functions. Studying curvature is the central theme of Riemannian geometry. The idea of a Riemannian metric having curvature, while intuitively appealing and natural, is for most people the stumbling block for further progress into the realm of geometry.

In the last section of the chapter we shall study what we call the fundamental equations of Riemannian geometry. These equations relate curvature to the Hessian of certain geometrically defined functions (Riemannian submersions onto intervals). These formulae hold all the information that we shall need when computing curvatures in new examples and also for studying Riemannian geometry in the abstract.

Surprisingly, the idea of a connection postdates Riemann's introduction of the curvature tensor. Riemann discovered the Riemannian curvature tensor as a second-order term in the Taylor expansion of a Riemannian metric at a point, where coordinates are chosen such that the zeroth-order term is the Euclidean metric and the first-order term is zero. Lipschitz, Killing, and Christoffel introduced the connection in various ways as an intermediate step in computing the curvature. Also, they found it was a natural invariant for what is called the invariance problem in Riemannian geometry. This problem, which seems rather odd nowadays (although it really is important), comes out of the problem one faces when writing the same metric in two different coordinates. Namely, how is one to know that they are the same or equivalent. The idea is to find invariants of the metric that can be computed in coordinates and then try to show that two metrics are equivalent if their invariant expressions are equal. After this early work by the above-mentioned German mathematicians, an Italian school around Levi-Civita, Ricci, et al. began systematically to study Riemannian metrics and tensor analysis. They eventually defined parallel translation and through that clarified the use of the connection. Hence the name Levi-Civita connection for the Riemannian connection. Most of their work was still local in nature and mainly centered on developing tensor analysis as a tool for describing many physical phenomena, such as stress, torque, and divergence. At the beginning of the twentieth century, Minkowski started developing the geometry of space-time with the hope of using it for Einstein's new special relativity theory. It was this work that eventually enabled Einstein to give a geometric formulation of general relativity theory. Since then, tensor calculus, connections, and curvature have become an indispensable language for many theoretical physicists.

We shall here take the approach to connections developed by Koszul. There is another very efficient and elegant development using forms invented by Cartan, called the Cartan formalism. (See Appendix B for more on this.)

2.1 Connections

2.1.1 *Directional Differentiation*

First we shall introduce some important notation. There are many ways of denoting the *directional derivative* of a function on a manifold. Given a function $f : M \to \mathbb{R}$ and a vector field X on M we will use the following ways of writing the directional derivative of f in the direction of X : $\nabla_X f = D_X f = df(X) = X(f)$.

If we have a function $f : M \to \mathbb{R}$ on a manifold, then the differential $df : TM \to \mathbb{R}$ measures the change in the function. In local coordinates, $df = \partial_i(f)dx^i$. If, in addition, M is equipped with a Riemannian metric g, then we also have the *gradient* of f, denoted by grad $f = \nabla f$, which is the vector field satisfying $g(v, \nabla f) = df(v)$ for all $v \in TM$. In local coordinates this reads, $\nabla f = g^{ij}\partial_i(f)\partial_j$, where g^{ij} is the inverse of the matrix g_{ij}. Defined in this way, the gradient clearly depends on the metric. But is there a way of defining a gradient vector field of a function without using Riemannian metrics? The answer is no and can be understood as follows. On $\mathbb{R}^n$ the gradient is defined as $\nabla f = \delta^{ij}\partial_i(f)\partial_j = \sum_{i=1}^n \partial_i(f)\partial_i$. But this formula depends on the fact that we used Cartesian coordinates. If instead we had used polar coordinates on $\mathbb{R}^2$, say, then it is not true that $\nabla f = \partial_r(f)\partial_r + \partial_\theta(f)\partial_\theta$, because after change of coordinates, this does not equal $\partial_x(f)\partial_x + \partial_y(f)\partial_y$. Now we do not wish to work with concepts that do not have an invariant description (i.e., coordinate-independent description). One rule of thumb for items that are invariantly defined is that they should satisfy the Einstein summation convention, where one sums over identical super- and subscripts. Thus, $df = \partial_i(f)dx^i$ is invariantly defined, while

$\nabla f = \partial_i (f) \partial_i$ is not. The metric $g = g_{ij} dx^i dx^j$ and gradient $\nabla f = g^{ij} \partial_i (f) \partial_j$ are invariant expressions that also depend on our choice of metric.

2.1.2 *Covariant Differentiation*

Having decided that ∇f is a Riemannian notion, rather than a differential topological one, we come to the question of attaching a meaning to the change of a vector field. The change in ∇f should obviously be the Hessian $\nabla^2 f$ of f. It turns out that this concept also depends on the Riemannian metric we use. If X is a vector field on $\mathbb{R}^n$, then $\nabla X = \nabla a^i \partial_i = d(a^i) \partial_i$ defines the change in X by measuring how the coefficients change. Thus, a vector field with constant coefficients does not change. This formula again depends on the fact that we used Cartesian coordinates (having constant coefficients with respect to Cartesian coordinates is clearly not the same as having constant coefficients with respect to polar coordinates) and is not invariant under change of coordinates (although it looks like we have used Einstein convention?). But the assignment $X \to \nabla X$ does have some important properties that we can replicate on a Riemannian manifold. First, note that ∇X is a (1,1)-tensor. The evaluation on a vector is denoted

$$\nabla X(v) = \nabla_v X = \left(D_v a^i\right) \partial_i,$$

where $D_v a^i$ is the directional derivative. If, therefore, Y is a vector field, we get a vector field $\nabla_Y X$ by defining

$$(\nabla_Y X)(p) = \nabla_{Y(p)} X.$$

With this is mind we can prove

Theorem 1.1 (The Fundamental Theorem of Riemannian Geometry) *The assignment $X \to \nabla X$ on $\mathbb{R}^n$ is uniquely defined by the following properties:*
(1) ∇X *is a (1,1)-tensor*

$$\nabla_{\alpha v + \beta w} X = \alpha \nabla_v X + \beta \nabla_w X.$$

(2) $X \to \nabla X$ *is a derivation*

$$\nabla (X + Y) = \nabla X + \nabla Y,$$
$$\nabla (f X) = d(f) X + f \nabla X$$

for functions $f : \mathbb{R}^n \to \mathbb{R}$.
(3) $X \to \nabla X$ *is torsion free*

$$\nabla_X Y - \nabla_Y X = [X, Y].$$

(4) $X \to \nabla X$ *is metric*

$$d(g(X, Y)) = g(\nabla X, Y) + g(X, \nabla Y),$$

or more precisely

$$\nabla_Z g(X, Y) = D_Z g(X, Y) = g(\nabla_Z X, Y) + g(X, \nabla_Z Y)$$

where g is the canonical metric on $\mathbb{R}^n$.

Proof. It is easily checked that $\nabla X = d(a^i)\partial_i$ satisfies these properties. On the other hand, if $X \to \bar{\nabla} X$ is any assignment satisfying these properties, then we can show

$$\begin{aligned} 2g\left(\bar{\nabla}_X Y, Z\right) = {} & D_X g(Y, Z) + D_Y g(Z, X) - D_Z g(X, Y) \\ & + g([X, Y], Z) - g([Y, Z], X) + g([Z, X], Y). \end{aligned}$$

This formula is called the *Koszul formula* and has the advantage that the right-hand side depends only on the metric and differential-topological notions. We must therefore have that $g(\nabla_X Y, Z) = g(\bar{\nabla}_X Y, Z)$ for all vector fields X, Y, Z on $\mathbb{R}^n$. Whence $\nabla_X Y = \bar{\nabla}_X Y$. □

Any assignment on a manifold that satisfies (1) and (2) is called an *affine connection*. If (M, g) is a Riemannian manifold and we have a connection which in addition also satisfies (3) and (4), then we call it a *Riemannian connection*. The fundamental theorem of Riemannian geometry asserts that on $(\mathbb{R}^n, \text{can})$ there is only one such connection. On a Riemannian manifold any Riemannian connection clearly must also satisfy the Koszul formula. Thus, the Riemannian connection is uniquely determined by the metric. The Koszul formula also gives us a way of defining a Riemannian connection. Namely, it can be used to compute $\nabla_X Y$ without knowing ∇. Some tedious calculations show that this way of defining ∇ actually gives us a Riemannian connection. Thus we have

Theorem 1.2 *On a Riemannian manifold (M, g) there is one and only one Riemannian connection.*

Before proceeding we need to discuss how $\nabla_X Y$ depends on X and Y. Since $\nabla_X Y$ is tensorial in X, we see that the value of $\nabla_X Y$ at $p \in M$ depends only on $X(p)$; but in what way does it depend on Y? Since $Y \to \nabla_X Y$ is a derivation, it is definitely not tensorial in Y. We can therefore not expect that $\nabla_X Y(p)$ depends only on $X(p)$ and $Y(p)$. The next two lemmas explore how $\nabla_X Y(p)$ depends on Y.

Lemma 1.3 *Let M be a manifold and ∇ a connection on M. If $p \in M$, $v \in T_pM$, and X, Y are vector fields on M such that $X = Y$ in a neighborhood $U \ni p$, then $\nabla_v X = \nabla_v Y$.*

Proof. Choose $\varphi : M \to \mathbb{R}$ such that $\varphi \equiv 0$ on $M - U$ and $\varphi \equiv 1$ in a neighborhood of p. Then we clearly have that $\varphi X = \varphi Y$ on M. Note that

$$\nabla_v \varphi X = \varphi(p)\nabla_v X + d\varphi(v) \cdot X(p) = \nabla_v X$$

since $d\varphi(p) = 0$ and $\varphi(p) = 1$. Thus

$$\begin{aligned}\nabla_v X &= \nabla_v \varphi X \\ &= \nabla_v \varphi Y \\ &= \nabla_v Y.\end{aligned}$$

□

For a Riemannian connection we could also have used the Koszul formula to prove this since the right hand side of this formula can be localized. This lemma tells us an important thing. Namely, if a vector field Y is defined only on an open subset of M, then ∇Y still makes sense on this subset. We could, therefore, potentially use coordinate vector fields or more generally frames to compute ∇ locally.

Lemma 1.4 *Let M be a manifold, ∇ a connection on M. If Y is a vector field on M and $\gamma : I \to M$ a smooth curve with $\dot{\gamma}(0) = v \in T_pM$, then $\nabla_v Y$ depends only on the values of Y along γ, i.e., if $X \circ \gamma = Y \circ \gamma$, then $\nabla_{\dot{\gamma}} X = \nabla_{\dot{\gamma}} Y$.*

Proof. Choose a framing $\{Z_1, \ldots, Z_n\}$ in a neighborhood of p and write $Y = \sum \alpha^i \cdot Z_i$, $X = \sum \beta^i Z_i$ on this neighborhood. From the assumption that $X \circ \gamma = Y \circ \gamma$ we get that $\alpha^i \circ \gamma = \beta^i \circ \gamma$. Thus,

$$\begin{aligned}\nabla_v Y &= \nabla_v \alpha^i Z_i \\ &= \alpha^i(p) \nabla_v Z_i + Z_i(p) d\alpha^i(v) \\ &= \beta^i(p) \nabla_v Z_i + Z_i(p) d\beta^i(v) \\ &= \nabla_v X.\end{aligned}$$

□

Thus, $\nabla_v Y$ makes sense as long as Y is prescribed along some curve (or submanifold) that has v as a tangent.

It will occasionally be useful to use orthonormal frames with certain nice properties. We say that an orthonormal frame E_i is *normal* at $p \in M$ if $\nabla E_i(p) = 0$ for all $i = 1, \ldots, n$. It is an easy exercise to show that such frames always exist.

2.1.3 Derivatives of Tensors

The connection is incredibly useful in generalizing many of the well-known concepts (such as Hessian, Laplacian, divergence) from multivariable calculus to the Riemannian setting.

If S is a $(0, r)$- or $(1, r)$-tensor field then we can define a *covariant derivative* ∇S that we interpret as a $(0, r + 1)$- or $(1, r + 1)$-tensor field. (Remember that a vector field X is a (1,0)-tensor field and ∇X is a (1,1) tensor field.) The main idea is to make sure that Leibniz rule holds. So if S is a (1,1) tensor then we want to have

$$\nabla_X (S(Y)) = (\nabla_X S)(Y) + S(\nabla_X Y).$$

Thus, it seems reasonable to define ∇S as

$$\begin{aligned}\nabla S(X, Y) &= (\nabla_X S)(Y) \\ &= \nabla_X (S(Y)) - S(\nabla_X Y).\end{aligned}$$

More generally, we define

$$\begin{aligned}\nabla S(X, Y_1, \ldots, Y_r) &= (\nabla_X S)(Y_1, \ldots, Y_r) \\ &= \nabla_X(S(Y_1, \ldots, Y_r)) - \sum_{i=1}^{r} S(Y_1, \ldots, \nabla_X Y_i, \ldots, Y_r),\end{aligned}$$

where ∇_X is interpreted as the directional derivative when applied to a function, while we use it as covariant differentiation on vector fields. It is easy to check that this indeed defines a tensor as promised.

A tensor is said to be *parallel* if $\nabla S \equiv 0$. In $(\mathbb{R}^n, \text{can})$ one can easily see that if a tensor is written in Cartesian coordinates, then it is parallel iff it has constant coefficients. Thus $\nabla Y \equiv 0$ for constant vector fields. On a Riemannian manifold (M, g) we always have that $\nabla g \equiv 0$ since

$$(\nabla g)(X, Y_1, Y_2) = \nabla_X (g(Y_1, Y_2)) - g(\nabla_X Y_1, Y_2) - g(Y_1, \nabla_X Y_2) = 0$$

from property (4) of the connection.

If $f : M \to \mathbb{R}$ is smooth, then we already have ∇f defined as the vector field satisfying $g(\nabla f, v) = D_v f = df(v)$. Thus, there is some confusion here, with ∇f now also being defined as df. In any given context it will generally be clear what we mean. The *Hessian* $\nabla^2 f$ is defined as the (1,1)-tensor $\nabla(\nabla f)$. This tensor is self-adjoint, or symmetric, since

$$\begin{aligned}g(\nabla^2 f(X), Y) &= g(\nabla_X \nabla f, Y) \\ &= D_X df(Y) - df(\nabla_X Y) \\ &= X(Y(f)) - df(\nabla_X Y) \\ &= X(Y(f)) - df(\nabla_Y X) - df([X, Y]) \\ &= X(Y(f)) - X(Y(f)) + Y(X(f)) - df(\nabla_Y X) \\ &= g(\nabla^2 f(Y), X).\end{aligned}$$

Thus, $\nabla^2 f$ can also be interpreted as the symmetric $(0, 2)$-tensor $\nabla^2 f(X, Y) = g(\nabla^2 f(X), Y)$, which might be a more familiar way of thinking about it. We shall, however, always use the $(1, 1)$ interpretation. One easily checks that ∇f and $\nabla^2 f$ coincide with our usual definitions on $\mathbb{R}^n$.

Sometimes $\nabla^2 f$ is actually used as a notation for the trace of $\nabla(\nabla f)$. This is, of course, the *Laplacian*, and we will use the notation $\Delta f = \text{tr}(\nabla^2 f)$. On $\mathbb{R}^n$ this is also written as $\Delta f = \text{div}\nabla f$. The *divergence* of a vector field, $\text{div}X$, on (M, g) is defined as

$$\text{div}X = \text{tr}(\nabla X) = \sum_{i=1}^{n} g\left(\nabla_{e_i} X, e_i\right)$$

if $\{e_i\}$ is an orthonormal basis. Thus, also $\Delta f = \mathrm{tr}(\nabla^2 f) = \mathrm{tr}(\nabla(\nabla f)) = \mathrm{div}(\nabla f)$.

In analogy with our definition of $\mathrm{div}X$ we can also define the divergence of a $(1, r)$-tensor S to be the $(0, r)$-tensor

$$\begin{aligned}(\mathrm{div}S)(v_1, \ldots, v_r) &= \mathrm{tr}(w \to (\nabla_w S)(v_1, \ldots, v_r)) \\ &= \sum_{i=1}^{n} g\left(\left(\nabla_{e_i} S\right)(v_1, \ldots, v_r), e_i\right).\end{aligned}$$

For a $(\cdot, r)$-tensor field S we can now also define the *second covariant derivative* $\nabla^2 S$ as the $(\cdot, r+2)$-tensor field

$$\begin{aligned}\left(\nabla^2_{X_1,X_2} S\right)(Y_1, \ldots, Y_r) &= \left(\nabla_{X_1}(\nabla S)\right)(X_2, Y_1, \ldots, Y_r) \\ &= \left(\nabla_{X_1}\nabla_{X_2} S\right)(Y_1, \ldots, Y_r) - \left(\nabla_{\nabla_{X_1}X_2} S\right)(Y_1, \ldots, Y_r).\end{aligned}$$

With this we get the $(0, 2)$ version of the Hessian of a function defined as

$$\begin{aligned}\nabla^2_{X,Y} f &= \nabla_X \nabla_Y f - \nabla_{\nabla_X Y} f \\ &= \nabla_X g(Y, \nabla f) - g(\nabla_X Y, \nabla f) \\ &= g(Y, \nabla_X \nabla f) \\ &= g\left(\nabla^2 f(X), Y\right).\end{aligned}$$

This second covariant derivative is symmetric in X and Y. In general, however this will not be the case. The defect in the second covariant derivative not being symmetric is a central feature in Riemannian geometry and is at the heart of the difference between Euclidean geometry and all other geometries.

From the new formula for the Hessian we see that the Laplacian can be written as

$$\Delta f = \sum_{i=1}^{n} \nabla^2_{E_i,E_i} f.$$

2.2 Curvature

Having now developed the idea of covariant derivatives and explained their relation to the classical concepts of gradient, Hessian, and Laplacian, one might hope that somehow these concepts carry over to tensors. As we have seen, this is true with one important exception, namely, the most important tensor for us, the Riemannian metric g. This tensor is parallel and therefore has no gradient, etc. Instead, we think of the connection itself as a sort of gradient of the metric. The next question then is, what should the Laplacian and Hessian be? The answer is, curvature.

Any connection on a manifold gives rise to a *curvature operator*. This operator measures in some sense how far away the connection is from being our standard connection on $\mathbb{R}^n$, which we assume is our canonical curvature-free, or flat, space. If we are on a Riemannian manifold, then it is possible to take traces of this curvature operator to obtain various averaged curvatures.

2.2.1 The Curvature Tensor

We shall work exclusively in the Riemannian setting. So let (M, g) be a Riemannian manifold and ∇ the Riemannian connection. The curvature tensor is a $(1, 3)$-tensor defined as

$$\begin{aligned} R(X, Y)Z &= \nabla^2_{X,Y}Z - \nabla^2_{Y,X}Z \\ &= \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z \end{aligned}$$

on vector fields X, Y, Z. Of course, it needs to be proved that this is indeed a tensor. Since both of the second covariant derivatives are tensorial in X and Y, we need only check that R is tensorial in Z. This is easily done:

$$\begin{aligned} R(X, Y)fZ &= \nabla^2_{X,Y}(fZ) - \nabla^2_{Y,X}(fZ) \\ &= f\nabla^2_{X,Y}(Z) - f\nabla^2_{Y,X}(Z) \\ &\quad + \left(\nabla^2_{X,Y}f\right)Z - \left(\nabla^2_{Y,X}f\right)Z \\ &\quad + (\nabla_Y f)\nabla_X Z + (\nabla_X f)\nabla_Y Z \\ &\quad - (\nabla_X f)\nabla_Y Z - (\nabla_Y f)\nabla_X Z \\ &= f\left(\nabla^2_{X,Y}(Z) - \nabla^2_{Y,X}(Z)\right) \\ &= fR(X, Y)Z. \end{aligned}$$

Notice that X, Y appear antisymmetrically in $R(X, Y)Z$, while Z plays its own role on top of the line, hence the unusual notation. Using the metric g we can change this to a $(0, 4)$-tensor as follows: $R(X, Y, Z, W) = g(R(X, Y)Z, W)$. The variables are now treated on a more equal footing, which is perhaps explained by the next proposition.

Proposition 2.1 *The Riemannian curvature tensor $R(X, Y, Z, W)$ satisfies the following properties:*

(1) *R is antisymmetric in the first two and last two entries:*

$$R(X, Y, Z, W) = -R(Y, X, Z, W) = R(Y, X, W, Z).$$

(2) *R is symmetric between the first two and last two entries:*

$$R(X, Y, Z, W) = R(Z, W, X, Y).$$

(3) *R satisfies a cyclic permutation property called Bianchi's first identity:*

$$R(X, Y)Z + R(Z, X)Y + R(Y, Z)X = 0.$$

(4) *∇R satisfies a cyclic permutation property called Bianchi's second identity:*

$$(\nabla_Z R)(X, Y)W + (\nabla_X R)(Y, Z)W + (\nabla_Y R)(Z, X)W = 0.$$

Proof. The first part of (1) has already been established. For part two of (1) observe that $[X, Y]$ is the unique vector field defined by $D_X D_Y f - D_Y D_X f - D_{[X,Y]} f = 0$. In other words, $R(X, Y) f = 0$. This is the idea behind the calculations that follow:

$$\begin{aligned}
g(R(X,Y)Z,Z) &= g(\nabla_X \nabla_Y Z, Z) - g(\nabla_Y \nabla_X Z, Z) - g(\nabla_{[X,Y]} Z, Z) \\
&= D_X g(\nabla_Y Z, Z) - g(\nabla_Y Z, \nabla_X Z) \\
&\quad - D_Y g(\nabla_X Z, Z) + g(\nabla_X Z, \nabla_Y Z) - \frac{1}{2} D_{[X,Y]} g(Z,Z) \\
&= \frac{1}{2} D_X D_Y g(Z,Z) - \frac{1}{2} D_Y D_X g(Z,Z) - \frac{1}{2} D_{[X,Y]} g(Z,Z) \\
&= \frac{1}{2} R(X,Y) g(Z,Z) = 0.
\end{aligned}$$

Now (1) follows by polarizing the identity $R(X, Y, Z, Z) = 0$ in Z. Part (3) is most easily proved by assuming $[X, Y] = [Y, Z] = [Z, X] = 0$. This is actually sufficient for the proof since R is a tensor and any three vectors can be extended to vector fields that mutually commute:

$$\begin{aligned}
R(X,Y)Z + R(Z,X)Y + R(Y,Z)X &= \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z \\
&\quad + \nabla_Z \nabla_X Y - \nabla_X \nabla_Z Y \\
&\quad + \nabla_Y \nabla_Z X - \nabla_Z \nabla_Y X \\
&= \nabla_X (\nabla_Y Z - \nabla_Z Y) \\
&\quad + \nabla_Z (\nabla_X Y - \nabla_Y X) \\
&\quad + \nabla_Y (\nabla_Z X - \nabla_X Z) \\
&= \nabla_X [Y, Z] + \nabla_Z [X, Y] + \nabla_Y [Z, X] \\
&= 0.
\end{aligned}$$

Part (2) is a purely algebraic consequence of (1) and (3):

$$\begin{aligned}
R(X,Y,Z,W) &= -R(Z,X,Y,W) - R(Y,Z,X,W) \\
&= R(Z,X,W,Y) + R(Y,Z,W,X) \\
&= -R(W,Z,X,Y) - R(X,W,Z,Y) \\
&\quad - R(W,Y,Z,X) - R(Z,W,Y,X) \\
&= 2R(Z,W,X,Y) + R(X,W,Y,Z) + R(W,Y,X,Z) \\
&= 2R(Z,W,X,Y) - R(Y,X,W,Z) \\
&= 2R(Z,W,X,Y) - R(X,Y,Z,W),
\end{aligned}$$

which implies $2R(X, Y, Z, W) = 2R(Z, W, X, Y)$.

Now for part (4). Assume again that all Lie brackets are 0. Then in particular, we have

$$R(X,Y)Z = [\nabla_X, \nabla_Y] Z - \nabla_{[X,Y]} Z = [\nabla_X, \nabla_Y] Z.$$

Observe that

$$\begin{aligned}(\nabla_Z R)(X,Y)W &= \nabla_Z(R(X,Y)W) - R(\nabla_Z X,Y)W \\ &\quad - R(X,\nabla_Z Y)W - R(X,Y)\nabla_Z W \\ &= [\nabla_Z, R(X,Y)]W - R(\nabla_Z X,Y)W - R(X,\nabla_Z Y)W.\end{aligned}$$

So

$$\begin{aligned}&(\nabla_Z R)(X,Y)W + (\nabla_X R)(Y,Z)W + (\nabla_Y R)(Z,X)W \\ &\quad = [\nabla_Z, R(X,Y)]W + [\nabla_X, R(Y,Z)]W + [\nabla_Y, R(Z,X)]W \\ &\qquad - R(\nabla_Z X,Y)W - R(X,\nabla_Z Y)W \\ &\qquad - R(\nabla_X Y,Z)W - R(Y,\nabla_X Z)W \\ &\qquad - R(\nabla_Y Z,X)W - R(Z,\nabla_Y X)W \\ &\quad = [\nabla_Z, R(X,Y)]W + [\nabla_X, R(Y,Z)]W + [\nabla_Y, R(Z,X)]W \\ &\qquad + R([X,Z],Y)W + R([Z,Y],X)W + R([Y,X],Z)W \\ &\quad = [\nabla_Z,[\nabla_X,\nabla_Y]]W + [\nabla_X,[\nabla_Y,\nabla_Z]]W + [\nabla_Y,[\nabla_Z,\nabla_X]]W \\ &\quad = 0,\end{aligned}$$

by Jacobi's identity for commutators. □

Notice that part (1) is related to the fact that ∇ is metric (i.e., $d(g(X,Y)) = g(\nabla X, Y) + g(X, \nabla Y)$), while part (3) follows from ∇ being torsion free (i.e., $\nabla_X Y - \nabla_Y X = [X,Y]$).

Example 2.2 $(\mathbb{R}^n, \text{can})$ has $R \equiv 0$ since $\nabla_{\partial_i}\partial_j = 0$ for the standard connection.

More generally for any tensor field S of type $(\cdot, r)$ we can define the curvature as the new $(\cdot, r)$ tensor field

$$R(X,Y)S = \nabla^2_{X,Y}S - \nabla^2_{Y,X}S.$$

Again one needs to check that this is indeed a tensor. This is done in the same way we checked that $R(X,Y)Z$ was tensorial in Z. Clearly, $R(X,Y)S$ is also tensorial and skew symmetric in X and Y.

From the curvature tensor R we can derive several different curvature concepts.

2.2.2 *The Curvature Operator*

The curvature operator that we define first is our Hessian of g. First recall that we have the space $\Lambda^2 M$ of bivectors. If e_i is an orthonormal basis for T_pM, then the inner product on $\Lambda^2_p M$ is such that the bivectors $e_i \wedge e_j$, $i < j$ will form an orthonormal basis. The inner product that $\Lambda^2 M$ inherits in this way is also denoted by g. Alternatively, we can define the inner product g on $\Lambda_p^2 M$ by the formula

$$g(x \wedge y, v \wedge w) = g(x, v)\, g(y, w) - g(x, w)\, g(y, v) = \det \begin{pmatrix} g(x, v) & g(x, w) \\ g(y, v) & g(y, w) \end{pmatrix}$$

and then extend it by linearity to all of $\Lambda_p^2 M$.

From the antisymmetry properties from (1) of Proposition 2.1 we see that R actually defines a map $R : \Lambda^2 M \times \Lambda^2 M \to \mathbb{R}$ by declaring $R(X \wedge Y, V \wedge W) = R(X, Y, W, V)$. Note the reversal of V and W! The relation $g(\mathfrak{R}(X \wedge Y), V \wedge W) = R(X \wedge Y, V \wedge W)$ therefore defines an operator $\mathfrak{R} : \Lambda^2 M \to \Lambda^2 M$, which by property (2) is symmetric. This operator is called the *curvature operator*. It is clearly just a different manifestation of the curvature tensor. The switch between V and W is related to our definition of the next curvature concept.

2.2.3 Sectional Curvature

For any $v \in T_p M$ let $R_v(w) = R(w, v)v : T_p M \to T_p M$ be the *directional curvature operator*. This operator is also known as the *tidal force operator*. The latter name accurately describes in physical terms the meaning of the tensor. The above conditions imply that this operator is selfadjoint and that v is always a zero eigenvector. The normalized quadratic form

$$\begin{aligned} \sec(v, w) &= \frac{g(R_v(w), w)}{g(v, v)g(w, w) - g(v, w)^2} \\ &= \frac{g(R(w, v)v, w)}{(\text{area}\square(v, w))^2} \\ &= \frac{g(\mathfrak{R}(v \wedge w), v \wedge w)}{(\text{area}\square(v, w))^2} \end{aligned}$$

is called the *sectional curvature* of (v, w). Since the denominator is the square of the area of the parallelogram $\{tv + sw : 0 \le t,\ s \le 1\}$, we can easily check that $\sec(v, w)$ depends only on the plane $\pi = \text{span}\{v, w\}$. One of the important relationships between directional and sectional curvature is the following observation by Riemann.

Proposition 2.3 (Riemann, 1854) *The following properties are equivalent:*

(1) $\sec(\pi) = k$ *for all 2-planes in* $T_p M$.

(2) $R(v_1, v_2)v_3 = k \cdot (g(v_2, v_3)v_1 - g(v_1, v_3)v_2)$ *for all* $v_1, v_2, v_3 \in T_p M$.

(3) $R_v(w) = k \cdot (w - g(w, v)v) = k \cdot pr_{v^\perp}(w)$ *for all* $w \in T_p M$ *and* $|v| = 1$.

(4) $\mathfrak{R}(\omega) = k \cdot \omega$ *for all* $\omega \in \Lambda_p^2 M$.

Proof. (2) $\Rightarrow$ (3) $\Rightarrow$ (1) are easy. For (1) $\Rightarrow$ (2) we introduce the multilinear maps:

$$R_k(v_1, v_2)v_3 = k\,(g(v_2, v_3)v_1 - g(v_1, v_3)v_2),$$

$$R_k(v_1, v_2, v_3, v_4) = k(g(v_2, v_3)\, g(v_1, v_4) - g(v_1, v_3)\, g(v_2, v_4)).$$

The first observation is that this "tensor" behaves exactly like the curvature tensor in that it satisfies properties 1, 2, and 3 of Proposition 2.1. Thus, we have a "tensor"

$$T(v_1, v_2, v_3, v_4) = R(v_1, v_2, v_3, v_4) - R_k(v_1, v_2, v_3, v_4)$$

that satisfies those same properties, and we know from our assumption that $\sec = k$ that it satisfies

$$T(v, w, w, v) = 0$$

for all $v, w \in T_pM$. Using polarization $w = w_1 + w_2$ we then have

$$\begin{aligned} 0 &= T(v, w_1 + w_2, w_1 + w_2, v) \\ &= T(v, w_1, w_2, v) + T(v, w_2, w_1, v) \\ &= 2T(v, w_1, w_2, v) \\ &= -2T(v, w_1, v, w_2). \end{aligned}$$

Using properties 1 and 2 of the curvature tensor we now see that T is alternating in all four variables. That, however, is in violation of Bianchi's first identity unless $T = 0$, which is exactly what we wish to prove.

To see why (2) $\Rightarrow$ (4), choose an orthonormal basis e_i for T_pM; then $e_i \wedge e_j$, $i < j$, is a basis for $\Lambda^2_p M$. Using (2) we see that

$$\begin{aligned} g\left(\mathfrak{R}\left(e_i \wedge e_j\right), e_t \wedge e_s\right) &= R(e_i, e_j, e_s, e_t) \\ &= k \cdot \left(g(e_j, e_s)g(e_i, e_t) - g(e_i, e_s)g(e_j, e_t)\right) \\ &= k \cdot g\left(e_i \wedge e_j, e_t \wedge e_s\right). \end{aligned}$$

But this implies that

$$\mathfrak{R}\left(e_i \wedge e_j\right) = k \cdot \left(e_i \wedge e_j\right).$$

For (4) $\Rightarrow$ (1) just observe that if $\{v, w\}$ are orthogonal unit vectors, then $k = g(\mathfrak{R}(v \wedge w), v \wedge w) = \sec(v, w)$. □

A Riemannian manifold (M, g) that satisfies either of the four conditions for all $p \in M$ and the same $k \in \mathbb{R}$ for all $p \in M$ is said to have *constant curvature* k. So far we only know that $(\mathbb{R}^n, \text{can})$ has curvature zero. Soon we shall see that $dr^2 + \text{sn}_k^2(r)ds_{n-1}^2$ has constant curvature k. When $k > 0$, recall that these represent $\left(S^n(1/\sqrt{k}), \text{can}\right)$, while when $k < 0$ we still don't have a good picture yet. Later we will devote a whole section to these constant negative curvature metrics.

2.2.4 *Ricci Curvature*

Our next curvature is the Ricci curvature, and this should be thought of as the Laplacian of g.

The *Ricci curvature* Ric is a trace of R. If $e_1, \ldots, e_n \in T_pM$ is an orthonormal basis, then

$$\mathrm{Ric}(v, w) = \sum_{i=1}^{n} g\left(R\left(e_i, v\right) w, e_i\right) = \sum_{i=1}^{n} g\left(R\left(v, e_i\right) e_i, w\right).$$

Thus Ric is a symmetric bilinear form. It could also be defined as a symmetric (1,1)-tensor

$$\mathrm{Ric}(v) = \sum_{i=1}^{n} R\left(v, e_i\right) e_i.$$

We can therefore adopt the language that $\mathrm{Ric} \geq k$ (or $\leq k$) if all eigenvalues of $\mathrm{Ric}(v)$ are $\geq k$ (or $\leq k$). If (M, g) satisfies $\mathrm{Ric}(v) = k \cdot v$, or equivalently $\mathrm{Ric}(v, w) = k \cdot g(v, w)$, then (M, g) is said to be an *Einstein manifold* with *Einstein constant* k. If (M, g) has constant curvature k, then (M, g) is also Einstein with Einstein constant $(n-1)k$.

We shall pretty soon be able to find interesting Einstein metrics that do not have constant curvature. Three basic types are

(1) $(S^n(1) \times S^n(1),\ ds_n^2 + ds_n^2)$ with Einstein constant $n-1$;

(2) The Fubini-Study metric on $\mathbb{CP}^n$ with Einstein constant $2n+2$; and

(3) The Schwarzschild metric on $\mathbb{R}^2 \times S^2$, which is a doubly warped product metric: $dr^2 + \varphi^2(r)d\theta^2 + \psi^2(r)ds_2^2$ with Einstein constant 0.

If $v \in T_pM$ is a unit vector and we complete it to an orthonormal basis $\{v, e_2, \ldots, e_n\}$ for T_pM, then

$$\mathrm{Ric}(v, v) = g\left(R\left(v, v\right) v, v\right) + \sum_{i=2}^{n} g\left(R\left(e_i, v\right) v, e_i\right) = \sum_{i=2}^{n} \sec\left(v, e_i\right).$$

Thus, when $n = 2$, there is no difference from an informational point of view in knowing R or Ric. This is actually also true in dimension $n = 3$, because if $\{e_1, e_2, e_3\}$ is an orthonormal basis for T_pM, we have

$$\begin{aligned} \sec\left(e_1, e_2\right) + \sec\left(e_1, e_3\right) &= \mathrm{Ric}\left(e_1, e_1\right), \\ \sec\left(e_1, e_2\right) + \sec\left(e_2, e_3\right) &= \mathrm{Ric}\left(e_2, e_2\right), \\ \sec\left(e_1, e_3\right) + \sec\left(e_2, e_3\right) &= \mathrm{Ric}\left(e_3, e_3\right). \end{aligned}$$

In other words:

$$\begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \sec\left(e_1, e_2\right) \\ \sec\left(e_2, e_3\right) \\ \sec\left(e_1, e_3\right) \end{pmatrix} = \begin{pmatrix} \mathrm{Ric}\left(e_1, e_1\right) \\ \mathrm{Ric}\left(e_2, e_2\right) \\ \mathrm{Ric}\left(e_3, e_3\right) \end{pmatrix}.$$

Here, the matrix has $\det = 2$, and we can therefore compute any sectional curvature from Ric. In particular, we see that (M^3, g) is Einstein iff (M^3, g) has constant sectional curvature. The search for Einstein metrics should therefore begin in dimension 4.

2.2.5 *Scalar Curvature*

The last curvature quantity we wish to mention is the *scalar curvature*: $\text{scal} = \text{trRic} = 2 \cdot \text{tr}\mathfrak{R}$. Notice that scal depends only on $p \in M$ and is therefore a function, $\text{scal} : M \to \mathbb{R}$. In an orthonormal basis $e_1, \dots, e_n$ for T_pM we have

$$\begin{aligned}
\text{scal} &= \text{trRic} \\
&= \sum_{j=1}^{n} g\left(\text{Ric}\left(e_j\right), e_j\right) \\
&= \sum_{j=1}^{n}\sum_{i=1}^{n} g\left(R\left(e_i, e_j\right) e_j, e_i\right) \\
&= \sum_{i,j=1}^{n} g\left(\mathfrak{R}\left(e_i \wedge e_j\right), e_i \wedge e_j\right) \\
&= 2\sum_{i<j} g\left(\mathfrak{R}\left(e_i \wedge e_j\right), e_i \wedge e_j\right) \\
&= 2\text{tr}\mathfrak{R} \\
&= 2\sum_{i<j} \sec\left(e_i, e_j\right).
\end{aligned}$$

When $n = 2$ we see that $\text{scal}(p) = 2 \cdot \sec(T_pM)$. Thus, all curvature information is contained in scal. We shall see, however, that already when $n = 3$ there are metrics with constant scalar curvature that are not Einstein. When $n \geq 3$ there is also another interesting phenomenon occurring, which is related to the scalar curvature.

Lemma 2.4 (Schur, 1886) *Suppose that a Riemannian manifold (M, g) of dimension $n \geq 3$ satisfies one of the following two conditions:*

(a) $\sec(\pi) = f(p)$ *for all 2-planes* $\pi \subset T_pM$, $p \in M$; *or*

(b) $\text{Ric}(v) = (n-1) \cdot f(p) \cdot v$ *for all* $v \in T_pM$, $p \in M$.

Then in either case f must be constant. In other words, the metric has constant curvature or is Einstein, respectively.

Proof. It clearly suffices to show (b), as the conditions for (a) imply that (b) holds. To show (b) we need the important identity:

$$d\text{scal} = 2\text{divRic}.$$

Let us see how this implies (b). First we have

$$\begin{aligned}
d\text{scal} &= d\text{trRic} \\
&= d\left(n \cdot (n-1) \cdot f\right) \\
&= n \cdot (n-1) \cdot df.
\end{aligned}$$

Now,

$$\begin{aligned}
2\mathrm{divRic}(v) &= 2\sum g\left(\left(\nabla_{e_i}\mathrm{Ric}\right)(v), e_i\right)\\
&= 2\sum g\left(\left(\nabla_{e_i}((n-1)f\cdot I)\right)(v), e_i\right)\\
&= 2\sum g\left((n-1)\left(\nabla_{e_i} f\right)v, e_i\right) + 2\sum g\left((n-1)f\left(\nabla_{e_i} I\right)(v), e_i\right)\\
&= 2(n-1)g\left(v, \sum\left(\nabla_{e_i} f\right)e_i\right)\\
&= 2(n-1)g(v, \nabla f)\\
&= 2(n-1)df(v).
\end{aligned}$$

Thus, we have shown that $n\cdot df = 2\cdot df$, but this is impossible unless $n = 2$ or $df \equiv 0$ (i.e., f is constant). □

Proposition 2.5

$$d\mathrm{trRic} = 2\mathrm{divRic}.$$

Proof. The identity is proved by a long and uninspired calculation that uses the second Bianchi identity. Choose a normal orthonormal frame E_i at $p \in M$ and let W be a vector field near p with constant coefficients in the chosen frame; then

$$\begin{aligned}
(d\mathrm{trRic})(W)(p) &= D_W\sum g\left(\mathrm{Ric}(E_i), E_i\right)\\
&= D_W\sum g\left(R\left(E_i, E_j\right)E_j, E_i\right)\\
&= \sum g\left(\nabla_W\left(R\left(E_i, E_j\right)E_j\right), E_i\right)\\
&= \sum g\left((\nabla_W R)\left(E_i, E_j\right)E_j, E_i\right)\\
&= -\sum g\left(\left(\nabla_{E_j} R\right)(W, E_i)E_j, E_i\right) - \sum g\left(\left(\nabla_{E_i} R\right)\left(E_j, W\right)E_j, E_i\right)\\
&= -\sum\left(\nabla_{E_j} R\right)\left(W, E_i, E_j, E_i\right) - \sum\left(\nabla_{E_i} R\right)\left(E_j, W, E_j, E_i\right)\\
&= \sum\left(\nabla_{E_j} R\right)\left(E_j, E_i, E_i, W\right) + \sum\left(\nabla_{E_i} R\right)\left(E_i, E_j, E_j, W\right)\\
&= 2\sum\left(\nabla_{E_j} R\right)\left(E_j, E_i, E_i, W\right)\\
&= 2\sum\nabla_{E_j}\left(R\left(E_j, E_i, E_i, W\right)\right)\\
&= 2\sum\nabla_{E_j} g\left(\mathrm{Ric}\left(E_j\right), W\right)\\
&= 2\sum\nabla_{E_j} g\left(\mathrm{Ric}(W), E_j\right)\\
&= 2\sum g\left(\nabla_{E_j}(\mathrm{Ric}(W)), E_j\right)\\
&= 2\sum g\left(\left(\nabla_{E_j}\mathrm{Ric}\right)(W), E_j\right)\\
&= 2\mathrm{divRic}(W)(p).
\end{aligned}$$

Here we used in several places that all covariant derivatives of the form $\nabla_V E_k = 0$ at p, and we used the second Bianchi identity in the fifth equality. □

Corollary 2.6 *An $n\,(>2)$-dimensional Riemannian manifold (M, g) is Einstein iff*

$$\mathrm{Ric} = \frac{\mathrm{scal}}{n}.$$

2.3 The Fundamental Curvature Equations

In this section we are going to study how curvature comes up "naturally" in the investigation of certain types of functions. This will lead us to various formulae that make it possible to calculate the curvature tensor on all of the rotationally symmetric and doubly warped product metrics from Chapter 1. With this information we can then exhibit the above mentioned examples. This latter work will be done in the next chapter.

2.3.1 Distance Functions

The functions we wish to look into are the *distance functions*. Since we don't have a concept of distance yet, we will say that $f : U \to \mathbb{R}$, where $U \subset (M, g)$ is open, is a *distance function* if $|\nabla f| \equiv 1$ on U. Distance functions are therefore simply solutions to the Hamilton-Jacobi equation

$$|\nabla u| = 1.$$

This is a nonlinear first-order PDE. The theory for solving such equations can be found in [4]. For now we shall assume that solutions exist and investigate their properties. Later, when we have developed the theory of geodesics, we shall show the existence of such functions and also show that their name is appropriate.

Example 3.1 On $(\mathbb{R}^n, \mathrm{can})$ define $f(x) = |x - y|$. Then f is smooth on $\mathbb{R}^n - \{y\}$ and has $|\nabla f| \equiv 1$. If we have two different points $\{y, z\}$, then $f(x) = d(x, \{y, z\}) = \min\{d(x, y),\ d(x, z)\}$ is smooth and has $|\nabla f| \equiv 1$ away from $\{y, z\}$ and the hyperplane $\{x \in \mathbb{R}^n : |x - y| = |x - z|\}$, which is equidistant from y and z.

Example 3.2 More generally if $M \subset \mathbb{R}^n$ is a submanifold, then it can be shown that $f(x) = d(x, M) = \inf\{d(x, y) : y \in M\}$ is a "distance function" on some open set $U \subset \mathbb{R}^n$. If M is an orientable hypersurface, then we can see this as follows. Since M is orientable, we can choose a unit normal vector field N on M. Now "coordinatize" $\mathbb{R}^n$ as $x = tN + y$, where $t \in \mathbb{R},\ y \in M$. In some neighborhood U of M these "coordinates" are actually well-defined. In other words, there is some function $\varepsilon(y) : M \to (0, \infty)$ such that any point in $U = \{tN + y : y \in M,\ |t| < \varepsilon(y)\}$ has unique coordinates (t, y). We can now define $f(x) = t$ on U or $g(x) = d(x, M) = |t|$ on $U - M$. Both functions will then define distance functions on their respective domains. Here f is usually referred to as the signed distance to M, while g is just the regular distance. Figure 2.1

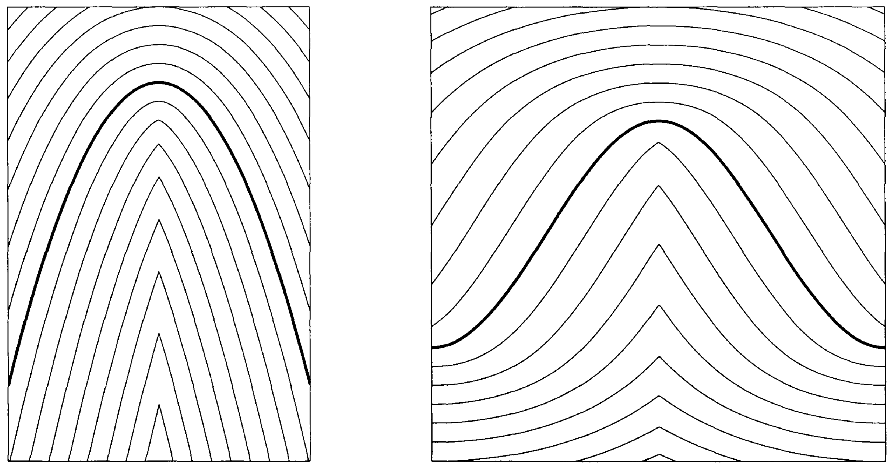


FIGURE 2.1.

shows some pictures of the level sets of a distance function together with the orthogonal trajectories that form the integral curves for the gradient of the distance function.

Example 3.3 On $I \times M$, where $I \subset \mathbb{R}$, is an interval we have metrics of the form $dr^2 + g_r$, where dr^2 is the standard metric on I and g_r is a metric on $\{r\} \times M$ that depends on r. In this case the projection $I \times M \to I$ is a distance function. Special cases of this situation are rotationally symmetric metrics, doubly warped products, and our submersion metrics on $I \times S^{\text{odd}}$.

Lemma 3.4 *Given $f : U \to I \subset \mathbb{R}$, then f is a distance function iff f is a Riemannian submersion.*

Proof. In general, we have $df(v) = g(\nabla f, v)$, so $Df(v) = df(v)\,\partial_t = 0$ iff $v \perp \nabla f$. Thus, v is perpendicular to the kernel of Df iff it is proportional to ∇f. For such $v = \alpha \nabla f$ we therefore have that

$$Df(v) = \alpha Df(\nabla f) = \alpha g(\nabla f, \nabla f)\,\partial_t.$$

Now ∂_t has length 1 in I, so

$$|v| = |\alpha|\,|\nabla f|,$$
$$|Df(v)| = |\alpha|\,|\nabla f|^2.$$

Thus, f is a Riemannian submersion iff $|\nabla f| = 1$. □

Before going on, let us introduce some simplifying notation. A distance function $f : U \to \mathbb{R}$ is fixed and $U \subset (M, g)$ is an open subset of a Riemannian manifold.

The gradient ∇f will usually be denoted by $\partial_r = N = \nabla f$. The ∂_r notation comes from our warped product metrics $dr^2 + \cdots$, while the N notation refers to the fact that ∇f is a unit normal vector field to the level sets $f^{-1}(r)$, which are smooth hypersurfaces in U. These hypersurfaces are renamed $U_r = f^{-1}(r)$, and the induced metric on U_r is g_r. In this spirit ∇^r, R^r are the Riemannian connection and curvature on (U_r, g_r). The Hessian of f is denoted by $S = \nabla^2 f$, where S stands for second derivative or *shape operator* or *second fundamental form*, depending on the point of view of the observer. The last two terms are more or less synonymous and refer to the shape of (U_r, g_r) in $(U, g) \subset (M, g)$. The idea is that $S = \nabla N$ measures how the induced metric on U_r changes by computing how the unit normal to U_r changes.

Example 3.5 Let $M \subset \mathbb{R}^n$ be an orientable hypersurface and N the unit normal, S the shape operator. If $S \equiv 0$ on M then N must be a constant vector field on M, and hence M must be an open subset of the hyperplane $H = \{x + p \in \mathbb{R}^n : x \cdot N = 0 \text{ and } p \in M \text{ is a fixed point}\}$. As an explicit example of this, recall our isometric immersion or embedding $(\mathbb{R}^{n-1}, \text{can}) \to (\mathbb{R}^n, \text{can})$ from Chapter 1 defined by $(x^1, \ldots, x^{n-1}) \to (\gamma(x^1), x^2, \ldots, x^{n-1})$, where γ is a unit speed curve $\gamma : \mathbb{R} \to \mathbb{R}^2$. In this case, $N = (n(x^1), 0, \ldots, 0)$ is a unit normal, where $n(x^1)$ is the unit normal to γ in $\mathbb{R}^2$. We can write this as $N = (-\dot{\gamma}^2(x^1), \dot{\gamma}^1(x^1), 0, \ldots, 0)$ in Cartesian coordinates. So

$$\begin{aligned}
S &= \nabla N \\
&= -d(\dot{\gamma}^2)\partial_1 + d(\dot{\gamma}^1)\partial_2 \\
&= -\ddot{\gamma}^2 dx^1 \partial_1 + \ddot{\gamma}^1 dx^1 \partial_2 \\
&= (-\ddot{\gamma}^2 \partial_1 + \ddot{\gamma}^1 \partial_2) dx^1.
\end{aligned}$$

Thus, $S \equiv 0$ iff $\ddot{\gamma}^1 = \ddot{\gamma}^2 = 0$ iff γ is a straight line iff M is an open subset of a hyperplane. The shape operator therefore really captures the idea that the hypersurface bends in $\mathbb{R}^n$, even though $\mathbb{R}^{n-1}$ of course cannot be seen to bend inside itself.

We have seen here the difference between *extrinsic* and *intrinsic geometry*. Intrinsic geometry is everything we can do on a Riemannian manifold (M, g) that does not depend on how (M, g) might be isometrically immersed in some other Riemannian manifold, while extrinsic geometry is the study of how an isometric immersion $(M, g) \to (Q, h)$ bends (M, g) inside (Q, h). Thus, the curvature tensor on (M, g) measures how the space bends intrinsically, while the shape operator measures extrinsic bending.

2.3.2 Curvature Equations

We are now ready to prove our *first fundamental equation*.

Theorem 3.6 (The Radial Curvature Equation) *If $U \subset (M, g)$ is an open set and $f : U \to \mathbb{R}$ a distance function, then*

$$\nabla_N S + S^2 = -R_N.$$

Proof. We proceed by straightforward computation. If X is a vector field on U, then

$$\begin{aligned}(\nabla_N S)(X) + S^2(X) &= \nabla_N(S(X)) - S(\nabla_N X) + S(S(X)) \\ &= \nabla_N \nabla_X N - \nabla_{\nabla_N X} N + \nabla_{\nabla_X N} N \\ &= \nabla_N \nabla_X N - \nabla_{\nabla_N X - \nabla_X N} N \\ &= \nabla_N \nabla_X N - \nabla_{[N,X]} N.\end{aligned}$$

In order for this to equal $-R(X, N)N$ we only need to see what happened to $-\nabla_X \nabla_N N$. However, since $N = \nabla f$ is unit, we see that for any vector field Y on U:

$$\begin{aligned}g(\nabla_N N, Y) &= g(S(N), Y) \\ &= g(N, S(Y)), \quad \text{by symmetry of } S \\ &= g(N, \nabla_Y N) \\ &= \frac{1}{2} D_Y g(N, N) \\ &= \frac{1}{2} D_Y 1 = 0.\end{aligned}$$

In particular, $\nabla_N N = S(N) = 0$ on all of U. □

This result tells us two things: First, that N is always a zero eigenvector for S and secondly how certain "radial curvatures" relate to the Hessian of f. The Hessian of a generic function cannot, of course, exhibit such predictable behavior (namely, being a solution to a PDE). It is only geometrically relevant functions that behave so nicely.

Even on $(\mathbb{R}^n, \text{can})$ we have arrived at a "new" result, that is, one that is not part of standard multivariable calculus. The most interesting thing is that while we now know that there are many different-looking distance functions on $(\mathbb{R}^n, \text{can})$, they all satisfy this same equation. This will become an important point later on.

The *second and third fundamental equations* are also known as the *Gauss equations* and *Codazzi-Mainardi equations,* respectively. They will be proved simultaneously but stated separately.

Theorem 3.7 (The Tangential Curvature Equation)

$$\begin{aligned}\tan R(X, Y)Z &= R^r(X, Y)Z - \mathrm{II}(Y, Z)S(X) + \mathrm{II}(X, Z)S(Y) \\ &= R^r(X, Y)Z - g(S(Y), Z)S(X) + g(S(X), Z)S(Y).\end{aligned}$$

Here X and Y are vector fields which are tangent to the level sets U_r and R^r is the curvature tensor of (U_r, g_r), $\tan(W) = W - g(W, N)N$ *is the projection of W onto TU_r, and* $\mathrm{II}(U, V) = g(S(U), V)$.

Theorem 3.8 (The Normal or Mixed Curvature Equation)

$$\mathrm{nor} R(X, Y)Z = g(-(\nabla_X S)(Y) + (\nabla_Y S)(X), Z) \cdot N,$$

where X, Y, Z are vector fields tangent to the level sets U_r, and $\mathrm{nor}(W) = g(W, N) \cdot N$ *is the projection of W onto N.*

Proof. The proof hinges on the important fact that if X, Y are vector fields that are tangent to the level sets U_r, then:

$$\begin{aligned} \nabla^r_X Y &= \tan(\nabla_X Y) \\ &= \nabla_X Y - g(\nabla_X Y, N) N \\ &= \nabla_X Y + g(S(X), Y)N \\ &= \nabla_X Y + \mathrm{II}(X, Y)N. \end{aligned}$$

Here the first equality is a consequence of the uniqueness of the Riemannian connection on (U_r, g_r). One can check either that $\tan(\nabla_X Y)$ satisfies properties 1 to 4 of a Riemannian connection or alternatively that it satisfies the Koszul formula. The latter task is almost immediate. The fourth equality is obvious. The third follows since $Y \perp N$ implies

$$\begin{aligned} 0 &= \nabla_X g(Y, N) \\ &= g(\nabla_X Y, N) + g(Y, S(X)), \end{aligned}$$

whence $g(S(X), Y) = -g(\nabla_X Y, N)$.

Both of the curvature equations are now verified by calculating $R(X, Y)Z$ using $\nabla_X Y = \nabla^r_X Y - g(S(X), Y) \cdot N$, and here is the calculation in all its glory:

$$\begin{aligned} R(X, Y)Z &= \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z \\ &= \nabla_X(\nabla^r_Y Z - g(S(Y), Z) \cdot N) - \nabla_Y(\nabla^r_X Z - g(S(X), Z) \cdot N) \\ &\quad - \nabla^r_{[X,Y]} Z + g(S([X, Y]), Z) \cdot N \\ &= \nabla_X \nabla^r_Y Z - \nabla_Y \nabla^r_X Z - \nabla^r_{[X,Y]} Z \\ &\quad - \nabla_X (g(S(Y), Z) \cdot N) \\ &\quad + \nabla_Y (g(S(X), Z) \cdot N) \\ &\quad + g(S([X, Y]), Z) \cdot N \\ &= R^r(X, Y)Z \\ &\quad - g(S(X), \nabla^r_Y Z) \cdot N + g(S(Y), \nabla^r_X Z) \cdot N + g(S([X, Y]), Z) \cdot N \\ &\quad - g(\nabla_X S(Y), Z) \cdot N - g(S(Y), \nabla_X Z) \cdot N - g(S(Y), Z)S(X) \\ &\quad + g(\nabla_Y S(X), Z) \cdot N + g(S(X), \nabla_Y Z) \cdot N + g(S(X), Z)S(Y) \end{aligned}$$

$$
\begin{aligned}
&= R^r(X,Y)Z \\
&\quad - g(S(Y),Z)S(X) + g(S(X),Z)S(Y) \\
&\quad + (-g(\nabla_X S(Y),Z) + g(\nabla_Y S(X),Z)) \cdot N \\
&\quad + (g(S(\nabla_X Y),Z) - g(S(\nabla_Y X),Z)) \cdot N \\
&= R^r(X,Y)Z - g(S(Y),Z)S(X) + g(S(X),Z)S(Y) \\
&\quad + g(-(\nabla_X S)(Y) + (\nabla_Y S)(X),Z) \cdot N.
\end{aligned}
$$

Notice that the important step in the fifth equality is the observation $g(S(Y),\nabla^r_X Z) = g(S(Y),\nabla_X Z)$ and $g(S(X),\nabla^r_Y Z) = g(S(X),\nabla_Y Z)$. These identities follow from the fact that $S(Y) \in TU_r$, which is a consequence of $g(S(Y),N) = g(\nabla_Y N, N) = \frac{1}{2}D_Y g(N,N) = 0$. In other words, S maps TU into TU_r. This can also be seen from our knowledge that S is self adjoint and $S(N) = \nabla_N N = 0$. □

The three fundamental equations give us a way of computing curvature tensors by induction on dimension, for the distance function f foliates U into hypersurfaces U_r. If, therefore, for some reason we know how to do computations on U_r and we also know how to compute S, then we can compute anything on U. We shall clarify and exploit this philosophy in subsequent chapters.

Before doing so, recall that the three curvature quantities sec, Ric, and scal had some special relationships between them in dimensions 2 and 3. Curiously enough this also manifests itself in our three fundamental equations.

If M has dimension 1 then there aren't too many distance functions. Our equations don't even seem to apply here since the level sets are points. This is related to the fact that $R \equiv 0$ on all 1-dimensional spaces.

If M has dimension 2 then any distance function $f : U \subset M \to \mathbb{R}$ has 1-dimensional level sets. Thus $R^r \equiv 0$ and the three vectors X, Y, and Z are proportional. Our equations therefore reduce to the single equation: $\nabla_N S + S^2 = -R_N$. Actually, since $S(N) = 0$, we know that S depends only on its value on a unit vector $v \in TU_r$ thus $S(v) = \alpha v$, where $\alpha = \mathrm{tr} S = \Delta f$. The radial curvature equation can therefore be reduced to: $D_N(\Delta f) + (\Delta f)^2 = -(\mathrm{scal}/2)$. To be even more concrete, we have that g_r on U_r can be written: $g_r = \varphi^2(r,\theta)d\theta^2$; so

$$g = dr^2 + \varphi^2(r,\theta)d\theta^2,$$

and since

$$
\begin{aligned}
\varphi \partial_r \varphi &= \frac{1}{2}\partial_r g(\partial_\theta, \partial_\theta) \\
&= g\left(\nabla_{\partial_r}\partial_\theta, \partial_\theta\right) \\
&= g(S(\partial_\theta), \partial_\theta) \\
&= \alpha\, |\partial_\theta|^2 \\
&= \alpha\varphi^2,
\end{aligned}
$$

we have

$$\operatorname{tr} S = \frac{\partial_r \varphi}{\varphi},$$

which implies

$$-\frac{\text{scal}}{2} = -\sec(T_p M) = \frac{\partial_r^2 \varphi}{\varphi}.$$

When M has dimension 3, the level sets of f are 2-dimensional. The radial curvature equation therefore doesn't reduce, but in the other two equations we have that one of the three vectors X, Y, Z is a linear combination of the other two. We might as well assume that $X \perp Y$ and $Z = X$ or Y. So, if $\{X, Y, N\}$ represents an orthonormal framing, then the complete curvature tensor depends on the quantities: $g(R(X, N)N, Y)$, $g(R(X, N)N, X)$, $g(R(Y, N)N, Y)$, $g(R(X, Y)Y, X)$, $g(R(X, Y)Y, N)$, $g(R(Y, X)X, N)$. The first three quantities can be computed from the radial curvature equation, the fourth from the tangential curvature equation, and the last two from the mixed curvature equation.

In the special case where $M^3 = \mathbb{R}^3$, $R = 0$, the only interesting equation is the tangential curvature equation:

$$\begin{aligned} \sec(T_p U_r) &= R^r(X, Y, Y, X) \\ &= g(S(X), X) g(S(Y), Y) - g(S(X), Y) g(S(X), Y) \\ &= \det S. \end{aligned}$$

This was *Gauss's wonderful observation*! namely, that the extrinsic quantity $\det S$ for U_r is actually the intrinsic quantity, $\sec(T_p U_r)$.

Finally, in dimension 4 everything reaches its most general level. We can start with an orthonormal framing $\{X, Y, Z, N\}$, and there will be twenty curvature quantities to compute. One must therefore incorporate some extra symmetry on U_r if one wants to compute anything.

2.4 The Equations of Riemannian Geometry

In this section we shall investigate the connection between the metric tensor and curvature. This is done by using the radial curvature equation together with some new formulae. Having established these fundamental equations, we shall introduce some useful coordinate systems that make it possible to see how the curvature influences the metric in some unexpected ways.

Note from the end of the last section that have we arrived at a very nice formula for the relationship between the metric and curvature on a surface, namely, if $g = dr^2 + \varphi^2(r, \theta) d\theta^2$, then $\partial_r^2 \varphi = -\sec \cdot \varphi$. This formula can be used not only to compute curvatures from knowledge of the metric, but also in reverse to conclude things about the metric from the curvature. This relationship, which is classical for surfaces, will be generalized in this section to manifolds of any dimension and then extensively used throughout the entire text as a universal tool for understanding the relationship between the metric and curvature.

2.4.1 The Coordinate-Free Equations

Proposition 4.1 *If we have a smooth distance function* $f : (U, g) \to \mathbb{R}$ *and denote* $\nabla f = \partial_r$ *and* $S = \nabla^2 f$*, then*

(1) $\nabla_{\partial_r} S + S^2 = -R_{\partial_r}$,

(2) $\left(L_{\partial_r} g\right)(X, Y) = 2g\left(S\left(X\right), Y\right)$, *and*

(3) $\nabla_{\partial_r} S = L_{\partial_r} S$.

Proof. (1) is just the radial curvature equation.

(2) We simply compute using the definition of the Lie derivative:

$$\begin{aligned}
\left(L_{\partial_r} g\right)(X, Y) &= \partial_r \left(g\left(X, Y\right)\right) - g\left(\left[\partial_r, X\right], Y\right) - g\left(X, \left[\partial_r, Y\right]\right) \\
&= g\left(\nabla_{\partial_r} X, Y\right) + g\left(X, \nabla_{\partial_r} Y\right) \\
&\quad - g\left(\nabla_{\partial_r} X - \nabla_X \partial_r, Y\right) - g\left(X, \nabla_{\partial_r} Y - \nabla_Y \partial_r\right) \\
&= g\left(\nabla_X \partial_r, Y\right) + g\left(X, \nabla_Y \partial_r\right) \\
&= g\left(S\left(X\right), Y\right) + g\left(X, S\left(Y\right)\right) \\
&= 2g\left(S\left(X\right), Y\right).
\end{aligned}$$

(3) Again it is a simple calculation:

$$\begin{aligned}
\left(L_{\partial_r} S\right)(X) &= \left[\partial_r, S\left(X\right)\right] - S\left(\left[\partial_r, X\right]\right) \\
&= \nabla_{\partial_r} \left(S\left(X\right)\right) - \nabla_{S(X)} \partial_r - S\left(\nabla_{\partial_r} X - \nabla_X \partial_r\right) \\
&= \nabla_{\partial_r} \left(S\left(X\right)\right) - S\left(\nabla_{\partial_r} X\right) - \nabla_{S(X)} \partial_r + S\left(\nabla_X \partial_r\right) \\
&= \left(\nabla_{\partial_r} S\right)(X) - S^2\left(X\right) + S^2\left(X\right) \\
&= \left(\nabla_{\partial_r} S\right)(X).
\end{aligned}$$

□

The first equation shows how curvature influences S and the second that S influences g through a simple linear equation. The last equation is important because we don't have a good way of writing covariant derivatives in coordinates. Lie derivatives, on the other hand, are merely directional derivatives in the right coordinates.

2.4.2 The Equations in the Correct Coordinates

Given an integral curve $\gamma : (a, b) \to U$ for ∇f, one can always find coordinates $\left(x^1, x^2, \ldots, x^n\right)$ on some neighborhood V of γ, where $r = f = x^1$ and the other coordinates are tangent to the level sets for f. First we should observe that $\nabla f = \partial_r = \partial_1$. This is because ∂_1 is characterized as the vector field with the property that

$$dx^i\left(\partial_1\right) = \delta^i_1.$$

On the other hand,

$$\begin{aligned} dx^i(\partial_r) &= g\left(\nabla x^i, \partial_r\right) \\ &= \begin{cases} 0 & \text{if} \quad i \geq 2, \\ 1 & \text{if} \quad i = 1, \end{cases} \end{aligned}$$

since we assumed that $\nabla x^1 = \nabla f = \partial_r$ and that the other coordinates are perpendicular to ∂_r. This calculation justifies our coordinate notation, ∂_r, for the gradient of f. We call such a coordinate system *adapted coordinates* with respect to f.

Now write the important tensors in adapted coordinates:

$$\begin{aligned} g_{ij} &= g\left(\partial_i, \partial_j\right), \\ R_{\partial_r}(\partial_i) &= R(\partial_i, \partial_r)\,\partial_r = \sum R_i^j \partial_j, \\ S(\partial_i) &= \nabla_{\partial_i}\partial_r = \sum S_i^j \partial_j. \end{aligned}$$

Then we have, first of all,

$$\left(g_{ij}\right) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & g_{22} & \cdots & g_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & g_{n2} & \cdots & g_{nn} \end{pmatrix},$$

$$\left(R_i^j\right) = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & R_2^2 & \cdots & R_2^n \\ \vdots & \vdots & \ddots & \vdots \\ 0 & R_n^2 & \cdots & R_n^n \end{pmatrix},$$

$$\left(S_i^j\right) = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & S_2^2 & \cdots & S_2^n \\ \vdots & \vdots & \ddots & \vdots \\ 0 & S_n^2 & \cdots & S_n^n \end{pmatrix}.$$

It therefore suffices to look at the components of the tensors for indices $2 \leq i, j \leq n$. Moreover, all of these matrices are symmetric, and the metric is, in addition, positive definite. The coordinate-free equations from the previous subsection now yield the following two matrix equations, where summation is over k:

$$\partial_r\left(S_i^j\right) + \left(S_i^k\right)\cdot\left(S_k^j\right) = -\left(R_i^j\right),$$

$$\partial_r\left(g_{ij}\right) = 2\left(S_i^k\right)\cdot\left(g_{kj}\right).$$

Since the first columns and rows of these matrix equations are trivially satisfied, we can eliminate them and consider the systems as being *decoupled* in the sense that $r = x^1$ is an independent variable and the matrices are dependent variables that can then be analyzed through the above equations. More precisely, the matrix (g_{ij}) is a matrix of functions depending on $(x^1, x^2, \ldots, x^n)$, but we can fix $(x^2, \ldots, x^n)$ and study how this matrix varies with respect to x^1, as that is the only derivative occurring. By relabeling x^1 as r, it then seems that we have systems of ODEs rather than PDEs. This method of reducing PDEs to ODEs is called the *method of characteristics*. It is often used to solve first-order PDEs of the form $L_X S = T$.

We can also get information about the volume form through the fundamental equations. In the exercises to this chapter it is shown that the volume form can be written in the form

$$\begin{aligned} d\text{vol} &= \sqrt{\det (g_{ij})} dx^1 \wedge \cdots \wedge dx^n \\ &= \sqrt{\det (g_{ij})} dr \wedge dx^2 \wedge \cdots \wedge dx^n. \end{aligned}$$

The equation $\partial_r (g_{ij}) = 2(S_i^k) \cdot (g_{kj})$, together with the fact that $\det (g_{ij})$ can be viewed as an alternating map in the rows of (g_{ij}), then yields

$$\begin{aligned} \partial_r \sqrt{\det (g_{ij})} &= \frac{\partial_r \det (g_{ij})}{2\sqrt{\det (g_{ij})}} \\ &= \frac{2\operatorname{tr} (S_i^k) \cdot \det (g_{ij})}{2\sqrt{\det (g_{ij})}} \\ &= \operatorname{tr} (S_i^k) \sqrt{\det (g_{ij})} \\ &= \Delta f \cdot \sqrt{\det (g_{ij})} \\ &= m \cdot \sqrt{\det (g_{ij})}. \end{aligned}$$

Thus, the volume density can be computed via

$$\sqrt{\det (g_{ij} (r))} = \sqrt{\det (g_{ij} (a))} \exp \left(\int_a^r m(t)\, dt \right).$$

Taking traces in the radial curvature equation yields

$$\begin{aligned} \operatorname{tr} \left(\partial_r \left(S_i^j \right) + (S_i^k) \cdot \left(S_k^j \right) \right) &= \operatorname{tr} \partial_r \left(S_i^j \right) + \operatorname{tr} \left((S_i^k) \cdot \left(S_k^j \right) \right) \\ &= \partial_r m + \operatorname{tr} \left((S_i^k) \cdot \left(S_k^j \right) \right) \\ &= -\operatorname{tr} \left(R_i^j \right) = -\operatorname{Ric} (\partial_r, \partial_r). \end{aligned}$$

Using the Cauchy-Schwartz inequality for matrices with the standard inner product $\langle A, B \rangle = \operatorname{tr} (AB^t)$, we see that

$$\operatorname{tr} \left((S_i^k) \cdot \left(S_k^j \right) \right) \geq \frac{m^2}{n-1}$$

with equality holding iff all the eigenvalues of S are equal. Thus we have

$$\partial_r m + \frac{m^2}{n-1} \leq -\mathrm{Ric}(\partial_r, \partial_r).$$

If we denote the volume density by

$$\lambda = \sqrt{\det(g_{ij})},$$

we see that

$$\partial_r \lambda = m\lambda,$$
$$\partial_r m + \frac{m^2}{n-1} \leq -\mathrm{Ric}(\partial_r, \partial_r).$$

These two equations together imply

$$\partial_r^2 \sqrt[n-1]{\lambda} \leq -\frac{\mathrm{Ric}(\partial_r, \partial_r)}{n-1} \cdot \sqrt[n-1]{\lambda}.$$

Let us list the above results in a collection of formulae that we refer to as *The Fundamental Equations of Riemannian Geometry*:

(1) $\partial_r \left(S_i^j\right) + \left(S_i^k\right) \cdot \left(S_k^j\right) = -\left(R_i^j\right).$

(2) $\partial_r \left(g_{ij}\right) = 2\left(S_i^k\right) \cdot \left(g_{kj}\right).$

(tr1) $\partial_r m + \frac{m^2}{n-1} \leq -\mathrm{Ric}(\partial_r, \partial_r).$

(tr2) $\partial_r \sqrt{\det(g_{ij})} = m \cdot \sqrt{\det(g_{ij})}.$

2.4.3 *Rotationally Symmetric Metrics*

Before explaining what these equations might tell us, let us look at what happens on a rotationally symmetric metric

$$dr^2 + \varphi^2 ds_{n-1}^2.$$

Clearly, the metric is diagonalized and looks like

$$\left(g_{ij}\right) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \varphi^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \varphi^2 \end{pmatrix}.$$

Equations (1) and (2) then compute the shape operator and the directional curvature:

$$\left(S_i^j\right) = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & \frac{\partial_r \varphi}{\varphi} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\partial_r \varphi}{\varphi} \end{pmatrix},$$

$$\left(R_i^j\right) = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & -\frac{\partial_r^2 \varphi}{\varphi} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -\frac{\partial_r^2 \varphi}{\varphi} \end{pmatrix}.$$

Note that as $r \to 0$ the metric degenerates in these coordinates if we assume that $\varphi(0) = 0$. We know from Chapter 1 that depending on the initial values of φ this can be resolved by changing the coordinates. It will be important to study such degenerations of the metric, as they actually tell us something. In this case note that even though the metric is smooth at $r = 0$, we can't in some sense go any further. We can only go back in the direction of increasing r.

2.4.4 Conjugate Points

In general, we might think of the curvatures $\left(R_i^j\right)$ as being given in some sense. They could be constant or merely satisfy some inequality. We then wish to investigate how the curvature influences the metric. Equation (2) is linear. We therefore know that the metric can't blow up in finite time unless the shape operator also blows up. However, if we assume that the curvature is bounded, then equation (1) tells us that, if the shape operator blows up, then it must be decreasing in r, hence it can only go to $-\infty$. Going back to (2), we then conclude that the only degeneration which can occur along an integral curve for ∂_r is that the metric stops being positive definite. This is obviously equivalent to saying that the volume density $\sqrt{\det\left(g_{ij}\right)}$ goes to zero. We say that the distance function f develops a *conjugate, or focal, point* along this integral curve if this occurs. Below we have some pictures of how conjugate points can develop. Note that as the metric itself is Euclidean, these singularities exist only in the coordinates, not in the metric.

It is worthwhile investigating equation (1) in its own right. If we rewrite it as

$$\partial_r \left(S_i^j\right) = -\left(R_i^j\right) - \left(S_i^k\right) \cdot \left(S_k^j\right),$$

then we can think of the curvatures as representing fixed *external forces,* while $-\left(S_i^k\right) \cdot \left(S_k^j\right)$ describes an *internal reaction (or interaction).* The reaction term is always negative and, it will try to make $\left(S_i^j\right)$ go to $-\infty$ in finite time (an explosion, or perhaps implosion, better describes this as the metric shrinks). If, for instance,

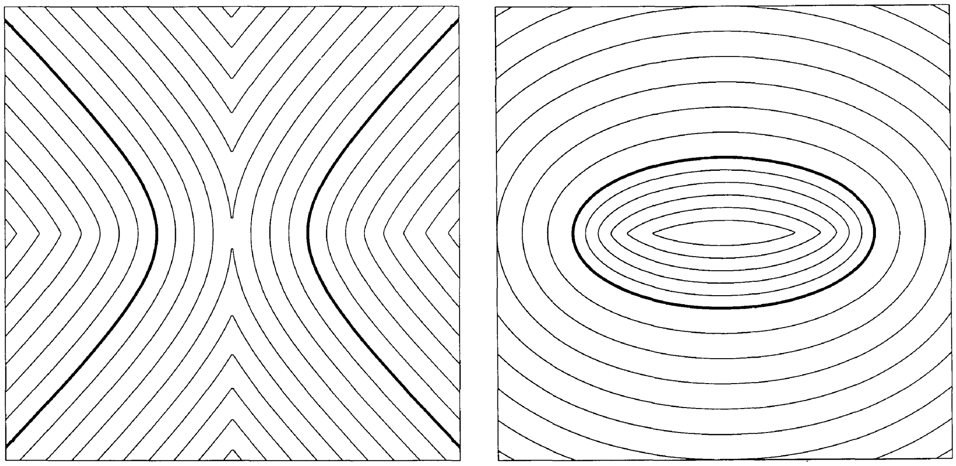


FIGURE 2.2.

the curvature is zero (or more generally) nonnegative, then $\partial_r\left(S_i^j\right)$ is negative (see also the next section for clarification of this). Therefore, if $\left(S_i^j\right)$ is negative at some point, then it will in future time decrease to $-\infty$ in finite time. This can be counteracted by the curvature term if, for instance, it is zero (or nonpositive) and we assume that $\left(S_i^j\right)$ is positive. Then it might still be true that $\left(S_i^j\right)$ decreases, but it can never become negative, and thus focal points can never develop.

2.4.5 Curvature and the Riemannian Metric

Let us list some important properties that we can easily derive. These properties will form the basis for many of our developments in *comparison geometry*. Comparison geometry is the study of how curvature inequalities influence the geometry and topology of the underlying manifold. In all of these results we assume that a smooth distance function is given on some open subset of a Riemannian manifold. Moreover, we start at some point p in this domain and consider what happens to the metric or Hessian along the integral curve for the gradient through this point. The curvature inequalities are meant to hold either on all of the manifold or simply for the directional curvature in the direction of the gradient.

(1) (Prototype for the Hadamard-Cartan theorem, see Chapter 6) Suppose that $\sec \leq 0$ and that $\left(S_i^j\right)$ is nonnegative at p, then it will remain nonnegative in future time.

(2) (Prototype for Synge's lemma and the Soul theorem, see Chapters 6 and 11) Suppose that $\sec \geq 0$ and that $\left(S_i^j\right)$ is nonpositive at p, then it will

remain nonpositive in future time. Moreover if either $\left(S_i^j\right)$ is negative at p or $\sec \geq k^2 > 0$, then a focal point will develop in finite future time.

(3) (Prototype for Myers' diameter bound, see Chapter 9) Suppose that $\text{Ric} \geq (n-1)k^2$, then (tr1) and (tr2) imply that the volume density can only stay positive on intervals of length $\leq 2\pi/k$. Thus focal points must develop in both future and past finite time.

(4) (Prototype for the splitting theorem, see Chapter 9) Suppose $\sec \geq 0$ and that no focal points develop in future and past infinite time; then the metric must be constant along this integral curve.

In order to prove these results and others to come, it is necessary to discuss in more detail what matrix inequalities mean.

2.4.6 *Matrix Inequalities*

Suppose we have an inner product g on a vector space E and a linear map $S : E \to E$. We say that S is self adjoint with respect to g if

$$g(S(v), w) = g(v, S(w)).$$

In case S is self adjoint, we know from the spectral theorem that there exists an orthonormal basis of eigenvectors and that all eigenvalues are real. Thus, an inequality like: $S \geq \lambda$ means that all eigenvalues of S are $\geq \lambda$. This holds iff

$$\begin{aligned} &g(S(v), v) \geq \lambda g(v, v) \text{ for all } v \in E, \\ \text{or } &g((S - \lambda I)(v), v) \geq 0 \text{ for all } v \in E. \end{aligned}$$

We can then also say that two self adjoint operators satisfy $S \geq T$ iff $S - T \geq 0$.

In the above situations we can think of the vector space as being $\mathbb{R}^{n-1}$, but the inner products change with respect to a parameter r and so do the operators. Let us fix families of $(n-1) \times (n-1)$ matrices that depend on r : $(g_{ij}(r))$ symmetric and positive definite with respect to the Euclidean metric, curvature $\left(R_i^j(r)\right)$ and shape operator $\left(S_i^j(r)\right)$ matrices that are self adjoint with respect to the inner product $(g_{ij}(r))$.

Let us now assume that these matrices satisfy the fundamental equations (1) and (2) and hence also (tr1) and (tr2). Suppose we have that $\left(S_i^j\right) \geq 0$ for all r in some interval (a, b). Then, we see immediately from equation (2) that $\partial_r (g_{ij}) \geq 0$ with respect to the Euclidean metric on $\mathbb{R}^{n-1}$, and hence we have that $(g_{ij}(t)) \geq (g_{ij}(s))$ for $t \geq s$ with respect to the Euclidean metric.

Suppose now that the curvature satisfies $\left(R_i^j\right) \geq 0$; then equation (1) shows that

$$\begin{aligned} \partial_r \left(S_i^j\right) &= -\left(R_i^j\right) - \left(S_i^k\right) \cdot \left(S_k^j\right) \\ &\leq 0. \end{aligned}$$

Hence if $\left(S_i^j\right)$ is nonpositive at some value for r, this will perpetuate for larger values of r. Using the above, we can then conclude that the metric will also decrease. This analysis establishes the spirit in which the above mentioned results can be proved. The analysis is slightly more complicated when the curvature is nonpositive or has a nonzero lower bound and will be deferred to Chapter 6.

The last comment we have to make now is that a sectional curvature inequality like $\sec \geq \lambda$ implies that the directional curvatures satisfy $\left(R_i^j\right) \geq \lambda$.

2.5 Some Tensor Concepts

In this section we shall collect together some notational baggage that we shall occasionally need.

2.5.1 Type Change

The inner product structure on the tangent spaces to a Riemannian manifold makes it possible to view tensors in different ways. We saw this with the Hessian and the Ricci tensor. This is nothing but the elementary observation that a bilinear map can be interpreted as a linear map when one has an inner product present.

If, in general, we have an (s,t)-tensor T, we view it as a section in the bundle

$$\underbrace{TM \otimes \cdots \otimes TM}_{s\text{times}} \otimes \underbrace{T^*M \otimes \cdots \otimes T^*M}_{t\text{times}}.$$

Then given a Riemannian metric g on M, we can make it into an $(s-k, t+k)$-tensor for any $k \in \mathbb{Z}$ such that both $s-k$ and $t+k$ are nonnegative. Abstractly, this is done as follows: On a Riemannian manifold TM is naturally isomorphic to T^*M; the isomorphism is given by sending $v \in TM$ to the linear map $(w \to g(v,w)) \in T^*M$. Using this isomorphism we can therefore replace TM by T^*M or vice versa and thus change the type of the tensor.

At a more concrete level what happens is this: We select a frame $E_1, \ldots, E_n$ and construct the coframe $\sigma^1, \ldots, \sigma^n$. The vectors and covectors (in T^*M) can be written as

$$\begin{aligned} v &= v^i E_i = \sigma^i(v) E_i, \\ \omega &= \alpha_j \sigma^j = \omega\left(E_j\right)\sigma^j. \end{aligned}$$

The tensor T can now be written as

$$T = T_{j_1 \cdots j_t}^{i_1 \cdots i_s} E_{i_1} \otimes \cdots \otimes E_{i_s} \otimes \sigma^{j_1} \otimes \cdots \otimes \sigma^{j_t}.$$

Now we need to know how we can change E_i into a covector and σ^j into a vector. As before, the dual to E_i is the covector $w \to g(E_i, w)$, which can be written as

$$g(E_i, w) = g\left(E_i, E_j\right)\sigma^j(w) = g_{ij}\sigma^j(w).$$

Conversely, we have to find the vector v corresponding to the covector σ^j. The defining property is

$$g(v, w) = \sigma^j(w).$$

Thus, we have

$$g(v, E_i) = \delta_i^j.$$

If we write $v = v^k E_k$, this gives

$$g_{ki} v^k = \delta_i^j.$$

Letting g^{ij} denote the ijth entry in the inverse of (g_{ij}), we therefore have

$$v = v^i E_i = g^{ij} E_i.$$

Thus,

$$E_i \to g_{ij}\sigma^j,$$
$$\sigma^j \to g^{ij} E_i.$$

Note that using Einstein notation properly will help keep track of the correct way of doing things as long as the inverse of g is given with superscript indices. With this formula one can easily change types of tensors by replacing E's with σ's and vice versa. Note that if we used coordinate vector fields in our frame, then one really needs to invert the metric, but if we had chosen an orthonormal frame, then one simply moves indices up and down as the metric coefficients satisfy $g_{ij} = \delta_{ij}$.

Let us list some examples:

The Ricci tensor: We write the Ricci tensor as a $(1, 1)$-tensor: $\mathrm{Ric}(E_i) = \mathrm{Ric}_i^j E_j$; thus

$$\mathrm{Ric} = \mathrm{Ric}_j^i \cdot E_i \otimes \sigma^j.$$

As a $(0, 2)$-tensor it will look like

$$\mathrm{Ric} = \mathrm{Ric}_{kj} \cdot \sigma^j \otimes \sigma^k = g_{ki}\mathrm{Ric}_j^i \cdot \sigma^j \otimes \sigma^k,$$

while as a $(2, 0)$-tensor acting on covectors it will be

$$\mathrm{Ric} = \mathrm{Ric}^{ik} \cdot E_i \otimes E_k = g^{kj}\mathrm{Ric}_j^i \cdot E_i \otimes E_k.$$

The curvature tensor: We start with the $(1, 3)$-curvature tensor $R(X, Y)Z$, which we write as

$$R = R_{ijk}^l \cdot E_l \otimes \sigma^i \otimes \sigma^j \otimes \sigma^k.$$

As a $(0, 4)$-tensor we get

$$\begin{aligned} R &= R_{ijkl} \cdot \sigma^i \otimes \sigma^j \otimes \sigma^k \otimes \sigma^l \\ &= R_{ijk}^s g_{sl} \cdot \sigma^i \otimes \sigma^j \otimes \sigma^k \otimes \sigma^l, \end{aligned}$$

while as a (2, 2)-tensor we have:

$$\begin{aligned} R &= R_{ij}^{kl} \cdot E_k \otimes E_l \otimes \sigma^i \otimes \sigma^j \\ &= R_{ijs}^{l} g^{sk} \cdot E_k \otimes E_l \otimes \sigma^i \otimes \sigma^j. \end{aligned}$$

Here, however, we must watch out, because there are several different ways of doing this. We choose to raise the last index, but we could also have chosen any other index, thus yielding different (2, 2)-tensors. The way we did it gives essentially the curvature operator.

2.5.2 *Contractions*

Contractions are simply traces of tensors. Thus, the contraction of a (1, 1)-tensor $T = T_j^i \cdot E_i \otimes \sigma^j$ is simply its trace:

$$C(T) = \mathrm{tr} T = T_i^i.$$

If instead we had a (0, 2)-tensor T, then we could, using the Riemannian structure, first change it to a (1, 1)-tensor and then take the trace

$$\begin{aligned} C(T) &= C\left(T_{ij} \cdot \sigma^i \otimes \sigma^j\right) \\ &= C\left(T_{ik} g^{kj} \cdot E_k \otimes \sigma^j\right) \\ &= T_{ik} g^{ki}. \end{aligned}$$

In this way the Ricci tensor becomes a contraction:

$$\begin{aligned} \mathrm{Ric} &= \mathrm{Ric}_j^i \cdot E_i \otimes \sigma^j \\ &= R_{ik}^{kj} \cdot E_i \otimes \sigma^j \\ &= R_{iks}^{j} g^{sk} \cdot E_i \otimes \sigma^j, \end{aligned}$$

or

$$\begin{aligned} \mathrm{Ric} &= \mathrm{Ric}_{ij} \cdot \sigma^i \otimes \sigma^j \\ &= g^{kl} R_{iklj} \cdot \sigma^i \otimes \sigma^j, \end{aligned}$$

which after type change can be seen to give the same expressions. The scalar curvature can be expressed as:

$$\begin{aligned} \mathrm{scal} &= \mathrm{tr}\mathrm{Ric} \\ &= \mathrm{Ric}_i^i \\ &= R_{iks}^{i} g^{sk} \\ &= \mathrm{Ric}_{ik} g^{ki} \\ &= R_{ijkl} g^{jk} g^{il}. \end{aligned}$$

Again, it is necessary to be careful to specify over which indices one contracts in order to get the right answer.

Note that the divergence of a $(1, k)$-tensor S is nothing but a contraction of the covariant derivative ∇S of the tensor. Here one contracts against the new variable introduced by the covariant differentiation.

2.5.3 Norms of Tensors

There are several conventions in Riemannian geometry for how one should measure the norm of a linear map. Essentially, there are two different norms in use, the *operator norm* and the *Euclidean norm*. The former is defined for a linear map $L : V \to W$ between inner product spaces as

$$|L| = \sup_{|v|=1} |Lv| .$$

The Euclidean norm, in contrast, is given by

$$|L| = \sqrt{\operatorname{tr}(L^* \circ L)} = \sqrt{\operatorname{tr}(L \circ L^*)},$$

where $L^* : W \to V$ is the adjoint. Despite the fact that we use the same notation for these norms, they are almost never equal. If, for instance, $L : V \to V$ is self adjoint and $\lambda_1 \leq \cdots \leq \lambda_n$ the eigenvalues of L counted with multiplicities, then the operator norm is $\max\{|\lambda_1|, |\lambda_n|\}$, while the Euclidean norm is $\sqrt{\lambda_1^2 + \cdots + \lambda_n^2}$. The Euclidean norm also has the advantage of actually coming from an inner product:

$$\langle L_1, L_2 \rangle = \operatorname{tr} L_1 \circ L_2^* = \operatorname{tr} L_2 \circ L_1^*.$$

As a general rule we shall always use the Euclidean norm.

It is worthwhile to see how the Euclidean norm of some simple tensors can be computed on a Riemannian manifold. Note that this computation uses type changes to compute adjoints and contractions to take traces.

Let us start with a $(1, 1)$-tensor $T = T_j^i \cdot E_i \otimes \sigma^j$. We think of this as a linear map $TM \to TM$. Then the adjoint is first of all the dual map $T^* : T^*M \to T^*M$, which we then change to $T^* : TM \to TM$. This means that

$$T^* = T_i^j \cdot \sigma^i \otimes E_j,$$

which after type change becomes

$$T^* = T_l^k g^{lj} g_{ki} \cdot E_j \otimes \sigma^i .$$

Finally,

$$|T|^2 = T_j^i T_l^k g^{lj} g_{ki} .$$

If the frame is orthonormal, this takes the simple form of

$$|T|^2 = T_j^i T_i^j .$$

For a (0, 2)-tensor $T = T_{ij} \cdot \sigma^i \otimes \sigma^j$ we first have to change type and then proceed as above. In the end one gets the nice formula

$$|T|^2 = T_{ij} T^{ij}.$$

2.5.4 *Positional Notation*

A final remark is in order. Many of the above notations could be streamlined even further so as to rid ourselves of some of the notational problems we have introduced by the way in which we write tensors in frames. Namely, tensors $TM \to TM$ (section of $TM \otimes T^*M$) and $T^*M \to T^*M$ (section of $T^*M \otimes TM$) seem to be written in the same way, and this causes some confusion when computing their Euclidean norms. That is, the only difference between the two objects $\sigma \otimes E$ and $E \otimes \sigma$ is in the ordering, not in what they actually do. We simply interpret the first as a map $TM \to TM$ and then the second as $T^*M \to T^*M$, but the roles could have been reversed, and both could be interpreted as maps $TM \to TM$. This can indeed cause great confusion.

One way to at least keep the ordering straight when writing tensors out in coordinates is to be even more careful with our indices and how they are written down. Thus, a tensor T that is a section of $T^*M \otimes TM \otimes T^*M$ should really be written as

$$T = T_i{}^j{}_k \cdot \sigma^i \otimes E_j \otimes \sigma^k.$$

Our standard (1, 1)-tensor (section of $TM \otimes T^*M$) could therefore be written

$$T = T^i{}_j \cdot E_i \otimes \sigma^j,$$

while the adjoint (section of $T^*M \otimes TM$) before type change is

$$\begin{aligned} T^* &= T_k{}^l \cdot \sigma^k \otimes E_l \\ &= T^i{}_j g_{ki} g^{lj} \cdot \sigma^k \otimes E_l. \end{aligned}$$

Thus, we have the nice formula

$$|T|^2 = T^i{}_j T_i{}^j.$$

In the case of the curvature tensor one would normally write

$$R = R^l{}_{ijk} \cdot E_l \otimes \sigma^i \otimes \sigma^j \otimes \sigma^k,$$

and when changing to the (2, 2) version we have

$$\begin{aligned} R &= R^{kl}{}_{ij} \cdot E_k \otimes E_l \otimes \sigma^i \otimes \sigma^j \\ &= R^l{}_{ijs} g^{sk} \cdot E_k \otimes E_l \otimes \sigma^i \otimes \sigma^j. \end{aligned}$$

It is then clear how to keep track of the other (2, 2) versions by writing

$$R_i{}^{jk}{}_l = R_{ist}{}^u g^{js} g^{kt} g_{lu}.$$

Nice as this notation is, it is not used consistently in the literature, probably due to typesetting problems. It would be convenient to use it, but in most cases one can usually keep track of things anyway. Most of this notation can of course also be avoided by using invariant (coordinate-free) notation, but often it is necessary to do coordinate or frame computations both in abstract and concrete situations.

To this we can add yet another piece of notation that is often seen. Namely, if S is a $(1, k)$-tensor written in a frame as:

$$S = S^i_{j_1 \cdots j_k} \cdot E_i \otimes \sigma^{j_1} \otimes \cdots \otimes \sigma^{j_k},$$

then the covariant derivative is a $(1, k+1)$-tensor that can be written as

$$\nabla S = S^i_{j_1 \cdots j_k, j_{k+1}} \cdot E_i \otimes \sigma^{j_1} \otimes \cdots \otimes \sigma^{j_k} \otimes \sigma^{j_{k+1}}.$$

The coefficient $S^i_{j_1 \cdots j_k, j_{k+1}}$ can be computed via the formula

$$\begin{aligned} \nabla_{E_{j_{k+1}}} S = {} & D_{E_{j_{k+1}}} \left(S^i_{j_1 \cdots j_k}\right) \cdot E_i \otimes \sigma^{j_1} \otimes \cdots \otimes \sigma^{j_k} \\ & + S^i_{j_1 \cdots j_k} \cdot \nabla_{E_{j_{k+1}}} \left(E_i \otimes \sigma^{j_1} \otimes \cdots \otimes \sigma^{j_k}\right), \end{aligned}$$

where one must find the expression for

$$\begin{aligned} \nabla_{E_{j_{k+1}}} \left(E_i \otimes \sigma^{j_1} \otimes \cdots \otimes \sigma^{j_k}\right) = {} & \left(\nabla_{E_{j_{k+1}}} E_i\right) \otimes \sigma^{j_1} \otimes \cdots \otimes \sigma^{j_k} \\ & + E_i \otimes \left(\nabla_{E_{j_{k+1}}} \sigma^{j_1}\right) \otimes \cdots \otimes \sigma^{j_k} \\ & \cdots \\ & + E_i \otimes \sigma^{j_1} \otimes \cdots \otimes \left(\nabla_{E_{j_{k+1}}} \sigma^{j_k}\right) \end{aligned}$$

by writing each of the terms $\left(\nabla_{E_{j_{k+1}}} E_i\right), \left(\nabla_{E_{j_{k+1}}} \sigma^{j_1}\right), \ldots, \left(\nabla_{E_{j_{k+1}}} \sigma^{j_k}\right)$ in terms of the frame and coframe and substitute back into the formula.

2.6 Further Study

It is still too early to give useful references. In the upcoming chapters we shall mention several other books on geometry that the reader might wish to consult. At this stage we shall only list the authoritative guide [53]. Every differential geometer must have a copy of these tomes, but their effective usefulness has probably passed away. In a way, it is the Bourbaki of differential geometry and should be treated as such.

2.7 Exercises

1. Show that the connection on Euclidean space is the only connection such that $\nabla X = 0$ for all constant vector fields X.

2. Let ∇ be a connection on a manifold M. If $f : M \to M$ is a diffeomorphism, recall that the push-forward of a vector field is defined as $f_*X(p) = DfX\left(f^{-1}(p)\right)$. A diffeomorphism is said to affine with respect to ∇ if $f_*(\nabla_X Y) = \nabla_{f_*X} f_*Y$ for all vector fields. Show that the affine transformations form a group. Show that the affine transformations on Euclidean space are of the form $f(x) = Ax + b$, where $A \in Gl(n, \mathbb{R})$ and $b \in \mathbb{R}^n$.

3. A manifold is said to be affinely flat if it admits an atlas where the transition functions are affine, i.e., look like $Ax + b$, $A \in Gl_n$, $b \in \mathbb{R}^n$. Show that an affinely flat manifold inherits a natural connection.

4. Let G be a Lie group. Show that there is a unique connection such that $\nabla X = 0$ for all left-invariant vector fields. Show that this connection is torsion free iff the Lie algebra is Abelian.

5. Show that if X is a vector field of constant length on a Riemannian manifold, then $\nabla_v X$ is always perpendicular to X.

6. For any $p \in (M, g)$ and orthonormal basis $e_1, \ldots, e_n$ for T_pM, show that there is an orthonormal frame $E_1, \ldots, E_n$ in a neighborhood of p such that $E_i = e_i$ and $\nabla E_i = 0$ at p. Hint: Fix an orthonormal frame $\bar{E}_i$ near $p \in M$ with $\bar{E}_i(p) = v_i$. Then observe that if we define $E_i = \alpha_i^j \bar{E}_j$, where $\left(\alpha_i^j(x)\right) \in SO(n)$ and $\alpha_i^j(p) = \delta_i^j$, then this will yield the desired frame provided that the $D_{v_k}\alpha_i^j$ are appropriately prescribed.

7. (Riemann) As in the previous problem, but now show that there are coordinates $x^1, \ldots, x^n$ such that $\partial_i = e_i$ and $\nabla \partial_i = 0$ at p. These conditions imply that the metric coefficients satisfy $g_{ij} = \delta_{ij}$ and $\partial_k g_{ij} = 0$ at p. Such coordinates are called normal coordinates at p.

8. Let (M, g) be oriented and define the Riemannian volume form ω as follows:

$$\omega(v_1, \ldots, v_n) = \det\left(g\left(v_i, e_j\right)\right),$$

where $e_1, \ldots, e_n$ is a positively oriented orthonormal basis for T_pM. Show that if $v_1, \ldots, v_n$ is positively oriented, then

$$\omega(v_1, \ldots, v_n) = \sqrt{\det\left(g\left(v_i, v_j\right)\right)}.$$

Show that the volume form is parallel. Show that in coordinates,

$$\omega = \sqrt{\det\left(g_{ij}\right)}dx^1 \wedge \cdots \wedge dx^n.$$

If X is a vector field, show that

$$L_X\omega = \operatorname{div}(X)\,\omega.$$

9. Show that in coordinates, the Laplacian has the formula

$$\Delta u = \frac{1}{\sqrt{\det(g_{ij})}} \partial_k \left(\sqrt{\det(g_{ij})} g^{kl} \partial_l u \right).$$

Given that the coordinates are normal at p, we then get as in Euclidean space that

$$\Delta f(p) = \sum_{i=1}^{n} \partial_i \partial_i f.$$

10. Using Stokes' theorem show that if f has compact support,

$$\int_M \Delta f = 0.$$

Here the integral is taken with respect to the Riemannian volume element constructed above. Show that

$$\Delta(f_1 \cdot f_2) = (\Delta f_1) \cdot f_2 + 2g(\nabla f_1, \nabla f_2) + f_1 \cdot (\Delta f_2).$$

Show that

$$\operatorname{div}(f \cdot X) = g(\nabla f, X) + f \cdot \operatorname{div} X.$$

Show the integration by parts formula for functions with compact support:

$$\int_M f_1 \cdot \Delta f_2 = -\int_M g(\nabla f_1, \nabla f_2).$$

Conclude that if f is sub- or superharmonic (i.e., $\Delta f \geq 0$ or $\Delta f \leq 0$) then f is constant. (Hint: first show $\Delta f = 0$; then use integration by parts on $f \cdot \Delta f$.) This result is known as the *weak maximum principle*. More generally, one can show that any subharmonic (respectively superharmonic) function that has a global maximum (respectively minimum) must be constant. For this one does not need f to have compact support. This result is usually referred to as the *strong maximum principle*.

11. A *Killing field* is a vector field such that $L_X g = 0$. Show that X is a Killing field iff the local flows it generates act by isometries.

12. A vector field or flow is said to be incompressible if $\operatorname{div} X = 0$. Show that X is incompressible iff the local flows it generates are volume preserving (i.e., leave the Riemannian volume form invariant).

13. Find a unit vector field X on $\mathbb{R}^3$ that is incompressible but where $\nabla X \neq 0$. Show that this is not possible on $\mathbb{R}^2$.

14. In coordinates on (M, g) we have the metric coefficients $g_{ij} = g\left(\partial_i, \partial_j\right)$, Christoffel symbols defined by $\nabla_{\partial_i}\partial_j = \Gamma_{ij}^k \partial_k$, and curvature components R_{ijk}^l defined by $R\left(\partial_i, \partial_j\right)\partial_k = R_{ijk}^l \partial_l$. Using the Koszul formula compute the Christoffel symbols in terms of the metric coefficients, and then compute the curvatures. Next, do these computations at p using that the coordinates are normal at p.

    Observe that ∇ at p, in normal coordinates at p, looks like the Euclidean connection. The fact that curvature enters naturally at the second derivative level shows that one cannot find normal coordinates where also $\partial_k \partial_l g_{ij} = 0$ at p. Thus the curvatures in some way measure the second-order variation for the metric to be Euclidean at a point.

15. Given an orthonormal frame $E_1, \ldots, E_n$ on (M, g), define the structure constants c_{ij}^k by $\left[E_i, E_j\right] = c_{ij}^k E_k$. Then define the Γ's and R's as before and compute them in terms of the c's. Notice that on Lie groups with left-invariant metrics the structure constants can be assumed to be constant. In this case, computations simplify considerably.

16. We should mention at least one other effective method for computing the connection and curvatures, namely, the *Cartan formalism*. Let (M, g) be a Riemannian manifold. Given a frame $E_1, \ldots, E_n$, the connection can be written

    $$\nabla E_i = \omega_i^j E_j,$$

    where ω_i^j are 1-forms. Thus,

    $$\nabla_v E_i = \omega_i^j (v) E_j.$$

    Suppose now that the frame is orthonormal and let θ^i be the dual coframe, i.e., $\theta^i\left(E_j\right) = \delta_j^i$. Show that the *connection forms* satisfy

    $$\omega_i^j = -\omega_j^i,$$
    $$d\theta^i = \theta^j \wedge \omega_j^i.$$

    These two equations can, conversely, be used to compute the connection forms given the orthonormal frame. Therefore, if the metric is given by declaring a certain frame to be orthonormal, then this method can be very effective in computing the connection.

    If we think of $\left(\omega_i^j\right)$ as a matrix, then it represents a 1-form with values in the skew-symmetric $n \times n$ matrices, or in other words, with values in the Lie algebra $\mathfrak{so}(n)$ for $O(n)$.

    The *curvature forms* Ω_i^j are 2-forms with values in $\mathfrak{so}(n)$. They are defined as

    $$R(\cdot, \cdot) E_i = \Omega_i^j E_j.$$

Show that they satisfy

$$d\omega_i^j = \omega_i^k \wedge \omega_k^j + \Omega_i^j.$$

When reducing to Riemannian metrics on surfaces we obtain for an orthonormal frame E_1, E_2 with coframe θ^1, θ^2

$$\begin{aligned} d\theta^1 &= \theta^2 \wedge \omega_2^1, \\ d\theta^2 &= -\theta^1 \wedge \omega_2^1, \\ d\omega_2^1 &= \Omega_2^1, \\ \Omega_2^1 &= \sec \cdot d\mathrm{vol}. \end{aligned}$$

17. Show that a Riemannian manifold with parallel Ricci tensor has constant scalar curvature. In Chapter 3 it will be shown that the converse is not true, and also that a metric with parallel curvature tensor doesn't have to be Einstein.

18. Show that if we use for R the $(1, 3)$-tensor and for Ric the $(0, 2)$-tensor, then

$$(\mathrm{div} R)(X, Y, Z) = (\nabla_X \mathrm{Ric})(Y, Z) - (\nabla_Y \mathrm{Ric})(X, Z).$$

Use this to show that $\mathrm{div} R = 0$ if $\nabla \mathrm{Ric} = 0$. Then show that $\mathrm{div} R = 0$ iff the $(1, 1)$ Ricci tensor satisfies:

$$(\nabla_X \mathrm{Ric})(Y) = (\nabla_Y \mathrm{Ric})(X) \text{ for all } X, Y.$$

19. Suppose a Lie group G has a bi-invariant metric, and let ∇ be the associated Riemannian connection. Show using the properties established in the exercises to Chapter 1, that if $X, Y, Z, W \in \mathfrak{g}$, then

(a) $\nabla_X Y = \frac{1}{2}[X, Y]$;
(b) $R(X, Y) Z = \frac{1}{4}[Z, [X, Y]]$; and
(c) $g\big(R(X, Y)Z, W\big) = -\frac{1}{4}(g([X, Y], [Z, W]))$.

Conclude that the sectional curvatures are nonnegative. Show that the curvature operator is also nonnegative by showing that:

$$g\left(\mathfrak{R}\left(\sum_{i=1}^k X_i \wedge Y_i\right), \left(\sum_{i=1}^k X_i \wedge Y_i\right)\right) = \frac{1}{4}\left|\sum_{i=1}^k [X_i, Y_i]\right|^2.$$

Show that $\mathrm{Ric}(X) = 0$ iff X commutes with all other left-invariant vector fields. Thus G has positive Ricci curvature if the center of G is discrete.

There is a linear map $\Lambda^2 \mathfrak{g} \to [\mathfrak{g}, \mathfrak{g}]$ that sends $X \wedge Y$ to $[X, Y]$. Show that the sectional curvature is positive iff this map is an isomorphism. Conclude that this can only happen if $n = 3$ and $\mathfrak{g} = \mathfrak{su}(2)$.

20. It is illustrative to use the Cartan formalism in the above problem and compute all quantities in terms of the structure constants for the Lie algebra. Given that the metric is bi-invariant, it follows that with respect to an orthonormal basis they satisfy

$$c_{ij}^{k} = -c_{ji}^{k} = c_{jk}^{i}.$$

The first equality is antisymmetry of the Lie bracket, and the second is bi-invariance of the metric.

21. (O'Neill) Suppose we have a Riemannian submersion $\varphi : M \to N$. If X is a vector field in N, show that it admits a unique *basic lift* $\tilde{X}$ to M, i.e., $D\varphi\left(\tilde{X}\right) = X$ and $\left|\tilde{X}\right| = |X|$ (or $\tilde{X}$ is perpendicular to the fibers). If v is a vector in TM let $v^{\perp}$ denote the orthogonal projection onto $\ker \varphi$. Show that $A(X, Y) = \frac{1}{2} \cdot [X, Y]^{\perp}$ is a tensor on M.

Show that for two vector fields X, Y on N we have $\widetilde{\nabla_X Y} + A\left(\tilde{X}, \tilde{Y}\right) = \nabla_{\tilde{X}} \tilde{Y}$.

If X and Y are perpendicular unit vector fields in N, establish the O'Neill formula

$$\sec(X, Y) = \sec\left(\tilde{X}, \tilde{Y}\right) + 3 \cdot \left|A\left(\tilde{X}, \tilde{Y}\right)\right|^2.$$

Conclude that N has positive sectional curvature if M does. It is not true that Riemannian submersions increase the curvature operator. As an example we have $S^5 \to \mathbb{C}P^2$. The above formula shows that all sectional curvatures on $\mathbb{C}P^2$ lie in the closed interval $[1, 4]$, however, we shall see that the curvature operator on $\mathbb{C}P^2$ has a zero eigenvalue.

One can find many examples of manifolds with nonnegative or positive curvature using this idea. They all come about by having (M, g) with a free compact group action G by isometries and using $N = M/G$. Examples are: $\mathbb{C}P^n = S^{2n+1}/S^1$, $TS^n = \left(SO\left(n+1\right) \times \mathbb{R}^n\right)/SO(n)$, $SU(3)/T^2$.

22. Suppose we have two Riemannian manifolds (M, g) and (N, h). Then the product has a natural product metric $(M \times N, g + h)$. If X is a vector field on M and Y one on N, show that if we regard these as vector fields on $M \times N$, then $\nabla_X Y = 0$. Conclude that $\sec(X, Y) = 0$. This means that product metrics always have many curvatures that are zero.

23. Suppose that we have two distributions E and F on (M, g), which are orthogonal complements of each other in TM. In addition, assume that the distributions are parallel, i.e., if two vector fields X and Y are tangent to, say, E, then $\nabla_X Y$ is also tangent to E. Show that the distributions are integrable. Show that around any point in M there is a product neighborhood $U = V_E \times V_F$ such that $(U, g) = \left(V_E \times V_F, g_{|E} + g_{|F}\right)$, where $g_{|E}$ and $g_{|F}$ are the restrictions of g to the two distributions. In other words, M is locally a product metric.

24. Let X be a parallel vector field on (M, g). Show that X has constant length. Show that X generates parallel distributions, one that contains X and the other that is the orthogonal complement to X. Conclude that locally the metric is a product with an interval $(U, g) = \left(V \times I, g_{|TV} + dt^2\right)$.

25. For 3-dimensional manifolds, show that if the curvature operator in diagonal form looks like
$$\begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \gamma \end{pmatrix},$$
then the Ricci curvature has a diagonal form like
$$\begin{pmatrix} \alpha+\beta & 0 & 0 \\ 0 & \beta+\gamma & 0 \\ 0 & 0 & \alpha+\gamma \end{pmatrix}.$$
Moreover, the numbers α, β, γ are actually sectional curvatures.

26. The Einstein tensor on a Riemannian manifold is defined as
$$G = \mathrm{Ric} - \frac{\mathrm{scal}}{2} \cdot I.$$
Show that $G = 0$ in dimension 2 and that $\mathrm{div} G = 0$ in higher dimensions. This tensor is supposed to measure the mass/energy distribution. The fact that it is divergence free tells us that energy and angular momentum are conserved. In a vacuum, one therefore imagines that $G = 0$. Show that this happens in dimensions > 2 iff the metric is Ricci flat.

27. This exercise will give you a way of finding the curvature tensor from the sectional curvatures. Using the Bianchi identity show that
$$\begin{aligned} -6R(X, Y, Z, W) = \\ & \left.\frac{\partial^2}{\partial s \partial t}\right|_{s=t=0} \{R(X + sZ, Y + tW, Y + tW, X + sZ) \\ & - R(X + sW, Y + tZ, Y + tZ, X + sW)\}. \end{aligned}$$

28. Using polarization show that the norm of the curvature operator on $\Lambda^2 T_p M$ is bounded by
$$\left|\mathfrak{R}_p\right| \leq c(n) \left|\sec\right|_p$$
for some constant $c(n)$ depending on dimension, and where $|\sec|_p$ denotes the largest absolute value for any sectional curvature of a plane in $T_p M$.

29. We can artificially complexify the tangent bundle to a manifold: $T_{\mathbb{C}}M = TM \otimes \mathbb{C}$. If we have a Riemannian structure, we can extend all the accompanying tensors to this realm. The metric tensor, in particular, gets extended as follows:

$$g_{\mathbb{C}}(v_1 + iv_2, w_1 + iw_2) = g(v_1, w_1) - g(v_2, w_2) + i(g(v_1, w_2) + g(v_2, w_1)).$$

This means that a vector can have "length" zero without being trivial. Such vectors are called isotropic. Clearly, they must have the form $v_1 + iv_2$, where $|v_1| = |v_2|$ and $g(v_1, v_2) = 0$. More generally, we can have isotropic subspaces, i.e., those subspace on which $g_{\mathbb{C}}$ vanishes. If, for instance, a plane is generated by two isotropic vectors $v_1 + iv_2$ and $w_1 + iw_2$, where v_1, v_2, w_1, w_2 are orthogonal, then the plane is isotropic. Note that one must be in dimension ≥ 4 to have isotropic planes. We now say that the isotropic curvatures are positive, if "sectional" curvatures on isotropic planes are positive. This means that if $v_1 + iv_2$ and $w_1 + iw_2$ span the plane and v_1, v_2, w_1, w_2 are orthogonal, then

$$0 < R(v_1 + iv_2, w_1 + iw_2, w_1 + iw_2, v_1 + iv_2).$$

(a) Show that the expression $R(v_1 + iv_2, w_1 + iw_2, w_1 + iw_2, v_1 + iv_2)$ is always a real number.

(b) Show that if the original metric is strictly quarter pinched, i.e., all sectional curvatures lie in an open interval of the form $\left(\frac{1}{4}k, k\right)$, then the isotropic curvatures are positive.

(c) Show that if the sum of the two smallest eigenvalues of the original curvature operator is positive, then the isotropic curvatures are positive.

30. Consider a Riemannian metric (M, g). Now *scale* the metric by multiplying it by a number λ^2. Then we get a new Riemannian manifold $\left(M, \lambda^2 g\right)$. Show that the new connection and (1,3)-curvature tensor remain the same, but that sec, Ric, scal, and $\mathfrak{R}$ all get multiplied by λ^{-2}.

31. Prove properties 1 through 4 from 2.4.5 for rotationally symmetric metrics.

32. For a $(1, 1)$-tensor T on a Riemannian manifold, show that if E_i is an orthonormal basis, then

$$|T|^2 = \sum |T(E_i)|^2.$$

33. If we have two tensors S, T of the same type (r, s), $r = 0, 1$, define the inner product

$$g(S, T)$$

and show that

$$D_X g(S, T) = g(\nabla_X S, T) + g(S, \nabla_X T).$$

If S is symmetric and T skewsymmetric show that $g(S, T) = 0$.

34. Recall that complex manifolds have complex tangent spaces. Thus we can multiply vectors by $\sqrt{-1}$. As a generalization of this we can define an *almost complex* structure. This is a (1,1)-tensor J such that $J^2 = -I$. Show that the *Nijenhuis tensor:*

$$N(X, Y) = [J(X), J(Y)] - J([J(X), Y]) - J([X, J(Y)]) - [X, Y]$$

is indeed a tensor. If J comes from a complex structure then $N = 0$, conversely Newlander-Nirenberg have shown that J comes from a complex structure if $N = 0$.

An *Hermitian structure* on a Riemannian manifold (M, g) is an almost complex structure J such that

$$g(J(X), J(Y)) = g(X, Y).$$

The *Kähler form* of a Hermitian structure is

$$\omega(X, Y) = g(J(X), Y).$$

Show that ω is a 2-form. Show that $d\omega = 0$ iff $\nabla J = 0$. If the Kähler from is closed, then we call the metric a Kähler metric.

3

Examples

We are now ready to compute the covariant derivative and curvature tensors on the examples we constructed earlier. After computing these quantities in general, we will try to find examples of manifolds with constant sectional, Ricci, and scalar curvature. In particular, we shall look at the standard product metrics on spheres and also construct the Riemannian version of the Schwarzschild metric.

The examples we present here, with the exception of the Berger spheres, were all understood by the early part of this century. Elementary as they are, they still form the foundation for many constructions in Riemannian geometry. There are two additional important constructions we do not cover in detail: left-invariant metrics and submersion metrics. We give some examples of both but do not develop the entire theory.

3.1 Computational Simplifications

Before we present the examples it will be useful to have some general results that deal with how one finds the range of the various curvatures.

Proposition 1.1 *Let e_i be an orthonormal basis for T_pM. If $e_i \wedge e_j$ diagonalize the curvature operator*

$$\mathfrak{R}\left(e_i \wedge e_j\right) = \lambda_{ij} e_i \wedge e_j$$

then for any plane π in T_pM we have $\sec(\pi) \in \left[\min \lambda_{ij}, \max \lambda_{ij}\right]$.

Proof. If v, w form an orthonormal basis for π, then we have $\sec(\pi) = g(\mathfrak{R}(v \wedge w), (v \wedge w))$, so the result is immediate. □

Proposition 1.2 *Let e_i be an orthonormal basis for T_pM and suppose that $R(e_i, e_j)e_k = 0$ if the indices are mutually distinct; then $e_i \wedge e_j$ diagonalize the curvature operator.*

Proof. If we use

$$g(\mathfrak{R}(e_i \wedge e_j), (e_k \wedge e_l)) = -g(R(e_i, e_j)e_k, e_l)$$
$$= g(R(e_i, e_j)e_l, e_k),$$

then we see that this expression is 0 when i, j, k are mutually distinct or if i, j, l are mutually distinct. Thus, the expression can only be nonzero when $\{k, l\} = \{i, j\}$. This gives the result. □

Combining these two results we get a very good criterion for when it is easy to find the range of the sectional curvatures. We shall see that in all rotationally symmetric and doubly warped products we can find e_i such that $R(e_i, e_j)e_k = 0$. In this case, the curvature operator can then be computed by finding the expressions $R(e_i, e_j, e_j, e_i)$. In general, however, this will not happen. But there is still a way to find the range of the Ricci curvatures:

Proposition 1.3 *Let e_i be an orthonormal basis for T_pM and suppose that*

$$g(R(e_i, e_j)e_k, e_l) = 0$$

if three of the indices are mutually distinct, then e_i diagonalizes Ric.

Proof. Recall that

$$g(\text{Ric}(e_i), e_j) = \sum_{k=1}^{n} g(R(e_i, e_k)e_k, e_j),$$

so if we assume that $i \neq j$, then $g(R(e_i, e_k)e_k, e_j) = 0$ unless k is either i or j. However, if $k = i, j$, then the expression is zero from the symmetry properties. Thus, e_i must diagonalize Ric. □

We shall at the very end of this section give an example where Proposition 1.2 cannot be used, but where Proposition 1.3 can be used.

The procedure for finding $\mathfrak{R}$, Ric, and scal uses the fundamental equations. Since Ric and scal can be computed from the curvature tensor, we focus our energies on $\mathfrak{R}$. The procedure uses the following guidelines: We have a fixed Riemannian manifold (M, g) and a smooth distance function f on some open subset U. Note that if U is dense in M and we can compute the curvatures on U, then we know the curvatures on all of M using continuity. Given the distance function, we have the setup as in Section 2.4 from the previous chapter, with $N = \nabla f = \partial_r$, $U_r = f^{-1}(r)$, $g_r = g|_{U_r}$, ∇^r, R^r etc. Now choose an orthonormal frame $\{F_1, \ldots, F_n\}$ such that $N = F_n$, and consequently $\{F_1, \ldots, F_{n-1}\}$ is an orthonormal frame on (U_r, g_r). We shall in addition assume that the sets U_r are the same and that g_r depends only on r in the sense that the frame $\{F_1, \ldots, F_{n-1}\}$ can be written as $\left\{(\varphi_1(r))^{-1}E_1, \ldots, (\varphi_{n-1}(r))^{-1}E_{n-1}\right\}$ where $\{E_1, \ldots, E_{n-1}\}$ are vector fields on U_r, i.e., they don't depend on r. Because of this we have in addition that $[N, E_i] = 0$. This extra condition on the metric g_r, while certainly not always satisfied, does hold in all of the warped product type of examples from Chapter 1. We could instead use the coordinates we introduced in the last section of the previous chapter. It is, however, not always possible to find coordinates that diagonalize the metric, and in such cases one would have to invert the metric tensor to compute the curvatures. So while this approach is much easier in some situations, it would make some more complicated situations even more complicated. Such a situation occurs in the last section of this chapter.

To compute things on U we first need the Hessian S of f. This is found from the formula

$$(L_N g)(F_i, F_j) = 2g(S(F_i), F_j).$$

If either i or j is n, then we know that $2g(S(F_i), F_j) = 0$ since $S(N) = 0$. For other i, j we can use

$$\begin{aligned}
(L_N g)(F_i, F_j) &= L_N(g(F_i, F_j)) - g([N, F_i], F_j) - g(F_i, [N, F_j]) \\
&= -g([N, F_i], F_j) - g(F_i, [N, F_j]) \\
&= -g\left(-\frac{\dot{\varphi}_i}{\varphi_i}F_i, F_j\right) - g\left(F_i, -\frac{\dot{\varphi}_j}{\varphi_j}F_j\right) \\
&= \left(\frac{\dot{\varphi}_i}{\varphi_i} + \frac{\dot{\varphi}_j}{\varphi_j}\right) g(F_i, F_j) \\
&= \left(\frac{\dot{\varphi}_i}{\varphi_i} + \frac{\dot{\varphi}_j}{\varphi_j}\right) \delta_{ij}.
\end{aligned}$$

Thus,

$$\begin{aligned}
g(S(F_i), F_j) &= 0 \text{ when } i \neq j, \\
g(S(F_i), F_i) &= \frac{\dot{\varphi}_i}{\varphi_i},
\end{aligned}$$

from which we see that

$$\nabla_{F_i} N = S(F_i) = \frac{\dot{\varphi}_i}{\varphi_i} F_i = \lambda_i F_i.$$

We can then compute

$$\nabla_N E_i = \nabla_{E_i} N = \varphi_i \nabla_{F_i} N$$

using that $[N, E_i] = 0$. For the other connection terms $\nabla_{F_i} F_j$, where $i, j < n$, we first observe that to compute the curvature tensor it suffices to know $\nabla_{E_i} E_j$. We can now use

$$\nabla_{E_i} E_j = \nabla^r_{E_i} E_j - g\left(S(E_i), E_j\right) N.$$

From the Koszul formula we observe that

$$\begin{aligned} 2g_r\left(\nabla^r_{E_i} E_j, E_k\right) &= D_{E_i} g_r\left(E_j, E_k\right) + D_{E_j} g_r\left(E_i, E_k\right) - D_{E_k} g_r\left(E_j, E_i\right) \\ &\quad + g_r\left(\left[E_i, E_j\right], E_k\right) - g_r\left(\left[E_j, E_k\right], E_i\right) + g_r\left(\left[E_k, E_i\right], E_j\right) \\ &= D_{E_i}\left(\varphi_j \varphi_k \delta_{jk}\right) + D_{E_j}\left(\varphi_i \varphi_k \delta_{ik}\right) - D_{E_k}\left(\varphi_j \varphi_i \delta_{ij}\right) \\ &\quad + g_r\left(\left[E_i, E_j\right], E_k\right) - g_r\left(\left[E_j, E_k\right], E_i\right) + g_r\left(\left[E_k, E_i\right], E_j\right) \\ &= g_r\left(\left[E_i, E_j\right], E_k\right) - g_r\left(\left[E_j, E_k\right], E_i\right) + g_r\left(\left[E_k, E_i\right], E_j\right), \end{aligned}$$

since the φ's only depend on r. Here it is easy to compute the last line when one knows the relevant Lie brackets.

Now, for the curvature operator $\mathfrak{R}$ we must compute

$$\mathfrak{R}\left(F_i \wedge F_j\right), \quad \text{where} \quad i < j.$$

For this we define $E_n = N$ and actually compute $R\left(E_i, E_j\right) E_k$ and then use tensoriality to find $R\left(F_i, F_j\right) F_k$. Finally, we use

$$\mathfrak{R}\left(F_i \wedge F_j\right) = \sum_{l<k} g\left(R\left(F_i, F_j\right) F_k, F_l\right) F_l \wedge F_k.$$

Using the radial curvature equation from Chapter 2 we can compute, for $i < n$,

$$\begin{aligned} R(E_i, E_n) E_n &= R(E_i, N) N \\ &= -(\nabla_N S)(E_i) - S^2(E_i) \\ &= -\nabla_N(S(E_i)) + S(\nabla_N E_i) - S^2(E_i) \\ &= -\nabla_N(S(E_i)) + S\left(\nabla_{E_i} N\right) - S^2(E_i) \\ &= -\nabla_N(S(E_i)) \\ &= -\nabla_N\left(\frac{\dot\varphi_i}{\varphi_i} E_i\right) \\ &= -\left(\frac{d}{dr}\frac{\dot\varphi_i}{\varphi_i}\right) E_i - \frac{\dot\varphi_i}{\varphi_i} \nabla_N(E_i) \\ &= -\frac{\ddot\varphi_i \varphi_i - \dot\varphi_i \dot\varphi_i}{\varphi_i^2} E_i - \frac{\dot\varphi_i}{\varphi_i} \nabla_{E_i} N \\ &= -\frac{\ddot\varphi_i}{\varphi_i} E_i. \end{aligned}$$

From the tangential curvature equation we get, for $i, j, k, l < n$,

$$g\left(R\left(F_i, F_j\right) F_k, F_l\right) = g_r\left(R^r\left(F_i, F_j\right) F_k, F_l\right)$$

$$
\begin{aligned}
&+ g\left(S\left(F_i\right), F_k\right) g\left(S\left(F_j\right), F_l\right) - g\left(S\left(F_j\right), F_k\right) g\left(S\left(F_i\right), F_l\right) \\
&= g_r\left(R^r\left(F_i, F_j\right) F_k, F_l\right) \\
&\quad + \frac{\dot{\varphi}_i \dot{\varphi}_j}{\varphi_i \varphi_j} g\left(F_i, F_k\right) g\left(F_j, F_l\right) - \frac{\dot{\varphi}_i \dot{\varphi}_j}{\varphi_i \varphi_j} g\left(F_j, F_k\right) g\left(F_i, F_l\right) \\
&= g_r\left(R^r\left(F_i, F_j\right) F_k, F_l\right) \\
&\quad + \frac{\dot{\varphi}_i \dot{\varphi}_j}{\varphi_i \varphi_j}\left(\delta_{ik}\delta_{jl} - \delta_{jk}\delta_{il}\right).
\end{aligned}
$$

We shall always be in a situation where $R^r\left(F_i, F_j\right) F_k = 0$ if the indices are distinct. Using that we are only looking for terms with $i < j$ and $l < k$, we see that the only nonzero term is:

$$
g\left(R\left(F_i, F_j\right) F_j, F_i\right) = g_r\left(R^r\left(F_i, F_j\right) F_j, F_i\right) - \frac{\dot{\varphi}_i \dot{\varphi}_j}{\varphi_i \varphi_j}.
$$

Finally, the normal curvature equation from Chapter 2 together with our definition $\lambda_i = \dot{\varphi}_i / \varphi_i$ yields, for $i, j, k < n$,

$$
\begin{aligned}
g\left(R\left(E_i, E_j\right) E_k, E_n\right) &= -g\left(\left(\nabla_{E_i} S\right)\left(E_j\right), E_k\right) + g\left(\left(\nabla_{E_j} S\right)\left(E_i\right), E_k\right) \\
&= -g\left(\nabla_{E_i}\left(\lambda_j E_j\right), E_k\right) + g\left(\nabla_{E_j}\left(\lambda_i E_i\right), E_k\right) \\
&\quad + g\left(S\left(\nabla_{E_i} E_j\right), E_k\right) - g\left(S\left(\nabla_{E_j} E_i\right), E_k\right) \\
&= -g\left(\lambda_j \nabla_{E_i} E_j, E_k\right) + g\left(\lambda_i \nabla_{E_j} E_i, E_k\right) \\
&\quad + g\left(S\left(\left[E_i, E_j\right]\right), E_k\right) \\
&= -\lambda_j g_r\left(\nabla_{E_i} E_j, E_k\right) + \lambda_i g_r\left(\nabla_{E_j} E_i, E_k\right) \\
&\quad + g\left(\left[E_i, E_j\right], S\left(E_k\right)\right) \\
&= -\lambda_j g_r\left(\nabla_{E_i} E_j, E_k\right) + \lambda_i g_r\left(\nabla_{E_j} E_i, E_k\right) \\
&\quad + \lambda_k g_r\left(\nabla_{E_i} E_j, E_k\right) - \lambda_k g_r\left(\nabla_{E_j} E_i, E_k\right) \\
&= \left(\lambda_k - \lambda_j\right) g_r\left(\nabla^r_{E_i} E_j, E_k\right) + \left(\lambda_i - \lambda_k\right) g_r\left(\nabla^r_{E_j} E_i, E_k\right).
\end{aligned}
$$

Observe that we can replace E's by F's, as both sides are tensorial in all variables. While things could definitely look better, we have achieved some serious simplifications. Namely, if $i < n$, then

$$
\begin{aligned}
\mathfrak{R}\left(F_i \wedge F_n\right) &= \sum_{l<k} g\left(R\left(F_i, F_n\right) F_k, F_l\right) F_l \wedge F_k \\
&= \sum_{l<k<n} g\left(R\left(F_i, F_n\right) F_k, F_l\right) F_l \wedge F_k \\
&\quad + \sum_{l<n} g\left(R\left(F_i, F_n\right) F_n, F_l\right) F_l \wedge F_n \\
&= \sum_{l<k<n} g\left(R\left(F_k, F_l\right) F_i, F_n\right) F_l \wedge F_k \\
&\quad + g\left(R\left(F_i, F_n\right) F_n, F_i\right) F_i \wedge F_n
\end{aligned}
$$

$$= \sum_{l<k<n} \left((\lambda_i - \lambda_l)\, g_r \left(\nabla^r_{F_k} F_l, F_i \right) + (\lambda_k - \lambda_i)\, g_r \left(\nabla^r_{F_l} F_k, F_i \right) \right) F_l \wedge F_k$$
$$- \frac{\ddot{\varphi}_i}{\varphi_i} F_i \wedge F_n.$$

And when $i < j < n$, we have

$$\begin{aligned}
\mathfrak{R}\left(F_i \wedge F_j\right) &= \sum_{l<k} g\left(R\left(F_i, F_j\right) F_k, F_l\right) F_l \wedge F_k \\
&= \sum_{l<k<n} g\left(R\left(F_i, F_j\right) F_k, F_l\right) F_l \wedge F_k \\
&\quad + \sum_{l<n} g\left(R\left(F_i, F_j\right) F_n, F_l\right) F_l \wedge F_n \\
&= \sum_{l<k<n} g\left(R\left(F_i, F_j\right) F_k, F_l\right) F_l \wedge F_k \\
&\quad - \sum_{l<n} g\left(R\left(F_i, F_j\right) F_l, F_n\right) F_l \wedge F_n \\
&= \sum_{l<k<n} \left(g_r\left(R^r\left(F_i, F_j\right) F_k, F_l\right) - \lambda_k \lambda_l\right) F_l \wedge F_k \\
&\quad - \sum_{l<n} \Big((\lambda_l - \lambda_j)\, g_r \left(\nabla^r_{F_i} F_j, F_l\right) \\
&\qquad\qquad + (\lambda_i - \lambda_l)\, g_r \left(\nabla^r_{F_j} F_i, F_l\right) \Big) F_l \wedge F_n.
\end{aligned}$$

Proposition 1.4 *Suppose $M = I \times \Omega$ and $g = dt^2 + g_r$, where g_r is a family of metrics on Ω. Suppose that g_r admits an orthonormal frame of the form $\{F_1, \ldots, F_{n-1}\} = \left\{(\varphi_1)^{-1} E_1, \ldots, (\varphi_{n-1})^{-1} E_{n-1}\right\}$, where the φ's depend only on r and the E's are vector fields on Ω, and in addition that the curvature R^r of g_r satisfies $R^r\left(F_i, F_j\right) F_k = 0$ if the three indices are distinct. Then*

$$\begin{aligned}
\mathfrak{R}\left(F_i \wedge F_n\right) &= -\frac{\ddot{\varphi}_i}{\varphi_i} F_i \wedge F_n \\
&\quad + \sum\nolimits_{l<k<n} (\lambda_i - \lambda_l)\, g_r \left(\nabla^r_{F_k} F_l, F_i\right) F_l \wedge F_k \\
&\quad + \sum\nolimits_{l<k<n} (\lambda_k - \lambda_i)\, g_r \left(\nabla^r_{F_l} F_k, F_i\right) F_l \wedge F_k,
\end{aligned}$$

$$\begin{aligned}\mathfrak{R}\left(F_i \wedge F_j\right) = &\left(g_r\left(R^r\left(F_i, F_j\right) F_j, F_i\right) - \lambda_i \lambda_j\right) F_i \wedge F_j \\ &-\sum\nolimits_{l<n} \left(\lambda_l - \lambda_j\right) g_r\left(\nabla^r_{F_i} F_j, F_l\right) F_l \wedge F_n \\ &-\sum\nolimits_{l<n} \left(\lambda_i - \lambda_l\right) g_r\left(\nabla^r_{F_j} F_i, F_l\right) F_l \wedge F_n,\end{aligned}$$

where $\lambda_i = \dot{\varphi}_i / \varphi_i$.

3.2 Warped Products

So far, all we know about curvature is that $(\mathbb{R}^n, \text{can})$ has $R \equiv 0$. Using this, let us figure out what R is on $(S^{n-1}(r), \text{can})$.

3.2.1 Spheres

On $\mathbb{R}^n$ we have the distance function $r(x) = |x|$. The level sets are $U_r = S^{n-1}(r)$ with the usual induced metric. The gradient $\partial_r = \nabla r = N = \frac{1}{r} x^i \partial_i$, and hence,

$$\begin{aligned} S &= \nabla N \\ &= d\left(\frac{x^i}{r}\right) \partial_i \\ &= \left(\frac{1}{r} dx^i - \left(\frac{x^i}{r^3} \sum_{j=1}^n x^j dx^j\right)\right) \partial_i \\ &= \frac{1}{r} dx^i \partial_i - \sum_{j=1}^n \frac{x^i x^j}{r^3} dx^j \partial_i \\ &= \frac{1}{r} dx^i \partial_i - \sum_{j=1}^n \frac{x^j dx^j}{r^2} N, \end{aligned}$$

which means that $S(v) = \frac{1}{r} \cdot v - \frac{1}{r} g(v, N) N$, or in other words, $S(v) = \frac{1}{r} v$ if $v \in TU_r = TS^{n-1}(r)$ and $S(N) = 0$. The tangential curvature equation now tells us that

$$R^r(X, Y)Z = r^{-2}(g_r(Y, Z)X - g_r(X, Z)Y),$$

since R, the curvature on $\mathbb{R}^n$, is zero. In particular, if e_i is any orthonormal basis, we see that $R^r\left(e_i, e_j\right) e_k = 0$ when the indices are mutually distinct. Therefore, $(S^{n-1}(r), \text{can})$ has constant curvature r^{-2}, provided that $n \geq 3$. This, in particular, justifies our notation $S^n_k =$ the rotational symmetric metric $dr^2 + \text{sn}^2_k(r) ds^2_{n-1}$ when $k \geq 0$, since these metrics have curvature k in this case. We shall see in a second that this is also true when $k < 0$.

3.2.2 Product Spheres

Let us next compute the curvatures on the product spheres $S^n_a \times S^m_b = S^n(1/\sqrt{a}) \times S^m(1/\sqrt{b})$. First, notice that the metric g_r on $S^n(r)$ is the same as $g_r = r^2 g_1$, so we can write $S_a^n \times S_b^m = \left(S^n \times S^m, \frac{1}{a}ds_n^2 + \frac{1}{b}ds_m^2\right)$. We can therefore fix orthonormal framings $\{E_1, \ldots, E_n\}$ and $\{E_{n+1}, \ldots, E_{m+n}\}$ on $\left(S^n, ds_n^2\right)$ and $\left(S^m, ds_m^2\right)$ to get an orthonormal framing $\left\{\sqrt{a}E_1, \ldots, \sqrt{a}E_n, \sqrt{b}E_{n+1}, \ldots, \sqrt{b}E_{n+m}\right\}$ on $S_a^n \times S_b^m$. Since this is an orthonormal frame, we see that the Koszul formula reduces to

$$\begin{aligned} 2g\left(\nabla_{\sqrt{a}E_i}\sqrt{b}E_j, \sqrt{?}E_k\right) &= g\left(\left[\sqrt{a}E_i, \sqrt{b}E_j\right], \sqrt{?}E_k\right) \\ &\quad + g\left(\left[\sqrt{?}E_k, \sqrt{a}E_i\right]\sqrt{b}E_j,\right) \\ &\quad - g\left(\left[\sqrt{b}E_j, \sqrt{?}E_k\right], \sqrt{a}E_i\right). \end{aligned}$$

When $i \leq n$ and $j > n$, we have that $\left[\sqrt{a}E_i, \sqrt{b}E_j\right] = 0$, and in the two other terms the Lie bracket is either zero or perpendicular to the other term, depending on whether $k \leq n$ or $k > n$. Thus, $\nabla_{\sqrt{a}E_i}\sqrt{b}E_j = \nabla_{\sqrt{b}E_j}\sqrt{a}E_i = 0$. In particular, if we write $\{F_1, \ldots, F_{n+m}\} = \left\{\sqrt{a}E_1, \ldots, \sqrt{a}E_n, \sqrt{b}E_{n+1}, \ldots, \sqrt{b}E_{n+m}\right\}$, then

$$R\left(F_i, F_j\right)F_k = 0$$

unless $1 \leq i, j, k \leq n$ or $n + 1 \leq i, j, k \leq n + m$. In the case $1 \leq i, j, k \leq n$ we reproduce the answer from S_a^n, while if $n + 1 \leq i, j, k \leq m + n$, we recapture S_b^m. Thus, $R\left(F_i, F_j\right)F_k = 0$ unless $i = k$ or $j = k$. With this information we can compute

$$\begin{aligned} \mathfrak{R}(F_i \wedge F_j) &= 0 \quad \text{if} \quad i \leq n, \ j \geq n+1, \\ \mathfrak{R}(F_i \wedge F_j) &= aF_i \wedge F_j \quad \text{if} \quad i, j \leq n, \\ \mathfrak{R}(F_i \wedge F_j) &= bF_i \wedge F_j \quad \text{if} \quad i, j \geq n+1. \end{aligned}$$

In particular, all sectional curvatures lie in the interval $[0, \max\{a, b\}]$. From this we see

$$\begin{aligned} \operatorname{Ric}(F_i) &= (n-1)aF_i, \quad i \leq n, \\ \operatorname{Ric}\left(F_j\right) &= (m-1)bF_j, \quad j \geq n+1, \\ \operatorname{scal} &= n(n-1)a + m(m-1)b. \end{aligned}$$

We therefore conclude that $S_a^n \times S_b^m$ always has constant scalar curvature, is an Einstein manifold exactly when $(n-1)a = (m-1)b$ (which requires $n, m \geq 2$ or $n = m = 1$), but never has constant sectional curvature. Note that the curvature tensor on $S_a^n \times S_b^m$ is always parallel, but the metric doesn't have to be Einstein.

3.2.3 *Rotationally Symmetric Metrics*

We shall now compute the curvatures on $(M, g) = \left(I \times S^{n-1}, dr^2 + \varphi^2(r)ds_{n-1}^2\right)$. We know that $f(x) = f(r, y) = r$ is a distance function. On $\left(S^{n-1}, ds_{n-1}^2\right)$

choose an orthonormal frame $\{E_1, \ldots, E_{n-1}\}$; then $\{F_1, \ldots, F_n\} = \left\{\frac{1}{\varphi}E_1, \ldots, \frac{1}{\varphi}E_{n-1}, N = \partial_r = \nabla f\right\}$ will be an orthonormal frame on (M, g). We can therefore use the approach that we set up in Section 3.1, with the simplification that $\varphi_1 = \cdots = \varphi_{n-1}$. This implies first of all that all the mixed curvatures vanish. Thus only the radial and tangential curvatures are relevant. Using that g_r is the metric of curvature φ^{-2} on the sphere, we see that

$$g_r\left(R^r\left(F_i, F_j\right)F_j, F_i\right) = \frac{1}{\varphi^2}.$$

From this we can conclude

$$\mathfrak{R}(F_i \wedge F_n) = -\frac{\ddot{\varphi}}{\varphi}F_i \wedge F_n \quad \text{if} \quad i < n,$$

$$\mathfrak{R}(F_i \wedge F_j) = \frac{1-\dot{\varphi}^2}{\varphi^2}F_i \wedge F_j \quad \text{if} \quad i, j < n.$$

In particular, the $F_i \wedge F_j$'s diagonalize $\mathfrak{R}$, and hence all sectional curvatures lie between the two values $-(\ddot{\varphi}/\varphi)$ and $(1-\dot{\varphi}^2)\varphi^{-2}$. Furthermore,

$$\begin{aligned}
\operatorname{Ric}(F_i) &= \sum_{j=1}^{n} R\left(F_i, F_j\right)F_j \\
&= \sum_{j=1}^{n-1} R\left(F_i, F_j\right)F_j + R\left(F_i, F_n\right)F_n \\
&= \left((n-2)\frac{1-\dot{\varphi}^2}{\varphi^2} - \frac{\ddot{\varphi}}{\varphi}\right)F_i \quad \text{if} \quad i < n, \\
\operatorname{Ric}(F_n) &= -(n-1)\frac{\ddot{\varphi}}{\varphi}F_n \\
\operatorname{scal} &= -(n-1)\frac{\ddot{\varphi}}{\varphi} + (n-1)\left((n-2)\frac{1-\dot{\varphi}^2}{\varphi^2} - \frac{\ddot{\varphi}}{\varphi}\right) \\
&= -2(n-1)\frac{\ddot{\varphi}}{\varphi} + (n-1)(n-2)\frac{1-\dot{\varphi}^2}{\varphi^2}.
\end{aligned}$$

Notice that when $n = 2$, we have $\sec = -(\ddot{\varphi}/\varphi)$, because there are no tangential curvatures. This makes for quite a difference between rotationally symmetric metrics in dimension 2 and those of higher dimension.

Constant curvature: First, we should compute the curvature of: $(S_k^n, dr^2 + \operatorname{sn}_k^2(r)ds_{n-1}^2)$. Since $\varphi = \operatorname{sn}_k$ solves $\ddot{\varphi} + k\varphi = 0$ we see that $\sec(E_i, N) = k$. To compute $\sec(E_i, E_j) = \frac{1-\dot{\varphi}^2}{\varphi^2}$, just recall that $\operatorname{sn}_k(r) = \frac{1}{\sqrt{k}}\sin(\sqrt{k}\,r)$ (even when $k < 0$), so $\dot{\varphi} = \cos(\sqrt{k}\,r)$ and $1 - \dot{\varphi}^2 = \sin^2(\sqrt{k}\,r) = k\varphi^2$. Thus, all sectional curvatures are equal to k, just as promised.

Next let us see if we can find any interesting Ricci flat or scalar flat examples.

Ricci flat metrics: A Ricci flat metric must satisfy $\frac{\ddot{\varphi}}{\varphi} = 0$ and $(n-2)\frac{1-\dot{\varphi}^2}{\varphi^2} - \frac{\ddot{\varphi}}{\varphi} = 0$. Hence, if $n > 2$, we must have $\ddot{\varphi} \equiv 0$ and $\dot{\varphi}^2 \equiv 1$. Thus, $\varphi(r) = a \pm r$. In case $n = 2$ we only need $\ddot{\varphi} = 0$. In any case, the only Ricci flat rotationally symmetric metrics are, in fact, flat.

Scalar flat metrics: To find scalar flat metrics we need to solve

$$2(n-1)\left[-\frac{\ddot{\varphi}}{\varphi} + \frac{n-2}{2} \cdot \frac{1-\dot{\varphi}^2}{\varphi^2}\right] = 0$$

when $n \geq 3$. We rewrite this equation as

$$-\varphi\ddot{\varphi} + \frac{n-2}{2}(1-\dot{\varphi}^2) = 0.$$

This is an autonomous second-order equation, so the change of variables $\dot{\varphi} = G(\varphi)$, $\ddot{\varphi} = G'\dot{\varphi} = G'G$ will yield a first-order equation:

$$-\varphi G'G + \frac{n-2}{2}(1-G^2) = 0.$$

Using separation of variables, we see that G must satisfy

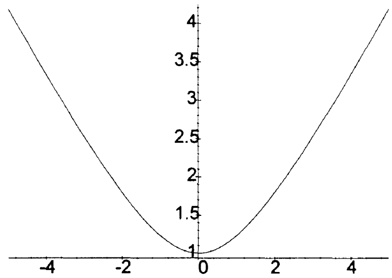
$$\dot{\varphi}^2 = G^2 = 1 + C\varphi^{2-n},$$

which after differentiation becomes:

$$2\ddot{\varphi}\dot{\varphi} = (2-n)C\varphi^{1-n}\dot{\varphi}.$$

To analyze the solutions to this equation that are positive and therefore yield Riemannian metrics, we need to study the cases $C > 0$, $C = 0$, $C < 0$ separately. But first, notice that if $C \neq 0$, then we cannot have that $\varphi(a) = 0$, as this would imply $\ddot{\varphi}(a) = \infty$.

- $C = 0$: In this case, we have $\ddot{\varphi} \equiv 0$ and $\dot{\varphi}^2(0) = 1$. Thus, $\varphi = a + r$ is the only solution and the metric is the standard Euclidean metric.
- $C > 0$: First, observe that from the equation $2\ddot{\varphi} = (2-n)C\varphi^{1-n}$ we get that φ is concave. Thus, if φ is extended to its maximal interval, it must cross the "x-axis," but as pointed out above this means that $\ddot{\varphi}$ becomes undefined, and therefore we don't get any metrics on the sphere this way.
- $C < 0$: This time the solution is convex and doesn't cross the "x-axis" as before. Thus we can assume that it is positive wherever defined. We claim that in fact φ must exist for all time. Otherwise, we could find $a \in \mathbb{R}$ where $\varphi(t) \to \infty$ as $t \to a$ from the left or right. But then $\dot{\varphi}^2(t) \to 1$, which is clearly impossible. Next, observe that $\varphi \to \infty$ as $t \to \pm\infty$. Since $\dot{\varphi}^2(t)$ doesn't converge to 0, the only other possibility is that φ converges to some



Graph of φ when $n = 3$ and $\alpha = 1$

FIGURE 3.1.

finite value. But then $\dot{\varphi}$ will also converge to a finite nonzero value, which is impossible. Finally, we can then conclude that φ must have a unique positive minimum. Using translational invariance of the solutions, we can assume that this minimum is achieved at $t = 0$.

So assume that $\varphi(0) = \alpha$ and in addition $\dot{\varphi}(0) = 0$. We get the relation $0 = \dot{\varphi}^2(0) = 1 + C \cdot \alpha^{2-n}$, which tells us that $C = -\alpha^{n-2}$. Let $\varphi_\alpha(r)$ denote this solution. Thus, we have a scalar flat rotationally symmetric metric on $\mathbb{R} \times S^{n-1}$. Notice that φ_α is also even, and so $(r, x) \to (-r, -x)$ is an isometry on $\left(\mathbb{R} \times S^{n-1}, dr^2 + \varphi_\alpha^2(r) ds_{n-1}^2\right)$. We therefore get a Riemannian covering map $\mathbb{R} \times S^{n-1} \to \tau(\mathbb{R}P^{n-1})$ and a scalar flat metric on $\tau(\mathbb{R}P^{n-1})$, the tautological line bundle over $\mathbb{R}P^{n-1}$. One can prove that $\varphi_\alpha(r) \geq |r|$ for all $r \in \mathbb{R}$ and that $\varphi_\alpha(r) \cdot |r|^{-1} \to 1$ as $r \to \infty$. Thus $\dot{\varphi}^2 = 1 - \alpha^{n-1}\varphi^{2-n} \simeq 1 - \alpha^{n-1}|r|^{2-n}$ and $\ddot{\varphi} \simeq (n-2)\alpha^{n-2}|r|^{1-n}$ as $r \to \infty$. This means, in particular, that all sectional curvatures are $\simeq |r|^{-n}$ as $r \to \infty$. The rotationally symmetric metric $dr^2 + \varphi_\alpha^2(r) ds_{n-1}^2$ therefore looks very much like $dr^2 + r^2 ds_{n-1}^2$ at ∞. Figure 3.1 shows a picture of the warping function when $\alpha = 1$ and $n = 3$.

We shall in Chapter 6 be able to show that $\mathbb{R} \times S^{n-1}$, $n \geq 3$, does not admit a constant curvature metric. Later in Chapter 9, we will see that if $\mathbb{R} \times S^{n-1}$ has $\mathrm{Ric} \equiv 0$, then S^{n-1} also has a metric with $\mathrm{Ric} \equiv 0$. When $n = 3$ or 4 this means that S^2 and S^3 have flat metrics, and we shall see in Chapter 6 that this is not possible. Thus we have found a manifold with a nice scalar flat metric that does not carry any Ricci flat or constant curvature metrics.

3.2.4 Doubly Warped Products

We wish to compute the curvatures on

$$\left(I \times S^p \times S^q, dr^2 + \varphi^2(r) ds_p^2 + \psi^2(r) ds_q^2\right).$$

To do this choose an orthonormal frame $E_1, \ldots, E_{p+q}$ on the product sphere $(S^p \times S^q,\ ds_p^2 + ds_q^2)$ as we did previously. Then $F_0 = N,\ F_1 = 1/\varphi E_1, \ldots, F_p = 1/\varphi E_p,\ F_{p+1} = 1/\psi E_{p+1}, \ldots, F_{p+q} = 1/\psi E_{p+q}$ is an orthonormal frame on our Riemannian manifold. In the notation of Section 3.1 we therefore have that:

$$\begin{aligned} \varphi_i &= \varphi \quad \text{for} \quad i \leq p. \\ \varphi_i &= \psi \quad \text{for} \quad p < i \leq p+q. \end{aligned}$$

Thus, the mixed curvatures can be written

$$\begin{aligned} g\left(R\left(E_i, E_j\right) E_k, N\right) &= \left(\lambda_k - \lambda_j\right) g_r\left(\nabla^r_{E_i} E_j, E_k\right) + \left(\lambda_i - \lambda_k\right) g_r\left(\nabla^r_{E_j} E_i, E_k\right) \\ &= \left(\lambda_k - \lambda_i\right) g_r\left(\left[E_i, E_j\right], E_k\right) \end{aligned}$$

if either $i, j \leq p$ or $p < i, j \leq p+q$. So if $0 < i, j, k \leq p$ or $p < i, j, k \leq p+q$, then the mixed curvatures vanish since $\lambda_i = \lambda_j = \lambda_k$. If $0 < i, j \leq p$ and $k > p$ or $p < i, j \leq p+q$ and $k \leq p$, then $g_r\left(\left[E_i, E_j\right], E_k\right) = 0$, and again the mixed curvatures vanish. Finally, if $i \leq p < j$, we have from our product sphere calculations that $\nabla^r_{E_i} E_j = 0$, so in this case as well the mixed curvatures vanish.

Using our curvature calculations from the rotationally symmetric case and the product sphere case we then obtain

$$\begin{aligned} \mathfrak{R}\left(F_0 \wedge F_j\right) &= \frac{-\ddot{\varphi}}{\varphi} F_0 \wedge F_j \quad \text{if} \quad 0 < j \leq p, \\ \mathfrak{R}\left(F_0 \wedge F_j\right) &= \frac{-\ddot{\psi}}{\psi} F_0 \wedge F_j \quad \text{if} \quad j > p, \\ \mathfrak{R}\left(F_i \wedge F_j\right) &= \frac{1 - \dot{\varphi}^2}{\varphi^2} F_i \wedge F_j \quad \text{if} \quad 0 < i < j \leq p, \\ \mathfrak{R}\left(F_i \wedge F_j\right) &= \frac{1 - \dot{\psi}^2}{\psi^2} F_i \wedge F_j \quad \text{if} \quad p < i < j, \\ \mathfrak{R}\left(F_i \wedge F_j\right) &= \frac{-\dot{\varphi}\dot{\psi}}{\varphi\psi} F_i \wedge F_j \quad \text{if} \quad i \leq p < j. \end{aligned}$$

From this we can see that all sectional curvatures are convex linear combinations of $-\ddot{\varphi}/\varphi$, $-\ddot{\psi}/\psi$, $1 - \dot{\varphi}^2/\varphi^2$, $1 - \dot{\psi}^2/\psi^2$, and $-\dot{\varphi}\dot{\psi}/\varphi\psi$; and that

$$\begin{aligned} \operatorname{Ric}(F_0) &= \left(-p\frac{\ddot{\varphi}}{\varphi} - q\frac{\ddot{\psi}}{\psi}\right) F_0, \\ \operatorname{Ric}(F_i) &= \left(\frac{-\ddot{\varphi}}{\varphi} + (p-1)\frac{1 - \dot{\varphi}^2}{\varphi^2} - q \cdot \frac{\dot{\varphi}\dot{\psi}}{\varphi\psi}\right) F_i, \quad 0 < i \leq p, \\ \operatorname{Ric}(F_j) &= \left(\frac{-\ddot{\psi}}{\psi} + (q-1)\frac{1 - \dot{\psi}^2}{\psi^2} - p \cdot \frac{\dot{\varphi}\dot{\psi}}{\varphi\psi}\right) F_j, \quad j > p. \end{aligned}$$

3.2.5 The Schwarzschild Metric

We wish to find a Ricci flat metric on $\mathbb{R}^2 \times S^2$, so let $p = 1$ and $q = 2$ in the above doubly warped product case. This means we have to solve the following three equations simultaneously:

$$\frac{-\ddot{\varphi}}{\varphi} - 2\frac{\ddot{\psi}}{\psi} = 0,$$
$$\frac{-\ddot{\varphi}}{\varphi} - 2\frac{\dot{\varphi}\dot{\psi}}{\varphi\psi} = 0,$$
$$\frac{-\ddot{\psi}}{\psi} + \frac{1-\dot{\psi}^2}{\psi^2} - \frac{\dot{\varphi}\dot{\psi}}{\varphi\psi} = 0.$$

Subtracting the first two equations gives $\ddot{\psi}/\dot{\psi} = \dot{\varphi}\dot{\psi}/\varphi\psi$. This is equivalent to $(\dot{\psi}/\varphi) = \alpha$, for some constant α. Thus, $\dot{\psi} = \alpha\varphi$ and $\ddot{\psi} = \alpha\dot{\varphi}$. Inserting this into the three equations we get

$$-\frac{\ddot{\varphi}}{\varphi} - 2\frac{\alpha\dot{\varphi}}{\psi} = 0,$$
$$-\frac{\ddot{\varphi}}{\varphi} - 2\frac{\alpha\dot{\varphi}}{\psi} = 0,$$
$$-\frac{\alpha\dot{\varphi}}{\psi} + \frac{1-\alpha^2\varphi^2}{\psi^2} - \frac{\alpha\dot{\varphi}}{\psi} = 0,$$
$$\dot{\psi} = \alpha\varphi,$$

which reduces to

$$-\frac{\ddot{\varphi}}{\varphi} - 2\frac{\alpha\dot{\varphi}}{\psi} = 0,$$
$$\frac{1-\alpha^2\varphi^2}{\psi^2} - 2\frac{\alpha\dot{\varphi}}{\psi} = 0,$$
$$\dot{\psi} = \alpha\varphi,$$

which implies

$$\frac{1-\alpha^2\varphi^2}{2\alpha\dot{\varphi}} = \psi,$$
$$-\frac{\ddot{\varphi}}{\varphi} - \frac{4\alpha^2\dot{\varphi}^2}{1-\alpha^2\varphi^2} = 0,$$
$$\dot{\psi} = \alpha\varphi,$$

which implies

$$2\psi\ddot{\psi} - (1-\dot{\psi}^2) = 0,$$
$$-\frac{\ddot{\varphi}}{\varphi} - \frac{4\alpha^2\dot{\varphi}^2}{1-\alpha^2\varphi^2} = 0,$$
$$\dot{\psi} = \alpha\varphi.$$

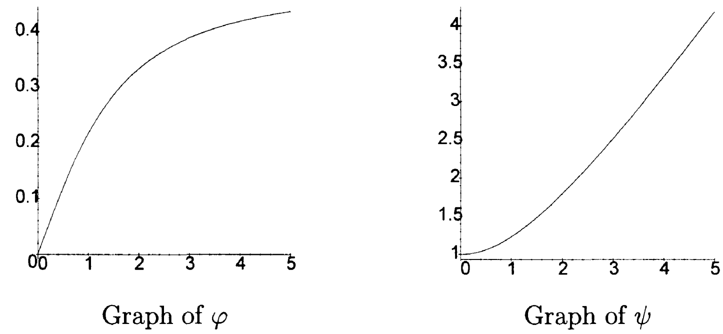


FIGURE 3.2.

Now, $\psi = r$ solves the first equation. This means that $\varphi = 1/\alpha$, which also "solves" the second equation. The metric, however, lives on $S^1 \times \mathbb{R}^3$ rather than $\mathbb{R}^2 \times S^2$, and it is the standard flat metric on this space. To get more complicated solutions, assume $\dot{\psi}^2 = G(\psi)$, $2\ddot{\psi} = G'$. Then the first equation becomes

$$\psi G' + G = 1,$$

so

$$G = 1 + C\psi^{-1}, \quad C \in \mathbb{R}.$$

Translating back we get

$$\begin{aligned} \dot{\psi}^2 &= 1 + C\psi^{-1}, \\ 2\ddot{\psi} &= -C\psi^{-2}, \\ \dot{\psi} &= \alpha\varphi. \end{aligned}$$

Note that the equation $-\ddot{\varphi}/\varphi - (4\alpha^2\dot{\varphi}^2)/(1-\alpha^2\varphi^2) = 0$ is redundant. Also, since we want a metric on $\mathbb{R}^2 \times S^2$, we may assume that $\varphi(0) = 0$, $\dot{\varphi}(0) = 1$, and $\psi(0) = \beta > 0$. This actually gives all the requirements for a smooth metric, since ψ is automatically even if it solves the above equation, and hence φ is odd (see Chapter 1). The constants α, β, and C are related through $0 = \dot{\psi}^2(0) = 1 + C \cdot \beta^{-1}$, so $C = -\beta$ and $2\alpha = 2\alpha\dot{\varphi}(0) = 2\ddot{\psi}(0) = -C\beta^{-2} = \beta^{-1}$. For given $\beta > 0$, let the solutions be denoted by φ_β and ψ_β. Since $\psi_\beta(0) = \beta > 0$ and $\ddot{\psi}_\beta = \beta/2\psi_\beta^{-2}$, we have that ψ_β is convex as long as it is positive. We can then prove as in the scalar flat case that ψ is defined for all r and that $\psi(r) \sim |r|$ as $r \to \pm\infty$.

Thus, the metric looks like $S^1 \times \mathbb{R}^3$ at infinity, where the metric on S^1 is multiplied by $(2 \cdot \beta)^2$. Thus, the Schwarzschild metric is a Ricci flat metric on $\mathbb{R}^2 \times S^2$ that at infinity looks approximately like the flat metric on $S^1 \times \mathbb{R}^3$. Both warping functions are sketched in Figure 3.2 in the case where $\alpha = 1$.

3.3 Hyperbolic Space

We have a pretty good picture of spheres and Euclidean space as models for constant curvature spaces. We even know what the symmetry groups are. We don't have a similarly good picture for spaces of constant negative curvature. This is partly because these metrics are not hypersurface metrics in Euclidean space. There are, however, a couple of good pictorial models. To explain them we also need to expand our general knowledge a little.

3.3.1 *The Rotationally Symmetric Model*

We define H^n to be the rotationally symmetric metric $dr^2 + \sinh^2(r)ds_{n-1}^2$ on $\mathbb{R}^n$ of constant curvature -1. As with all rotationally symmetric metrics, we see that $O(n)$ acts by isometries in a natural way. But it is not clear that H^n is homogeneous from this description because the origin seems to be singled out as being fixed by the $O(n)$ action.

3.3.2 *The Upper Half Plane Model*

Let $M = \{(x^1, \ldots, x^n) \in \mathbb{R}^n : x^n > 0\}$ and let $ds^2 = g = (1/x^n)^2 \left((dx^1)^2 + \cdots + (dx^n)^2\right)$. Thus $1/x^n dx^1, \ldots, 1/x^n dx^n$ is an orthonormal coframing on M. This can be used to check that the curvature is $\equiv -1$. Another way is to notice that $g = dr^2 + (e^{-r})^2 \left((dx^1)^2 + \cdots + (dx^{n-1})^2\right)$, where $r = \log(x^n)$, and then to use the fundamental equations; but then the metric is on $\mathbb{R}^n = \mathbb{R} \times \mathbb{R}^{n-1}$. In this case Iso $\left(\mathbb{R}^{n-1}\right) = \mathbb{R}^{n-1} \rtimes O(n-1)$ (this is a semi direct product) acts by isometries on M, so there is no fixed point for the action, and it acts transitively on the hypersurfaces $r =$ constant.

3.3.3 *The Riemann Model*

If (M, g) is a Riemannian manifold and φ is positive on M, then we can get a new Riemannian manifold $(M, h = \varphi^2 g)$. Such a change in metric is called a *conformal change,* and φ^2 is referred to as the *conformal factor* "$\varphi^2 = h/g$." The upper half plane model is a conformal change of the Euclidean metric on H. Here we wish to find $\varphi^2(x)$ on subsets of $\mathbb{R}^n$ such that $\varphi^2 \cdot ((dx^1)^2 + \cdots + (dx^n)^2)$ has constant curvature. Clearly, $\varphi \cdot dx^1, \ldots, \varphi \cdot dx^n$ is an orthonormal coframing, and $\frac{1}{\varphi}\partial_1, \ldots, \frac{1}{\varphi}\partial_n$ is an orthonormal framing. We can use the Koszul formula to compute $\nabla_{\partial_i}\partial_j$ and hence the curvature tensor. This tedious task is done in [76, vols. II and IV]. Using $\varphi = \left(1 + k/4r^2\right)^{-1}$ gives a metric of constant curvature k on $\mathbb{R}^n$ if $k \geq 0$ and on $B(0, -4 \cdot k^{-1})$ if $k < 0$.

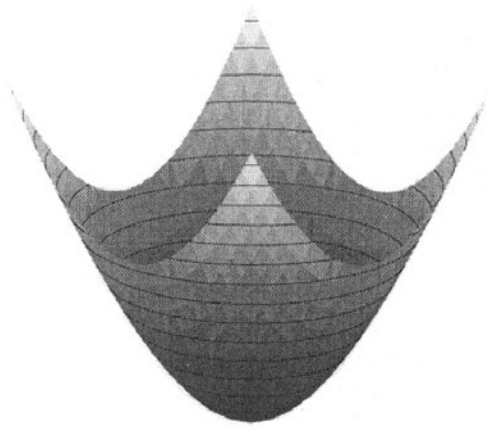


FIGURE 3.3.

3.3.4 The Imaginary Unit Sphere Model

Our last model exhibits H^n as a hypersurface in Minkowski space by analogy with $S^n(1) \subset \mathbb{R}^{n+1}$. A discussion of this model can also be found in Chapter 8. Minkowski space is the physicists' model for space-time. Topologically, the space is $\mathbb{R}^{n+1}$, but we have a different sort of metric on it. If $(x^0, x^1, \ldots, x^n)$ are Cartesian coordinates on $\mathbb{R}^{n+1}$, then we have the indefinite metric:

$$g = -(dx^0)^2 + (dx^1)^2 + \cdots + (dx^n)^2.$$

In other words, the framing $\partial_0, \partial_1, \ldots, \partial_n$ consists of orthogonal vectors where $|\partial_0|^2 = -1$ and $|\partial_i|^2 = 1, i = 1, \ldots, n$. The zeroth coordinate is singled out as having imaginary norm this is the physicists' time variable. One can more generally define Minkowski inner product spaces as vector spaces with an inner product of this type, and then develop a theory for Minkowski manifolds. This theory is in many ways analogous to what we have done for Riemannian manifolds. These types of spaces are exactly the ones used in general relativity. Here we'll confine ourselves to studying just the Minkowski space $\mathbb{R}^{1,n} = (\mathbb{R}^{n+1}, g)$ as defined above. The "spheres" in this space of radius $i \cdot r$ must satisfy the equation

$$-(x^0)^2 + (x^1)^2 + \cdots + (x^n)^2 = -r^2.$$

A picture of the hyperbolic plane in Minkowski 3-space is given in Figure 3.3.

Thus, we should study the "distance" function

$$r(x) = \left|-(x^0)^2 + (x^1)^2 + \cdots + (x^n)^2\right|^{1/2}$$

on $U = \left\{x \in \mathbb{R}^{n+1} : x^0 > 0 \text{ and } -(x^0)^2 + (x^1)^2 + \cdots + (x^n)^2 < 0\right\}$. The level sets $H(r) \subset U$ are diffeomorphic to $\mathbb{R}^n$ and look like hyperbolae of revolution.

Furthermore, if we restrict the Minkowski metric g to these level sets, they induce Riemannian metrics on $H(r)$. This is because

$$dr = \frac{1}{r}(x^0 dx^0 - x^1 dx^1 - \cdots - x^n dx^n),$$

so any tangent vector $v \in TH(r)$ satisfies

$$0 = (-x^0 v^0 + x^1 v^1 + \cdots + x^n v^n) = g(x, v);$$

and therefore, for any such v,

$$\begin{aligned} g(v, v) &= -(v^0)^2 + (v^1)^2 + \cdots + (v^n)^2 \\ &= -\frac{(x^1 v^1 + \cdots + x^n v^n)^2}{(x^0)^2} + (v^1)^2 + \cdots + (v^n)^2 \\ &\geq -\frac{\left((x^1)^2 + \cdots + (x^n)^2\right)\left((v^1)^2 + \cdots + (v^n)^2\right)}{(x^0)^2} + (v^1)^2 + \cdots + (v^n)^2 \\ &= \left(-1 + \frac{r^2}{(x^0)^2}\right)\left((v^1)^2 + \cdots + (v^n)^2\right) + (v^1)^2 + \cdots + (v^n)^2 \\ &= \frac{r^2}{(x^0)^2}\left((v^1)^2 + \cdots + (v^n)^2\right) \geq 0 \quad \text{and} = 0 \text{ only if} \quad v = 0. \end{aligned}$$

We have therefore shown that g is positive definite on $H(r)$. Our claim is that $H(r)$ with the induced metric has constant curvature $-r^{-2}$. There are several ways to check this. One way is to observe that $(t, x) \to r(\cosh(t), \sinh(t) \cdot x)$ defines a Riemannian isometry from $dt^2 + r^2 \sinh^2\left(\frac{t}{r}\right) ds_{n-1}^2$ to $H(r)$, where $x \in S^{n-1} \subset \mathbb{R}^n$ is viewed as a vector in $\mathbb{R}^n$. This also shows that at least two of our models are equal. Finally, we can also compute gradients, etc., and use the tangential curvature equation as we did for the sphere. This works out as follows. The Minkowski gradient $\nabla r = \alpha^i \partial_i$ must satisfy

$$\begin{aligned} g(\nabla r, v) &= dr(v), \\ -\alpha^0 v^0 + \alpha^1 v^1 + \cdots + \alpha^n v^n &= \frac{1}{r}(x^0 v^0 - x^1 v^1 - \cdots - x^n v^n), \end{aligned}$$

or equivalently, $\nabla r = g^{ij} \partial_i(r) \partial_j$, so $\nabla r = \frac{-1}{r} x^i \partial_i$. This is clearly not the same as the Euclidean gradient, but aside from the minus sign it corresponds exactly to the gradient for the distance function in $\mathbb{R}^{n+1}$ that has $S^n(1)$ as level sets. Also, $g(\nabla r, \nabla r) = 1/r^2\left(-(x^0)^2 + (x^1)^2 + \cdots + (x^n)^2\right) = -1$, so we are working with an (imaginary) distance function, which aside from the sign should satisfy all of the fundamental equations we have already established. The Minkowski connection on $\mathbb{R}^{1,n}$ of course satisfies all of the same properties as the Riemannian connection and can in particular be found using the Koszul formula. But since $g(\partial_i, \partial_j)$ is always constant and $[\partial_i, \partial_j] = 0$, we see that $\nabla_{\partial_i} \partial_j = 0$. Hence, we get just the standard Euclidean connection and therefore the curvature tensor $R = 0$ as well.

With all this information one can easily compute $\nabla^2 r = S$ and check using the tangential curvature equation that $H(r)$ indeed has constant curvature $-r^{-2}$.

Finally, we should compute $\mathrm{Iso}(H(r))$. On $\mathbb{R}^{1,n}$ the linear isometries that preserve the Minkowski metric are denoted by

$$O(1,n) = \left\{\varphi : \mathbb{R}^{1,n} \to \mathbb{R}^{1,n} : g(\varphi v, \varphi v) = g(v,v)\right\}.$$

One can, as in the case of the sphere, see that these are exactly the isometries on $H(r)$. The isotropy group that preserves $(r, 0, \ldots, 0)$ can be identified with $O(n)$ (isometries we get from the metric being rotationally symmetric). One can easily check that $O(1,n)$ acts transitively on $H(r)$.

With all this we now have a fairly complete picture of all the space forms S^n_k, i.e., our models for constant curvature. We shall later prove that in a suitable sense these are the only simply connected Riemannian manifolds of constant curvature $k \in \mathbb{R}$.

3.4 More Left-Invariant Metrics

We will give two examples of left-invariant metrics. The first represents H^2, and the other is the Berger sphere. In the next section the Berger spheres will be used to make computations on $\mathbb{C}P^2$.

3.4.1 *Hyperbolic Space as a Lie Group*

Let G be the 2-dimensional Lie group

$$G = \left\{ \begin{pmatrix} \alpha & \beta \\ 0 & 1 \end{pmatrix} : \alpha > 0, \beta \in \mathbb{R} \right\}.$$

Notice that the first row can be identified with the upper half plane. The Lie algebra of G is

$$\mathfrak{g} = \left\{ \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} : a, b \in \mathbb{R} \right\}.$$

If we define $X = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $Y = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, then $[X,Y] = XY - YX = Y$. Now declare $\{X, Y\}$ to be an orthonormal frame on G. Then use the Koszul formula to compute

$$\nabla_X X = 0, \nabla_Y Y = X, \nabla_X Y = 0, \nabla_Y X = \nabla_X Y - [X,Y] = -Y.$$

Hence,

$$R(X,Y)Y = \nabla_X \nabla_Y Y - \nabla_Y \nabla_X Y - \nabla_{[X,Y]} Y = \nabla_X X - 0 - \nabla_Y Y = -X,$$

which implies that G has constant curvature -1.

This example can be generalized to higher dimensions. Thus, the upper half plane is in a natural way a Lie group that has a left-invariant metric of constant curvature -1. This is in sharp contrast to the sphere, where only $S^3 = SU(2)$ and $S^1 = SO(2)$ are Lie groups.

3.4.2 Berger Spheres

On $SU(2)$ we have the left-invariant metric where $\{\varepsilon^{-1}X_1, X_2, X_3\}$ is an orthonormal frame and $[X_i, X_{i+1}] = 2X_{i+2}$ (indices are mod 3), as discussed in Chapter 1. Using the Koszul formula we can compute

$$\nabla_{X_1}X_2 = (2-\varepsilon^2)X_3,\ \nabla_{X_2}X_3 = X_1,\ \nabla_{X_3}X_1 = \varepsilon^2 X_2,\ \nabla_{X_i}X_i = 0,$$

and

$$\nabla_{X_2}X_1 = \nabla_{X_1}X_2 + [X_2, X_1] = -\varepsilon^2 X_3,\ \nabla_{X_3}X_2 = -X_1,\ \nabla_{X_1}X_3 = (\varepsilon^2 - 2)X_2.$$

Thus,

$$R(X_1, X_2)X_3 = \nabla_{X_1}\nabla_{X_2}X_3 - \nabla_{X_2}\nabla_{X_1}X_3 - \nabla_{[X_1,X_2]}X_3 = 0 - 0 - 0,$$

and in a similar way, all curvatures between three distinct vectors are zero. Finally,

$$\begin{aligned} R(X_1, X_2)X_2 &= \nabla_{X_1}\nabla_{X_2}X_2 - \nabla_{X_2}\nabla_{X_1}X_2 - \nabla_{[X_1,X_2]}X_2 \\ &= 0 - \nabla_{X_2}\nabla_{X_1}X_2 - 2\nabla_{X_3}X_2 = \varepsilon^2 X_1, \end{aligned}$$

so we get

$$\mathfrak{R}\left(\left(\varepsilon^{-1}X_1\right)\wedge X_2\right) = \varepsilon^2\left(\varepsilon^{-1}X_1\right)\wedge X_2.$$

In a similar fashion one computes

$$\begin{aligned} \mathfrak{R}(X_2\wedge X_3) &= \left(4-3\varepsilon^2\right)X_2\wedge X_3, \\ \mathfrak{R}\left(X_3\wedge\left(\varepsilon^{-1}X_1\right)\right) &= \varepsilon^2 X_3\wedge\left(\varepsilon^{-1}X_1\right). \end{aligned}$$

Thus, all sectional curvatures must lie in the interval $\left[\varepsilon^2, 4-3\varepsilon^2\right]$. Notice that as $\varepsilon \to 0$ the sectional curvature perpendicular to the Hopf fiber $\to 4$. Thus, as we collapse the Hopf fiber, the curvatures will converge to what they are on the base space $S^2\left(\frac{1}{2}\right)$.

3.5 Complex Projective Space

Recall that $\mathbb{C}P^n = \left(\mathbb{C}^{n+1} - \{0\}\right)/\,\mathbb{C} = S^{2n+1}/\,S^1$, where S^1 and $\mathbb{C}$ act by complex scalar multiplication. If we write the metric $ds^2_{2n+1} = dr^2 + \sin^2(r)ds^2_{2n-1} + \cos^2(r)d\theta^2$, then we can think of the S^1 action on S^{2n+1} as acting separately on

S^{2n-1} and S^1. Then $\mathbb{C}P^n = \left[0, \frac{\pi}{2}\right] \times \left(\left(S^{2n-1} \times S^1\right)/ S^1\right)$, and the metric can be written as

$$dr^2 + \sin^2(r)\left(g + \cos^2(r)h\right).$$

We will restrict our attention to the case where $n = 2$. For a more general discussion see also Chapter 8. In this case the metric can be written as

$$dr^2 + \sin^2(r)\left(\cos^2(r)(\sigma^1)^2 + (\sigma^2)^2 + (\sigma^3)^2\right).$$

More generally, we can consider metrics on $I \times S^3$ of the form

$$dr^2 + \varphi^2(r)\left(\psi^2(r)(\sigma^1)^2 + (\sigma^2)^2 + (\sigma^3)^2\right).$$

If we define f = projection onto the I-factor, then f is a distance function, and we can use the fundamental equations to compute curvatures. We have a natural orthogonal frame $\{X_1, X_2, X_3, \partial_r\}$ that yields an orthonormal frame $\{F_1, F_2, F_3, F_4\} = \left\{\frac{1}{\varphi \cdot \psi} X_1, \frac{1}{\varphi} X_2, \frac{1}{\varphi} X_3, \partial_r\right\}$. From section 3.1 we have

$$\begin{aligned} S(F_1) &= \frac{\dot{\mu}}{\mu} F_1 = \lambda_1 F_1, \quad \mu = \varphi \cdot \psi, \\ S(F_i) &= \frac{\dot{\varphi}}{\varphi} F_i = \lambda_i F_i, \quad i = 2, 3. \end{aligned}$$

The $(1, 3)$-curvature tensor R^r on S^3 with the metric

$$g_r = \mu^2(\sigma^1)^2 + \varphi^2(\sigma^2)^2 + \varphi^2(\sigma^3)^2 = \varphi^2(r)(\psi^2(r)(\sigma^1)^2 + (\sigma^2)^2 + (\sigma^3)^2)$$

has already been computed in the case where $\varphi = 1$ and $\psi = \varepsilon$ (the Berger spheres), so all we need to observe is that multiplying a Riemannian metric by a number cannot change the connection or curvature tensor $R(X, Y)Z$. Thus,

$$\begin{aligned} \nabla^r_{X_i} X_j &= \nabla_{X_i} X_j, \\ R^r(F_1, F_i) F_i &= \frac{\psi^2}{\varphi^2} F_1, \quad i = 2, 3, \\ R^r(F_2, F_3) F_3 &= \frac{4 - 3\psi^2}{\varphi^2} F_3, \\ R^r\left(F_i, F_j\right) F_k &= 0 \quad \text{if the indices are distinct.} \end{aligned}$$

We now have to compute the terms

$$\begin{aligned} &(\lambda_i - \lambda_l)\, g_r\left(\nabla^r_{F_k} F_l, F_i\right) + (\lambda_k - \lambda_i)\, g_r\left(\nabla^r_{F_l} F_k, F_i\right), \\ &\left(\lambda_l - \lambda_j\right) g_r\left(\nabla^r_{F_i} F_j, F_l\right) + (\lambda_i - \lambda_l)\, g_r\left(\nabla^r_{F_j} F_i, F_l\right), \end{aligned}$$

which for the first time will not all be zero. It is easier to compute this when the F's are replaced by X's and then use tensoriality to find the above terms. Having done this, the next important reduction is that

$$g_r\left(\nabla^r_{X_i} X_j, X_k\right) = 0$$

unless the indices are distinct. Using that $\lambda_2 = \lambda_3$ we can finally compute

$$\begin{aligned}
(\lambda_3 - \lambda_2)\, g_r\left(\nabla^r_{X_1} X_2, X_3\right) + (\lambda_1 - \lambda_3)\, g_r\left(\nabla^r_{X_2} X_1, X_3\right) &\\
= (\lambda_1 - \lambda_3)\, g_r\left(\nabla^r_{X_2} X_1, X_3\right) &\\
= -(\lambda_1 - \lambda_3)\, \psi^2 \varphi^2, &\\
(\lambda_2 - \lambda_3)\, g_r\left(\nabla^r_{X_1} X_3, X_2\right) + (\lambda_1 - \lambda_2)\, g_r\left(\nabla^r_{X_3} X_1, X_2\right) &\\
= (\lambda_1 - \lambda_2)\, g_r\left(\nabla^r_{X_3} X_1, X_2\right) &\\
= -(\lambda_1 - \lambda_2)\, \psi^2 \varphi^2, &\\
(\lambda_1 - \lambda_3)\, g_r\left(\nabla^r_{X_2} X_3, X_1\right) + (\lambda_2 - \lambda_1)\, g_r\left(\nabla^r_{X_3} X_2, X_1\right) &\\
= (\lambda_1 - \lambda_2)\, g_r\left([X_2, X_3], X_1\right) &\\
= 2(\lambda_1 - \lambda_2)\, \psi^2 \varphi^2. &
\end{aligned}$$

After normalizing we then get

$$\begin{aligned}
(\lambda_3 - \lambda_2)\, g_r\left(\nabla^r_{F_1} F_2, F_3\right) + (\lambda_1 - \lambda_3)\, g_r\left(\nabla^r_{F_2} F_1, F_3\right) &= -(\lambda_1 - \lambda_3)\frac{\psi}{\varphi},\\
(\lambda_2 - \lambda_3)\, g_r\left(\nabla^r_{F_1} F_3, F_2\right) + (\lambda_1 - \lambda_2)\, g_r\left(\nabla^r_{F_3} F_1, F_2\right) &= -(\lambda_1 - \lambda_3)\frac{\psi}{\varphi},\\
(\lambda_1 - \lambda_3)\, g_r\left(\nabla^r_{F_2} F_3, F_1\right) + (\lambda_2 - \lambda_1)\, g_r\left(\nabla^r_{F_3} F_2, F_1\right) &= 2(\lambda_1 - \lambda_3)\frac{\psi}{\varphi}.
\end{aligned}$$

We are now ready to compute all the curvatures:

$$\begin{aligned}
\mathfrak{R}(F_1 \wedge F_4) &= -\frac{\ddot{\mu}}{\mu} F_1 \wedge F_4 - 2\left(\frac{\dot{\mu}}{\mu} - \frac{\dot{\varphi}}{\varphi}\right)\frac{\psi}{\varphi} F_2 \wedge F_3,\\
\mathfrak{R}(F_2 \wedge F_4) &= -\frac{\ddot{\varphi}}{\varphi} F_2 \wedge F_4 + \left(\frac{\dot{\mu}}{\mu} - \frac{\dot{\varphi}}{\varphi}\right)\frac{\psi}{\varphi} F_1 \wedge F_3,\\
\mathfrak{R}(F_3 \wedge F_4) &= -\frac{\ddot{\varphi}}{\varphi} F_3 \wedge F_4 + \left(\frac{\dot{\mu}}{\mu} - \frac{\dot{\varphi}}{\varphi}\right)\frac{\psi}{\varphi} F_1 \wedge F_2,\\
\mathfrak{R}(F_1 \wedge F_2) &= \left(\frac{\psi^2}{\varphi^2} - \frac{\dot{\mu}}{\mu}\frac{\dot{\varphi}}{\varphi}\right) F_1 \wedge F_2 = \left(\frac{\dot{\mu}}{\mu} - \frac{\dot{\varphi}}{\varphi}\right)\frac{\psi}{\varphi} F_3 \wedge F_4,\\
\mathfrak{R}(F_1 \wedge F_3) &= \left(\frac{\psi^2}{\varphi^2} - \frac{\dot{\mu}}{\mu}\frac{\dot{\varphi}}{\varphi}\right) F_1 \wedge F_3 + \left(\frac{\dot{\mu}}{\mu} - \frac{\dot{\varphi}}{\varphi}\right)\frac{\psi}{\varphi} F_2 \wedge F_4,\\
\mathfrak{R}(F_2 \wedge F_3) &= \left(\frac{4 - 3\psi^2}{\varphi^2} - \frac{\dot{\varphi}}{\varphi}\frac{\dot{\varphi}}{\varphi}\right) F_2 \wedge F_3 - 2\left(\frac{\dot{\mu}}{\mu} - \frac{\dot{\varphi}}{\varphi}\right)\frac{\psi}{\varphi} F_1 \wedge F_4,\\
\operatorname{Ric}(F_4) &= \left(-\frac{\ddot{\mu}}{\mu} - 2\frac{\ddot{\varphi}}{\varphi}\right) F_4,\\
\operatorname{Ric}(F_1) &= \left(-\frac{\ddot{\mu}}{\mu} + 2\left(\frac{\psi^2}{\varphi^2} - \frac{\dot{\mu}\dot{\varphi}}{\mu\varphi}\right)\right) F_1,\\
\operatorname{Ric}(F_i) &= \left(-\frac{\ddot{\varphi}}{\varphi} + \left(\frac{\psi^2}{\varphi^2} - \frac{\dot{\mu}\dot{\varphi}}{\mu\varphi}\right) + \frac{4 - 3\psi^2}{\varphi^2} - \frac{\dot{\varphi}^2}{\varphi^2}\right) F_i, \quad i = 2, 3.
\end{aligned}$$

In particular, if $\varphi = \sin(r)$, $\psi = \cos(r)$, $\mu = \sin(r) \cdot \cos(r) = \frac{1}{2}\sin(2r)$, then we see that $\text{Ric}(F_i) = 6 \cdot F_i$, $i = 1, 2, 3, 4$. Thus, $\mathbb{C}P^2$ is an Einstein metric with Einstein constant $6 = 2 \cdot n + 2$.

For the curvatures we have

$$\begin{aligned}
\mathfrak{R}(F_1 \wedge F_4) &= 4F_1 \wedge F_4 + 2F_2 \wedge F_3, \\
\mathfrak{R}(F_2 \wedge F_4) &= F_2 \wedge F_4 - F_1 \wedge F_3, \\
\mathfrak{R}(F_3 \wedge F_4) &= F_3 \wedge F_4 - F_1 \wedge F_2, \\
\mathfrak{R}(F_1 \wedge F_2) &= F_1 \wedge F_2 - F_3 \wedge F_4, \\
\mathfrak{R}(F_1 \wedge F_3) &= F_1 \wedge F_3 - F_2 \wedge F_4, \\
\mathfrak{R}(F_2 \wedge F_3) &= 4F_2 \wedge F_3 + 2F_1 \wedge F_4.
\end{aligned}$$

Some tedious algebraic manipulations will convince you that all sectional curvatures lie in the interval $[1, 4]$. On the other hand, if we use the basis

$$\frac{1}{\sqrt{2}}(F_1 \wedge F_2 \pm F_3 \wedge F_4), \frac{1}{\sqrt{2}}(F_1 \wedge F_3 \pm F_4 \wedge F_2), \frac{1}{\sqrt{2}}(F_1 \wedge F_4 \pm F_2 \wedge F_3)$$

for $\Lambda^2 M$, we will diagonalize the curvature operator, and the eigenvalues will lie in $[0, 6]$. Thus, while all sectional curvatures are positive, there are zero eigenvalues for the curvature operator. Observe that since the eigenvalues of the curvature operator are constant, it must follow that the curvature tensor is parallel. Nevertheless, the metric does not have constant curvature.

Spaces with parallel curvature tensor are called *locally symmetric*. One can show that such spaces have many isometries. They must therefore still be fairly nice.

3.6 Further Study

The book by O'Neill [65] gives an excellent account of Minkowski geometry and also studies in detail the real Schwarzschild metric, which was the first nontrivial solution to the vacuum Einstein field equations. There is also a good introduction to locally symmetric spaces and their properties. This book is probably the most comprehensive elementary text and is good for a first encounter with most of the concepts in differential geometry.

Another book, which contains more (actually almost all) advanced examples, is [11]. This is a tremendously good reference on Riemannian geometry in general.

3.7 Exercises

1. Show that the Schwarzschild metric doesn't have parallel curvature tensor.

2. Show that the Berger spheres ($\varepsilon \neq 1$) do not have parallel curvature tensor.

3. Show that $\mathbb{C}P^2$ has parallel curvature tensor.

4. The Heisenberg group with its Lie algebra is

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

$$\mathfrak{g} = \left\{ \begin{pmatrix} 0 & x & z \\ 0 & 0 & y \\ 0 & 0 & 0 \end{pmatrix} : a, b, c \in \mathbb{R} \right\}.$$

A basis for the Lie algebra is:

$$X = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

(a) Show that the only nonzero brackets are

$$[X, Y] = -[Y, X] = Z.$$

Now introduce a left-invariant metric on G such that X, Y, Z form an orthonormal frame.

(b) Show that the Ricci tensor has both negative and positive eigenvalues.

(c) Show that the scalar curvature is constant.

(d) Show that the Ricci tensor is not parallel.

5. (a) Show that there is a family of Ricci flat metrics on TS^2 of the form $dr^2 + \varphi^2(r)\left(\psi^2(r)(\sigma^1)^2 + (\sigma^2)^2 + (\sigma^3)^2\right)$,

$$\begin{aligned} \dot{\varphi} &= \psi, \\ \dot{\varphi}^2 &= 1 - k\varphi^{-4}, \\ \varphi(0) &= k^{\frac{1}{4}}, \dot{\varphi}(0) = 0, \\ \psi(0) &= 0, \dot{\psi}(0) = 2. \end{aligned}$$

(b) Show that $\varphi(r) \sim r$, $\dot{\varphi}(r) \sim 1$, $\ddot{\varphi}(r) \sim -4\varphi^{-5}$ as $r \to \infty$. Conclude that all curvatures are of order r^{-6} as $r \to \infty$ and that the metric looks like $(0, \infty) \times \mathbb{R}P^3 = (0, \infty) \times SO(3)$ at infinity. Moreover, show that scaling one of these metrics corresponds to changing k. Thus, we really have only one Ricci flat metric; it is called the *Eguchi-Hanson metric*.

6. For the general metric $dr^2 + \varphi^2(r)\left(\psi^2(r)(\sigma^1)^2 + (\sigma^2)^2 + (\sigma^3)^2\right)$ show that the $(1, 1)$-tensor, which in the orthonormal frame looks like

$$\begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

yields an Hermitian structure. Show that this structure is Kähler iff $\dot{\varphi} = \psi$. Find the scalar curvature for such metrics. Show that there are scalar flat metrics on all the 2-dimensional vector bundles over S^2. The one on TS^2 is the Eguchi-Hanson metric, and the one on $S^2 \times \mathbb{R}^2$ is the Schwarzschild metric.

7. Show that $\tau\left(\mathbb{R}P^{n-1}\right)$ admits rotationally symmetric metrics

$$dr^2 + \varphi^2(r)\, ds_{n-1}^2$$

such that $\varphi(r) = r$ for $r > 1$ and the Ricci or scalar curvatures are nonpositive. Thus, the Euclidean metric can be topologically perturbed to have nonpositive Ricci curvature. It is not possible to perturb the Euclidean metric in this way to have nonnegative scalar curvature or nonpositive sectional curvature. Try to convince yourself of that by looking at rotationally symmetric metrics on $\mathbb{R}^n$ and $\tau\left(\mathbb{R}P^{n-1}\right)$.

8. A Riemannian manifold (M, g) is said to be locally conformally flat if for every $p \in M$ there is a neighborhood U around and coordinates on U such that $g = \varphi^2\left(\left(dx^1\right)^2 + \cdots + (dx^n)^2\right)$. Show that the space forms are locally conformally flat. With some help from the literature, show that any 2-dimensional Riemannian manifold is locally conformally flat. In fact, any metric on a closed surface is conformal to a metric of constant curvature. This is called the *uniformization theorem*.

9. We say that (M, g) admits orthogonal coordinates around $p \in M$ if we have coordinates on some neighborhood of p, where

$$g_{ij} = 0 \quad \text{for} \quad i \neq j,$$

i.e., the coordinate vector fields are perpendicular. Show that such coordinates always exist in dimension 2, while they may not exist in dimension > 3. For the latter counterexample, you may want to show that in such coordinates the curvatures $R_{ijk}^{l} = 0$ if all indices are distinct. What about 3 dimensions?

10. There is a strange curvature quantity we have not yet mentioned. Its definition is somewhat cumbersome and unintuitive. First, for two symmetric (0, 2)-tensors h, k define the *Kulkarni-Nomizu product* as the (0, 4)-tensor

$$\begin{aligned} h \circ k\,(v_1, v_2, v_3, v_4) = {} & h(v_1, v_3) \cdot k(v_2, v_4) + h(v_2, v_4) \cdot k(v_1, v_3) \\ & - h(v_1, v_4) \cdot k(v_2, v_3) - h(v_2, v_3) \cdot k(v_1, v_4). \end{aligned}$$

Note that (M, g) has constant curvature c iff the (0, 4)-curvature tensor satisfies $R = c \cdot (g \circ g)$. If we use the (0, 2) form of the Ricci tensor, then we can decompose the (0, 4)-curvature tensor as follows in dimensions $n \geq 4$:

$$R = \frac{\text{scal}}{2n(n-1)} g \circ g + \frac{1}{n-2}\left(\text{Ric} - \frac{\text{scal}}{n} \cdot g\right) \circ g + W.$$

When $n = 3$ we have instead

$$R = \frac{\text{scal}}{12} g \circ g + \left(\text{Ric} - \frac{\text{scal}}{3} \cdot g\right) \circ g.$$

The (0, 4)-tensor W defined for $n > 3$ is called the *Weyl tensor*.

(a) Show that these decompositions are orthogonal, in particular:

$$|R|^2 = \left|\frac{\text{scal}}{2n(n-1)} g \circ g\right|^2 + \left|\frac{1}{n-2}\left(\text{Ric} - \frac{\text{scal}}{n} \cdot g\right) \circ g\right|^2 + |W|^2.$$

(b) Show that if we conformally change the metric $g' = f \cdot g$, then $W' = f \cdot W$.
If (M, g) has constant curvature, then $W = 0$. If (M, g) is locally conformally equivalent to the Euclidean metric, i.e., locally we can always find coordinates where: $g = f \cdot \left((dx^1)^2 + \cdots + (dx^n)^2\right)$, then $W = 0$. The converse is also true but much harder to prove.

(c) Show that the Weyl tensors for the Schwarzschild metric and the Eguchi-Hanson metrics are not zero.

(d) Show that (M, g) has constant curvature iff $W = 0$ and $\text{Ric} = \frac{\text{scal}}{n}$.

11. In this problem we shall see that even in dimension 4 the curvature tensor has some very special properties. Throughout we let (M, g) be a 4-dimensional oriented Riemannian manifold. The bi-vectors $\Lambda^2 TM$ come with a natural endomorphism called the Hodge $*$ operator. It is defined as follows: for any oriented orthonormal basis e_1, e_2, e_3, e_4 we define $*(e_1 \wedge e_2) = e_3 \wedge e_4$.

(a) Show that this gives a well-defined linear endomorphism which satisfies: $** = I$. (Extend the definition to a linear map: $* : \Lambda^p TM \to \Lambda^q TM$, where $p + q = n$. When $n = 2$, we have: $* : TM \to TM = \Lambda^1 TM$ satisfies: $** = -I$, thus yielding an almost complex structure on any surface.)

(b) Now decompose $\Lambda^2 TM$ into $+1$ and -1 eigenspaces $\Lambda^+ TM$ and $\Lambda^- TM$ for $*$. Show that if e_1, e_2, e_3, e_4 is an oriented orthonormal basis, then

$$e_1 \wedge e_2 \pm e_3 \wedge e_4 \in \Lambda^\pm TM,$$
$$e_1 \wedge e_3 \pm e_4 \wedge e_2 \in \Lambda^\pm TM,$$
$$e_1 \wedge e_4 \pm e_2 \wedge e_3 \in \Lambda^\pm TM.$$

Thus, any linear map $L : \Lambda^2 TM \to \Lambda^2 TM$ has a block decomposition

$$L = \begin{pmatrix} A & D \\ B & C \end{pmatrix},$$
$$A : \Lambda^+ TM \to \Lambda^+ TM,$$
$$D : \Lambda^+ TM \to \Lambda^- TM,$$
$$B : \Lambda^- TM \to \Lambda^+ TM,$$
$$C : \Lambda^- TM \to \Lambda^- TM.$$

In particular, we can decompose the curvature operator $\mathfrak{R} : \Lambda^2 TM \to \Lambda^2 TM$:

$$\mathfrak{R} = \begin{pmatrix} A & D \\ B & C \end{pmatrix}.$$

(c) Since $\mathfrak{R}$ is symmetric, we get that A, C are symmetric and that $D = B^*$ is the adjoint of B. One can furthermore show that

$$A = W^+ + \frac{\text{scal}}{12} I,$$
$$C = W^- + \frac{\text{scal}}{12} I,$$

where the Weyl tensor can be written

$$W = \begin{pmatrix} W^+ & 0 \\ 0 & W^- \end{pmatrix}.$$

Find these decompositions for both of the doubly warped metrics:

$$I \times S^1 \times S^2, dr^2 + \varphi^2(r) d\theta^2 + \psi^2(r) ds_2^2,$$
$$I \times S^3, dr^2 + \varphi^2(r) \left(\psi^2(r)(\sigma^1)^2 + (\sigma^2)^2 + (\sigma^3)^2\right).$$

Use as basis for TM the natural frames in which we computed the curvature tensors. Now find the curvature operators for the Schwarzschild metric, the Eguchi-Hanson metric, $S^2 \times S^2$, S^4, and $\mathbb{C}P^2$.

(d) Show that (M, g) is Einstein iff $B = 0$ iff for every plane π and its orthogonal complement $\pi^\perp$ we have: $\sec(\pi) = \sec\left(\pi^\perp\right)$.

4
Hypersurfaces

In this chapter we shall explain some of the classical results for hypersurfaces in Euclidean space. First we introduce the Gauss map and show that convex immersions are embeddings of spheres. We then establish a connection between convexity and positivity of the intrinsic curvatures. This connection will enable us to see that $\mathbb{C}P^2$ and the Berger spheres are not even locally hypersurfaces in Euclidean space. We give a brief description of some classical existence results for isometric embeddings. Finally, a description of the Gauss-Bonnet theorem and its generalizations is given. One thing one might hope to get out of this chapter is the feeling that positively curved objects somehow behave like convex hypersurfaces, and might therefore have a very restricted topological type.

In this chapter we develop the theory of hypersurfaces in general as opposed to just presenting surfaces in 3-space. The reason is that there are some differences depending on the ambient dimension. Essentially, there are three different categories of hypersurfaces that behave very differently from a geometric point of view: curves, surfaces, and hypersurfaces of dimension > 2. We shall see that as the dimension increases, the geometry becomes more and more rigid. Strangely enough, this is a phenomenon that is rarely discussed in books, even though people often use hypersurfaces as a starting point for more general investigations.

The study of hypersurfaces started as the study of surfaces in Euclidean 3 space. Even before Gauss, both Euler and Meusner made contributions to this area. It was with Gauss, however, that things really picked up speed. One of his most amazing discoveries was that one can detect curvature by measuring angels in polygons. This immediately sent him out to the Lüneburg Heath to check whether the space we live in is flat. Much to its credit the German government has dedicated the 10 mark bill to Gauss. On it one finds a picture of his triangulation of the heath and

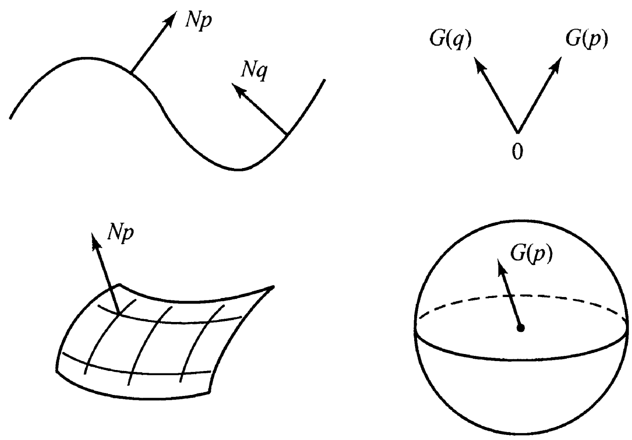


FIGURE 4.1.

also a graph of the normal distribution, two of his most important discoveries. The story of triangles will be studied in Chapter 11.

4.1 The Gauss Map

We shall suppose that we have a Riemannian manifold (M, g) with $\dim M = n$, and in addition a Riemannian immersion $\varphi : (M, g) \looparrowright (\mathbb{R}^{n+1}, \text{can})$. Locally we therefore have a Riemannian embedding, whence we can find a smooth distance function on some open subset of $\mathbb{R}^{n+1}$ that has the image of M as a level set. Using this we can define the shape operator $S : TM \to TM$ as a locally defined $(1, 1)$-tensor, which is well-defined up to sign (we just restrict the Hessian of the distance function to TM). If there is a globally defined normal field for M in $\mathbb{R}^{n+1}$, then we also get a globally defined shape operator. However, it still depends on our choice of normal and is therefore still only well-defined up to sign. Observe that such a global normal field exists exactly when M is orientable. By possibly passing to the orientation cover of M we can therefore assume that such a normal field exists globally (we can even assume that M is simply connected, although we won't do this). Let $N : M \to T\mathbb{R}^{n+1}$ be such a choice for a unit normal field. Using the trivialization $T\mathbb{R}^{n+1} = \mathbb{R}^{n+1} \times \mathbb{R}^{n+1}$ we then obtain the *Gauss map* $M \to S^n(1) \subset \mathbb{R}^{n+1}$, $G(x) = N(x)$ that to each point $x \in M$ assigns our choice of a normal to M at x in $\mathbb{R}^{n+1}$. A picture of the Gauss map for curves and surfaces is presented in Figure 4.1.

Our first important observation is that if we think of TM as a subset of $T\mathbb{R}^{n+1}$ then

$$DG(v) = S(v).$$

This is because $S(v) = \nabla_v N$, and since ∇ here is the Euclidean connection we know that this corresponds to our usual notion of the differential of a map. Now, N is the map that we differentiate, and hence the desired formula holds.

With our first definition of the shape operator it becomes clear that the hypersurface is locally convex (i.e., it lies locally on one side of its tangent space) provided that the shape operator is positive. Below we shall see how positivity of S is actually something that can be measured intrinsically by saying that some curvatures are positive. Before doing this let us use the above interpretation of the shape operator to show

Theorem 1.1 (Hadamard) *Let $\varphi : (M, g) \looparrowright (\mathbb{R}^{n+1}, \mathrm{can})$ be an isometric immersion, where $n > 1$ and M is a closed manifold. If the shape operator is always positive, then M is diffeomorphic to a sphere via the Gauss map. Thus, local convexity implies global convexity.*

Proof. If the shape operator is positive, then it is in particular nonsingular. The Gauss map $G : M \to S^n(1)$ is therefore a local diffeomorphism. When M is closed, it must therefore be a covering map. In case $n = 1$ the degree of this map is the winding number of the curve, while if $n > 1$, then $S^n(1)$ is simply connected, and hence G must be a diffeomorphism. □

4.2 Existence of Hypersurfaces

Let us recall the Tangential and Normal curvature equations. The curvature of $\mathbb{R}^{n+1}$ is simply zero everywhere, so if the curvature tensor of M is denoted R, then we have that R is related to S as follows:

$$\begin{aligned} 0 &= R(X, Y)Z - g(S(Y), Z)S(X) + g(S(X), Z)S(Y), \\ 0 &= g(-(\nabla_X S)(Y) + (\nabla_Y S)(X), Z), \end{aligned}$$

where X, Y, Z are vector fields on M. We can rewrite these equations as

$$\begin{aligned} R(X, Y)Z &= g(S(Y), Z)S(X) - g(S(X), Z)S(Y), \\ (\nabla_X S)(Y) &= (\nabla_Y S)(X). \end{aligned}$$

Here, the first equation is the *Gauss equation,* and the latter is the *Codazzi-Mainardi equation*. Thus, R can be computed if we know S. In the Codazzi-Mainardi equations there is of course a question of which connection we use. However, we know that the Euclidean connection when projected down to M gives the Riemannian connection for (M, g), so it doesn't matter which connection is used.

We are now ready to show that positive curvature is equivalent to positive shape operator, as mentioned above.

Proposition 2.1 *Suppose we have a Riemannian immersion $\varphi : (M, g) \looparrowright (\mathbb{R}^{n+1}, \mathrm{can})$, and we fix $x \in M$. If $\{e_1, \ldots, e_n\}$ is an orthonormal eigenbasis for $S : T_xM \to T_xM$ with eigenvalues λ_i, $i = 1, \ldots, n$, then $\{e_i \wedge e_j : i < j\}$ is an eigenbasis for the curvature operator $\mathfrak{R} : \Lambda^2(T_xM) \to \Lambda^2(T_xM)$ with* *eigenvalues* $\lambda_i \lambda_j$. *In particular, if all sectional curvatures are* $\geq \varepsilon^2 \geq 0$, *then the curvature operator is also* $\geq \varepsilon^2$.

Proof. Suppose we have an orthonormal eigenbasis $\{e_i\}$ for $T_x M$ with respect to S. Then $S(e_i) = \lambda_i e_i$. Using the Gauss equations we obtain

$$\begin{aligned} g\left(\mathfrak{R}\left(e_i \wedge e_j\right), e_k \wedge e_l\right) &= g\left(R\left(e_i, e_j\right) e_l, e_k\right) \\ &= g\left(S\left(e_j\right), e_l\right) g\left(S\left(e_i\right), e_k\right) \\ &\quad - g\left(S\left(e_i\right), e_l\right) g\left(S\left(e_j\right), e_k\right) \\ &= \lambda_i \lambda_j \left(g\left(e_j, e_l\right) g\left(e_i, e_k\right) - g\left(e_i, e_l\right) g\left(e_j, e_k\right)\right) \\ &= \lambda_i \lambda_j g\left(e_i \wedge e_j, e_k \wedge e_l\right). \end{aligned}$$

Thus we have diagonalized the curvature operator and shown that the eigenvalues are $\lambda_i \lambda_j$, $1 \leq i < j \leq n$. For the last statement we need only observe that the eigenvalues for the curvature operator satisfy

$$\lambda_i \lambda_j = g\left(\mathfrak{R}\left(e_i \wedge e_j\right), e_i \wedge e_j\right) = \sec\left(e_i, e_j\right) \geq \varepsilon^2. \qquad \square$$

This proposition shows that hypersurfaces have positive curvature operator iff they have positive sectional curvatures. In particular, the standard metric on $\mathbb{C}P^2$ cannot even locally be realized as a hypersurface metric.

There is a more holistic way of stating the above proposition and with it the Gauss equation. Given a linear map $L : V \to W$ we can construct $L \wedge L : \Lambda^2 V \to \Lambda^2 W$, by saying that: $(L \wedge L)(v_1 \wedge v_2) = L(v_1) \wedge L(v_2)$. The proposition can now be stated as

$$\mathfrak{R} = S \wedge S.$$

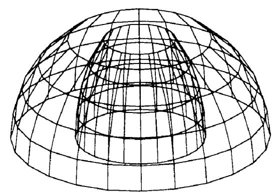
The shape operator is therefore a "square root" of the curvature operator. From this interpretation it is tempting to believe that the shape operator can somehow be computed from curvatures. This is always false for surfaces $(n = 2)$, as we shall see below, but to some extent true when $n \geq 3$.

Example 2.2 Consider a surface $dt^2 + (a \sin(t))^2 d\theta^2$. We know that this can be represented as a surface of revolution in $\mathbb{R}^3$ when $|a| \leq 1$. Such a surface certainly has constant curvature 1. Now it only remains to see how one can represent it as a surface of revolution. We know from Chapter 1 that such surfaces look like $\left(\dot{x}^2 + \dot{y}^2\right) dt^2 + y^2 d\theta^2$. In our case we therefore have to solve

$$\begin{aligned} \left(\dot{x}^2 + \dot{y}^2\right) &= 1, \\ y &= a \sin t, \end{aligned}$$

which implies:

$$\begin{aligned} x &= \int \sqrt{1 - (a \cos t)^2} dt, \\ y &= a \sin t. \end{aligned}$$



Two surfaces of constant curvature

FIGURE 4.2.

The embedding is written as:

$$\varphi(t, \theta) = (x(t), y(t)\cos\theta, y(t)\sin\theta),$$

where

$$D\varphi(\partial_t) = (\dot{x}(t), \dot{y}(t)\cos\theta, \dot{y}(t)\sin\theta),$$
$$D\varphi\left(\frac{1}{y}\partial_\theta\right) = (0, -\sin\theta, \cos\theta)$$

are unit vectors perpendicular to the surface. Then the normal can be computed as

$$N = D\varphi(\partial_t) \times D\varphi\left(\frac{1}{y}\partial_\theta\right) = (\dot{y}(t), -\dot{x}(t)\cos\theta, -\dot{x}(t)\sin\theta).$$

Since the curvature is $1 = \det S$, either $S = I$ or S has two eigenvalues $\lambda > 1$ and $\lambda^{-1} < 1$. However, if we choose $y = a\sin t$ with $0 < a < 1$ then, for example, $S(\partial_t) \neq \partial_t$. Thus, we must be in the second case. The shape operator is therefore really an extrinsic invariant for surfaces. It is not hard to picture these surfaces together with the sphere, although one can't of course see that they actually have the same curvature. In Figure 4.2 we have a picture of the unit sphere together with one of these surfaces.

It turns out that this phenomenon occurs only for surfaces. Having codimension 1 for a surface leaves enough room to bend the surface without changing the metric intrinsically. In higher dimensions, however, we have

Proposition 2.3 *Suppose we have a Riemannian immersion* $\varphi : (M, g) \looparrowright (\mathbb{R}^{n+1}, \mathrm{can})$, *where* $n \geq 3$. *Fix* $x \in M$ *and suppose the curvature operator* $\mathfrak{R} : \Lambda^2(T_xM) \to \Lambda^2(T_xM)$ *is positive. Then* $S : T_xM \to T_xM$ *is intrinsic, i.e., we can compute* S *from information about* (M, g) *alone without knowledge of* φ.

Proof. We shall assume for simplicity that $n = 3$. If $\{e_1, e_2, e_3\}$ is an orthonormal basis for T_xM, then it suffices to compute the matrix $(s_{ij}) = (g(S(e_i), e_j))$.

We already know that S is invertible from the above proposition and that all the eigenvalues have the same sign which we can assume to be positive. Thus, it suffices to determine the cofactor matrix (c_{ij}) defined by:

$$c_{ij} = (-1)^{i+j} \left(s_{i+1,j+1} s_{i+2,j+2} - s_{i+2,j+1} s_{i+1,j+2} \right).$$

The Gauss equations tell us that

$$\begin{aligned} g\left(\mathfrak{R}\left(e_i \wedge e_j\right), e_k \wedge e_l\right) &= g\left(R\left(e_i, e_j\right) e_l, e_k\right) \\ &= g\left(S\left(e_j\right), e_l\right) g\left(S\left(e_i\right), e_k\right) \\ &\quad - g\left(S\left(e_i\right), e_l\right) g\left(S\left(e_j\right), e_k\right). \end{aligned}$$

Index manipulation will therefore enable one to find c_{ij} from the curvature operator. We also need to find the determinant of S in order to compute S^{-1} from the cofactor matrix. But this can be done using

$$\det\left(c_{ij}\right) = (\det S)^{n-1}. \qquad \square$$

In case the curvature operator is only nonnegative we can still extract square roots, but they won't be unique. One can find more general conditions under which the shape operator is uniquely defined. As the cofactor matrices can always be found, the only important condition is that $\det S \neq 0$. This will be taken care of below and used for some very interesting purposes.

This information can be used to rule out even more candidates for hypersurfaces than did the previous result. Namely, when a space has positive curvature operator, then one can find the potential shape operator. However, this shape operator must also satisfy the Codazzi-Mainardi equations. It turns out that in dimensions > 3, the Codazzi equations are a consequence of the Gauss equations provided the shape operator has nonzero determinant. This was proved by T.Y. Thomas in [80] (see also the exercises to Chapter 7). For dimension 3, however, the following example shows that the Codazzi equations cannot follow from the Gauss equations:

Example 2.4 Let (M, g) be the Berger sphere $\left(S^3, \varepsilon^2\sigma_1^2 + \sigma_2^2 + \sigma_3^2\right)$ with $\{Y_1, Y_2, Y_3\} = \left\{\varepsilon^{-1}X_1, X_2, X_3\right\}$ as an orthonormal left-invariant frame on $SU(2)$. We computed in Chapter 3 that the 2-frame $\{Y_1 \wedge Y_2, Y_2 \wedge Y_3, Y_3 \wedge Y_1\}$ diagonalizes the curvature operator with eigenvalues ε^2, $\left(4 - 3\varepsilon^2\right)$, ε^2. It now follows from our calculations above that if this metric can be locally embedded in $\mathbb{R}^4$, then the shape operator can be computed using this information. If $\left(s_{ij}\right) = g\left(S\left(Y_i\right), Y_j\right)$, then it is easily seen that S must be diagonal, with

$$S(Y_1) = \frac{\varepsilon^2}{\sqrt{4 - 3\varepsilon^2}} Y_1,$$

$$S(Y_2) = \sqrt{4 - 3\varepsilon^2}\, Y_2,$$

$$S(Y_3) = \sqrt{4 - 3\varepsilon^2}\, Y_3.$$

We can now get a contradiction by showing that some of the Codazzi-Mainardi equations are not satisfied. For instance, we must have that $\left(\nabla_{Y_2} S\right)(Y_3) = \left(\nabla_{Y_3} S\right)(Y_2)$. However, these two quantities are not equal

$$\begin{aligned}
\left(\nabla_{Y_2} S\right)(Y_3) &= \sqrt{4-3\varepsilon^2}\nabla_{Y_2} Y_3 - S\left(\nabla_{Y_2} Y_3\right) \\
&= \sqrt{4-3\varepsilon^2}\varepsilon Y_1 - \varepsilon \frac{\varepsilon^2}{\sqrt{4-3\varepsilon^2}} Y_1 \\
&= \left(\sqrt{4-3\varepsilon^2} - \frac{\varepsilon^2}{\sqrt{4-3\varepsilon^2}}\right)\varepsilon Y_1; \\
\left(\nabla_{Y_3} S\right)(Y_2) &= \sqrt{4-3\varepsilon^2}\nabla_{Y_3} Y_2 - S\left(\nabla_{Y_3} Y_2\right) \\
&= -\sqrt{4-3\varepsilon^2}\varepsilon Y_1 + \frac{\varepsilon^2}{\sqrt{4-3\varepsilon^2}}\varepsilon Y_1 \\
&= \left(-\sqrt{4-3\varepsilon^2} + \frac{\varepsilon^2}{\sqrt{4-3\varepsilon^2}}\right)\varepsilon Y_1.
\end{aligned}$$

Now for some positive results.

Theorem 2.5 (Fundamental Theorem of Hypersurface Theory) *Suppose we have a Riemannian manifold (M, g) and a symmetric (1,1)-tensor S on M that satisfies both the Gauss and the Codazzi-Mainardi Equations on M. Then for every $x \in M$, we can find an isometric embedding $\varphi : (U, g) \hookrightarrow \left(\mathbb{R}^{n+1}, \text{can}\right)$ on some neighborhood $U \ni x$ with the property that S becomes the shape operator for this embedding.*

Proof. We shall give a short outline of the proof. Our first claim is that we can find a flat metric on $(-\varepsilon, \varepsilon) \times U$, where $U \subset M$ is relatively compact and ε is smaller than $\left|\lambda_i^{-1}\right|$ for any eigenvalue λ_i of S on U. It will then follow from material in Chapter 5 that any flat metric is locally isometric to a subset of $\left(\mathbb{R}^{n+1}, \text{can}\right)$. This will then finish the proof.

To construct the metric h on $(-\varepsilon, \varepsilon) \times U$ let us assume that it is of the type where $f(r, x) = r$ is a distance function, or in other words that $h = dr^2 + g_r$. Then if $x = \left(x^2, \ldots, x^{n+1}\right)$ are coordinates on U, we have that $\left(r, x^2, \ldots, x^{n+1}\right)$ are adapted coordinates. Now write down the metric $g(r, x)$ and shape operator $S(r, x)$ (we ignore indices and think of g and S as the whole matrix). At $r = 0$ these are given as the metric on U and the potential shape operator S. In addition, we wish the curvature to be zero, so we must have

$$\partial_r S(r, x) + S^2(r, x) = 0.$$

For given x we can solve this, as $S(0, x)$ is given to us. Specifically,

$$\begin{aligned}
S(r, x) &= S(0, x) \cdot (I + r \cdot S(0, x))^{-1} \\
&= \partial_r \log(I + r \cdot S(0, x))
\end{aligned}$$

solves this equation with the desired initial value. The equation

$$\partial_r g(r, x) = 2S(r, x)\, g(r, x),$$

together with the fact the the initial data $g(0, x)$ is already specified, now completely determines the metric.

We now need to prove that this metric is flat. As the metric is already flat in the direction of ∂_r, we need to show that the tangential and mixed curvature equations reduce to

$$\begin{aligned} R^r(X, Y)Z &= h(S(Y), Z)S(X) - h(S(X), Z)S(Y), \\ (\nabla_X S)(Y) &= (\nabla_Y S)(X), \end{aligned}$$

where R^r is the intrinsic curvature of g_r on $\{r\} \times U$, and X, Y, Z are tangent to U. At $r = 0$ this is certainly true, since we assumed that S was a solution to these equations. It suffices to check the above equations for coordinate vector fields. Both the metric and S are given to us explicitly in the chosen coordinates. A direct but nasty calculation will then show that equality holds for all r. □

We have already seen that positively curved manifolds of dimension $n > 2$ cannot necessarily be represented as hypersurfaces. When $n = 2$, the situation is drastically different.

Theorem 2.6 *If (M, g) is a 2-dimensional Riemannian manifold with positive curvature, then one can locally isometrically embed (M, g) into $\mathbb{R}^3$. Moreover if M is closed then a global embedding exists.*

The proof is beyond what we can cover here, but the previous theorem gives us an idea. Namely, one could simply try to find an appropriate shape operator. This would at least establish the local result. The global result is known as Weyl's problem and was established by Pogorelov and then later by Nirenberg.

4.3 The Gauss-Bonnet Theorem

To finish this chapter we give a description the Gauss-Bonnet Theorem and its generalizations. It was shown above that when a hypersurface has positive curvature then the shape operator is determined by intrinsic data . It turns out that the determinant of the shape operator is always intrinsic. This determinant is also called the *Gauss curvature.*

Lemma 3.1 *Let $(M, g) \looparrowright (\mathbb{R}^{n+1}, \mathrm{can})$ be an isometric immersion. If n is even, then* $\det S$ *is intrinsic, and if n is odd, then* $|\det S|$ *is intrinsic.*

Proof. Use an eigenbasis for $S : S(e_i) = \lambda_i e_i$; then of course $\det S = \lambda_1 \cdot \cdots \cdot \lambda_n$. In case $n = 2$ we therefore have $\det S = \sec$. Thus, $\det S$ is intrinsic. In higher dimensions the curvature operator is diagonalized by $e_i \wedge e_j$ with eigenvalues $\lambda_i \lambda_j$. Thus,

$$\begin{aligned}
\det \mathfrak{R} &= \prod_{i<j} \lambda_i \lambda_j \\
&= (\lambda_1 \cdot \cdots \cdot \lambda_n)^{n-1} \\
&= (\det S)^{n-1}.
\end{aligned}$$

This clearly proves the lemma. □

The importance of this lemma lies in the fact that $\det S$ is the Jacobian determinant of the Gauss map $G : M \to S^n(1) \subset \mathbb{R}^{n+1}$. When M is a closed manifold we therefore have

$$\begin{aligned}
\deg G &= \frac{1}{\mathrm{vol} S^n} \int_M \det S \cdot d\mathrm{vol} \\
&= \frac{1}{\mathrm{vol} S^n} \int_M \sqrt[n-1]{\det \mathfrak{R}} \cdot d\mathrm{vol}.
\end{aligned}$$

The degree of the Gauss map is therefore also intrinsic when n is even. This is perhaps less surprising, as H. Hopf has shown that closed even-dimensional hypersurfaces have the property that $\deg G$ is related to the Euler characteristic by the formula

$$\deg G = \frac{1}{2} \chi(M).$$

For an even-dimensional hypersurface we have therefore arrived at the important formula

$$\chi(M) = \frac{2}{\mathrm{vol} S^n} \int_M \sqrt[n-1]{\det \mathfrak{R}} \cdot d\mathrm{vol}.$$

As both sides of the formula are intrinsic quantities one might expect this formula to hold for all orientable even-dimensional closed Riemannian manifolds. When $n = 2$, this is the Gauss-Bonnet formula:

$$\chi(M) = \frac{1}{2\pi} \int_M \sec \cdot d\mathrm{vol}.$$

For higher dimensions, however, we run into trouble. First, observe that the above formula does not give the right answer for manifolds that are not hypersurfaces. A counterexample is $\mathbb{C}P^2$, which has two zero eigenvalues for the curvature operator, but the Euler characteristic is 3. Thus, a more complicated integrand is necessary. The correct expression is actually a generalized determinant of the curvature operator called the Pfaffian determinant. It is easiest to write it down in an oriented orthonormal frame $E_1, \ldots, E_n$ using the curvature forms defined by

$$R(\cdot, \cdot) E_i = \Omega_i^j E_j.$$

If we assume that the dimension is $n = 2m$, then the formula looks like this:

$$\begin{aligned}\chi(M) &= \frac{2}{\mathrm{vol} S^n}\int_M K \\ &= \frac{2(2m-1)!}{2^{2m}\pi^m(m-1)!}\int_M K,\end{aligned}$$

where K is defined as

$$\begin{aligned}K &= \frac{1}{n!}\sum \varepsilon^{i_1\cdots i_n}\cdot \Omega^{i_1}_{i_2}\wedge\cdots\wedge\Omega^{i_{n-1}}_{i_n}, \\ \varepsilon^{i_1\cdots i_n} &= \text{sign of the permutation } (i_1\cdots i_n).\end{aligned}$$

The generalized theorem was first proven for manifolds that lie isometrically in some Euclidean space, not necessarily of one higher dimension, independently by Allendoerfer and Fenchel. Allendoerfer and Weil then established the general case, using some interesting tricks about local isometric embedability (see [1]). Finally, Chern found a completely intrinsic proof, which makes no mention of isometric embeddings. The theorem is now called the Chern-Gauss-Bonnet Theorem despite the fact that Allendoerfer and Weil were the first to prove it in complete generality in higher dimensions.

Using more or less Chern's approach we can give a brief account of how the Gauss-Bonnet theorem can be proven for surfaces (M^2, g). First suppose that M is the torus, and pick your favorite nonzero vector field X. Using the metric, normalize it to have length 1, and then select another field such that we get an orientable orthonormal frame $\{E_1, E_2\}$. Let $\{\theta^1, \theta^2\}$ be the dual coframe and compute the connection form and curvature form as described in the exercises to Chapter 2:

$$\begin{aligned}d\theta^1 &= \theta^2\wedge\omega^1_2, \\ d\omega^1_2 &= \Omega^1_2 = \sec\cdot d\mathrm{vol}.\end{aligned}$$

Then we have

$$\begin{aligned}\int_M \sec\cdot d\mathrm{vol} &= \int_M \Omega^1_2 \\ &= \int_M d\omega^1_2 \\ &= \int_{\partial M}\omega^1_2 = 0.\end{aligned}$$

On other surfaces we can choose a vector field X with isolated zeros at $p_1, \ldots, p_k \in M$. Then we choose the frame $\{E_1, E_2\}$ as above on $M - \{p_1, \ldots, p_k\}$. On a neighborhood U_i around each p_i introduce normal coordinates such that

$$g = g_{\alpha\beta} = \delta_{\alpha\beta} + O\left(r^2\right).$$

Here, r is the Euclidean distance from p_i. We can then consider the manifold with boundary $M_\varepsilon = M - \bigcup_{i=1}^k B(p_i, \varepsilon)$, where $B(p_i, \varepsilon)$ is the Euclidean ball of radius ε around p_i. As before we still have

$$\begin{aligned}\int_{M_\varepsilon} \sec \cdot d\mathrm{vol} &= \int_{\partial M_\varepsilon} \omega_2^1 \\ &= \sum_{i=1}^{k} \int_{\partial B(p_i,\varepsilon)} \omega_2^1.\end{aligned}$$

Let us now analyze each of the integrals $\int_{\partial B(p_i,\varepsilon)} \omega_2^1$ on U_i. On U_i we could instead find an orientable orthonormal frame $\left\{F_1 = \frac{X}{|X|}, F_2\right\}$, but this time with respect to the Euclidean metric on U_i. If $\tilde{\omega}_2^1$ is the connection form for this frame, we can construct the integral

$$\int_{\partial B(p_i,\varepsilon)} \tilde{\omega}_2^1.$$

Using that the metric is Euclidean up to first order, we obtain that

$$\omega_2^1 - \tilde{\omega}_2^1 = O(r).$$

In particular, we must have

$$\lim_{\varepsilon \to 0} \int_{\partial B(p_i,\varepsilon)} \tilde{\omega}_2^1 = \lim_{\varepsilon \to 0} \int_{\partial B(p_i,\varepsilon)} \omega_2^1.$$

This proves that the integral $\int_{M_\varepsilon} \sec \cdot d\mathrm{vol}$ does not depend on the metric.

Let us now relate the term $\lim_{\varepsilon \to 0} \int_{\partial B(p_i,\varepsilon)} \tilde{\omega}_2^1$ to the vector field X. We can suppose that we are on a neighborhood $U \subset \mathbb{R}^2$ around the origin and that we have a vector field X that vanishes only at the origin. If we normalize X to have unit length $E = X/|X|$, then we get for each $\varepsilon > 0$ a map

$$\begin{aligned}\partial B(p_i, \varepsilon) &\to \partial B(p_i, \varepsilon), \\ x &\to \varepsilon \cdot E(x).\end{aligned}$$

The degree (see Appendix A) of this map is easily seen to be independent of ε. This degree is known as the *index* of the vector field at the origin and is denoted by $\mathrm{ind}_{\bar{0}} X$. The degree of this map can now be computed as

$$\frac{1}{\ell(\partial B(p_i, \varepsilon))} \int_{\partial B(p_i,\varepsilon)} D(\varepsilon \cdot E(x)) = \frac{1}{2\pi} \int_{\partial B(p_i,\varepsilon)} D(E(x)).$$

One can now easily check that

$$\int_{\partial B(p_i,\varepsilon)} D(E(x)) = \lim_{\varepsilon \to 0} \int_{\partial B(p_i,\varepsilon)} \tilde{\omega}_2^1.$$

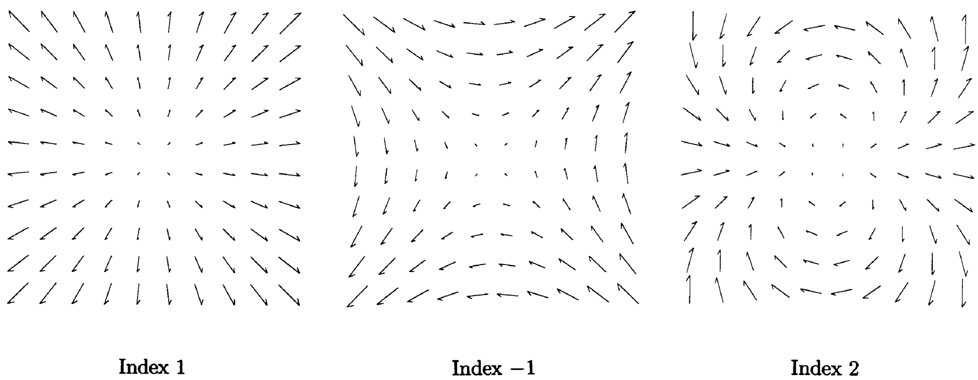


FIGURE 4.3.

All in all, we have therefore shown that

$$\frac{1}{2\pi}\int_M \sec \cdot d\mathrm{vol} = \sum_{i=1}^{k} \mathrm{ind}_{p_i}(X).$$

The left-hand side is therefore independent of the metric, while the right-hand side must now be independent of the chosen vector field. Knowing that the right-hand side is independent of the vector field, one can easily compute it as the Euler characteristic by choosing a particular vector field on each surface. Figure 4.3 shows a few pictures of vector fields in the plane.

4.4 Further Study

All of the results mentioned in this chapter and much more can be found in Spivak's [76, volume 5]. In fact we recommend all of his volumes as a good and thorough introduction to geometry. Spivak is also quite careful and complete with references to all the work mentioned here. The only fault Spivak's book has in reference to the generalized Gauss-Bonnet theorem is that he claims that Allendoerfer and Weil established this formula for analytic metrics. But of course if one knows the theorem for analytic metrics, then it is simple matter to prove it for all smooth metrics. For a very nice discussion of the Gauss-Bonnet theorem for surfaces see also [18].

We can also recommend Stoker's book [77]. This book goes from curves to surfaces and ends up with a discussion of general relativity. For the reader who likes the old-fashioned well-written book this is a must.

One defect here is that we haven't developed submanifold theory in general. This is done in [76].

4.5 Exercises

1. Consider the hypersurface $x^{n+1} = (x^n)^2$. Show that the shape operator is not zero but that the hypersurface is isometric to $\mathbb{R}^n$.

2. For $n \geq 3$, show that it is not possible for a Riemannian n-manifold to have negative curvature, if it admits a Riemannian immersion into $\mathbb{R}^{n+1}$. Give a counterexample when $n = 2$.

3. Let (M, g) be a closed Riemannian n-manifold, and suppose that there is a Riemannian embedding into $\mathbb{R}^{n+1}$. Show that there must be a point $p \in M$ where the curvature operator $\mathfrak{R} : \Lambda^2 T_pM \to \Lambda^2 T_pM$ is positive. (Hint: consider a sphere that circumscribes the submanifold, and consider points where this sphere touches the submanifold.)

4. Suppose (M, g) is immersed as a hypersurface in $\mathbb{R}^{n+1}$, with shape operator S.

   (a) Using the Codazzi-Mainardi equations, show that
   $$\mathrm{div} S = d\,(\mathrm{tr} S).$$

   (b) Show that if $S = f(x) \cdot I$ for some function f, then f must be a constant and the hypersurface must have constant curvature.

   (c) Show that $S = \lambda \cdot \mathrm{Ric}$ iff the metric has constant curvature. This problem is interesting because the Ricci flow (see [48])
   $$\partial_t g = -2\mathrm{Ric}$$
   resembles the equidistant hypersurface equation
   $$L_{\partial_r} g = 2g\,(S\cdot, \cdot).$$

5. Let g be a metric on S^2 with curvature ≤ 1. Use the Gauss-Bonnet formula to show that $\mathrm{vol}\left(S^2, g\right) \geq \mathrm{vol} S^2(1) = 4\pi$.

   Show that such a result cannot hold on S^3 by considering the Berger metrics.

6. Assume that we have an orientable Riemannian manifold with nonzero Euler characteristic and $|\mathfrak{R}| \leq 1$. Find a lower bound for $\mathrm{vol}(M, g)$. The one-sided curvature bound that we used on surfaces does not suffice in higher dimensions, as one-sided curvature bounds do not necessarily imply one sided bounds on K.

7. Show that in even dimensions, orientable manifolds with positive (or nonnegative) curvature operator have positive (nonnegative) Euler characteristic. Conclude that if in addition, such manifolds have bounded curvature operator, then they have volume bounded from below. What happens when the curvature operator is nonpositive or negative?

8. In dimension 4 show using the exercises from Chapter 3 that

$$\frac{1}{8\pi^2}\int_M \left(|R|^2 - \left|\mathrm{Ric} - \frac{\mathrm{scal}}{4}g\right|^2\right) = \frac{1}{8\pi^2}\int_M \mathrm{tr}\left(A^2 - 2BB^* + C^2\right).$$

It was shown by Allendoerfer and Weil that in dimension 4

$$\chi(M) = \frac{1}{8\pi^2}\int_M \left(|R|^2 - \left|\mathrm{Ric} - \frac{\mathrm{scal}}{4}g\right|^2\right).$$

You can try to prove this using the above definition of K. If the metric is Einstein, show that

$$\begin{aligned}\chi(M) &= \frac{1}{8\pi^2}\int_M \mathrm{tr}\left(A^2 - 2BB^* + C^2\right) \\ &= \frac{1}{8\pi^2}\int_M \left(\left|W^+\right|^2 + \left|W^-\right|^2 + \frac{\mathrm{scal}^2}{24}\right).\end{aligned}$$

5

Geodesics and Distance

We are now ready to introduce the important concepts of parallel transport and geodesics. This will help us to define and understand Riemannian manifolds as metric spaces. One is led to two types of completeness. The first is of standard metric completeness, and the other is what we call geodesic completeness, namely, when all geodesics exist for all time. We shall prove the Hopf-Rinow Theorem, which asserts that these types of completeness for a Riemannian manifold are equivalent. Using the metric structure we can define metric distance functions. We shall study when these distance functions are smooth and therefore show the existence of the kind of distance functions we worked with earlier. In the last section we give some metric characterizations of Riemannian isometries and submersions.

Parallel translation was, as already mentioned, introduced by Levi-Civita. The idea of thinking of a Riemannian manifold as a metric space must be old, but still it wasn't until the early 1930s that Hopf and Rinow began to understand the relationship between extendability of geodesics and completeness of the metric. Nonetheless, both Gauss and Riemann had a pretty firm grasp on local geometry, as is evidenced by their contributions. Gauss worked with geodesic polar coordinates and also isothermal coordinates. Riemann was able to give a local characterization of Euclidean space as the only manifold whose curvature tensor vanishes. Nevertheless, it wasn't until Klingenberg's work in the 1950s that one got a good understanding of the domain on which one has geodesic polar coordinates. This work led to the introduction of the two terms *injectivity radius* and *conjugate radius*. Many of our results will require a detailed analysis of these concepts. The metric characterization of Riemannian isometries strangely enough wasn't realized until the late 1930s with the work of Myers and Steenrod. Even more surprising is Berestovskii's very recent metric characterization of submersions. It seems that part of the lag effect for these two results is due to the fact that people have not paid much attention to distance coordinates but have instead been obsessed with doing everything in exponential coordinates.

5.1 The Connection Along Curves

Let $\gamma : I \to M$ be a curve in M. A vector field V along γ is by definition a function $V : I \to TM$ with $V(t) \in T_{\gamma(t)}M$ for all $t \in I$. We want to define the covariant derivative

$$\dot{V}(t) = \frac{d}{dt}V(t) = \nabla_{\dot{\gamma}} V$$

of V along γ when γ and V have appropriate smoothness. Whenever $\dot{\gamma}(t_0) \neq 0$, there is a natural way to do this: In this case, γ is locally (near t_0) an embedded curve, and every vector field V along γ is locally the restriction to γ of a vector field defined in a whole neighborhood of $\gamma(t_0)$, i.e., for t near t_0, $V(t) = \widehat{V}(\gamma(t))$, for some vector field $\widehat{V}$ defined in a neighborhood of $\gamma(t_0)$. Then we could just put $\dot{V}(t_0) = \nabla_{\dot{\gamma}(t_0)}\widehat{V}$.

On the other hand, when $\gamma(t_0) = 0$, this approach may not work. Consider, for example, a constant curve $\gamma : I \to M$, $\gamma(I) = \{p\}$. A vector field along γ is then merely a map $V : I \to T_pM$. In this case, one would like $\dot{V}(t_0)$ to be the usual derivative of V as a curve in a vector space.

To accommodate these two situations we do the following: Choose a neighborhood U around $\gamma(t_0)$, and vector fields $E_1, \ldots, E_n$ on M that form a basis for T_qM for all $q \in U$. For instance, we could take the E_i's to be coordinate vector fields $\partial/\partial x^i$ with U a coordinate chart. Write

$$V(t) = \sum \varphi^i(t) \cdot E_i \circ \gamma(t)$$

for t in some neighborhood of t_0. Then we define

$$\dot{V}(t_0) = \frac{dV}{dt}\Big|_{t=t_0} = \sum \dot{\varphi}^i(t_0) \cdot E_i \circ \gamma(t_0) + \varphi^i(t_0) \cdot \nabla_{\dot{\gamma}(t_0)} E_i.$$

It is easily seen that $\dot{V}(t_0)$ is independent of the choice of $E_1, \ldots, E_n$. Also, this definition generalizes the two special cases discussed above: it gives the "right" definition when $\dot{\gamma}(t_0) \neq 0$ and also when γ is a constant curve. In particular, if Z is a vector field defined in a neighborhood of $\gamma(t_0)$, then you can see directly that $\frac{d}{dt} Z(\gamma(t)) \,|_{t=t_0}$, in this just introduced sense of differentiation along curves, is equal to $\nabla_{\dot{\gamma}(t_0)}Z$. (This is related to the fact that $\nabla_{\dot{\gamma}(t_0)}Z$ in this case is determined by the values of Z along γ.) Using this definition and the fact that the connection is Riemannian, it is easy to show that

$$\frac{d}{dt} g(V, W) = g\left(\dot{V}, W\right) + g\left(V, \dot{W}\right)$$

for vector fields V, W along γ.

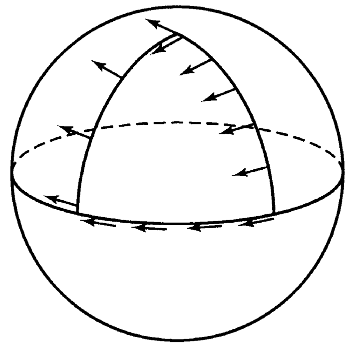


FIGURE 5.1.

A vector field V along γ is said to be *parallel along* γ provided that $\dot{V} \equiv 0$. If V, W are two parallel fields along γ, then we clearly have that $g(V, W)$ is constant along γ. Parallel fields along a curve therefore change neither their lengths nor their angles relative to each other, just as parallel fields in Euclidean space are of constant length and make constant angles. Parallel translation around a triangle on the sphere is pictorially described in Figure 5.1. The exercises to this chapter will cover some features of parallel translation for surfaces to aid the reader's geometric understanding.

Theorem 1.1 (Existence and Uniqueness of Parallel Fields) *If $t_0 \in I$ and $v \in T_{\gamma(t_0)}M$, then there is a unique parallel field $V(t)$ defined on all of I with $V(t_0) = v$.*

Proof. Choose vector fields $E_1(t), \ldots, E_n(t)$ along γ forming a basis for $T_{\gamma(t)}M$ for all $t \in I$. This is always possible. Any vector field $V(t)$ along γ can then be written $V(t) = \sum \varphi^i(t)E_i(t)$ for $\varphi^i \in F(I)$. Thus,

$$\begin{aligned}\dot{V} = D_{\dot{\gamma}}V &= \sum \dot{\varphi}^i(t)E_i(t) + \varphi^i(t)\nabla_{\dot{\gamma}}E_i\\ &= \sum \dot{\varphi}^j(t)E_j(t) + \sum_{i,j} \varphi^i(t)\cdot\alpha_i^j(t)E_j(t), \text{ where } \nabla_{\dot{\gamma}}E_i = \sum \alpha_i^j(t)E_j\\ &= \sum_j (\dot{\varphi}^j(t) + \varphi^i(t)\alpha_i^j(t))E_j(t).\end{aligned}$$

Hence, V is parallel iff $(\varphi^1(t), \ldots, \varphi^n(t))$ satisfies the first-order linear differential equation

$$\dot{\varphi}^j(t) = -\sum_{i=1}^n \varphi^i(t)\alpha_i^j(t) \qquad j = 1, \ldots, n.$$

Such differential equations, however, have the property that given an initial value $(\varphi^1(t_0), \ldots, \varphi^n(t_0))$, there is a unique solution defined on all of I with this value. □

The existence and uniqueness assertion that concluded this proof is a standard theorem in differential equations that we take for granted. The reader should recall that linearity of the equations is a crucial ingredient. Nonlinear equations can fail to have solutions over a whole given interval. This failure will be observed explicitly in our discussion, in subsequent sections, of the existence of geodesics.

Parallel translation can be used as a substitute for Cartesian coordinates. Namely, if we choose a parallel orthonormal framing $\{E_1(t), \dots, E_n(t)\}$ along the curve $\gamma(t) : I \to (M, g)$, then we've seen that any vector field $V(t)$ along γ has the property that

$$\begin{aligned}\frac{dV}{dt} &= \frac{d}{dt}\alpha^i(t)E_i(t)\\ &= \dot{\alpha}^i(t)E_i(t) + \alpha^i(t) \cdot \dot{E}_i(t)\\ &= \dot{\alpha}^i(t)E_i(t).\end{aligned}$$

So $d/(dt)V$, when represented in coordinates of the frame, is exactly what we would expect. We could more generally choose a tensor T along $\gamma(t)$ of type $(0, p)$ or $(1, p)$ and compute $d/(dt)T$. For the sake of simplicity, choose a $(1, 1)$ tensor S. Then write $S(E_i(t)) = \alpha_i^j(t)E_j(t)$. Thus S is represented by the matrix $(\alpha_i^j(t))$ along the curve. As before, we see that $d/(dt)S$ is represented by $(\dot{\alpha}_i^j(t))$.

This makes it possible to understand equations involving only one differentiation of the type ∇_X. Let φ^t be the local flow near some point $p \in M$ and H a hypersurface in M through p that is perpendicular to X. Next choose vector fields $E_1, \dots, E_n$ on H which form an orthonormal frame for the tangent space to M restricted to H. Finally, construct an orthonormal framing in a neighborhood of p by parallel translating $E_1, \dots, E_n$ along the integral curves for X. Thus, $\nabla_X E_i = 0$, $i = 1, \dots, n$. So if we have a vector field Y near p, we can write $Y = \alpha^i \cdot E_i$ and $\nabla_X Y = D_X(\alpha^i) \cdot E_i$. Similarly, if S is a $(1, 1)$-tensor, we have $S(E_i) = \alpha_i^j E_i$, and $\nabla_X S$ is represented by $(D_X(\alpha_i^j))$.

Thus, these parallel frames make covariant derivatives look like normal derivatives in just the same way one can use coordinates that make Lie derivatives look like normal derivatives.

5.2 Geodesics

A C^∞ curve $\gamma : I \to M$ is called a *geodesic* if $\dot{\gamma}(t)$ is parallel along γ, i.e., $\ddot{\gamma} = d/(dt)\dot{\gamma} \equiv 0$ on I. If γ is a geodesic, then $|\dot{\gamma}| = \sqrt{g(\dot{\gamma}, \dot{\gamma})}$ is constant, since parallel vector fields along a curve have constant length. So a geodesic is a constant-speed curve, or phrased differently, it is parametrized proportional to arc length. If $|\dot{\gamma}| \equiv 1$, one says that γ is parametrized by arc length.

If $f : U \to \mathbb{R}$ is a distance function, then we know that for $N = \nabla f$ we have $\nabla_N N = 0$. The integral curves for $\nabla f = N$ are therefore geodesics. We shall in the sequel develop a theory for geodesics independently of distance functions and then use this to show the existence of distance functions.

Geodesics are fundamental in the study of the geometry of Riemannian manifolds in the same way that straight lines are fundamental in Euclidean geometry.

But at first sight it is not even clear that there are going to be any nonconstant geodesics to study on a general Riemannian manifold. In this section we are going to establish that every Riemannian manifold has many nonconstant geodesics. Informally speaking, we can find a unique one at each point with a given tangent vector at that point. The question of how far it will extend from that point is subtle, however. To deal with the existence and uniqueness questions, we need to use some (more) information from differential equations. The first step toward applying this information is to figure out what the equation of geodesics $d/(dt)\dot{\gamma} = 0$ looks like in local coordinates.

For this, let $\varphi : U \to \mathbb{R}^n$ be a coordinate chart with coordinate vector fields $\partial_1, \ldots, \partial_n$, $\partial_i = \partial/(\partial x^i)$. As before, we set

$$\nabla_{\partial_i}\partial_j = \sum_{k=1}^{n} \Gamma_{ij}^k \partial_k.$$

Now consider a curve $\gamma : I \to U$ and write $(\varphi \circ \gamma)(t) = (\varphi^1(t), \ldots, \varphi^n(t))$. Then $\dot{\gamma}(t) = \sum_i \dot{\varphi}^i(t)\partial_i \mid_{\gamma(t)}$. From this and our definition of differentiation of vector fields along curves

$$\begin{aligned}
\ddot{\gamma} &= \sum_i \ddot{\varphi}^i \partial_i \mid_{\gamma(t)} + \sum_i \dot{\varphi}^i(t)\nabla_{\dot{\gamma}}\partial_i \\
&= \sum_i \ddot{\varphi}^i \partial_i \mid_{\gamma(t)} + \sum_i \dot{\varphi}^i(t)\nabla_{\sum_j \dot{\varphi}^j\partial_j}\partial_i \\
&= \sum_i \ddot{\varphi}^i \partial_i \mid_{\gamma(t)} + \sum_{i,j,k} \dot{\varphi}^i(t)\dot{\varphi}^j(t)\Gamma_{ij}^k \mid_{\gamma(t)} \partial_k \mid_{\gamma(t)} \\
&= \sum_k \left(\ddot{\varphi}^k + \sum_{i,j} \dot{\varphi}^i\dot{\varphi}^j\Gamma_{ij}^k \mid_{\gamma(t)}\right) \partial_k \mid_{\gamma(t)} .
\end{aligned}$$

Thus, the curve $\gamma : I \to U$ is a geodesic if and only if its coordinate representation $(\varphi^1(t), \ldots, \varphi^n(t))$ satisfies the second-order differential equation

$$\ddot{\varphi}^k(t) = -\sum_{i,j} \dot{\varphi}^i(t)\dot{\varphi}^j(t)\Gamma_{ij}^k \mid_{\varphi^{-1}(\varphi_1(t),\ldots,\varphi_n(t))}$$

for $k = 1, \ldots, n$. Because this is a second-order system of differential equations, we expect an existence and a uniqueness result for the initial value problem of specifying value and first derivative, i.e., $\varphi^1(0), \ldots, \varphi^n(0)$ and $\dot{\varphi}^1(0), \ldots, \dot{\varphi}^n(0)$. But because the system is nonlinear, we are not entitled to expect that solutions will exist for all t values.

The precise statements obtained from differential equations theory are a bit of a mouthful, but we might as well go for the whole thing right off the bat, since we shall need it all eventually. Still working in our coordinate situation, we get the following facts from standard general theorems on ordinary differential equations:

Theorem 2.1 (Existence and Uniqueness) *For each $\alpha \in \varphi(U) \subset \mathbb{R}^n$ and $\beta \in \mathbb{R}^n$, there is a neighborhood U_1 of α, $U_1 \subset U$; a neighborhood U_2 of β; and an $\varepsilon >$*

0 *such that for each* $\alpha' \in U_1$ *and* $\beta' \in U_2$, *there is a geodesic* $\gamma_{\alpha',\beta'} : (-\varepsilon, \varepsilon) \to U$ *with*

$$\gamma(0) = \varphi^{-1}(\alpha')$$

and

$$\dot{\gamma}(0) = D\varphi^{-1}|_{\alpha'}(\beta').$$

Moreover, the mapping

$$(\alpha', \beta', t) \to \gamma_{\alpha',\beta'}(t)$$

is C^∞ *on* $U_1 \times U_2 \times (-\varepsilon, \varepsilon)$.

Theorem 2.2 (Uniqueness 2) *If* I_1 *and* I_2 *are connected open subsets of* $\mathbb{R}$ *with* $0 \in I_1 \cap I_2$, *and* $\gamma_1 : I_1 \to U$ *and* $\gamma_2 : I_2 \to U$ *are geodesics; and if* $\gamma_1(0) = \gamma_2(0)$ *and* $\dot{\gamma}_1(0) = \dot{\gamma}_2(0)$; *then* $\gamma_1 \mid I_1 \cap I_2 = \gamma_2 \mid I_1 \cap I_2$.

It is worthwhile to consider what these assertions become in informal terms. The existence statement includes not only "small-time" existence of a geodesic with given initial point and initial tangent, it also asserts a kind of local uniformity for the interval of existence. If you vary the initial conditions but don't vary them too much, then there is a fixed interval $(-\varepsilon, \varepsilon)$ on which all the geodesics with the various initial conditions are defined. Some or all may be defined on larger intervals, but all are defined at least on $(-\varepsilon, \varepsilon)$.

The uniqueness assertion amounts to saying that geodesics cannot be tangent at one point without coinciding. Just as two straight lines that intersect and have the same tangent (at the point of intersection) must coincide, so two geodesics with a common point and equal tangent at that point must coincide.

Both the differential equations statements are for geodesics with image in a fixed coordinate chart. By relatively easy covering arguments these statements can be extended to geodesics not necessarily contained in a coordinate chart. Let us begin with the uniqueness question:

Lemma 2.3 *If* I_1 *and* I_2 *are connected open subsets of* $\mathbb{R}$ *with* $0 \in I_1 \cap I_2$ *and if* $\gamma_1 : I_1 \to M$ *and* $\gamma_2 : I_2 \to M$ *are geodesics with* $\gamma_1(0) = \gamma_2(0)$ *and* $\dot{\gamma}_1(0) = \dot{\gamma}_2(0)$, *then* $\gamma_1 \mid (I_1 \cap I_2) = \gamma_2 \mid (I_1 \cap I_2)$.

Proof. Define $A = \{t \in I_1 \cap I_2 : \gamma_1(t) = \gamma_2(t) \text{ and } \dot{\gamma}_1(t) = \dot{\gamma}(t)\}$. Then $0 \in A$. Also, A is closed in $I_1 \cap I_2$ by continuity of $\gamma_1, \gamma_2, \dot{\gamma}_1$, and $\dot{\gamma}_2$. Finally, A is open, by virtue of the uniqueness statement for geodesics in coordinate charts: if $\lambda \in A$, then choose a coordinate chart U around $\gamma_1(\lambda) = \gamma_2(\lambda)$. Then $(\lambda - \varepsilon, \lambda + \varepsilon) \subset I_1 \cap I_2$ and $\gamma_2 \mid (\lambda - \varepsilon, \lambda + \varepsilon)$ both have images contained in U. Then the coordinate uniqueness result shows that $\gamma_1 \mid (\lambda - \varepsilon, \lambda + \varepsilon) = \gamma_2 \mid (\lambda - \varepsilon, \lambda + \varepsilon)$, so that $(\lambda - \varepsilon, \lambda + \varepsilon) \subset A$. □

The coordinate-free existence picture is a little more subtle. The first, and easy, step is to notice that if we start with a geodesic, then we can enlarge its interval of definition to be maximal. This follows from the uniqueness assertions: If we look at all geodesics $\gamma : I \to M, 0 \in I, \gamma(0) = p, \dot{\gamma}(0) = v$, p and v fixed, then the union of all their domains of definition is a connected open subset of $\mathbb{R}$ on which such a geodesic is defined. And clearly its domain of definition is maximal.

The next observation, also straightforward, is that if $\widehat{K}$ is a compact subset of TM, then there is an $\varepsilon > 0$ such that for each $(q, v) \in \widehat{K}$, there is a geodesic $\gamma : (-\varepsilon, \varepsilon) \to M$ with $\gamma(0) = q$ and $\dot{\gamma}(0) = v$. This is an immediate application of the local uniformity part of the differential equations existence statement together with the usual covering-of-compact-set argument.

The next point to ponder is what happens when the maximal domain of definition is not all of $\mathbb{R}$. For this, let I be a connected open subset of $\mathbb{R}$ that is bounded above, i.e., I has the form $(-\infty, \lambda_1)$, $\lambda_1 \in \mathbb{R}$ or (λ_2, λ_1), $\lambda_2, \lambda_1 \in \mathbb{R}$. Suppose $\gamma : I \to M$ is a maximal geodesic. Then $\gamma(t)$ as t approaches λ_1 must have a specific kind of behavior: If K is a compact subset of M, then there must be a number $t_K < \lambda_1$ such that if $t \in I$ and $t > t_K$ then $\gamma(t) \in M - K$. We say that γ *leaves every compact set* as $t \to \lambda_1$.

To see why γ must leave every compact set, suppose K is a compact set it doesn't leave, i.e., suppose there is a sequence $t_1, t_2, \ldots \in I$ with $\lim t_j = \lambda_1$ and $\gamma(t_j) \in K$ for each j. Now $|\dot{\gamma}(t_j)|$ is independent of j, since geodesics have constant speed. So $\{\dot{\gamma}(t_j) : j = 1, \ldots\}$ belong to a compact subset of TM, namely,

$$\widehat{K} = \{(v_q) : q \in K, v \in T_qM, |v| \leq |\dot{\gamma}(t_i)|\}.$$

So there is an $\varepsilon > 0$ such that for each $(v_q) \in \widehat{K}$, there is a geodesic $\gamma : (-\varepsilon, \varepsilon) \to M$ with $\gamma(0) = q, \dot{\gamma}(0) = v$. Now choose t_j such that $\lambda_1 - t_j < \varepsilon/2$. Then $\gamma_{q,v}$ patches together with γ to extend γ: beginning at t_j we can continue γ by ε, which takes us beyond λ_1, since t_j is within $\varepsilon/2$ of λ_1. This contradicts the maximality of I.

One important consequence of these observations is what happens when M itself is compact:

Lemma 2.4 *If M is a compact Riemannian manifold, then for each $p \in M$ and $v \in T_pM$, there is a geodesic $\gamma : \mathbb{R} \to M$ with $\gamma(0) = p, \dot{\gamma}(0) = v$. In other words, geodesics exist for all time.*

In other informal words, all geodesics on a compact manifold are infinitely extendable. A Riemannian manifold where all geodesics exist for all time is called *geodesically complete*.

A slightly trickier point is the following: Suppose $\gamma : I \to M$ is a geodesic and $0 \in I$, where I is a bounded connected open subset of $\mathbb{R}$. Then we would like to say that for $q \in M$ near enough to $\gamma(0)$ and $v \in T_qM$ near enough to $\dot{\gamma}(0)$ there is a geodesic $\gamma_{q,v}$ with q, v as initial position and tangent, respectively, *and* with $\gamma_{q,v}$ defined on an interval almost as big as I. This is true, and it is worth putting in formal language:

Lemma 2.5 *Suppose $\gamma : I \to M, I = (\lambda_2, \lambda_1)$, $-\infty < \lambda_2 < 0 < \lambda_1 < \infty$, is a geodesic. Then, given $\varepsilon > 0$, there is a neighborhood U in TM of $(\gamma(0), \dot{\gamma}(0))$*

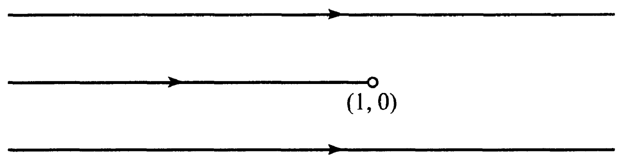


FIGURE 5.2.

such that if $(q, v) \in U$, *then there is a geodesic*

$$\gamma_{q,v} : (\lambda_2 + \varepsilon, \lambda_1 - \varepsilon) \to M$$

with $\gamma_{q,v}(0) = q$ *and* $\dot{\gamma}_{q,v}(0) = v$.

The proof of this statement comes from the understanding we have already developed of how a geodesic can stop being defined, namely, only by leaving every compact set. If we choose a neighborhood V of $\gamma([\lambda_2 + \varepsilon/2, \lambda_1 - \varepsilon/2])$ with compact closure $\overline{V}$, then altering the initial conditions of γ only slightly to get $\hat{\gamma}$ will by the continuity of dependence on initial conditions—make $\hat{\gamma}$ stay in V, hence in $\overline{V}$, essentially as long as γ did. Since $\overline{V}$ is compact, $\hat{\gamma}$ can be extended to be defined up to almost the same interval as γ itself. Of course, in outline this argument sounds circular! But working with a covering of $\overline{V}$ by a finite set of coordinate patches will make it possible to fill in the details.

All this seems a bit formal and pedantic and perhaps abstract as well, in the absence of explicitly computed examples. First, one can easily check that geodesics in Euclidean space are straight lines. Using this observation one can easily give examples of the above ideas by taking M's to be open subsets of $\mathbb{R}^2$ with its usual metric.

Example 2.6 In the plane $\mathbb{R}^2$ minus one point, say $\mathbb{R}^2 - \{(1, 0)\}$ the geodesic from $(0, 0)$ with tangent $(1, 0)$ is defined on $(-\infty, 1)$ only. But nearby geodesics from $(0, 0)$ with tangents $(1 + \varepsilon_1, \varepsilon_2)$, $\varepsilon_1, \varepsilon_2$ small, $\varepsilon_2 \neq 0$, are defined on $(-\infty, \infty)$. Thus maximal intervals of definition can jump up in size, but, as already noted, not down. See also Figure 5.2.

Example 2.7 On the other hand, for the region

$$\{(x, y) : |xy| < 1\},$$

the curve $t \to (t, 0)$ is a geodesic defined on all of $\mathbb{R}$ that is a limit of geodesics $t \to (t, +\varepsilon)$, $\varepsilon \to 0$, each which is defined only on a finite interval $\left(-\frac{1}{\varepsilon}, +\frac{1}{\varepsilon}\right)$. Note that as required, the endpoints of these intervals go to infinity (in both directions). See also Figure 5.3.

The reader should think through these examples and those in the exercises very carefully, since geodesic behavior is a fundamental topic in all that follows.

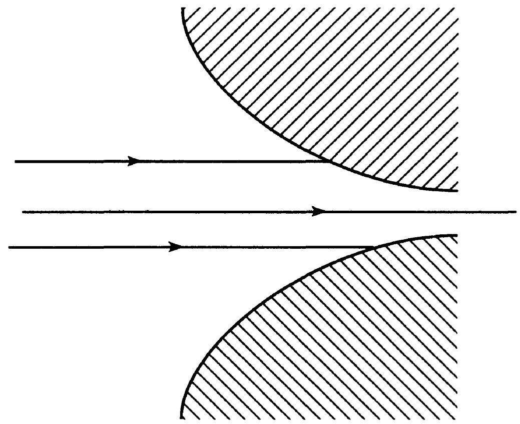


FIGURE 5.3.

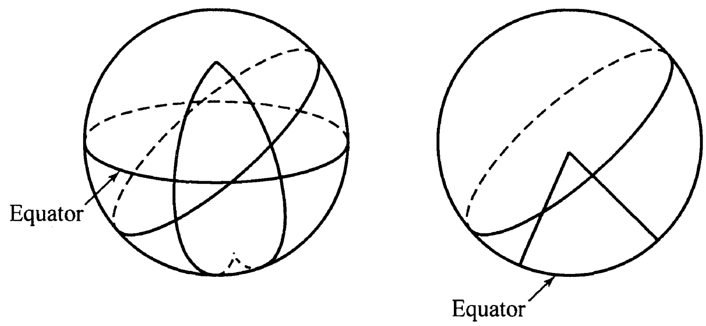


FIGURE 5.4.

Example 2.8 The spheres $(S^n(r), \text{can}) = S^n_{r^{-2}}$ we think of as being in $\mathbb{R}^{n+1}$. Then the connection on $S^n(r)$ is just the connection of $\mathbb{R}^{n+1}$ projected onto $S^n(r)$. The acceleration of a curve $\gamma : I \to S^n(r)$ is therefore the Euclidean acceleration projected onto $S^n(r)$. Thus γ is a geodesic iff $\ddot{\gamma}$ is normal to $S^n(r)$. This means that $\ddot{\gamma}$ and γ should be proportional as vectors in $\mathbb{R}^{n+1}$. Great circles $\gamma(t) = a\cos(\alpha t) + b\sin(\alpha t)$, where $a, b \in \mathbb{R}^{n+1}$, $|a| = |b| = r$, and $a \perp b$, clearly have this property. Furthermore, since $\gamma(0) = a \in S^n(r)$ and $\dot{\gamma}(0) = \alpha b \in T_a S^n(r)$, we see that we have a geodesic for each initial value problem.

We can easily picture great circles on spheres as depicted in Figure 5.4. Still, it is convenient to have a different way of understanding this. For this we project the sphere orthogonally onto the plane containing the equator. Thus the north and south poles are mapped to the origin. As all geodesics are great circles, they must project down to ellipses that have the origin as center and whose greater axis has length r. Of course, this simply describes exactly the way in which we draw three-dimensional pictures on paper.

Example 2.9 We think of $S^n_{-r^{-2}}$ as the imaginary hypersurface in Minkowski space $\mathbb{R}^{1,n}$. The connection in $\mathbb{R}^{1,n}$ is the same as the Euclidean connection, so

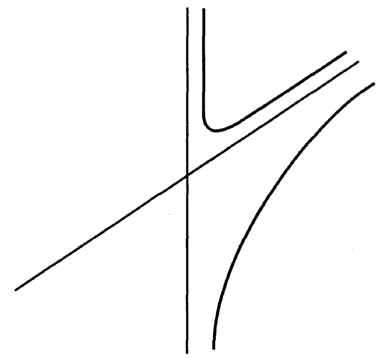


FIGURE 5.5.

the acceleration concept doesn't change except we have to find the Minkowski projection onto the hypersurface. By analogy with the sphere, one might guess that the hyperbolae $\gamma(t) = a\cosh(\alpha t) + b\sinh(\alpha t),\ a, b \in \mathbb{R}^{1,n},\ |a|^2 = -r^2,\ |b|^2 = r^2$, and $a \perp b$ all in the Minkowski sense, are our geodesics. And indeed this is true.

This time the geodesics are hyperbolae. Drawing several of them on the space itself as seen in Minkowski space is not so easy. However, as with the sphere we can resort to the trick of projecting hyperbolic space onto the plane containing the last n coordinates. The geodesics there can then be seen to be hyperbolae whose asymptotes are straight lines through the origin. See also Figure 5.5.

Example 2.10 On a Lie group G with a left-invariant metric one might suspect that the geodesics are the integral curves for the left-invariant vector fields. This in turn is equivalent to the assertion that $\nabla_X X \equiv 0$ for all left-invariant vector fields. On the Berger spheres this is for instance the case, but our Lie group model for the upper half plane does not satisfy this. In general, one can show that bi-invariant metrics (left and right invariant) have this property. Furthermore, all compact Lie groups admit bi-invariant metrics.

5.3 The Metric Structure of a Riemannian Manifold

The positive definite inner product structures on the tangent space of a Riemannian manifold automatically give rise to a concept of lengths of tangent vectors. From this one can obtain an idea of the length of a curve as the integral of the length of its tangent vector. This is a direct extension of the usual calculus concept of the length of curves in Euclidean space. Indeed, the definition of Riemannian manifolds is motivated from the beginning by lengths of curves. The situation is turned around a bit from that of $\mathbb{R}^n$, though: On Euclidean spaces, we have in advance a concept of distance between points. Thus, the definition of lengths of curves is justified by the fact that the length of a curve should be approximated by sums of distances for a fine subdivision (e.g., a fine polygonal approximation).

For Riemannian manifolds, there is no immediate idea of distance between points. Instead, we have a natural idea of (tangent) vector length, hence curve length, and we shall use the length-of-curve idea to define distance between points. The goal of this section is to carry out these constructions in detail.

First, recall that a mapping $\gamma : [a, b] \to M$ defined on a closed interval is said to be C^∞ if there is an open interval (c, d) with $[a, b] \subset (c, d)$ and a C^∞ curve $\hat{\gamma} : (c, d) \to M$ such that $\gamma = \hat{\gamma}[a, b]$. A mapping $\gamma : [a, b] \to M$ is a piecewise C^∞ curve if γ is continuous and if there is a partition $a = a_1 < a_2 < \ldots < a_k = b$ of $[a, b]$ such that $\gamma \mid [a_i, a_{i+1}]$ is C^∞ for $i = 1, \ldots, k-1$.

Suppose now $\gamma : [a, b] \to M$ is a piecewise C^∞ curve in a Riemannian manifold. Then the *length* $\ell(\gamma)$ is defined as follows:

$$\ell(\gamma) = \int_a^b |\dot{\gamma}(t)|\, dt = \int_a^b \sqrt{g\,(\dot{\gamma}(t), \dot{\gamma}(t))}dt.$$

It is clear from the definition of piecewise C^∞ that the function $t \to |\dot{\gamma}(t)|$ is integrable in the Riemann integral sense, so $\ell(\gamma)$ is a well-defined finite, nonnegative number. It is also easy to show that $\ell(\gamma)$ is invariant under reparametrization. A piecewise C^∞ curve $\gamma : [a, b] \to M$ is said to be parametrized by arc length if $\ell(\gamma|_{[a,\lambda]}) = \lambda - a$ for all $\lambda \in [a, b]$, or equivalently, if $|\dot{\gamma}(t)| = 1$ at all smooth points $t \in [a, b]$. A piecewise C^∞ curve $\gamma : [a, b] \to M$ with $|\dot{\gamma}(t)| > 0$ for all $t \in (a_i, a_{i+1})$, $i = 1, \ldots, k-1$, can be reparametrized by arc length without changing the length of the curve, i.e., the function

$$\varphi(s) = \int_a^s |\dot{\gamma}(t)|\, dt$$

is strictly increasing on $[a, b]$, and the curve $\gamma \circ \varphi^{-1} : [0, \ell(\gamma)] \to M$ is piecewise C^∞ and has tangent vectors of unit length at all its smooth points, as you can check by the chain rule. These considerations show that geometrically, we can concentrate on arc-length parametrized curves when it is convenient to do so, without any real loss of generality.

We are now ready to introduce the idea of distance between points. First, for each pair of points $p, q \in M$ we define the path space

$$\Omega(p, q) = \{\gamma : [0, 1] \to M : \gamma \ \text{ is piecewise } \ C^\infty \ \text{ and } \ \gamma(0) = p, \gamma(1) = q\}.$$

We can then define the distance $d(p, q)$ between points $p, q \in M$ as

$$d(p, q) = \inf\{\ell(\gamma) : \gamma \in \Omega(p, q)\}.$$

It follows immediately from this condition that $d(p, q) = d(q, p)$ and $d(p, q) \leq d(p, r) + d(r, q)$. We leave it to the reader to verify that $d(p, q)$ can be zero only when $p = q$. Thus, $d(\ ,\ )$ satisfies all the properties of a metric.

As for metric spaces, we have various metric balls defined via the metric

$$B\,(p, r) = \{x \in M : d\,(p, x) < r\}\,,$$
$$\bar{B}\,(p, r) = D\,(p, r) = \{x \in M : d\,(p, x) \leq r\}\,.$$

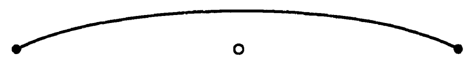


FIGURE 5.6.

More generally, we can define the distance between subsets $A, B \subset M$ as

$$d(A, B) = \inf\{d(p, q) : p \in A, q \in B\}.$$

With this we then have

$$B(A, r) = \{x \in M : d(A, x) < r\},$$
$$\bar{B}(A, r) = D(A, r) = \{x \in M : d(A, x) \le r\}.$$

The infimum of curve lengths in the definition of $d(p, q)$ can fail to be realized. This is illustrated, for instance, by the "punctured plane" $\mathbb{R}^2 - \{(0, 0)\}$ with the usual Riemannian metric of $\mathbb{R}^2$ restricted to $\mathbb{R}^2 - \{(0, 0)\}$. The distance $d((-1, 0), (1, 0)) = 2$, but this distance is not realized by any curve, since every curve of length 2 in $\mathbb{R}^2$ from $(-1, 0)$ to $(1, 0)$ passes through $(0, 0)$ (see Figure 5.6). In a sense that we shall explore later, $\mathbb{R}^2 - \{(0, 0)\}$ is incomplete. For the moment, we introduce some terminology for the cases where the infimum $d(p, q)$ is realized.

A curve $\sigma \in \Omega(p, q)$ is a *segment* if $\ell(\sigma) = d(p, q)$ and σ is parametrized proportional to arc length, i.e., $|\dot{\sigma}|$ is constant.

Example 3.1 In a Euclidean space $\mathbb{R}^n$, segments according to this definition are straight line segments parametrized proportional to arc length, i.e. curves of the form $t \to p_1 + t \cdot p_2$. In $\mathbb{R}^n$, each pair of points p, q is joined by a unique segment $t \to p + t(q - p)$.

Example 3.2 In $S^2(1) = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$ segments are portions of great circles with length $\le \pi$. (We assume for the moment some basic observations of spherical geometry: these will arise in detail later as special cases of more general results.) Every two points are joined by a segment, but there may be more than one segment joining a given pair if the pair are far enough apart, i.e., each pair of antipodal points is joined by infinitely many distinct segments.

Example 3.3 In $\mathbb{R}^2 - \{(0, 0)\}$, as already noted, not every pair of points is joined by a segment.

Later we shall show that segments are always geodesics. Moreover, geodesics are segments if they are short enough; precisely, if γ is a geodesic defined on an open interval containing 0, then $\gamma|[0, \varepsilon]$ is a segment for all sufficiently small $\varepsilon > 0$. Furthermore, we shall show that each pair of points in a Riemannian manifold can be joined by at least one segment provided that the Riemannian manifold is complete as a metric space in the metric just defined. This result explains what is "wrong" with the punctured plane. It also explains why spheres have to have segments between each pair of points: compact spaces are always complete in any metric compatible with the (compact) topology.

Some work needs to be done before we can prove these general statements. To start with, let us dispose of the question of compatibility of topologies:

Theorem 3.4 *The metric topology obtained from the distance $d(,)$ on a Riemannian manifold is the same as the manifold topology.*

Proof. Fix $p \in M$. We need to show that each open ball $B(p, \varepsilon)$, $\varepsilon > 0$, contains a neighborhood of p in the manifold topology, and that each open set in the manifold topology that contains p contains some $B(p, \varepsilon)$, $\varepsilon > 0$. To check these points, choose a coordinate chart $\varphi : U \to \mathbb{R}^n$, $p \in U$. Choose also $\rho > 0$ such that the Euclidean closed ball $\overline{B}(\varphi(p), \rho) = \{\varphi(p)+v : v \in \mathbb{R}^n,\ |v| \leq \rho\} \subset \varphi(U)$. Since φ is a homeomorphism of U onto $\varphi(U)$, the set $V = \varphi^{-1}(B(\varphi(p), \rho))$ is open in M with compact closure $\overline{V} = \varphi^{-1}\overline{B}(\varphi(p), \rho))$. Thus the set $W = \{v \in T_xM : x \in \overline{V}, |v| = 1\}$ is a compact subset of TM. Hence, we can find r and R with $0 < r < R < +\infty$ such that

$$r \leq |D\varphi(v)| \leq R$$

for all $v \in W$. By linearity,

$$r \cdot |v| \leq |D\varphi(v)| \leq R \cdot |v|$$

for all $v \in T_xM$, $x \in \overline{V}$. (Here $|v|$ is the norm in the Riemannian metric of M, $|D\varphi(v)|$ is the Euclidean norm.) Thus, for each piecewise smooth curve $\gamma : [a, b] \to \overline{V}$ we have

$$r\ell(\gamma) \leq \ell(\varphi \circ \gamma) \leq R\ell(\gamma).$$

Set $d'(p, x) = |\varphi(p) - \varphi(x)|$, $x \in \overline{V}$, i.e., the Euclidean distance from $\varphi(p)$ to $\varphi(x)$. Fix $q \in V$, and suppose $\gamma \in \Omega(p, q)$, $\gamma : [a, b] \to M$. If $\gamma([a, b]) \subset \overline{V}$, then

$$d'(p, q) \leq \ell(\varphi \circ \gamma) \leq R\ell(\gamma).$$

If $\gamma([a, b])$ doesn't lie entirely in $\bar{V}$, define $t = \max\{s : \gamma(\lambda) \subset \overline{V}, \text{for all } \lambda \leq s\}$. Then $t < b$. Also,

$$|\varphi(p) - \varphi(\gamma(t))| = \rho$$

and

$$d'(p, q) \leq \rho = \ell((\varphi \cdot \gamma)|_{[a,t]}) \leq R\ell(\gamma|_{[a,t]}) \leq R\ell(\gamma).$$

So in either case, $d'(p, q) \leq R\ell(\gamma)$. On the other hand, the curve $\gamma(t) = \varphi^{-1}(\varphi(p) + t(\varphi(q) - \varphi(p)))$ has image in V and so has length $\ell(\gamma)$ satisfying

$$r\ell(\gamma) \leq \ell(\varphi \circ \gamma) = |\varphi(q) - \varphi(p)| = d'(p, q).$$

Since $d(p,q) \le \ell(\gamma)$, $rd(p,q) \le r\ell(\gamma) \le d'(p,q)$, we obtain

$$r \cdot d(p,q) \le d'(p,q) \le R \cdot d(p,q).$$

From these inequalities and the fact that d' is the Euclidean distance on V, the comparability of the two kinds of neighborhoods of p follows. □

This proof shows in effect also that $d(p,q) \ne 0$ if $p \ne q$. This theorem also yields a corollary about completeness.

Corollary 3.5 *If M is a compact manifold and if g is a Riemannian metric on M, then (M,d) is a complete metric space, where d is the Riemannian distance function determined by g.*

Let us relate these new concepts to our old distance functions:

Lemma 3.6 *Supposing $f : U \to \mathbb{R}$ is a smooth distance function and $U \subset (M,g)$ is open, then the integral curves for ∇f are segments on (U,g).*

Proof. Fix $p, q \in U$ and let $\gamma(t) : [0,b] \to U$ be a curve from p to q. Then

$$\begin{aligned}
\ell(\gamma) &= \int_0^b |\dot\gamma| dt \\
&= \int_0^b |(\nabla f)\circ\gamma| \cdot |\dot\gamma| dt \text{since } |\nabla f| = 1 \\
&\ge \int_0^b |D(f\circ\gamma)| dt \\
&= f(q) - f(p).
\end{aligned}$$

Here the last inequality is the Cauchy-Schwartz inequality. In particular, if $\dot\gamma = \nabla f \circ \gamma$, equality holds in the Cauchy-Schwartz inequality and we have shown that

$$d(p,q) \le \ell(\gamma) = f(q) - f(p).$$

On the other hand: $f(q) - f(p) \le \ell(\gamma)$ for any γ, and $d(p,q) = \inf \ell(\gamma)$. Thus integral curves must be segments. Notice that we only considered curves in U, and therefore only established the result for (U,g) and not (M,g). □

Example 3.7 Let $M = S^1 \times \mathbb{R}$ and $U = S^1 - \{e^{i0}\} \times \mathbb{R}$. On U let $f(\theta, r) = \theta$, $\theta \in (0, 2\pi)$. The previous discussion shows that any curve $\gamma(t) = (e^{it}, r_0)$, $t \in I$, where I does not contain 0 is a segment in U. If, however, the length of I is $> \pi$, then such curves can clearly not be segments in M.

The *functional distance* d_F between points in a manifold is defined as

$$d_F(p,q) = \sup\{|f(p) - f(q)| : f : M \to \mathbb{R} \text{ has } |\nabla f| \le 1 \text{ on } M\}.$$

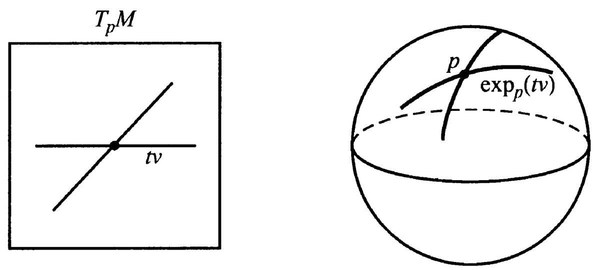


FIGURE 5.7.

This new distance is always smaller than the previously defined distance, and one can show as before that it generates the standard manifold topology. When we have established the existence of smooth distance functions, it will become clear that the two distances are equal if p and q are sufficiently close to each other.

5.4 The Exponential Map

For a tangent vector $v \in T_pM$, let γ_v be the unique geodesic with $\gamma(0) = p$ and $\dot{\gamma}(0) = v$, and let $[0, \ell_v)$ be the nonnegative part of the maximal interval containing 0 on which γ is defined. Notice that $\gamma_{\alpha v}(t) = \gamma_v(\alpha t)$ for all $\alpha > 0$ and $t < \ell_{\alpha v}$. In particular, $\ell_{\alpha v} = \alpha^{-1}\ell_v$. Let $O_p \subset T_pM$ be the set of vectors v such that $1 < \ell_v$, so that $\gamma_v(t)$ is defined on $[0, 1]$. Then define the *exponential map*[1] at p by

$$\exp_p(v) = \gamma_v(1),\ v \in O_p.$$

So $\exp_p$ maps O_p onto a subset of M. In Figure 5.7 we have shown how radial lines in the tangent space are mapped to radial geodesics in M via the exponential map.

The individual $\exp_p$ maps can of course be combined to form a map $\exp : \bigcup O_p \to M$ by setting $\exp | O_p = \exp_p$. This map exp is called just the *exponential map*.

The standard theory of ordinary differential equations that we have already discussed tells us that the set $O = \bigcup O_p$ is open in TM and that $\exp : O \to M$ is smooth (C^∞). Each of the sets O_p is thus open, and $\exp_p : O_p \to M$ is also C^∞ smooth. It is an important fact that $\exp_p$ is in fact a local diffeomorphism around $0 \in T_pM$. The details of this are given in the following:

Proposition 4.1 *If $p \in M$ and 0 denotes the zero vector in T_pM, then*

---

[1]The name "exponential map" comes from the fact that this mapping coincides with exponentiation of matrices when certain special metrics are assigned to Lie groups of matrices. This is, from our viewpoint at the moment, only of historical interest. See also the exercises to Chapters 1 and 5.

(1) $D\exp_p : T_0(T_pM) \to T_pM$ *is nonsingular. Consequently, there is an open neighborhood* U_p *of* 0 *in* T_pM *such that* $\exp_p : U_p \to M$ *is a diffeomorphism of* U_p *onto* $\exp_p(U_p)$*, which is an open subset of* M*, i.e.,* $\exp_p$ *is a local diffeomorphism.*

(2) *Define* $E : O \to M \times M$ *by* $E(v) = (\pi(v), \exp v)$*, where* $\pi(v)$ *is the unique point of* M *such that* $v \in T_{\pi(v)}M$*, i.e. the base point of* v*. Then for each* $p \in M$ *and with it the zero vector,* $0_p \in T_pM$*,* $DE : T_{(p,0_p)}(TM) \supset T_{(p,0_p)}O \to T_{(p,p)}(M\times M)$ *is nonsingular. Consequently,* E *is a diffeomorphism of a neighborhood of the zero section of* TM *onto an open neighborhood of the diagonal in* $M \times M$.

Proof. The proofs of both statements are an immediate application of the inverse function theorem, once a crucial observation has been made. This observation is as follows: Let $I_0 : T_pM \to T_0T_pM$ be the canonical isomorphism, i.e., $I_0(v) = \frac{d}{dt}(tv)|_{t=0}$. Now we recall that if $v \in O_p$, then $\gamma_v(t) = \gamma_{tv}(1)$ for all $t \in [0, 1]$. Thus,

$$\begin{aligned} D\exp_p(I_0(v)) &= \frac{d}{dt}\exp_p(tv)|_{t=0} \\ &= \frac{d}{dt}\gamma_{tv}(1)|_{t=0} \\ &= \frac{d}{dt}\gamma_v(t)|_{t=0} \\ &= \dot{\gamma}_v(0)|_{t=0} \\ &= v. \end{aligned}$$

So $D\exp_p \circ I_0$ = the identity map from T_pM to T_pM. In particular, $D\exp_p$ is nonsingular.

This looks quite formal, almost like sleight of hand. One should think through it carefully. Underlying the reasoning here is the "homogeneity property" $\gamma_v(t) = \gamma_{tv}(1)$. Given that, it is natural to think of $\exp_p(v)$ in a sort of polar coordinate representation: one goes from p in the "direction" $v/|v|$, supposing $v \neq 0$, and goes a distance of length $|v|$. This gives the point $\exp_p(v)$, since

$$\gamma_{v/|v|}(|v|) = \gamma_v(1).$$

Note, however, that this polar idea makes obscure whether the resulting map is smooth at 0, whereas setting $\exp_p v = \gamma_v(1)$ makes 0 no problem.

Once it is clear that $D\exp_p$ is nonsingular, and indeed is the identity up to the canonical identification of $T_0(T_pM)$ with T_pM, then the second statement of (1) follows from the inverse function theorem.

The proof of (2) is again an exercise in unraveling tangent spaces and identifications. The tangent space $T_{(p,p)}(M \times M)$ is naturally identified with $T_pM \times T_pM$. The tangent space $T_{(p,0)}(TM)$ is naturally identified to $T_pM \times T_{0_p}(T_pM) \simeq T_pM \times T_pM$.

Now, the linear map $DE : T_pM \times T_pM \to T_pM \times T_pM$ acts as follows at $(p, 0_p)$: it is the identity on the first factor to the first factor, and identically 0 from the second factor to the first, since the first factor image is the base point, which by definition does not vary with the second domain factor. The second to second map is $D\exp_p$, i.e., the "identity" (up to isomorphism). Schematically, the map DE looks like

$$\begin{pmatrix} I & 0 \\ ? & I \end{pmatrix}$$

This map is clearly nonsingular.

The notation here is a bit confusing, and it is worthwhile to pursue the matter intuitively. What is happening is that E takes (p, v), $v \in T_pM$, to $(p, \exp_p v)$. So varying p in the domain varies p in the image but does something unpredictable to $\exp_p v$. On the other hand, varying v in the domain does nothing to p and changes $\exp_p v$ only. When we vary v at 0, the differential of that is what we have already observed to be the identity map, up to canonical identifications. So we get the schematic matrix picture already given.

Now, the inverse function theorem gives (local) diffeomorphisms via E of neighborhoods of points in TM of the form $(p, 0_p)$ onto neighborhoods of (p, p) in $M \times M$. Since the map E is the identity on the first factor, it is easy to see that these local diffeomorphisms fit together to give a diffeomorphism of a neighborhood of the zero section in TM onto a neighborhood of the diagonal in $M \times M$. □

All this formalism yields some results with geometric meaning. First, we get a coordinate system around p; namely, by identifying T_pM with $\mathbb{R}^n$ via an isomorphism, we get

$$\exp_p^{-1} : \exp_p(U_p) \to T_pM \simeq \mathbb{R}^n .$$

Such a coordinate system is called *normal (exponential) coordinates* at p: they are unique up to how we choose to identify T_pM with $\mathbb{R}^n$. Requiring this identification to be a linear isometry gives uniqueness up to an orthogonal transformation of $\mathbb{R}^n$.

The second item of geometric interest is the following idea: Thinking about S^2 and great circles (which we know are geodesics), it is clear that we cannot say that two points that are close together are joined by a unique geodesic. On S^2 there will be a short geodesic connection, but there will be other, long ones, too. What might be hoped is that points that are close together would have only one geodesic connecting them that was short, and that all the others (if any) were a lot longer. This is exactly what (2) in the proposition says! As long as we keep q_1 and q_2 near p, there is only one way to go from q_1 to q_2 via a geodesic that isn't very long, i.e., has the form $\exp_{q_1} tv$, $v \in T_{q_1}M$, with $|v|$ near 0. This will be made more useful and clear in the next section, where we show that such short geodesics in fact are segments.

Suppose N is an embedded submanifold of M. Let

$$TN^\perp = \{v \in T_pM : p \in N, v \in (T_pN)^\perp \subset T_pM\} .$$

Here $(T_pN)^\perp =$ the orthogonal complement of T_pN in T_pM. So for each $p \in N$, $T_pM = T_pN \oplus (T_pN)^\perp$, an orthogonal direct sum. Define the *normal exponential* *map* $\exp^\perp$ by restricting exp to $O \cap TN^\perp$ so $\exp^\perp : O \cap TN^\perp \to M$. As in the previous proposition, you can show that $D\exp^\perp$ is nonsingular at 0_p, $p \in N$. Then it follows that there is an open neighborhood U of the zero section in $TN^\perp$ on which $\exp^\perp$ is a diffeomorphism onto its image in M. Such an image $\exp^\perp(U)$ is called a *tubular neighborhood* of N in M, because intuitively it looks like a solid tube around N, containing N.

5.5 Why Short Geodesics Are Segments

In the last section, we saw that points that are close together on a Riemannian manifold are connected by a short geodesic, and by exactly one short geodesic in fact. But so far, we don't have any real evidence that such short geodesics are segments (in the sense already defined, that their length equals the distance between their endpoints). In this section we shall prove that short geodesics are segments. Incidentally, several different ways of saying that a curve is a segment are in common use: "minimal geodesic," "minimizing curve," "minimizing geodesic," and even "minimizing geodesic segment."

The precise result we want to prove in this section is this:

Theorem 5.1 *Suppose M is a Riemannian manifold, $p \in M$, and $\varepsilon > 0$ is such that $\exp_p$ is defined on $\{v \in T_pM : |v| < \varepsilon\}$ and is a diffeomorphism of that set onto its image in M. Then*

(1) *For each $v \in T_pM$ with $|v| < \varepsilon$, the geodesic $\gamma_v : [0,1] \to M$ defined by*

$$\gamma_v(t) = \exp_p(tv)$$

is the unique segment from p to $\exp_p v$.

(2) *Furthermore, if a piecewise smooth curve $c : [0,1] \to M$ is a segment from p to $\exp_p v$, i.e., if $c(0) = p$ and $c(1) = \exp_p(v)$, then c is a reparametrization of γ_v. In other words, for some piecewise smooth, monotone nondecreasing function $h : [0,1] \to [0,1]$, $c = \gamma_v \circ h$.*

On $U = \exp_p(B(0,\varepsilon) \subset T_pM)$ we have the function $f(x) = |\exp_p^{-1}(x)|$. That is, f is simply the Euclidean distance function from the origin on $B(0,\varepsilon) \subset T_pM$ in exponential coordinates. This Euclidean distance is usually denoted by $r(v) = |v|$. We know that $\nabla r = \partial_r = \frac{1}{r}(x^i\partial_i)$ in Cartesian coordinates on T_pM. The goal here is to establish:

Lemma 5.2 (The Gauss Lemma) *$\nabla f = \partial_r$, where $\partial_r = D\exp_p(\partial_r)$.*

Let us see how this implies the theorem. First observe that in $B(0,r)$ the integral curves for ∂_r are the line segments $\gamma(s) = s \cdot \frac{v}{|v|}$ of unit speed. The integral curves for ∂_r on U are therefore the geodesics $\gamma(s) = \exp\left(s \cdot \frac{v}{|v|}\right)$. These geodesics clearly also have unit speed. Thus the Lemma implies that f is a distance function on U. In particular we then know that among curves from p to $q = \exp(x)$ in

U, the geodesic from p to q is the shortest curve. To see that this geodesic is a segment in M, we must, however, in addition see that there is no shortcut from p to q that leads us outside U. Supposing we had such a curve $\gamma(t) : [0, b]$ with $|\dot{\gamma}| \equiv 1$, define $t_0 < b$ to be the first value for which $\gamma(t_0) \notin U$. Then

$$\begin{aligned} t_0 &= \ell\left(\gamma|_{[0,t_0]}\right) \\ &= \int_0^{t_0} |\dot{\gamma}|\, dt \\ &= \int_0^{t_0} |\nabla f| \cdot |\dot{\gamma}|\, dt \\ &\geq \int_0^{t_0} |Df \circ \dot{\gamma}|\, dt \\ &= \lim_{t \to t_0} f(\gamma(t)) - f(\gamma(0)) \\ &= \varepsilon, \end{aligned}$$

since $f(p) = 0$ and the values of f converge to ε as we approach ∂U. Now, we already have a curve from p to q of length $< \varepsilon$, so γ cannot be a shortcut from p to q.

For the second part of the theorem we first note that the curve c must lie in U. We then obtain that

$$\begin{aligned} \ell(c) &= \int_0^1 |\dot{c}(t)|\, dt \\ &= \int_0^1 |\nabla f|\, |\dot{c}(t)|\, dt \\ &\geq \int_0^1 |Df \circ \dot{c}|\, dt \\ &= f(c(1)) - f(c(0)). \end{aligned}$$

As c was supposed to have length $d(p, c(1)) = f(c(1)) - f(c(0))$, it must follow that $\dot{c}$ and ∇f are proportional everywhere. This proves the assertion. Note that the discontinuity points of $\dot{c}$ do not affect this analysis as there are only finitely many of them.

Now to the proof of the Gauss Lemma: We assume that Cartesian coordinates have been chosen on T_pM, via choosing an orthonormal basis, and then transferred to U via $\exp_p$. They are as usual denoted by $(x^1, \ldots, x^n)$ with coordinate vector fields $\partial_1, \ldots, \partial_n$. Then we have $\partial_r = \frac{1}{r} x^i \partial_i$, $r^2 = (x^1)^2 + \cdots + (x^n)^2$. To show that this is the gradient for $f(x) = \sqrt{(x^1)^2 + \cdots + (x^n)}$ on (M, g), we must prove that $df(v) = g(\partial_r, v)$. We already have that $df = \frac{1}{r}(x^1 dx^1 + \cdots + x^n dx^n)$, but we have no knowledge of g, since it is just some abstract metric. First observe that the equality needs to be proven only for some convenient basis of vectors at each point.

Guided by the fact that we seem to be working with polar coordinates implicitly, let us consider the "frame" $\{\partial_r, X_{ij}\}$ where $X_{ij} = -x^i \partial_j + x^j \partial_i$, $i, j = 2, \ldots, n$, and $i < j$. Here X_{ij} is analogous to ∂_θ. Notice that there are too many vectors for them to be linearly independent. However, they always span the tangent space.

The important feature about this frame is that $[\partial_r, X_{ij}] = 0$ for $i, j = 1, \ldots, n$. Now notice that $df(\partial_r) = 1$, while $df(X_{ij}) = 0$, so we must show $g(\partial_r, \partial_r) = 1$ and $g(\partial_r, X_{ij}) = 0$. Since the integral curves to ∂_r are unit speed geodesics, we already know that $g(\partial_r, \partial_r) \equiv 1$, so we can concentrate on $g(\partial_r, X_{ij})$. Now compute

$$\begin{aligned}
\partial_r g(\partial_r, X_{ij}) &= g(\nabla_{\partial_r}\partial_r, X_{ij}) + g(\partial_r, \nabla_{\partial_r} X_{ij}) \\
&= 0 + g(\partial_r, \nabla_{\partial_r} X_{ij}), \text{ since integral curves for } \partial_r \text{ are geodesics,} \\
&= -g(\partial_r, \nabla_{X_{ij}}\partial_r), \text{ since } [\partial_r, X_{ij}] = 0, \\
&= -\frac{1}{2} D_{X_{ij}} g(\partial_r, \partial_r) = 0, \text{ since } g(\partial_r, \partial_r) \equiv 1.
\end{aligned}$$

Thus, $g(\partial_r, X_{ij}) = 0$ if we can show that $|g(\partial_r, X_{ij})| \le |\partial_r||X_{ij}| = |X_{ij}| \to 0$ as $r \to 0$. But we know that $X_{ij} = -x^i\partial_j + x^j\partial_i$. Here ∂_i, ∂_j are bounded on all of U by continuity of $D\exp_p$ and $x^i, x^j \to 0$ as $r \to 0$.

There is an equivalent statement of the Gauss Lemma that asserts that $\exp_p : T_pM \to M$ is a radial isometry $g(D\exp_p(\partial_r), D\exp_p(v)) = g_p(\partial_r, v)$ on T_pM. A careful translation process of the previous proof shows that this is exactly what we have proved.

Corollary 5.3 *If $x \in M$ and $\varepsilon > 0$ is such that $\exp_x : \{v \in T_xM : |v| < \varepsilon\} \to M$ is defined and is a diffeomorphism onto its image, then for each $\lambda \le \varepsilon$ such that $B(x, \lambda) \subset \exp_p(B(0, \varepsilon))$,*

$$\exp_x(B(0, \lambda)) = B(x, \lambda),$$

and moreover,

$$\exp_x(\bar{B}(0, \lambda)) = \bar{B}(x, \lambda).$$

The proof is straightforward and is left as an exercise for the reader. Note, however, that such $\lambda > 0$ exist by the equivalence of the metric topology and manifold topology. Theorem 5.1 has another important corollary, but first we need

Proposition 5.4 *Suppose $\sigma : [a, b] \to M$ is a segment. Then it must be true that $\sigma|[\alpha, \beta]$ is a segment for each $[\alpha, \beta] \subset [a, b]$.*

Proof. Otherwise, we could alter σ by replacing $\sigma|[\alpha, \beta]$ with a shorter (piecewise smooth) curve from $\sigma(\alpha)$ to $\sigma(\beta)$. The result would be a piecewise smooth curve from $\sigma(a)$ to $\sigma(b)$ that was shorter than σ. □

Theorem 5.1 can now be used to show

Corollary 5.5 *Suppose σ is a piecewise smooth segment. Then σ is a smooth geodesic.*

Proof. The reasoning to see this is slightly delicate. First, we have to recall that given $x \in M$, there is a neighborhood U of x and an $\varepsilon > 0$ such that for each $y \in U$, $\exp_y : \{v \in T_yM : |v| < \varepsilon\} \to M$ is defined and is a diffeomorphism onto its image. With this in mind let us consider the segment $\sigma : [a, b] \to M$ around a value $t_0 \in (a, b)$. Choose ε and U as in the first part of the paragraph for $x = \sigma(t_0)$. Then choose $\delta > 0$ so small that

$$(t_0 - \delta, t_0 + \delta) \subset (a, b),$$
$$\sigma([t_0 - \delta, t_0 + \delta]) \subset U,$$
$$\ell(\sigma|[t_0 - \delta, t_0 + \delta]) < \varepsilon.$$

With these choices, it follows from Theorem 5.1 that $\sigma|[t_0 - \delta, t_0 + \delta]$ is a (C^∞) geodesic; here we need only note that $\sigma|[t_0 - \delta, t_0 + \delta]$ is a segment (since σ is) and that $\sigma|[t_0 - \delta, t_0 + \delta]$ is short enough that Theorem 5.1 indeed applies. So now we know that in a neighborhood of each $t_0 \in (a, b)$ the segment σ is a C^∞ geodesic. Thus, $\sigma|(a, b)$ is a geodesic. The endpoint situation is established by similar (and easier) reasoning to complete the proof that $\sigma|[a, b]$ is a (C^∞) geodesic. □

5.6 Local Geometry in Constant Curvature

Let us restate what we have done in this chapter so far. Given $p \in (M, g)$ we found coordinates near p using the exponential map such that the distance function $f(x) = d(p, x)$ to p has the formula

$$f(x) = \sqrt{(x^1)^2 + \cdots + (x^n)^2}$$

in these coordinates. Furthermore we showed that

$$\nabla f = \partial_r = \frac{1}{r} x^i \partial_i = \frac{1}{f(x)} x^i \partial_i$$

and is a unit vector field. This was used to show that the integral curves for ∇f, which are exactly the geodesics emanating from p, are segments near p. Let us now see what happens when we use *polar coordinates* $r = f$ and $\theta = \left(\theta^2, \dots, \theta^n\right) \in S^{n-1}$ on a neighborhood of p. Actually we shall throughout use frames rather than coordinates. Thus, we pick an orthonormal frame E_α, $\alpha = 2, \dots, n$, on the unit sphere in T_pM and define $E_1 = \partial_r$. Now as in Chapter 3 we extend these vector fields radially to all of T_pM. If we denote the dual coframe by θ^α, $\alpha = 2, \dots, n$, then the Euclidean metric on T_pM can be written as

$$dr^2 + r^2 \sum_{\alpha=2}^{n} \left(\theta^\alpha\right)^2.$$

When we pull back the metric g on M to the tangent space via the exponential map, we can write it as

$$g = dr^2 + g_{\alpha\beta}\theta^\alpha\theta^\beta,$$
$$g_{\alpha\beta} = g(E_\alpha, E_\beta).$$

An important observation to make here is that the choice of polar coordinates (or frames) we make happens in Euclidean space and is therefore independent of the metric g. In Cartesian coordinates one can easily show that the metric coefficients satisfy

$$g_{ij} = \delta_{ij} + O\left(r^2\right).$$

Here, the fact that $g_{ij}(p) = \delta_{ij}$ simply follows from the fact that the exponential map is an isometry at the origin. To see that the first derivative $\partial_k g_{ij}$ equals 0 at p it clearly suffices to show that $\nabla_{\partial_i}\partial_j$ equals 0 at p. First, note that $\nabla_{\partial_i}\partial_i$ at p is simply the limit of $\nabla_{\partial_r}\partial_r$ along a geodesic emanating from p in the direction of ∂_i. Hence, that term is zero. For $i \neq j$ we use that away from p we have

$$\begin{aligned}
0 &= g\left(\nabla_{\partial_r}\partial_r, \partial_k\right) \\
&= \frac{x^i}{r} \cdot g\left(\nabla_{\partial_i}\partial_r, \partial_k\right) \\
&= \frac{x^i}{r} \cdot \partial_i\left(\frac{x^j}{r}\right) \cdot g\left(\partial_j, \partial_k\right) + \frac{x^i}{r} \cdot \frac{x^j}{r} \cdot g\left(\nabla_{\partial_i}\partial_j, \partial_k\right) \\
&= -\frac{x^i}{r} \cdot \frac{x^i \cdot x^j}{r^3} \cdot g\left(\partial_j, \partial_k\right) + \frac{x^i}{r} \cdot \frac{x^j}{r} \cdot g\left(\nabla_{\partial_i}\partial_j, \partial_k\right).
\end{aligned}$$

Here, the first term in the last line goes to zero as we approach p, thus forcing the last term also to converge to zero at p. However, this can happen only if in fact $g\left(\nabla_{\partial_i}\partial_j, \partial_k\right) \to 0$ as x gets close to p. This establishes our claim.

The equation

$$\partial_r\left(g_{ij}\right) = 2\left(S_i^k\right)\left(g_{kj}\right)$$

now implies that as $r \to 0$ the shape operator goes to $+\infty$. This can also be seen by computing the Hessian of $\frac{1}{2}f^2$, and noting that as this function has a critical point at p the Hessian is defined independently of the metric and must therefore be the identity map at p.

Theorem 6.1 (Riemann, 1854) *If a Riemannian n-manifold (M, g) has constant sectional curvature k, then every point in M has a neighborhood that is isometric to an open subset of the space form S_k^n.*

Proof. Using polar coordinates, rather than Cartesian coordiantes, around $p \in M$, we see that the metric and shape operator look like

$$\left(g_{\alpha\beta}(r,\theta)\right) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & r^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r^2 \end{pmatrix} + O\left(r^2\right),$$

$$\left(S_{\alpha\beta}(r,\theta)\right) = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{r} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{r} \end{pmatrix} + O(r).$$

Guided by the fact that the rotationally symmetric model $dr^2 + \operatorname{sn}_k^2(r)\, ds_{n-1}^2$ has constant curvature k, we see that the matrices

$$g_k(r,\theta) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \operatorname{sn}_k^2(r) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \operatorname{sn}_k^2(r) \end{pmatrix},$$

$$S_k(r,\theta) = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & \frac{\operatorname{sn}_k'(r)}{\operatorname{sn}_k(r)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\operatorname{sn}_k'(r)}{\operatorname{sn}_k(r)} \end{pmatrix}$$

satisfy the equations:

$$\partial_r g = 2S \cdot g,$$

$$\partial_r S + S^2 = -\begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & k \end{pmatrix}.$$

As these solutions have the same initial values as $(g_{\alpha\beta})$ and $(S_{\alpha\beta})$ when $r \to 0$, and clearly satisfy the same equations, we can conclude that they must be equal. But then the metric in polar coordinates must look like

$$g = dr^2 + \operatorname{sn}_k^2(r)\, ds_{n-1}^2.$$

This is the same as the metric for the space form S_k^n in polar coordinates. So we have found a local isometry from a neighborhood around $p \in M$ to the desired space form using polar coordinates. □

5.7 Completeness

One of the foundational centerpieces for Riemannian geometry is the Hopf-Rinow theorem. This theorem states that all concepts of completeness are equivalent. This should not be an unexpected result for those who have played around with constant-curvature spaces. For it seems that in these examples, geodesic and metric completeness break down in exactly the same places. As with most foundational theorems, the proof is fairly nasty, for in these matters one should always watch every turn and be careful not to take anything for granted.

Theorem 7.1 (H. Hopf-Rinow, 1931) *The following statements are equivalent:*

(1) *M is geodesically complete, i.e., all geodesics are defined for all time.*

(2) *M is geodesically complete at p, i.e., all geodesics through p are defined for all time.*

(3) *M satisfies the Heine-Borel property, i.e., every closed bounded set is compact.*

(4) *M is metrically complete.*

Proof. (1) $\Rightarrow$ (2), (3) $\Rightarrow$ (4) are trivial.

(4) $\Rightarrow$ (1) Recall that every geodesic $c : [0, \alpha) \to M$ defined on a maximal interval must leave every compact set if $\alpha < \infty$. This violates metric completeness ($c(t_i)$, $t_i \to \alpha$ is a Cauchy sequence).

(2) $\Rightarrow$ (3) Consider $\exp_p : T_pM \to M$. It suffices to show that $\exp_p(\overline{B}(0, r)) = \overline{B}(p, r)$ for all r (note that $\subset$ always holds). Consider $I = \{r : \exp(\overline{B}(0, r)) = \overline{B}(p, r)\}$.

(i) We have already seen that I contains all r close to zero.
(ii) I is closed: If $r_i \in I$ converge to r, then let $q \in \overline{B}(p, r)$ and find $q_i \in \overline{B}(p, r_i)$ converging to q. We can find $v_i \in \overline{B}(0, r_i)$ with $q_i = \exp_p(v_i)$. The v_i will subconverge to some $v \in \overline{B}(0, r)$, and continuity of $\exp_p$ implies that $\exp_p(v) = q$. (You should think about why it is possible to choose the q_i's.)
(iii) If $R \in I$ then $R + \varepsilon \in I$ for all small ε . First, choose a compact set K that contains $\overline{B}(p, R)$ in its interior. Then fix $\varepsilon > 0$ such that all points in K of distance $\leq \varepsilon$ can be joined by a unique geodesic segment. For $q \in \overline{B}(p, R + \varepsilon) - \overline{B}(p, R)$ select for each $\delta > 0$ a curve $\gamma_\delta : [0, 1] \to M$ with $\gamma_\delta(0) = q$, $\gamma_\delta(1) = p$, and $L(\gamma_\delta) \leq d(p, q) + \delta$. Suppose t_δ is the first value such that $\gamma_\delta(t_\delta) \in \partial\overline{B}(p, R)$. If x is an accumulation point for $\gamma_\delta(t_\delta)$, then we must have that $R + d(x, q) = d(p, x) + d(x, q) = d(p, q)$.

Now choose a segment from q to x and a segment from p to x of the form $\exp_p(tv)$; see also Figure 5.8. These two geodesics together form a curve from p to q of length $d(p, q)$; hence, it is a segment. Consequently, it is smooth and by uniqueness of geodesics is the continuation of $\exp_p(tv) : 0 \leq t \leq 1 + \varepsilon/|\dot\gamma|$. This shows that $q \in \exp_p\left(\overline{B}\,(0, R + \varepsilon)\right)$.

Statements i through iii together imply that $I = (0, \infty)$, which is what we wanted to prove. □

From part ii of (2) $\Rightarrow$ (3) we get the following result:

Corollary 7.2 *If M is complete in any of the above ways, then any two points in M can be joined by a segment.*

Corollary 7.3 *Suppose M admits a proper (preimages of compact sets are compact) Lipschitz function $f : M \to \mathbb{R}$. Then M is complete.*

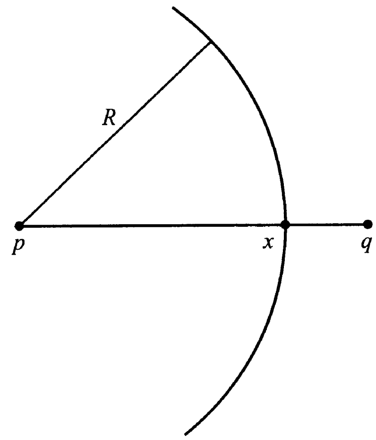


FIGURE 5.8.

Proof. Establish the Heine-Borel property. □

This corollary makes it easy to check completeness for all of our examples. In these examples, the distance function can be extended to a proper continuous function on the entire space.

From now on, all Riemannian manifolds will automatically be assumed to be connected and complete.

5.8 Characterization of Segments

In this section we will try to determine when a geodesic is a segment and then use this to find a maximal domain in T_pM on which the exponential map is an embedding. These issues can be understood through a systematic investigation of when distance functions to points are smooth.

5.8.1 The Segment Domain

Fix $p \in (M, g)$ and let $r = f(x) = d(x, p)$. We know that f is smooth near p and that the integral curves for f are geodesics emanating from p. Since M is complete, these integral curves can be continued indefinitely beyond the places where f is smooth. These geodesics could easily intersect after some time, so they do not generate a flow on M, but just having them at places where f might not be smooth helps us understand why f is not smooth at these places. We know form Chapter 2 that another obstruction to f being smooth is the possibility of conjugate points (we use the notation *conjugate points* for distance functions to a point).

Let us introduce some terminology: The *segment domain* is

$$\operatorname{seg}(p) = \{v \in T_pM : \exp_p(tv) : [0, 1] \to M \text{ is a segment}\}.$$

The Hopf-Rinow Theorem implies that $M = \exp_p(\mathrm{seg}(p))$. We see that $\mathrm{seg}(p)$ is a closed star-shaped subset of T_pM. The interior of $\mathrm{seg}(p)$ is denoted by $\mathrm{seg}^0(p)$ and is characterized as

$$\mathrm{seg}^0 = \{tv : t \in [0, 1),\ v \in \mathrm{seg}(p)\}.$$

This actually tells us that $\exp_p$ is one-to-one on $\mathrm{seg}^0(p)$ and that any $x \in \exp_p(\mathrm{seg}^0(p))$ is joined to p by a unique segment. This is because such x have the property that there is a segment $\sigma : [0, 1) \to M$ with $\sigma(0) = p$, $\sigma(t_0) = x$, $t_0 < 1$. Therefore, if $\hat{\sigma} : [0, t_0] \to M$ is another segment from p to x, we could construct a nonsmooth segment

$$\gamma(s) = \begin{cases} \hat{\sigma}(s), & s \in [0, t_0], \\ \sigma(s), & s \in [t_0, 1], \end{cases}$$

and we know that this is impossible. Let U_p denote $\exp_p(\mathrm{seg}^0(p))$. On this set we have the vector field $\partial_r = D\exp_p(\partial_r)$, which is, we hope, the gradient for $f(x) = d(x, p)$. Furthermore, $f(x) = |\exp_p^{-1}(x)|$. From our earlier observations we know that f would be smooth on U_p with gradient ∂_r if we could show that $\exp_p : \mathrm{seg}^0(p) \to U_p$ is a diffeomorphism. We already know that $\exp_p$ is one-to-one on $\mathrm{seg}^0(p)$, so we need only to show

Lemma 8.1 $\exp_p : \mathrm{seg}^0(p) \to U_p$ *is nonsingular everywhere, or, in other words, has no critical points.*

Proof. The set $\{v \in \mathrm{seg}^0(p) : D\exp_p \text{ is singular at } v\}$ is closed and does not contain $0 \in T_pM$. So we can find $v \in T_pM$ with the property that $\exp_p : B(0, |v|) \cap \mathrm{seg}^0(p) \to B(p, |v|)$ is a diffeomorphism and $D\exp_p$ is singular at $v \in T_pM$. Now use Cartesian coordinates on $B(p, |v|)$ via the exponential map and use the same notation as before for the metric on $B(p, |v|)$. Now observe that if we take a tangent vector $w \in TB(0, |v|) \cap \mathrm{seg}^0(p)$, then we can write it as $w = w^i\partial_i$. The image of w under $D\exp_p$ clearly has the same coordinate representation but its length in M is

$$\left|D\exp_p(w)\right|^2 = g_{ij}w^iw^j.$$

From this formula it follows in particular that $D\exp_p$ is singular iff $\exp_p(v)$ is a conjugate point for f, i.e., the determinant of the Jacobian goes to zero as we approach $\exp_p(v)$. This characterization of course presupposes that f is smooth on a region that approaches $\exp_p(v)$.

Now select $w = w^i\partial_i$ such that $w \perp v$ in $T_vT_pM \approx T_pM$ and $D\exp_p(w) = 0$. Then we know from the Gauss lemma that in M, w is perpendicular to ∂_r. Furthermore, if we switch to polar coordinates and use Greek indices we have (ignoring the spherical coordinate, as it is fixed)

$$\sum_{\alpha,\beta=2}^{n} g_{\alpha\beta}(r)\, w^\alpha w^\beta \to 0 \quad \text{as} \quad r \to |v|.$$

Thus, also,

$$g\left(w^{\alpha}\partial_{\alpha}, w^{\alpha}\partial_{\alpha}\right) = g_{\alpha\beta}(r)\, w^{\alpha} w^{\beta} \to 0 \quad \text{as} \quad r \to |v|\,.$$

But then

$$\log g\left(w^{\alpha}\partial_{\alpha}, w^{\alpha}\partial_{\alpha}\right) \to -\infty \quad \text{as} \quad r \to |v|\,.$$

There must therefore be a sequence of numbers $r_n \to |v|$ such that

$$\partial_r \log g_{r_n}\left(w^{\alpha}\partial_{\alpha}, w^{\alpha}\partial_{\alpha}\right) \to -\infty \quad \text{as} \quad n \to \infty.$$

Now use the identity

$$\begin{aligned}
\partial_r \log g\left(w^{\alpha}\partial_{\alpha}, w^{\alpha}\partial_{\alpha}\right) &= \frac{2g\left(w^{\alpha}\nabla_{\partial_r}\partial_{\alpha}, w^{\alpha}\partial_{\alpha}\right)}{g\left(w^{\alpha}\partial_{\alpha}, w^{\alpha}\partial_{\alpha}\right)} \\
&= \frac{2g\left(w^{\alpha}\nabla_{\partial_{\alpha}}\partial_r, w^{\alpha}\partial_{\alpha}\right)}{g\left(w^{\alpha}\partial_{\alpha}, w^{\alpha}\partial_{\alpha}\right)} \\
&= \frac{2g\left(S\left(w^{\alpha}\partial_{\alpha}\right), w^{\alpha}\partial_{\alpha}\right)}{g\left(w^{\alpha}\partial_{\alpha}, w^{\alpha}\partial_{\alpha}\right)}
\end{aligned}$$

to conclude that there is a sequence $r_n \to |v|$ such that the Hessian $\nabla^2 f = S_{r_n}$ satisfies

$$\frac{g\left(S_{r_n}\left(w^{\alpha}\partial_{\alpha}\right), w^{\alpha}\partial_{\alpha}\right)}{g\left(w^{\alpha}\partial_{\alpha}, w^{\alpha}\partial_{\alpha}\right)} \to -\infty \quad \text{as} \quad n \to \infty.$$

Now for the contradiction. The curve $\gamma(t) = \exp_p(tv)$ is a segment on some interval $[0, 1+\varepsilon]$, $\varepsilon > 0$. Choose ε so small that $\tilde{f}(x) = d(x, \gamma(1+\varepsilon))$ is smooth on a ball $B(\gamma(1+\varepsilon), 2\varepsilon)$ (which contains $\gamma(1)$). Then consider the function $h(x) = f(x) + \tilde{f}(x)$. From the triangle inequality, we know that $h(x) \geq 1+\varepsilon = d(p, \gamma(1+\varepsilon))$ everywhere. Furthermore, $h(x) = 1+\varepsilon$ whenever $x = \gamma(t)$, $t \in [0, 1+\varepsilon]$. Thus, h has an absolute minimum along $\gamma(t)$ and must therefore have nonnegative Hessian at all the points $\gamma(t)$. However,

$$\begin{aligned}
&\frac{g\left(\nabla^2 h\left(w^{\alpha}\partial_{\alpha}\right), w^{\alpha}\partial_{\alpha}\right)}{g\left(w^{\alpha}\partial_{\alpha}, w^{\alpha}\partial_{\alpha}\right)} \\
&\qquad = \frac{g\left(\nabla^2 f\left(w^{\alpha}\partial_{\alpha}\right), w^{\alpha}\partial_{\alpha}\right)}{g\left(w^{\alpha}\partial_{\alpha}, w^{\alpha}\partial_{\alpha}\right)} + \frac{g\left(\nabla^2 \tilde{f}\left(w^{\alpha}\partial_{\alpha}\right), w^{\alpha}\partial_{\alpha}\right)}{g\left(w^{\alpha}\partial_{\alpha}, w^{\alpha}\partial_{\alpha}\right)} \underset{r_n \to |v|}{\to} -\infty
\end{aligned}$$

since $\nabla^2 \tilde{f}$ is bounded in a neighborhood of $\gamma(1)$ and the term involving $\nabla^2 f$ goes to $-\infty$ as $r_n \to |v|$. □

We have now shown that $f(x) = d(x, p)$ is smooth on the open and dense subset $U_p - \{p\} \subset M$. To complete our investigation, we'll show that f is not smooth on $M - U_p$.

Lemma 8.2 *If $v \in \text{seg}(p) - \text{seg}^0(p)$, then either*
(1) $\exists w\,(\neq v) \in \text{seg}(p) : \exp_p(v) = \exp_p(w)$, *or*
(2) $D\exp_p$ *is singular at* v.

Notice that in the first case the gradient ∂_r on M becomes undefined at $x = \exp_p(v)$, since it would be either $D\exp_p(v)$ or $D\exp_p(w)$, while in the second case the Hessian S becomes undefined, since it will be forced to have $-\infty$ as an eigenvalue.

Proof. Let $\gamma(t) = \exp_p(tv)$. For $t > 1$ choose segments $\sigma_t(s) : [0,1] \to M$ with $\sigma_t(0) = p$, $\sigma_t(1) = \gamma(t)$. Since we have assumed that $\gamma : [0,t]$ is not a segment for $t > 1$ we see that $\dot\sigma_t(0)$ is never proportional to $\dot\gamma(0)$. Now choose $t_n \to 1$ such that $\dot\sigma_{t_n}(0) \to w \in T_pM$. We have that $\ell(\sigma_{t_n}) = |\dot\sigma_{t_n}(0)| \to \ell(\gamma : [0,1] \to M) = |\dot\gamma(0)|$. So either $w = \dot\gamma(0)$ or w is not proportional to $\dot\gamma(0)$. In the latter case, we have found the promised w from (1). If the former happens, we must show that $D\exp_p$ is singular at v. But if $D\exp_p$ is nonsingular at v, then $\exp_p$ must be an embedding near v. Now, $\dot\sigma_{t_n}(0) \to v = \dot\gamma(0)$ and $\exp_p(\dot\sigma_{t_n}(0)) = \exp_p(t_n\dot\gamma(0))$, so $\dot\sigma_{t_n}(0) = t_n \cdot v$, which implies that γ must be a segment on some interval $[0, t_n]$, $t_n > 1$. This, however, contradicts our choice of γ. □

The set $\text{seg}(p) - \text{seg}^0(p)$ is called the *cut locus* of p in T_pM. Thus, being inside the cut locus means that we are on the region where f is smooth. Going back to our characterization of segments, we have

Corollary 8.3 *Let $\gamma : [0,\infty) \to M$ be a geodesic with $\gamma(0) = p$. If $\text{cut}(\dot\gamma(0)) = \max\{t : \gamma|_{[0,t]}$ is a segment$\}$, then f is smooth at $\gamma(t)$, $t < \text{cut}(\dot\gamma(0))$, but not smooth at $x = \gamma(\text{cut}(\dot\gamma(0)))$, and the failure of f to be smooth at x is because $\exp_p : \text{seg}(p) \to M$ either fails to be one-to-one at x or has singular differential there.*

5.8.2 *The Injectivity Radius*

The largest radius ε for which $\exp_p : B(0,\varepsilon) \to B(p,\varepsilon)$ is a diffeomorphism is called the *injectivity radius* $\text{inj}(p)$ at p. If $v \in \text{seg}(p) - \text{seg}^0(p)$ is the closest point to 0 in this set, then we have that $\text{inj}(p) = |v|$. It turns out that such v can be characterized as:

Lemma 8.4 (Klingenberg) *Suppose $v \in \text{seg}(p) - \text{seg}^0(v)$ and that $|v| = \text{inj}(p)$. Then either*

(1) *There is only one other vector w with $\exp_p(w) = \exp_p(v)$ and $\frac{d}{dt}|_{t=1}\exp_p(tw) = -\frac{d}{dt}|_{t=1}\exp_p(tv)$, or*

(2) $x = \exp_p(v)$ *is a critical value for* $\exp_p : \text{seg}(p) \to M$.

In the first case, we therefore have exactly two segments from p to $x = \exp_p(v)$, and they fit smoothly together at x to form a geodesic loop.

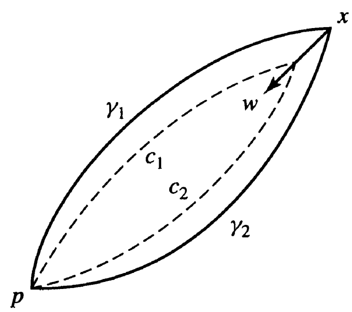


FIGURE 5.9.

Proof. Suppose x is a regular value for $\exp_p : \mathrm{seg}(p) \to M$ and that $\gamma_1, \gamma_2 : [0,1] \to M$ are segments from p to $x = \exp_p(v)$. If $\dot\gamma_1(1) \neq -\dot\gamma_2(1)$, then we can find $w \in T_xM$ such that $g(w, \dot\gamma_1(1)), g(w, \dot\gamma_2(1)) < 0$, i.e., w forms an angle $> \frac{\pi}{2}$ with both $\dot\gamma_1(1)$ and $\dot\gamma_2(1)$, whence if we take a curve $c(s)$ with $\dot c(0) = w$ we have $d(p, c(s)) < d(p, x)$ for $s > 0$.

To see this, note that since $D\exp_p$ is nonsingular at $\dot\gamma_i(0)$, we can find vectors v_i in T_pM close to $\dot\gamma_i(0)$ such that $\exp_p(v_i) = c(s)$ (see also Figure 5.9). But then the curves $t \to \exp_p(tv_i)$ have length

$$\begin{aligned} |v_i| &= d(p, c(s)) \\ &< d(p, x) \\ &= |v| . \end{aligned}$$

This implies that $\exp_p$ is not one-to-one on $\mathrm{seg}^0(p)$, a contradiction. □

5.9 Metric Characterization of Maps

As promised we shall in this section give some metric characterizations of Riemannian isometries and Riemannian submersions. For a Riemannian manifold (M, g) we let the corresponding metric space be denoted by (M, d_g) or simply (M, d) if only one metric is in play. It is natural to ask whether one can somehow recapture the Riemannian metric g from the distance d_g. If for instance $v, w \in T_pM$, then we would somehow like to be able to compute $g(v, w)$ form knowledge of d_g. One way of doing this is by taking two curves α, β such that $\dot\alpha(0) = v$ and $\dot\beta(0) = w$ and then observing that

$$\begin{aligned} |v| &= \lim_{t\to 0} \frac{d(\alpha(t), \alpha(0))}{t}, \\ |w| &= \lim_{t\to 0} \frac{d(\beta(t), \beta(0))}{t}, \\ \angle(v, w) &= \frac{g(v, w)}{|v|\,|w|} = \lim_{t\to 0} \frac{d(\alpha(t), \beta(t))}{t}. \end{aligned}$$

Thus, g can really be found from d given that we use the differentiable structure of M. It is perhaps then not so surprising that many of the Riemannian maps we consider have synthetic characterizations, that is, characterizations that involve only knowledge of the metric space (M, d).

Before proceeding with our investigations, let us introduce a new type of coordinates. Using geodesics we have already introduced one set of geometric coordinates via the exponential map. We shall now use the distance functions to construct *distance coordinates*. For a point $p \in M$ fix a neighborhood $U \ni p$ such that for each $x \in U$ we have that $B(q, \text{inj}(q)) \supset U$. Thus, for each $q \in U$ the distance function $f_q(x) = d(x, q)$ is smooth on $U - \{q\}$. Now choose $q_1, \ldots, q_n \in U - \{p\}$, where $n = \dim M$. If the vectors $\nabla f_{q_1}(p), \ldots, \nabla f_{q_n}(p) \in T_pM$ are linearly independent, the inverse function theorem tells us that $\varphi = \left(f_{q_1}, \ldots, f_{q_n}\right)$ can be used as coordinates on some neighborhood V of p. The size of the neighborhood will depend on how these gradients vary. Thus, an explicit estimate for the size of V can be gotten from bounds on the Hessians of the distance functions. Clearly, one can arrange for the gradients to be linearly independent or even orthogonal at any given point.

Theorem 9.1 (Myers-Steenrod, 1939) *If (M, g) and (N, h) are Riemannian manifolds and $\varphi : M \to N$ a bijection, then φ is a Riemannian isometry if φ is distance-preserving, i.e., $d_h(\varphi(p), \varphi(q)) = d_g(p, q)$ for all $p, q \in M$.*

Proof. Let φ be distance-preserving. First let us show that φ is differentiable. Fix $p \in M$ and let $q = \varphi(p)$. Near q introduce distance coordinates $\left(f_{q_1}, \ldots, f_{q_n}\right)$ and find p_i such that $\varphi(p_i) = q_i$. Now observe that

$$\begin{aligned} f_{q_i} \circ \varphi(x) &= d(\varphi(x), q_i) \\ &= d(\varphi(x), \varphi(p_i)) \\ &= d(x, p_i). \end{aligned}$$

Since $d(x, p_i) = d(q, q_i)$, we can assume that the q_i's and p_i's are chosen such that $g_{p_i}(x) = d(x, p_i)$ are smooth at p. Thus, $\left(f_{q_1}, \ldots, f_{q_n}\right) \circ \varphi$ is smooth at p, showing that φ must be smooth at p.

To show that φ is a Riemannian isometry it suffices to check that $|D\varphi(v)| = |v|$ for all tangent vectors $v \in TM$. For a fixed $v \in T_pM$ let $\gamma(t) = \exp_p(tv)$. For small t we know that γ is a constant speed segment. Thus, for small t, s we can conclude

$$|t - s| \cdot |v| = d_g(\gamma(t), \gamma(s)) = d_h(\varphi \circ \gamma(t), \varphi \circ \gamma(s)).$$

But since $\varphi \circ \gamma$ is smooth, we know that

$$\begin{aligned} |D\varphi(v)| &= \left|\frac{d(\varphi \circ \gamma)}{dt}\right|_{t=0} \\ &= \lim_{t \to 0} \frac{d_h(\varphi \circ \gamma(t), \varphi \circ \gamma(0))}{|t|} \end{aligned}$$

$$= \lim_{t \to 0} \frac{d_g(\gamma(t), \gamma(0))}{|t|}$$
$$= |\dot{\gamma}(0)|$$
$$= |v|.$$
□

It is an exercise that Riemannian isometries are distance-preserving. Our next goal is to find a characterization of Riemannian submersions. Unfortunately, the description only gives us functions that are C^1, but there doesn't seem to be a better formulation. Let $f : (M, g) \to (N, h)$ be a function. We call f a *submetry* if for every $p \in M$ we can find $r > 0$ such that for each $\varepsilon \leq r$ we have $f(B(p, \varepsilon)) = B(f(p), \varepsilon)$. Submetries are locally distance-nonincreasing and therefore also continuous. In addition, we have that the composition of submetries (or Riemannian submersions) are again submetries (or Riemannian submersions). We can now prove

Theorem 9.2 (Berestovski, 1995) *If $f : (M, g) \to (N, h)$ is a surjective submetry, then f is a C^1 Riemannian submersion.*

Proof. Fix points $q \in N$ and $p \in M$ with $f(p) = q$. Then select distance coordinates $(f_1, \ldots, f_k)$ around q. Now observe that all of the f_i's are Riemannian submersions and therefore also submetries. Then the compositions $f_i \circ f$ are also submetries. Thus, f is C^1 iff all the maps $f_i \circ f$ are C^1. It therefore suffices to prove the result in the case where $f : (U \subset M, g) \to ((a, b), \text{can})$.

Let $x \in M$. By restricting f to a small convex neighborhood of x, we can assume that the fibers of f are closed and that any two points in the domain are joined by a unique geodesic. We now wish to show that f has a continuous unit gradient field ∇f. We know that the integral curves for ∇f should be exactly the unit speed geodesics that are mapped to unit speed geodesics by f. Since f is distance-nonincreasing, it is clear that any piecewise smooth unit speed curve that is mapped to a unit speed geodesic must be a smooth unit speed geodesic. Thus, these integral curves are unique and vary continuously to the extent that they exist. To establish the existence of these curves we use the submetry property. First fix $p \in M$ and let $c(t) : [0, r] \to (a, b)$ be the unit speed segment with $c(0) = f(p)$. Let F_t denote the fiber of f above $c(t)$. Now select a unit speed segment $\gamma(t)$:$[0, r] \to M$ with $\gamma(0) = p$ and $\gamma(r) \in F_r$. This is possible since $f(B(p, r)) = B(c(0), r)$. It is now easy to check, again using the submetry property, that $c(t) = f \circ \gamma(t)$, as desired. □

5.10 Further Study

There are many textbooks on Riemannian geometry that treat all of the basic material included in this chapter. Some of the better texts are [19], [21], [37], [50], and [65]. All of these books, as is usual, emphasize the variational approach as being *the* basic technique used to prove every theorem. However, if one wishes to learn about variational calculus, one might as well see how it really should be used in connection with the loop space. For this we refer to the excellent text [60] on Morse theory by Milnor.

5.11 Exercises

1. A Riemannian manifold is said to be homogeneous if the isometry group acts transitively. Show that homogeneous manifolds are geodesically complete.

2. Show that if we have a vector field X on a Riemannian manifold (M, g) that vanishes at $p \in M$, then for any tensor T we have $L_X T = \nabla_X T$ at p. Conclude that the Hessian of a function is independent of the metric at a critical point. Can you find an interpretation of $L_X T$ at p?

3. Suppose we have a rotationally symmetric metric $dr^2 + \varphi^2(r)\, d\theta^2$. We wish to understand parallel translation along a latitude, i.e., a curve with $r = a$. To do this we construct a cone that is tangent to this surface at the latitude $r = a$: $dr^2 + (\varphi(a) + \dot{\varphi}(a)(r - a))^2\, d\theta^2$. In case the surface really is a surface of revolution, this cone is a real cone that is tangent to the surface along the latitude $r = a$.

   (a) Show that in the standard coordinates ∂_r and ∂_θ on these surfaces, the covariant derivative ∇_{∂_θ} is the same along the curve $r = a$. Conclude that parallel translation is the same along this curve on these two surfaces.

   (b) Now take a piece of paper and try to figure out what parallel translations along a latitude on a cone look like. If you unfold the paper it is flat; thus parallel translation is what it is in the plane. Now refold the paper and observe that parallel translation along a latitude does not generate a closed parallel field.

   (c) Show that in the above example the parallel field along $r = a$ closes up iff $\dot{\varphi}(a) = 0$.

4. Find the Riemannian (pullback) metrics one obtains by projecting the sphere and hyperbolic space onto a hyperplane as described at end of the section on geodesics.

5. Show that any Riemannian manifold admits a complete Riemannian metric.

6. On an open subset $U \subset \mathbb{R}^n$ we have the induced Riemannian metric from the Riemannian metric, and also the Euclidean metric restricted to this set. Show that the two agree iff U is convex.

7. Let $f : (M, g) \to \mathbb{R}$ be a smooth function on a Riemannian manifold. If $\gamma : (a, b) \to M$ is a geodesic, compute the first and second derivatives of $f \circ \gamma$. Use this to show that at a local maximum (or minimum) for f the gradient is zero and the Hessian nonpositive definite (or nonnegative definite). Show that f has everywhere nonnegative Hessian iff $f \circ \gamma$ is convex for all geodesics γ in (M, g).

8. Let $N \subset M$ be a submanifold of a Riemannian manifold (M, g).

   (a) The distance from N to $x \in M$ is defined as

   $$d(x, N) = \inf \{d(x, p) : p \in N\}.$$

   A unit speed curve $\sigma : [a, b] \to M$ with $\sigma(a) \in N, \sigma(b) = x$, and $\ell(\sigma) = d(x, N)$ is called a segment from x to N. Show that σ is also a segment from N to any $\sigma(t), t < b$. Show that σ is perpendicular to N. (Hint: find t close to a such that the distance function from $\sigma(t)$ is smooth near $\sigma(a)$, then use that this distance function when restricted to N has a minimum at $\sigma(a)$.)

   (b) Show that if N is a closed subspace of M and (M, g) is complete, then any point in M can be joined to x by a segment.

   (c) Show that in general there is an open neighborhood of N in M where all points are joined to N by segments.

   (d) Show that $d(\cdot, N)$ is smooth on a neighborhood of N and that the integral curves for its gradient are the geodesics that are perpendicular to N.

9. Compute the cut locus on a square torus $\mathbb{R}^2/\mathbb{Z}^2$. Compute the cut locus on a sphere and real projective space with the constant curvature metrics.

10. In a metric space (X, d) one can measure the length of continuous curves $\gamma : [a, b] \to M$ by

   $$\ell(\gamma) = \sup \left\{ \sum d(\gamma(t_i), \gamma(t_{i+1})) : a = t_1 \le t_2 \le \cdots \le t_{k-1} \le t_k = b \right\}.$$

   (a) Show that a curve has finite length iff it is absolutely continuous.

   (b) Show that this definition gives back our previous definition for smooth curves on Riemannian manifolds.

   (c) Let $\gamma : [a, b] \to M$ be an absolutely continuous curve whose length is $d(\gamma(a), \gamma(b))$. Show that $\gamma = \sigma \circ \varphi$ for some segment σ and reparametrization φ.

11. Show that in a Riemannian manifold,

   $$d\left(\exp_p(tv), \exp_p(tw)\right) = |t| \cdot |v - w| + O\left(t^2\right).$$

12. Show that for a complete manifold the functional distance is the same as the distance. What about incomplete manifolds?

13. Show, using the exercises on Lie groups from Chapters 1 and 2, that on a Lie group G with a bi-invariant metric the geodesics through the identity are exactly the homomorphisms $\mathbb{R} \to G$. Conclude that the Lie group exponential map coincides with the exponential map generated by the bi-invariant Riemannian metric.

14. Construct a Riemannian metric on the tangent bundle to a Riemannian manifold (M, g) such that $\pi : TM \to M$ is a Riemannian submersion and the metric restricted to the tangent spaces is the given Euclidean metric. What do geodesics look like in this metric?

15. For a Riemannian manifold (M, g) let FM be the frame bundle of M. This is a fiber bundle $\pi : FM \to M$ whose fiber over $p \in M$ consists of orthonormal bases for T_pM. Find a Riemannian metric on FM that makes π into a Riemannian submersion and such that the fibers are isometric to $O(n)$.

16. Show that a bijective Riemannian isometry is distance-preserving.

17. Show that a Riemannian submersion is a submetry.

18. (Hermann) Let $f : (M, g) \to (N, h)$ be a Riemannian submersion.

    (a) Show that (N, h) is complete if (M, g) is complete.

    (b) Show that f is a fibration if (M, g) is complete, i.e., for every $p \in N$ there is a neighborhood $p \in U$ such that $f^{-1}(U)$ is diffeomorphic to $U \times f^{-1}(p)$. Give a counterexample when (M, g) is not complete.

19. Show that on a complete manifold, $d(\cdot, p)$ is smooth at x iff $d(\cdot, x)$ is smooth at p.

20. Take any book on Riemannian geometry, other than this one, and study the first and second variation formulae and how they are used.

6

Sectional Curvature Comparison I

We shall first classify spaces with constant curvature. The real subject of this chapter is how one can compare manifolds to spaces with constant curvature. We shall for instance prove the Hadamard-Cartan theorem, which says that a simply connected manifold with sec ≤ 0 is diffeomorphic to $\mathbb{R}^n$. There are also some interesting restrictions on the topology in positive curvature that we shall investigate, notably, Synge's theorem, which says that an orientable even-dimensional manifold with positive curvature is simply connected. In Chapter 11 we shall deal with some more advanced topics in the theory of manifolds with lower sectional curvature bounds.

The results we present here all belong to what people would call the classical results for Riemannian manifolds. Aside from work on submanifolds, this body of work essentially comprises all that was known prior to 1943.

6.1 Constant Curvature Revisited

First let us discuss some general facts about isometries between Riemannian manifolds.

Theorem 1.1 *Suppose we have two isometries $\varphi, \psi : (M^n, g) \to (N^n, h)$. If $D\varphi_p = D\psi_p$. Then $\varphi \equiv \psi$ provided that M is connected. In other words: an isometry is uniquely determined by its differential at just one point.*

Proof. We use one of the standard closed-open arguments. Let $A = \{p \in M : D\varphi_p = D\psi_p\}$. Continuity of $D\varphi$ and $D\psi$ implies that A is closed, so we need to establish that A is open. Note that $A \neq \emptyset$ by assumption. Fix $q \in A$ and choose $\varepsilon > 0$ such that

$$\exp_q : B(0, \varepsilon) \to B(q, \varepsilon),$$
$$\exp_{\varphi(q)=\psi(q)} : B(0, \varepsilon) \to B(\varphi(q), \varepsilon)$$

are diffeomorphisms. We will show that

$$\varphi(x) = \exp_{\varphi(q)} \circ D\varphi_q \circ \exp_q^{-1} \text{on} B(q, \varepsilon).$$

Choose $\gamma(t) = \exp_q(tv)$, $v \in B(0, \varepsilon)$, $t \in [0, 1]$. Then we know that $\gamma(t)$ is a geodesic and a segment. The curve $\sigma(t) = \varphi \circ \gamma(t)$, $t \in [0, 1]$ must also be a segment. For otherwise, we could find a shorter curve $\tilde{\sigma} : [0, 1] \to N$ with $\tilde{\sigma}(0) = \sigma(0)$ and $\tilde{\sigma}(1) = \sigma(1)$. But then

$$\ell(\varphi^{-1} \circ \tilde{\sigma}) = \int_0^1 |D\varphi^{-1} \circ \frac{d}{dt}\tilde{\sigma}| = \int_0^1 |\frac{d}{dt}\tilde{\sigma}| = \ell(\tilde{\sigma}),$$

since $D\varphi^{-1}$preserves length of vectors, and we will have found a curve from $\gamma(0)$ to $\gamma(t)$ of length $\ell(\tilde{\sigma}) < \ell(\varphi \circ \gamma) = \ell(\gamma)$. Now that $\varphi \circ \gamma$ is a segment, it must also be a geodesic and therefore have the form

$$\begin{aligned}
\varphi \circ \gamma(t) &= \exp_{\varphi(q)}\left(t \cdot \frac{d}{dt}(\varphi \circ \gamma)|_{t=0}\right) \\
&= \exp_{\varphi(q)}(t \cdot D\varphi_q \cdot \dot{\gamma}(0)) \\
&= \exp_{\varphi(q)}(D\varphi_q(t \cdot \dot{\gamma}(0))) \\
&= \exp_{\varphi(q)}(D\varphi_q(\exp_q^{-1}(\gamma(1))).
\end{aligned}$$

Similar reasoning shows that

$$\psi(x) = \exp_{\psi(q)} \circ D\psi_q \circ \exp_q^{-1} \text{on} B(q, \varepsilon).$$

Thus, $\varphi = \psi$ on $B(q, \varepsilon)$, as we assumed that $D\psi_q = D\varphi_q$. □

This theorem is clearly false for Riemannian immersions and submersions in general, but it remains valid as long as $\dim M = \dim N$ and φ, ψ are Riemannian immersions.

What about the inverse problem? Given any linear isometry $L : T_pM \to T_qN$, is there an isometry $\varphi : M \to N$ such that $D\varphi_p = L$? If we let $M = N$, this would in particular mean that if π is a 2-plane in T_pM and $\tilde{\pi}$ a 2-plane in T_qM, then there should be an isometry $\varphi : M \to M$ such that $\varphi(\pi) = \tilde{\pi}$. But this would imply that M has constant sectional curvature. The above problem can therefore not be solved in general. If we go back and inspect our knowledge of $\mathrm{Iso}(S_k^n)$, we see that these spaces have enough isometries so that any linear isometry $L : T_pS_k^n \to T_qS_k^n$ can be extended to a global isometry $\varphi : S_k^n \to S_k^n$ with $D\varphi_p = L$. In some sense these are the only spaces with this property, as we shall see.

Theorem 1.2 *Suppose (M, g) is a Riemannian manifold of dimension n and constant curvature k. If M is simply connected and $L : T_pM \to T_qS^n_k$ is a linear isometry, then there is a unique isometric immersion called the* monodromy map *$\varphi : M \to S^n_k$ with $D\varphi_p = L$. Furthermore, this map is a diffeomorphism if (M, g) is complete.*

Before giving the proof, let us look at some examples.

Example 1.3 Suppose we have an immersion $M^n \to S^n_k$. Then φ will be one of the maps described in the theorem if we use the pullback metric on M. Such maps can be arbitrarily ugly when $n \geq 2$ and need not resemble covering maps in any way whatsoever.

Example 1.4 If $U \subset S^n_k$ is a contractible bounded open set with ∂U a nice hypersurface, then one can easily construct a diffeomorphism $\varphi : M = S^n_k - \{pt\} \to S^n_k - U$. Near the missing point in M the metric will necessarily look pretty awful, although it has constant curvature.

Example 1.5 If $M = \mathbb{R}P^n$ or $\mathbb{R}^n - \{0\}$/antipodal map, then M is not simply connected and does not admit an immersion into S^n_k.

Corollary 1.6 *If M is a closed simply connected manifold with constant-curvature k, then $k > 0$ and $(M, g) = S^n_k$. Thus, $S^p \times S^q, \mathbb{C}P^n$ do not admit any constant curvature metrics.*

Corollary 1.7 *If M is geodesically complete and noncompact with constant curvature k, then $k \leq 0$ and the universal covering is S^n_k. In particular, $S^2 \times \mathbb{R}^2$ and $S^n \times \mathbb{R}$ do not admit any geodesically complete metrics of constant curvature.*

Now for the proof of the theorem. A different proof is developed in the exercises to this chapter.

Proof. We know that M can be covered by sets U_α such that each U_α admits a Riemannian embedding $\varphi_\alpha : U_\alpha \to S^n_k$. Furthermore, if $p \in U_{\alpha_0}$, then we can choose φ_{α_0} such that $D\varphi_{\alpha_0|p}$ is any predetermined isometry. Also, each φ_α is well defined up to an element in $\mathrm{Iso}(S^n_k)$; in other words if $\varphi_\alpha, \psi_\alpha : U_\alpha \to S^n_k$ are isometries then $\varphi_\alpha \circ \psi_\alpha^{-1}$ is the restriction of an element in $\mathrm{Iso}(S^n_k)$. The construction of φ now proceeds in the same way one does analytic continuation on simply connected domains.

The geodesically complete situation is divided into two cases.

Case $k \leq 0$: From the proof of the uniqueness theorem we get that φ maps geodesics to geodesics. Completeness tells us that any two points $p, q \in M$ can be joined by a geodesic $\gamma : [0, 1] \to M$. Now $\varphi \circ \gamma$ is also a geodesic in S^n_k, $k \leq 0$. We know from earlier that geodesics in S^n_k never intersect themselves, so in particular we see that $\varphi(p) \neq \varphi(q)$. To see that φ is onto, choose $\hat{p} \in \mathrm{Im}(\varphi)$ and $\hat{q} \in S^n_k$. Then join $\hat{p}$ and $\hat{q}$ by a geodesic $\hat{\gamma} : [0, 1] \to S^n_k$. In M, let $\gamma : [0, 1] \to M$ be the unique geodesic with $\gamma(0) = \varphi^{-1}(\hat{p})$ and $D\varphi_{\varphi^{-1}(\hat{p})}(\dot{\gamma}(0)) = \frac{d}{dt}\hat{\gamma}(0)$. This geodesic exists by geodesic completeness. Now observe that $\varphi \circ \gamma$ is a geodesic with the same initial values as $\hat{\gamma}$. Thus, $\varphi \circ \gamma = \hat{\gamma}$ everywhere, and $\varphi(\gamma(1)) = \hat{q}$ in particular.

Case $k > 0$: Surjectivity is established by the same method, but since geodesics in S_k^n are closed curves, we can't use the same proof for injectivity. In the case where M is closed, φ must of course be a covering map. But then it is a diffeomorphism, since $\pi_1(S_k^n) = 0$. In the case where M is not closed, we can still show that φ is a covering map using that M is geodesically complete. We have to show that each $q \in S_k^n$ has a neighborhood which is evenly covered. Let $\varepsilon < \pi/\sqrt{k} = \operatorname{diam}(S_k^n)$, and $B(q, \varepsilon) = \{x \in S_k^n : d(q, x) < \varepsilon\}$. For $p \in \varphi^{-1}(q)$ let $U_p = \exp_p(B(0, \varepsilon) \subset T_pM)$. Arguing as above, we can prove that $\varphi : U_p \to B(q, \varepsilon)$ is surjective. To check injectivity, fix $x, y \in U_p$ and join them to p by geodesics $\gamma_x(t)$, $\gamma_y(t)$, $t \in [0, 1]$, $|\dot{\gamma}_x(0)| < \varepsilon$, and $|\dot{\gamma}_y(0)| < \varepsilon$. Then $\varphi \circ \gamma_x$ and $\varphi \circ \gamma_y$ are geodesics emanating from q that stay in $B(q, \varepsilon)$. Since geodesics in S_k^n either completely overlap or only intersect in antipodal points, we see that $\varphi(x) = \varphi(y)$ only when $\varphi \circ \gamma_x = \varphi \circ \gamma_y$. But then $D\varphi(\dot{\gamma}_x(0)) = D\varphi(\dot{\gamma}_y(0))$ which implies $\dot{\gamma}_x(0) = \dot{\gamma}_y(0)$ and $x = y$. □

Corollary 1.8 (Classification of Constant Curvature Spaces, H. Hopf, 1926) *If (M, g) is a connected, geodesically complete Riemannian manifold with constant curvature k, then the universal covering is isometric to S_k^n.*

This result shows how important the completeness of the metric is. A large number of open manifolds admit immersions into Euclidean space of the same dimension (e.g., $S^n \times \mathbb{R}^k$) and therefore carry incomplete metrics with zero curvature. Carrying a complete Riemannian metric of a certain type, therefore, often implies various topological properties of the underlying manifold. Riemannian geometry at its best tries to understand this interplay between metric and topological properties.

6.2 Basic Comparison Estimates

In this section we shall prove most of the comparison estimates that will be needed throughout the text. First we prove some abstract results on comparison of matrices. Then we extract a geometric corollary, which is used again and again in many different contexts.

Theorem 2.1 (First Comparison Estimate) *Suppose we have real numbers $k \leq K$ and a Lipschitz (or absolutely) continuous function $\lambda : (0, b) \to \mathbb{R}$ which satisfies*

$$-K \leq \dot{\lambda} + \lambda^2 \leq -k.$$

If the initial condition for λ *is* $\lambda(0) = \frac{1}{r} + O(r)$, *where* $a > 0$, *then we have*

$$\frac{\mathrm{sn}'_K(r)}{\mathrm{sn}_K(r)} \leq \lambda(r) \leq \frac{\mathrm{sn}'_k(r)}{\mathrm{sn}_k(r)}$$

for as long as $\mathrm{sn}_K(r) > 0$.

Proof. The two inequalities are proved in a similar manner, the only thing to watch for is that all functions in play should be defined. This is why we must assume that $\mathrm{sn}_K(r) > 0$. In case $K \leq 0$, this is always true, while if $K > 0$, one must assume that $r < \pi/\sqrt{K}$. The next thing to observe is that the functions we wish to compare with are the solutions to the initial value problems

$$\begin{aligned} \dot{\lambda} + \lambda^2 &= -k, \\ \lambda(r) &= \frac{1}{r} + O(r), \end{aligned}$$

$$\begin{aligned} \dot{\lambda} + \lambda^2 &= -K, \\ \lambda(r) &= \frac{1}{r} + O(r), \end{aligned}$$

Thus, we are simply comparing a function satisfying a differential inequality to the solution for the corresponding differential equation. The result should therefore not come as any surprise, at least as long as we assume that we work with differentiable functions. Now, the calculus for absolutely continuous functions is virtually the same as for differentiable functions and does not cause any problems here. Nevertheless, there are several different ways of approaching this problem and we shall mention a few of them. For convenience, we shall concentrate on showing only $\lambda(r) \leq (\mathrm{sn}'_k(r))/(\mathrm{sn}_k(r))$. One recurring problem with the proofs is that $\lambda(r)$ and $(\mathrm{sn}'_k(r))/(\mathrm{sn}_k(r))$ are not defined at $r = 0$. This can be partially averted by considering $\mu = \lambda^{-1}$ and $(\mathrm{sn}_k(r))/(\mathrm{sn}'_k(r))$ instead. Then we have:

$$\begin{aligned} \dot{\mu} - 1 &\geq k \cdot \mu^2, \\ \mu(0) &= 0. \end{aligned}$$

However, then we run into the problem that when $k > 0$ the function $(\mathrm{sn}_k(r))/(\mathrm{sn}'_k(r))$ becomes undefined at $\pi/(2\sqrt{k})$. This is not really a serious issue, but it means that some extra analysis is still necessary.

Method 1: This is the most general method. We consider the function

$$\psi(r) = \left(\frac{\mathrm{sn}'_k(r)}{\mathrm{sn}_k(r)} - \lambda(r)\right) \exp\left(\int_0^r \left(\frac{\mathrm{sn}'_k(t)}{\mathrm{sn}_k(t)} + \lambda(t)\right) dr\right).$$

This function is absolutely continuous with derivative

$$\begin{aligned}
\psi'(r) &= \left(\frac{\operatorname{sn}_k'(r)}{\operatorname{sn}_k(r)} - \lambda(r)\right)' \exp\left(\int_0^r \left(\frac{\operatorname{sn}_k'(t)}{\operatorname{sn}_k(t)} + \lambda(t)\right) dr\right) \\
&\quad + \left(\frac{\operatorname{sn}_k'(r)}{\operatorname{sn}_k(r)} - \lambda(r)\right)\left(\frac{\operatorname{sn}_k'(t)}{\operatorname{sn}_k(t)} + \lambda(t)\right) \exp\left(\int_0^r \left(\frac{\operatorname{sn}_k'(t)}{\operatorname{sn}_k(t)} + \lambda(t)\right) dr\right) \\
&= \left(\left(\frac{\operatorname{sn}_k'(r)}{\operatorname{sn}_k(r)}\right)' + \left(\frac{\operatorname{sn}_k'(t)}{\operatorname{sn}_k(t)}\right)^2 \right. \\
&\qquad \left. - \left(\lambda'(r) + \lambda^2(r)\right)\right) \exp\left(\int_0^r \left(\frac{\operatorname{sn}_k'(t)}{\operatorname{sn}_k(t)} + \lambda(t)\right) dr\right) \\
&= \left(-k - \left(\lambda'(r) + \lambda^2(r)\right)\right) \exp\left(\int_0^r \left(\frac{\operatorname{sn}_k'(t)}{\operatorname{sn}_k(t)} + \lambda(t)\right) dr\right) \\
&\geq 0,
\end{aligned}$$

thus, showing that $\psi(r) \geq \psi(0) = 0$. The last equality follows from the initial values. There is one slight problem with this, namely, the integral is divergent at $r = 0$. This can be handled by integrating from $\varepsilon > 0$ instead of integrating from 0. Then we get that $\psi_\varepsilon(r) \geq \psi_\varepsilon(\varepsilon)$ for all $\varepsilon > 0$ and $r > \varepsilon$, where

$$\psi_\varepsilon(r) = \left(\frac{\operatorname{sn}_k'(r)}{\operatorname{sn}_k(r)} - \lambda(r)\right) \exp\left(\int_\varepsilon^r \left(\frac{\operatorname{sn}_k'(t)}{\operatorname{sn}_k(t)} + \lambda(t)\right) dr\right).$$

Using that $\int_\varepsilon^r \left((\operatorname{sn}_k'(t))/(\operatorname{sn}_k(t)) + \lambda(t)\right) dr = -2\log(\varepsilon) + O(1)$ as $\varepsilon \to 0$, we can then conclude that

$$\begin{aligned}
\psi_\varepsilon(r) &= \left(\frac{\operatorname{sn}_k'(r)}{\operatorname{sn}_k(r)} - \lambda(r)\right)\left(\frac{2}{\varepsilon} + O(1)\right) \\
&\geq \psi_\varepsilon(\varepsilon) \\
&= \left(\frac{\operatorname{sn}_k'(\varepsilon)}{\operatorname{sn}_k(\varepsilon)} - \lambda(\varepsilon)\right) \\
&= O(\varepsilon),
\end{aligned}$$

which implies that

$$\left(\frac{\operatorname{sn}_k'(r)}{\operatorname{sn}_k(r)} - \lambda(r)\right) \geq \left(\frac{2}{\varepsilon} + O(1)\right)^{-1} O(\varepsilon).$$

As the right-hand side goes to zero as $\varepsilon \to 0$, we get the desired estimate.

Method 2: This approach will also be used later in Chapter 9. The idea is simply that the inequality

$$\dot{\lambda} + \lambda^2 \leq -k$$

can be separated to yield

$$\frac{\dot{\lambda}}{\lambda^2 + k} \leq -1,$$

and then integrated to get an inequality for λ. This inequality will then yield the desired inequality after inverting the functions. There is some slight nastiness, again coming form the problem with $\lambda(r)$ not being defined at $r = 0$.

Method 3: This time we consider the function

$$\phi(r) = \frac{\mathrm{sn}_k'(r)}{\mathrm{sn}_k(r)} - \lambda(r).$$

This function is again absolutely continuous and the derivative satisfies

$$\begin{aligned}\phi'(r) &= \left(\frac{\mathrm{sn}_k'(r)}{\mathrm{sn}_k(r)}\right)' - \lambda'(r) \\ &\geq -\left(\frac{\mathrm{sn}_k'(r)}{\mathrm{sn}_k(r)}\right)^2 - k + \lambda^2(r) + k \\ &= -\left(\frac{\mathrm{sn}_k'(r)}{\mathrm{sn}_k(r)} + \lambda(r)\right)\phi(r).\end{aligned}$$

Now, both λ and $(\mathrm{sn}_k'(r))/(\mathrm{sn}_k(r))$ start out being positive, so if ϕ becomes negative, then some derivative must be negative since

$$\begin{aligned}\phi(r) &= \phi(0) + \int_0^r \phi'(t)\,dt \\ &= \int_0^r \phi'(t)\,dt.\end{aligned}$$

But this contradicts the above inequality for ϕ'. The only problem that can occur with this argument is that eventually $\lambda + \frac{\mathrm{sn}_k'(r)}{\mathrm{sn}_k(r)}$ might become negative. By the time this happens we have already established that ϕ is positive, the above inequality for ϕ' then says that the derivative will remain positive and thus force ϕ to increase. □

Corollary 2.2 (First Comparison Estimate) *Suppose we have a smooth function $S(r)$ of symmetric matrices such that*

$$\begin{aligned}-K \cdot I \leq \dot{S} + S^2 \leq -k \cdot I, \\ S(r) = \frac{1}{r} \cdot I + O(r).\end{aligned}$$

Then

$$\frac{\mathrm{sn}_K'(r)}{\mathrm{sn}_K(r)} \cdot I \leq S(r) \leq \frac{\mathrm{sn}_k'(r)}{\mathrm{sn}_k(r)} \cdot I$$

for as long as $\mathrm{sn}_K(r) > 0$.

Proof. We shall prove each of these inequalities for the smallest and largest eigenvalues of $S(r)$. First note that the set of eigenvalues $\{\lambda_1(r), \ldots, \lambda_k(r)\}$

varies smoothly with respect to r. The largest and smallest eigenvalues which are defined by

$$\lambda_{\max}(r) = \max\{\lambda_1(r), \dots, \lambda_k(r)\},$$
$$\lambda_{\min}(r) = \min\{\lambda_1(r), \dots, \lambda_k(r)\}$$

must therefore be at least Lipschitz continuous. We now claim that

$$\begin{aligned}
\dot{\lambda}_{\max}(r) + \lambda_{\max}^2(r) &\leq -k, \\
\lambda_{\max}(r) &= \frac{1}{r} + O(r), \\
\dot{\lambda}_{\min}(r) + \lambda_{\min}^2(r) &\geq -K, \\
\lambda_{\min}(r) &= \frac{1}{r} + O(r).
\end{aligned}$$

The initial conditions are of course trivial from the initial condition on $S(r)$. To establish the first inequality at a point r_0 where $\lambda_{\max}(r)$ is differentiable, select a unit eigenvector v for $\lambda_{\max}(r)$. Then consider the function $\phi(r) = v^t S(r) v$. This function is less than $\lambda_{\max}(r)$ everywhere, and equal to $\lambda_{\max}(r)$ at $r = r_0$. Thus they must have the same derivative at $r = r_0$. This implies

$$\begin{aligned}
\dot{\lambda}_{\max}(r_0) + \lambda_{\max}^2(r_0) &= \phi'(r_0) + \phi^2(r_0) \\
&= v^t\left(\dot{S}(r_0) + S^2(r_0)\right)v \\
&\leq -v^t k v \\
&= -k.
\end{aligned}$$

The analysis is similar for the smallest eigevalue. The above theorem now implies the desired matrix inequalities. □

Note that all of the above results can be adjusted to the situation where λ is defined at $r = 0$. The comparison function is easy to find in this case, and the similar inequalities can be proved without much effort. The exact statement is as follows:

Theorem 2.3 (Second Comparison Estimate) *Suppose we have a smooth function $S(r)$ of symmetric matrices such that*

$$-K \cdot I \leq \dot{S} + S^2 \leq -k \cdot I,$$
$$S(0) = a \cdot I.$$

Then

$$\left(\frac{\mathrm{cs}_K'(r+c_1)}{\mathrm{cs}_K(r+c_1)}\right) \cdot I \leq S(r) \leq \left(\frac{\mathrm{cs}_k'(r+c_2)}{\mathrm{cs}_k(r+c_2)}\right) \cdot I,$$
$$\frac{\mathrm{cs}_K'(c_1)}{\mathrm{cs}_K(c_1)} = \frac{cs_k'(c_2)}{cs_k(c_2)} = a$$

for as long as $(\mathrm{cs}_K'(r+c_1))/(\mathrm{cs}_K(r+c_1)) > 0$.

Let us now apply these results to one of the most commonly occurring geometric situations. Suppose that on a Riemannian manifold (M, g) we have introduced polar coordinates (r, θ) around a point $p \in M$.

Corollary 2.4 *Assume that (M, g) satisfies $k \leq \sec \leq K$. If $(g_{\alpha\beta})$ represents the metric in the polar coordinates and (S_α^β) the Hessian of f, then we have*

$$\mathrm{sn}_K^2(r) \leq (g_{\alpha\beta}(r,\theta))_{2\leq\alpha,\beta\leq n} \leq \mathrm{sn}_k^2(r),$$
$$\sqrt{K}\mathrm{ct}_K(r) \leq (S_\alpha^\beta(r,\theta))_{2\leq\alpha,\beta\leq n} \leq \sqrt{k}\mathrm{ct}_k(r),$$

where

$$\sqrt{k}\mathrm{ct}_k(r) = \frac{\mathrm{sn}_k'(r)}{\mathrm{sn}_k(r)} = \sqrt{k}\frac{\mathrm{cs}_k(r)}{\mathrm{sn}_k(r)}.$$

Moreover, for the exponential map at p we have

$$|D\exp_p^{-1}| \leq \min\left\{1, \frac{\mathrm{sn}_K(r)}{r}\right\},$$
$$|D\exp_p| \leq \max\left\{1, \frac{\mathrm{sn}_k(r)}{r}\right\},$$

Proof. We first need to observe that $(S_{\alpha\beta}(r,\theta))$ has the initial values

$$(S_\alpha^\beta(r,\theta)) = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{r} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{r} \end{pmatrix} + O(r).$$

Restricting the range of indices to $2 \leq \alpha, \beta \leq n$ therefore gives us a family of matrices to which the previous corollary can be used. Thus, we obtain the desired estimate for the Hessian. For the metric itself we need to use the differential equation

$$\partial_r(g_{\alpha\beta}(r,\theta)) = 2(S_\alpha^\gamma(r,\theta))(g_{\gamma\beta}(r,\theta)).$$

Since $S_{\alpha\beta}$ is zero whenever one of the indices is 1, we can restrict this differential equation to the case where the indices are in the range $2 \leq \alpha, \beta \leq n$. Using the estimates for the Hessian, we then arrive at the differential inequalities

$$2\sqrt{K}\mathrm{ct}_K(r)(g_{\alpha\beta}(r,\theta)) \leq \partial_r(g_{\alpha\beta}(r,\theta)) \leq 2\sqrt{k}\mathrm{ct}_k(r)(g_{\alpha\beta}(r,\theta)).$$

Using the trick from earlier of considering the largest and smallest eigenvalues for $(g_{\alpha\beta}(r,\theta))$, we then arrive at the two differential inequalities

$$\dot{\mu}_{\max} \leq 2\sqrt{k}\mathrm{ct}_k(r)\mu,$$
$$\dot{\mu}_{\min} \geq 2\sqrt{K}\mathrm{ct}_K(r)\mu.$$

Using that $\left(g_{\alpha\beta}(r,\theta)\right)_{2\leq\alpha,\beta\leq n} = O\left(r^2\right)$, we get that these eigenvalues are both zero at $r = 0$. Integrating these differential inequalities then yields

$$\dot{\mu}_{\max} \leq \operatorname{sn}_k^2(r),$$
$$\dot{\mu}_{\min} \geq \operatorname{sn}_K^2(r).$$

This, in turn, gives the desired estimates for $\left(g_{\alpha\beta}(r,\theta)\right)_{2\leq\alpha,\beta\leq n}$.

Finally,

$$\left|D\exp_p\left(w = w^\alpha\partial_\alpha\right)\right|^2 = g_{\alpha\beta}w^\alpha w^\beta.$$

So using that $\left|D\exp_p(\partial_r)\right| = 1$ and that the Euclidean norm of $\sum_{\alpha=2}^n w^\alpha\partial_\alpha$ at distance r from the origin is given by $\left|\sum_{\alpha=2}^n w^\alpha\partial_\alpha\right| = r\cdot\sqrt{\sum_{\alpha=2}^n (w^\alpha)^2}$, we arrive at the estimate for $\left|D\exp_p\right|$. The estimate for $|D\exp_p^{-1}|$ is obtained in a similar manner. □

6.3 Riemannian Covering Maps

In this section we shall use everything we have learned so far, and then some, to show that the exponential map $\exp_p : T_pM \to M$ is a covering map, provided that (M, g) has nonpositive sectional curvature everywhere. This implies, in particular, that no compact simply connected manifold admits such metrics.

6.3.1 Manifolds Without Conjugate Points

First some generalities:

Theorem 3.1 *If $\varphi : (M, g) \to (N, h)$ is a Riemannian immersion between complete Riemannian manifolds of the same dimension, then φ is a Riemannian covering map.*

Proof. The proof follows what we did for geodesically complete manifolds of constant positive curvature. First observe that φ maps geodesics to geodesics and is therefore onto. Next fix $q \in N$ and $\varepsilon = \operatorname{inj}(p)$. Then $\varphi : B(p,\varepsilon) \to B(q,\varepsilon)$ is a diffeomorphism if $p \in \varphi^{-1}(q)$, since geodesics in $B(p,\varepsilon)$ emanating from p are mapped to segments in $B(p,\varepsilon)$ emanating from q, and we know that such segments can't intersect. Hence, $B(q,\varepsilon)$ is evenly covered. □

Corollary 3.2 *Suppose* $\exp_p : T_pM \to M$ *is nonsingular everywhere (i.e., has no critical points); then it is a covering map.*

Proof. By definition $\exp_p$ is an immersion, so on T_pM choose the pullback metric to make it into a Riemannian immersion. To apply our theorem, we must then check that this new metric on T_pM is complete. To see this, observe that the metric is geodesically complete at the origin, since straight lines through the origin are still geodesics. □

We can now prove our big result:

Theorem 3.3 (von Mangoldt, 1881, Hadamard, 1898, Cartan, 1926) *If (M, g) is complete, connected, and has* $\sec \leq 0$, *then the universal covering is diffeomorphic to* $\mathbb{R}^n$.

Proof. Let $O \subset T_pM$ be an open star-shaped set such that $\exp_p : O \to M$ has no critical values. We can then pull back the Riemannian metric g to O via the exponential map. We still have that the Euclidean distance function $f(x) = |x|$ is the distance function to the origin with gradient ∂_r. Using the first comparison estimate with $\sec \leq 0$, we then conclude

$$\left(g_{\alpha\beta}(r,\theta)\right)_{2\leq\alpha,\beta\leq n} \geq r^2 \cdot \left(\delta_{\alpha\beta}\right)_{2\leq\alpha,\beta\leq n}.$$

Thus, the pullback metric is always larger than the Euclidean metric on O. In particular, there can't be any conjugate points for f, and

$$\left|D\exp_p^{-1}\right| \leq 1 \quad \text{on} \quad O.$$

It then follows that $\exp_p$ has no critical points in the closure of O. By continuity it can't have any critical point in a neighborhood of this closure either. From this we can conclude that $\exp_p : T_pM \to M$ has no critical values. □

No similar theorem is true for Riemannian manifolds with $\mathrm{Ric} \leq 0$, $\mathrm{scal} \leq 0$, since we have Ricci flat metrics on $\mathbb{R}^2 \times S^2$ and scalar flat metrics on $\mathbb{R} \times S^p$, $p \geq 1$. This is interesting to ponder. For we know that the existence of conjugate points is related only to the determinant of the Jacobian. And this determinant is governed by the Ricci curvature through the equations tr1 and tr2 from Section 2.4.2. Note however that we can only get a inequality for the volume form in one direction:

$$\partial_r^2\left(\sqrt[n-1]{\sqrt{\det g_{\alpha\beta}(r,\theta)}}\right) \leq -\frac{\mathrm{Ric}(\partial_r,\partial_r)}{n-1}\sqrt[n-1]{\det\sqrt{g_{\alpha\beta}(r,\theta)}}.$$

So assuming upper Ricci curvature bounds doesn't give us any information on the volume form.

Note that we can use the second comparison estimate to conclude that the Hessian of f is always nonnegative. In particular, f must be convex. In fact, on any Riemannian manifold $\partial_r\left(g_{\alpha\beta}(r,\theta)\right)_{2\leq\alpha,\beta\leq n}$ is positive in a neighborhood of p. Thus, distance functions are always locally convex. Below we shall get a more exact estimate of the size of such neighborhoods.

6.3.2 The Conjugate Radius

Example 3.4 Consider S^n_K, $K > 0$. If we fix $p \in S^n_K$ and use polar coordinates, then the metric looks like $dr^2 + \mathrm{sn}^2_K ds^2_{n-1}$. At distance $\pi/\sqrt{K}$ from p we therefore hit a conjugate point no matter what direction we go in.

As a generalization of our result on no conjugate points from above, we can show

Theorem 3.5 *If (M, g) has* $\sec \leq K$, $K > 0$, *then* $\exp_p : B\left(0, \pi/\sqrt{K}\right) \to M$ *has no critical values.*

Proof. As before, pick an open star-shaped set $O \subset B\left(0, \pi/\sqrt{K}\right)$ that contains no critical points for $\exp_p$. The comparison estimate for the metric in polar coordinates with $\sec \leq K$ then yields

$$\left(g_{\alpha\beta}(r, \theta)\right)_{2 \leq \alpha, \beta \leq n} \geq \mathrm{sn}^2_K(r) \cdot \left(\delta_{\alpha\beta}\right)_{2 \leq \alpha, \beta \leq n} \quad \text{for} \quad r \in \left(0, \frac{\pi}{\sqrt{K}}\right).$$

Therefore if the closure of O is contained in $B\left(0, \frac{\pi}{\sqrt{K}}\right)$ we can argue as above. □

We can now show

Theorem 3.6 *Suppose r satisfies*

(1) $r \leq \frac{1}{2} \cdot \mathrm{inj}(x),\ x \in B(p, r)$,
(2) $r \leq \frac{1}{2} \cdot \frac{\pi}{\sqrt{K}},\ K = \sup\{\sec(\pi) : \pi \subset T_x M,\ x \in B(p, r)\}$.

Then $f(x) = d(x, p)$ is smooth and convex on $B(p, r)$, and any two points in $B(p, r)$ are joined by a unique segment that lies in $B(p, r)$.

Proof. The first condition tells us that any two points in $B(p, r)$ are joined by a unique segment, and that $f(x)$ is smooth on $B(p, 2 \cdot r)$. The second condition ensures us that $\nabla^2 f \geq 0$ on $B(p, r)$. It then remains to be shown that if $x, y \in B(p, r)$, and $\gamma : [0, 1] \to M$ is the unique segment joining them, then $\gamma \subset B(p, r)$. For fixed $x \in B(p, r)$, define C_x to be the set of y's for which this holds. Then $x \in C_x$ and C_x is open. If $y \in B(p, r) \cap \partial C_x$, then the geodesic $\gamma : [0, 1] \to M$ joining x to y must lie in $\overline{B(p, r)}$ by continuity. Now consider $\varphi(t) = f(\gamma(t))$. By assumption $\varphi(0), \varphi(1) < r$ and $\ddot{\varphi}(t) = g(\nabla^2 f \cdot \dot{\gamma}(t), \dot{\gamma}(t)) \geq 0$. Thus, φ is convex, and consequently $\max \varphi(t) \leq \max\{\varphi(0), \varphi(r)\} < r$, showing that $\gamma \subset B(p, r)$. □

The largest r such that $f(x)$ is convex on $B(p,r)$ and any two points in $B(p,r)$ are joined by unique segments in $B(p,r)$ is called the *convexity radius* at p. Globally, $\text{conv.rad}(M,g) = \inf_{p\in M} \text{conv.rad.}(p)$. The previous results tell us

$$\text{conv.rad.}(M,g) \geq \min\left\{\frac{\text{inj}(M,g)}{2}, \frac{\pi}{2\sqrt{K}}\right\}, K = \sup \sec(M,g).$$

In nonpositive curvature we simply have

$$\text{conv.rad.}(M,g) = \frac{\text{inj}(M,g)}{2}.$$

Now that we can control conjugate points, we also get estimates for the injectivity radius. For Riemannian manifolds with $\sec \leq 0$ the injectivity radius satisfies

$$\text{inj}(p) = \frac{1}{2} \cdot (\text{length of shortest geodesic loop based at } p).$$

This is, of course, because there are no conjugate points whatsoever. On a closed Riemannian manifold with $\sec \leq 0$ we get that

$$\text{inj}(M) = \inf_{p\in M} \text{inj}(p) = \frac{1}{2} \cdot (\text{length of shortest closed geodesic}).$$

Since M is closed, the infimum must be a minimum (this is not obvious, since we haven't shown that $p \to \text{inj}(p)$ is continuous, but you can prove this for yourself using that $\exp : TM \to M \times M$ is continuous). If $p \in M$ realizes this infimum, and $\gamma : [0,1] \to M$ is the geodesic loop realizing $\text{inj}(p)$, then we can split γ into two equal segments joining p and $\gamma\left(\frac{1}{2}\right)$. Thus, $\text{inj}\left(\gamma\left(\frac{1}{2}\right)\right) \leq \text{inj}(p)$, but this means that γ must also be a geodesic loop as seen from $\gamma\left(\frac{1}{2}\right)$. In particular, it is smooth at p and forms a closed geodesic.

More generally, we have that if (M,g) has $\sec \leq K$, where $K > 0$, then

$$\text{inj}(p) \geq \min\left\{\frac{\pi}{\sqrt{K}}, \frac{1}{2} \cdot (\text{length of shortest geodesic loop based at } p)\right\},$$

$$\text{inj}(M) = \inf_{p\in M} \text{inj}(p) \geq \min\left\{\frac{\pi}{\sqrt{K}}, \frac{1}{2} \cdot (\text{length of shortest closed geodesic})\right\}.$$

These estimates will be used in the next section.

6.3.3 The Fundamental Group in Nonpositive Curvature

There are two key results that we can prove. One actually provides a rather complete picture of fundamental groups in nonpositive curvature. The interested reader is referred to the survey by Eberlein-Hammenstad-Schroeder in [41] for more details.

First we need a little preparation. Let (M,g) be a complete simply connected Riemannian manifold of nonpositive curvature. The two key properties we shall use are that any two points in M lie on a unique geodesic, and that distance functions are everywhere smooth and convex.

If, as on Euclidean space, we consider the modified distance function

$$x \to f_{0,p}(x) = \frac{1}{2}(d(x,p))^2,$$

then the Hessian of this function must be bigger than I, which is the Hessian of this function on Euclidean space. Thus, the modified distance function is strictly convex. We can now generalize convexity slightly (see also Chapter 9) to mean that the function is convex or strictly convex when restricted to any geodesic. With this, one sees that the maximum of any number of convex functions is again convex (you only need to prove this in dimension 1, as we can restrict to geodesics). Given a finite collection of points $p_1, \ldots, p_k \in M$, we can then consider the strictly convex function

$$x \to \max\left\{f_{0,p_1}(x), \ldots f_{0,p_k}(x)\right\}.$$

In general, any nonnegative strictly convex proper function has a unique minimum. To see this, first observe that there must be a minimum. If there were two minima, then the function would be strictly convex when restricted to a geodesic joining these two minima. But the function would have smaller values on the interior of this segment than at the endpoints. The uniquely defined minimum for

$$x \to \max\left\{f_{0,p_1}(x), \ldots f_{0,p_k}(x)\right\}$$

is denoted by $\text{cm}\{p_1, \ldots, p_k\}$ and called the *center of mass* of $\{p_1, \ldots, p_k\}$. In Euclidean space we have the well-known formula

$$\text{cm}\{p_1, \ldots, p_k\} = \frac{1}{k}\sum_{i=1}^{k} p_i.$$

The first theorem is concerned with fixed points of isometries.

Theorem 3.7 (E. Cartan, 1926) *If (M,g) is a complete simply connected Riemannian manifold of nonpositive curvature, then any isometry $\varphi : M \to M$ of finite order has a fixed point.*

Proof. The idea, which is borrowed from Euclidean space, is that the center of mass of any orbit must be a fixed point. So first find the period k of φ, i.e., the smallest integer such that $\varphi^k = id$. Second, for any $p \in M$ consider the orbit $\left\{p, \varphi(p), \ldots, \varphi^{k-1}(p)\right\}$ of p. Then construct the center of mass

$$q = \text{cm}\left\{p, \varphi(p), \ldots, \varphi^{k-1}(p)\right\}.$$

We claim that $\varphi(q) = q$. This is because the function

$$x \to f(x) = \max\left\{f_{0,p}(x), \ldots f_{0,\varphi^{k-1}(p)}(x)\right\}$$

has not only q as a minimum, but also $\varphi(q)$. To see this just observe that since φ is an isometry, we have

$$\begin{aligned} f(\varphi(q)) &= \max\left\{f_{0,p}(\varphi(q)), \dots f_{0,\varphi^{k-1}(p)}(\varphi(q))\right\} \\ &= \frac{1}{2}\left(\max\left\{d(\varphi(q), p), \dots, d\left(\varphi(q), \varphi^{k-1}(p)\right)\right\}\right)^2 \\ &= \frac{1}{2}\left(\max\left\{d\left(\varphi(q), \varphi^{k}(p)\right), \dots, d\left(\varphi(q), \varphi^{k-1}(p)\right)\right\}\right)^2 \\ &= \frac{1}{2}\left(\max\left\{d\left(q, \varphi^{k-1}(p)\right), \dots, d\left(q, \varphi^{k-2}(p)\right)\right\}\right)^2 \\ &= f(q). \end{aligned}$$

Therefore, the uniqueness of minima for strictly convex functions implies that $\varphi(q) = q$. □

Corollary 3.8 *If (M, g) is a complete Riemannian manifold of nonpositive curvature, then the fundamental group is torsion free; i.e., all nontrivial elements have infinite order.*

The second theorem requires a little more preparation and more careful analysis of distance functions. Suppose again that (M, g) is complete, simply connected and of nonpositive curvature. Let us fix a modified distance function: $x \to f_{0,p}(x)$ and a unit speed geodesic $\gamma : \mathbb{R} \to M$. The Hessian estimate from above implies that

$$\frac{d^2}{dt^2}\left(f_{0,p}\circ\gamma\right) \geq 1.$$

Integrating this twice yields

$$\begin{aligned} (d(p, \gamma(t)))^2 &\geq (d(p, \gamma(0)))^2 + 2g\left(\nabla f_{0,p}, \dot{\gamma}(0)\right)\cdot t + t^2 \\ &= (d(p, \gamma(0)))^2 + (d(\gamma(0), \gamma(t)))^2 \\ &\quad - 2d(p, \gamma(0))\, d(\gamma(0), \gamma(t)) \cos\angle\left(\nabla f_{0,p}, \dot{\gamma}(0)\right). \end{aligned}$$

Thus, if we have a triangle in M with sides lengths a, b, c and where the angle opposite a is α, then

$$a^2 \geq b^2 + c^2 - 2bc\cos\alpha.$$

From this, one can conclude that the angle sum in any triangle is $\leq \pi$, and more generally that the angle sum in any quadrilateral is $\leq 2\pi$. See Figure 6.1.

Now suppose that (M, g) has negative curvature. Then it must follow that all of the above inequalities are strict, unless p lies on the geodesic γ. In particular, the angle sum in any (nondegenerate) quadrilateral is $< 2\pi$. With this we can now show

Theorem 3.9 (Preissmann, 1943) *If (M, g) is a compact manifold of negative curvature, then any Abelian subgroup of the fundamental group is cyclic. In particular, no compact product manifold $M \times N$ admits a metric with negative curvature.*

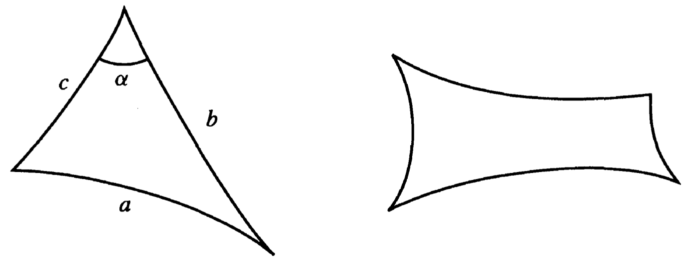


FIGURE 6.1.

Proof. We think of the fundamental group $\pi_1(M)$ as acting by isometries on the universal covering $\tilde{M}$, and fix $\alpha \in \pi_1(M)$. An *axis* for α is a geodesic $\gamma : \mathbb{R} \to \tilde{M}$ such that $\alpha(\gamma) \subset \gamma$. Since isometries map geodesics to geodesics, it must follow that

$$\alpha \circ \gamma(t) = \gamma(t + a).$$

Namely, α translates the geodesic either forward or backward. It is not possible for α to reverse the orientation of γ so that

$$\alpha \circ \gamma(t) = \gamma(-t + a),$$

because then we would have a fixed point

$$\alpha\left(\gamma\left(\frac{a}{2}\right)\right) = \gamma\left(\frac{a}{2}\right).$$

The uniquely defined number a is called the *period* of α along γ.

We now claim two things: first, that axes exist for the given α, and second, that they are unique when the curvature is negative.

To prove the first claim consider the *displacement function*

$$x \to d(x, \alpha(x)).$$

Since M is compact and α has no fixed points, there must be a point p where this function attains its minimum. We now claim that the unique geodesic γ going through p and $\alpha(p)$ is an axis with period $d(p, \alpha(p))$. Since geodesics on $\tilde{M}$ are uniquely determined by any two points on them, it suffices to check that $\alpha^2(p)$ lies on γ. If it doesn't, then take $x = \gamma(t)$, where $t \in (0, d(p, \alpha(p)))$, and observe that the following triangle inequality must be strict:

$$\begin{aligned} d(x, \alpha(x)) &< d(x, \alpha(p)) + d(\alpha(p), \alpha(x)) \\ &= d(x, \alpha(p)) + d(p, \gamma(t)) \\ &= d(p, \alpha(p)) - t + t \\ &= d(p, \alpha(p)), \end{aligned}$$

thus showing that p is not a minimum for the displacement function. We have in Figure 6.2 a pictorial proof, where x is chosen as the midpoint on the segment joining p and $\alpha(p)$.

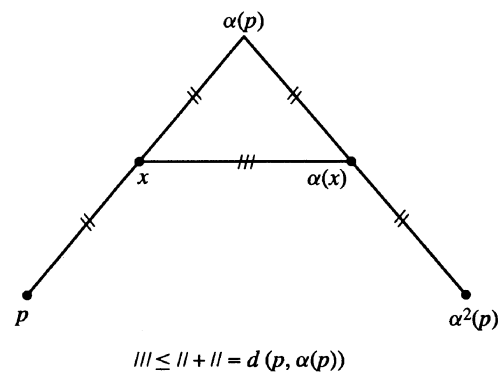


FIGURE 6.2.

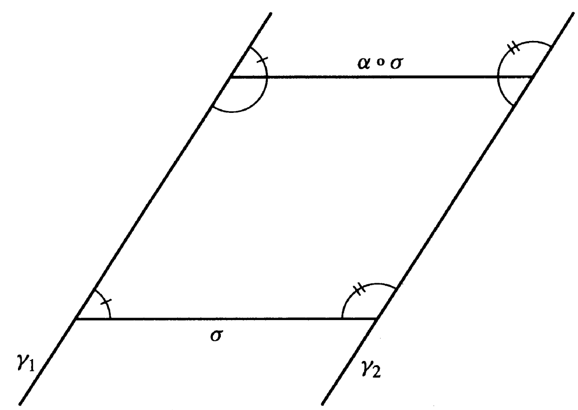


FIGURE 6.3.

To see that axes are unique in negative curvature, assume that we have two different axes γ_1 and γ_2 for α. If these intersect in one point, they must, by virtue of being invariant under α, intersect in at least two points. But then they must be equal. We can therefore assume that they do not intersect. Then pick $p_1 \in \gamma_1$ and $p_2 \in \gamma_2$, and join these points by a segment σ. Then $\alpha \circ \sigma$ is a segment from $\alpha(p_1)$ to $\alpha(p_2)$. Since α is an isometry that preserves γ_1 and γ_2, we see that the adjacent angles along the two axes formed by the quadrilateral $p_1, p_2, \alpha(p_1), \alpha(p_2)$ must add up to π (see also Figure 6.3). But then the angle sum is 2π, which is not possible unless the quadrilateral is degenerate. That is, all points lie on one geodesic.

Now pick an element $\beta \in \pi_1(M)$ that commutes with α. First, note that β preserves the unique axis γ for α, since

$$\begin{aligned}\beta(\gamma) &= \beta(\alpha(\gamma)) \\ &= \alpha(\beta(\gamma))\end{aligned}$$

implies that $\beta \circ \gamma$ is an axis for α, and must therefore be γ itself. Then, consider the group H generated by α, β. Any element in this group has γ as an axis. Thus we get a map

$$H \to \mathbb{R}$$

that sends an isometry to its uniquely defined period. Clearly, this map is a homomorphism with no kernel. Now, it is easy to check that any subgroup of $\mathbb{R}$ must either be cyclic or dense (like $\mathbb{Q}$). In the present case $H \subset \mathbb{R}$ must be discrete, since it acts discretely on $\tilde{M}$. □

It should be noted that the above techniques are, in a way, simply adapted from Euclidean space to nonpositive curvature, using the crucial fact that any two points lie on a unique "line" and the convexity of distance functions. As already pointed out, one can refine these techniques to get much stronger results.

6.4 Positive Sectional Curvature

In this section we shall prove some of the classical results for manifolds with positive sectional curvature. In contrast with the previous section, it is not possible to carry Euclidean geometry over to this setting. So while we try to imitate the results, new techniques are necessary.

6.4.1 The Diameter Estimate

Our first restriction on the topology of positively curved manifolds is

Theorem 4.1 *Suppose (M, g) is complete and satisfies* $\sec \geq k > 0$. *Then M is compact and satisfies* $\operatorname{diam}(M, g) \leq \frac{\pi}{\sqrt{k}} = \operatorname{diam} S_k^n$. *In particular, M has finite fundamental group.*

Proof. In polar coordinates the metric satisfies

$$\left(g_{\alpha\beta}\right)_{2\leq\alpha,\beta\leq n} \leq \operatorname{sn}_k^2(r) \cdot I$$

as long as we are inside the cutlocus. This implies that a conjugate point will always develop along any geodesic before r becomes $\pi/\sqrt{k}$. Thus, the diameter cannot exceed $\pi/\sqrt{k}$. Now use that the universal cover has the same curvature condition to conclude that it must also be compact. Thus, the fundamental group is finite. □

The history of this result is quite complicated. Bonnet proved it for convex surfaces in 1855; Synge in 1925 found the above estimate for the conjugate points, but failed to make the diameter bound conclusion partly because the Hopf-Rinow theorem wasn't yet available. Myers without giving Synge any credit then in 1935 established the above result using the Hopf-Rinow theorem.

6.4.2 *Hypersurfaces in Riemannian Manifolds*

For our next applications some discussion on the existence of various hypersurfaces is necessary.

Suppose we have an embedded hypersurface $H \hookrightarrow (M, g)$. We say that $H \hookrightarrow M$ is *orientable* if H has a normal field defined on all of H. Note that all hypersurfaces are locally orientable in this sense, regardless of whether H and/or M are orientable as manifolds. Given an orientable hypersurface $H \hookrightarrow (M, g)$, we can find an open subset $U \subset M$ and a smooth distance function $f : U \to \mathbb{R}$ such that $f^{-1}(0) = H$. This is done by coordinatizing a neighborhood of U using the normal exponential map. Namely, fix a unit normal field N to H in M (there are only two such fields). Then consider the exponential map

$$\exp : \mathbb{R} \times H \to M,$$
$$\exp(r, x) = \exp_x (r \cdot N(x)).$$

We know that at points $(0, x)$ the differential of this map is nonsingular and that $\exp_{|\{0\} \times H}$ is a diffeomorphism. There are therefore neighborhoods V around $H \times \{0\} \subset H \times \mathbb{R}$ and U around $H \subset M$ such that $\exp : V \to U$ is a diffeomorphism. On U we can therefore use coordinates $(r, x) \in V$. As was done earlier in the Gauss lemma, we can now show that on U the function $f(r, x) = r$ is a distance function with gradient ∂_r. Observe that $\partial_r = N$ on H. The *shape operator* S for H is the Hessian of f restricted to TH. If we introduce coordinates $x^2, \ldots, x^n$ on some part of H, then we have our usual coordinates $(x^1, x^2, \ldots, x^n) = (r, x^2, \ldots, x^n)$.

The hypersurface is said to be convex (respectively concave), with respect to our choice of normal N, if the shape operator is nonnegative (respectively nonpositive). By changing the normal one can obviously change convexity to concavity. Our picture is this: If we stand on the surface of Earth and look up into space, then the surface of Earth is convex. Usually, the normal that makes H convex is called the outward pointing normal, and the opposite direction the downward, or inward pointing, normal.

We know that the level sets of f are equidistant. They are therefore called the *equidistant hypersurfaces* to H.

The difference between nonnegative and nonpositive curvature in this context can now be stated as follows: Suppose we have an orientable hypersurface $H \hookrightarrow (M, g)$. The following statements are completely dual:

(1) If $\sec \geq 0$ and H is concave, then the level sets $H_t = f^{-1}(t)$ are concave for $t \geq 0$.

(2) If $\sec \leq 0$ and H is convex, then the level sets $H_t = f^{-1}(t)$ are convex for $t \geq 0$.

This does not mean that studying these two classes of Riemannian manifolds is dual in any way. Recall that distance functions to a point always have convex level sets near the point. This is clearly important in nonpositive curvature and was the key point in the Hadamard-Cartan theorem. In nonnegative curvature, however, we gain no information. One of the challenges in nonnegative curvature is to find concave surfaces that can be used in a nontrivial way.

It is, in general, very hard to find hypersurfaces that are everywhere concave. Instead, we shall be a little more modest and consider hypersurfaces H that are *concave on some connected subset* $A \subset H$. This just means that with our choice of normal N the shape operator for H is nonpositive on T_xH for all $x \in A$.

Suppose we have $p \in M$ and a unit vector $e \in T_pM$. Then there is an orientable hypersurface H containing p such that the normal at p is e and the shape operator at p is identically zero. Let $V \subset T_pM$ denote the orthogonal complement to e, and consider the exponential map $\exp_p : V \to M$. In a neighborhood of $0 \in T_pM$ this is an embedding whose image will be the desired hypersurface H. Clearly, $p \in H$, and e is normal to H. If we select a basis $e_1, \ldots, e_n$ for T_pM, where $e_1 = e$ and $e_2, \ldots, e_n \in V$, then the exponential map introduces a coordinate system $\varphi\left(x^1, \ldots, x^n\right) = \exp_p\left(\sum x^i e_i\right)$. The Gauss lemma tells us that the first coordinate vector field ∂_1 is a vector field that is perpendicular to H everywhere (H consists of geodesics going through p that are tangent to V). Furthermore, $\nabla_v \partial_1 = 0$ for all v. The unit normal field is, of course $N = \partial_1 / |\partial_1|$. Using that $\partial_k g_{ij} = 0$ at p, we see also that $\nabla N = 0$ at p, as desired.

Suppose now we have a geodesic $\gamma : (a, b) \to M$. We shall assume that it is either an embedding or a closed geodesic, i.e., can be considered as an embedding $\gamma : S^1 \to M$. Suppose also we have a unit parallel field E that is perpendicular to γ. Then we can find an orientable hypersurface H containing γ such that the normal on γ is E and such that the shape operator is identically zero on γ. The construction is, of course, similar. Namely, let $\nu(\gamma) = \left\{v \in T_{\gamma(t)}M : g(v, \dot{\gamma}(t)) = 0\right\}$ be the *normal bundle to* γ in M. Define

$$\begin{aligned} \exp &: \nu(\gamma) \to M, \\ \exp(v) &= \exp_{\gamma(t)} v. \end{aligned}$$

This is, as usual, a diffeomorphism on a neighborhood V of the zero section in $\nu(\gamma)$, onto some neighborhood U of γ in M. Now we have a section E of the normal bundle, so let $E^\perp$ denote the set of vectors in $\nu(\gamma)$ that are perpendicular to E. This is again a vector bundle over γ. Then define the hypersurface $H = \exp\left(E^\perp \cap V\right) \subset U$. We can then extend E to a normal field N on H. We now have to show that $\nabla_v N = 0$ for all $v \in T_{\gamma(t)}M$ that are perpendicular to E. First, note that as in the construction above, we have by the Gauss lemma that $\nabla_v N = 0$ if v is also perpendicular to $\dot{\gamma}(t)$. We then have to show that $\nabla_{\dot{\gamma}(t)} N = 0$, but this is true by definition, since the restriction of N to γ is the parallel field E. This finishes the construction. Note that it would not work if E were not parallel.

In the case of an embedded geodesic, such a parallel field always exists, but for a closed geodesic this is not so clear. However, if we fix $p = \gamma(t)$ on γ and consider parallel translation around γ, then we get a linear isometry $P : T_pM \to T_pM$. Since γ is a closed geodesic, we have that $P(\dot{\gamma}(t)) = \dot{\gamma}(t)$. Thus, P preserves the orthogonal complement to $\dot{\gamma}(t)$ in T_pM. Now recall that

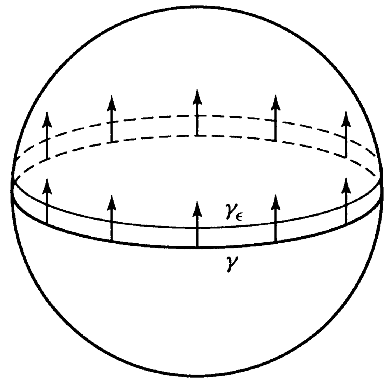


FIGURE 6.4.

if $L : \mathbb{R}^k \to \mathbb{R}^k$ is a linear isometry with $\det L = (-1)^{k+1}$, then L has 1 as an eigenvalue $\left(L(v) = v \text{forsome} v \in \mathbb{R}^k\right)$. We can use this to construct a closed parallel field around γ. Namely,

(1) If M is orientable and even-dimensional, then parallel translation around a closed geodesic preserves orientation and therefore has det $= 1$. Since the complement to $\dot{\gamma}(t)$ in T_pM is odd-dimensional we can therefore find a closed parallel field around γ.

(2) If M is not orientable, has odd dimension, and furthermore, γ is a nonorientable loop (this means that the orientation changes as we go around this loop), then parallel translation around γ is orientation reversing and therefore has det $= -1$. Now, the complement to $\dot{\gamma}(t)$ in T_pM is even-dimensional, and since $P(\dot{\gamma}(t)) = \dot{\gamma}(t)$, we have that the restriction of P to this even-dimensional subspace still has det $= -1$. Thus, we get a closed parallel field in this case as well.

In Figure 6.4 we have sketched what happens when the closed geodesic is the equator on the standard sphere. In this case there is only one choice for the parallel field, and the shorter curves are the latitudes close to the equator.

6.4.3 *The Fundamental Group in Even Dimensions*

Lemma 4.2 (Length comparison) *Suppose $H \hookrightarrow (M, g)$ is an orientable hypersurface, with normal N, which is concave on $A \subset H$, and suppose* $\sec M \geq 0$ *(respectively > 0). Then, for each $x \in A$ and sufficiently small $\varepsilon > 0$ we have*

$$\left(g_{ij}(\varepsilon, x)\right) \leq \left(g_{ij}(0, x)\right),$$
$$\left(g_{ij}(\varepsilon, x)\right) < \left(g_{ij}(0, x)\right).$$

In particular, if $\gamma : [a, b] \to A$ is a curve, then the nearby curves $\gamma_\varepsilon = (\varepsilon, \gamma(t))$ have shorter (or strictly shorter) length.

Proof. The metric satisfies

$$\begin{aligned}\partial_r \left(g_{ij}(r, x)\right) &= 2 \cdot \left(S_i^k\right) \cdot \left(g_{kj}(0, x)\right) \\ &\leq 0 \quad (\text{or} < 0)\end{aligned}$$

as the Hessian becomes negative (or strictly negative) by the hypothesis on sectional curvature. This implies our claim. □

We shall now prove an analogue to the Hadamard-Cartan theorem for positively curved manifolds.

Lemma 4.3 (Synge, 1936) *Let M be a compact manifold with* $\sec > 0$.

(1) *If M is even-dimensional and orientable, then M is simply connected.*
(2) *If M is odd-dimensional, then M is orientable.*

Proof. First let us show that any free homotopy class of loops contains a shortest loop, and that this loop is a closed geodesic. We can consider all curves $\gamma : S^1 \to M$ in a given homotopy class, which are parametrized proportionally to arclength and with length $\leq K$ for some large K. This family is then equicontinuous, and we can use the Arzela-Ascoli lemma to extract a sequence $\gamma_i \to \gamma$ that converges uniformly to some curve in this homotopy class, and such that $\ell(\gamma_i)$ converges to the infimum ℓ of the lengths of all curves in this class. First of all, this infimum is positive, because very short curves lie in coordinate charts and are therefore contractible. Secondly, the limit curve γ is a closed geodesic. For if γ is not smooth at t or $\ddot{\gamma}(t) \neq 0$, then we can find a shorter curve from $\gamma(t-\varepsilon)$ to $\gamma(t+\varepsilon)$ for small ε. Thus we could make γ shorter within the same homotopy class. On the same interval $(t-\varepsilon, t+\varepsilon)$ we could then also alter the γ_i's, and get a sequence of curves whose lengths are shorter by some fixed amount. This violates that $\ell(\gamma_i)$ converges to ℓ. Finally, γ has a well-defined length, which we claim is ℓ. To see this, partition S^1 into a union of intervals $[s_i, t_i]$ that overlap only on the end points, and such that $\ell\left(\gamma_{|[s_i,t_i]}\right) = d\left(\gamma(s_i), \gamma(t_i)\right)$. We can now use that

$$\begin{aligned}\ell(\gamma_k) &\geq \sum \ell\left(\gamma_{k|[s_i,t_i]}\right) \\ &\geq \sum d\left(\gamma_k(s_i), \gamma_k(t_i)\right) \\ &\underset{k\to\infty}{\to} \sum d\left(\gamma(s_i), \gamma(t_i)\right) \\ &= \ell(\gamma).\end{aligned}$$

Thus, we certainly have $\ell(\gamma) \leq \ell$. The other inequality is trivial.

The two parts of the proof are similar and go by contradiction. Namely, either take the shortest loop in some free homotopy class or find the shortest orientation-reversing loop in some free homotopy class. In both cases our assumptions are such that these loops are closed geodesics, which have perpendicular parallel fields by our discussion above. In positive curvature, however, we know from our length comparison result that such closed geodesics have nearby curves that are strictly shorter in length. As these nearby curves lie in the same homotopy class, we have arrived at a contradiction with the minimality of the length of the closed geodesic in this homotopy class. □

The first important conclusion we get from this result is that while $\mathbb{R}P^2 \times \mathbb{R}P^2$ has positive Ricci curvature (its universal cover $S^2 \times S^2$ has positive Ricci curvature), it cannot support a metric of positive sectional curvature. It is, however, completely unknown whether $S^2 \times S^2$ admits a metric of positive sectional curvature. This is known as the Hopf problem (there is also the other Hopf problem from Chapter 4 about the Euler characteristic). In Chapter 7 we study these issues in greater detail. Recall that above we showed, using fundamental group considerations, that no product manifold admits negative curvature. In this case, fundamental group considerations did not take us as far, since positively curved manifolds are often simply connected, something that never happens for compact negatively curved manifolds.

6.4.4 The Injectivity Radius in Even Dimensions

We get another interesting restriction on the geometry of positively curved manifolds.

Lemma 4.4 (Klingenberg, 1959) *Suppose* (M, g) *is an orientable even-dimensional manifold with* $0 < \sec \leq 1$. *Then* $\operatorname{inj}(M, g) \geq \pi$. *If* M *is not orientable, then* $\operatorname{inj}(M, g) \geq \pi/2$.

Proof. The nonorientable case follows from the orientable case, as the orientation cover has $\operatorname{inj}(M, g) \geq \pi$. From our previous discussion on the injectivity radius, we know that the upper curvature bound implies that if $\operatorname{inj} M < \pi$, then it must be realized by a closed geodesic. So let us assume that we have a closed geodesic $\gamma : [0, 2\operatorname{inj} M] \to M$ parametrized by arclength, where $2\operatorname{inj} M < 2\pi$. Since M is orientable and even-dimensional, we know that for all small $\varepsilon > 0$ there are curves $\gamma_\varepsilon : [0, 2\operatorname{inj} M] \to M$ that converge to γ as $\varepsilon \to 0$ and with $\ell(\gamma_\varepsilon) < \ell(\gamma) = 2\operatorname{inj} M$. The latter condition implies that $d = d(\cdot, \gamma_\varepsilon(0))$ is smooth when restricted to γ_ε, since $\gamma_\varepsilon \subset B(\gamma_\varepsilon(0), \operatorname{inj} M)$. Thus, if $\gamma_\varepsilon(t_\varepsilon)$ is the point at maximal distance from $\gamma_\varepsilon(0)$ on γ_ε, we have that there is a unique segment σ_ε from $\gamma_\varepsilon(0)$ to $\gamma_\varepsilon(t_\varepsilon)$, and this segment is perpendicular to γ_ε at $\gamma_\varepsilon(t_\varepsilon)$. As $\varepsilon \to 0$ we have that $t_\varepsilon \to \operatorname{inj} M$, and thus the segments σ_ε must subconverge to a segment from $\gamma(0)$ to $\gamma(\operatorname{inj} M)$, which is perpendicular to γ at $\gamma(\operatorname{inj} M)$. However, as the conjugate radius is $\geq \pi > \operatorname{inj} M$, and γ is a geodesic loop realizing the injectivity radius at $\gamma(0)$, we know that there can be only two segments from $\gamma(0)$ to $\gamma(\operatorname{inj} M)$. Thus, we have a contradiction with our assumption $\pi > \operatorname{inj} M$. □

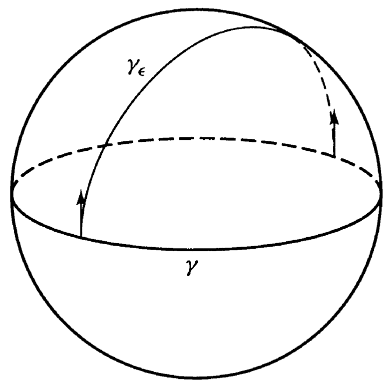


FIGURE 6.5.

In Figure 6.5 we have pictured a fake situation, which still gives the correct idea of the proof. The closed geodesic is the equator on the standard sphere, and σ_ε converges to a segment going through the north pole.

A similar result can clearly not hold for odd-dimensional manifolds. In dimension 3 we have the quotients of spheres $S^3/\mathbb{Z}_k$ for all positive integers k. Here the image of the Hopf fiber via the covering map $S^3 \to S^3/\mathbb{Z}_k$ is a closed geodesic of length $(2\pi)/k$ which goes to 0 as $k \to \infty$. Also, the Berger spheres $\left(S^3, g_\varepsilon\right)$ give counterexamples, as the Hopf fiber is a closed geodesic of length $2\pi\varepsilon$. In this case the curvatures lie in $\left[\varepsilon^2, 4 - 3\varepsilon^2\right]$. So if we rescale the upper curvature bound to be 1, the length of the Hopf fiber becomes $2\pi\varepsilon\sqrt{4 - 3\varepsilon^2}$ and the curvatures will lie in the interval $\left[\varepsilon^2/(4 - 3\varepsilon^2), 1\right]$. When $\varepsilon < \sqrt{3}/3$, the Hopf fibers have length $< 2\pi$. In this case the lower curvature bound becomes smaller than $1/9$.

A much deeper result by Klingenberg asserts that if a simply connected manifold has all its sectional curvatures in the interval $(\frac{1}{4}, 1]$, then the injectivity radius is still $\geq \pi$. This result has been improved first by Klingenberg-Sakai and Cheeger-Gromoll to allow for the curvatures to be in $\left[\frac{1}{4}, 1\right]$. Recently, Abresch-Meyer showed that the injectivity radius estimate still holds if the curvatures are in $\left[\frac{1}{4} - 10^{-6}, 1\right]$. The Berger spheres show that such an estimate will not hold if the curvatures are allowed to be in $\left[\frac{1}{9} - \varepsilon, 1\right]$. Notice that the simply connected hypothesis is necessary in order to eliminate all the constant-curvature spaces with small injectivity radius.

6.5 Further Study

Several textbooks treat the material mentioned in this chapter, and they all use variational calculus. We especially recommend [19] and [50]. The latter also discusses in more detail closed geodesics and, more generally, minimal maps and surfaces in Riemannian manifolds.

As we won't have recourse to discuss manifolds of nonpositive curvature again some references for this subject should be mentioned here (see, however, the holonomy discussion in Chapter 8). With the knowledge we have right now, it shouldn't be too hard to read the books [8] and [6]. For a more advanced account we recommend the survey by Eberlein-Hammenstad-Schroeder in [41].

6.6 Exercises

1. Show that in even dimensions only the sphere and real projective space have constant positive curvature.

2. Let (M, g) be a complete n-manifold of constant curvature k. Select a linear isometry $L : T_pM \to T_{\bar{p}}S^n_k$. When $k \leq 0$ show that $\exp_p \circ L^{-1} \circ \exp_{\bar{p}}^{-1} : S^n_k \to M$ is a Riemannian covering map. When $k > 0$ show that $\exp_p \circ L^{-1} \circ \exp_{\bar{p}}^{-1} : S^n_k - \{-\bar{p}\} \to M$ extends to a Riemannian covering map $S^n_k \to M$. (Hint: Use that the differential of the exponential maps is controlled by the metric, which in turn can be computed when the curvature is constant. You should also use the conjugate radius ideas presented in connection with the Hadamard-Cartan theorem.)

3. A Riemannian manifold is said to be *k-point homogeneous* if for all pairs of points $(p_1, \ldots, p_k)$ and $(q_1, \ldots, q_k)$ with $d(p_i, p_j) = d(q_i, q_j)$ there is an isometry φ with $\varphi(p_i) = q_i$. When $k = 1$ we simply say that the space is homogeneous.

   (a) Show that a homogenous space has constant scalar curvature.

   (b) Show that if $k > 2$ and (M, g) is k-point homogeneous, then M is also $(k-1)$-point homogeneous.

   (c) Show that if (M, g) is two-point homogeneous, then (M, g) is an Einstein metric.

   (d) Show that if (M, g) is three-point homogeneous, then (M, g) has constant curvature. Classify all three-point homogeneous spaces.

4. Show that if G is an Abelian group that is the subgroup of the fundamental group of a manifold with constant curvature, then either the manifold is flat or G is cyclic.

5. Let $M \to N$ be a Riemannian k-fold covering map. Show $\mathrm{vol}M = k \cdot \mathrm{vol}N$.

6. Starting with a geodesic on a two-dimensional space form, discuss how the equidistant curves change as they move away from the original geodesic.

7. Introduce polar coordinates $(r, \theta) \in (0, \infty) \times S^{n-1}$ on a neighborhood U around a point $p \in (M, g)$. If (M, g) has $\sec \geq 0$ ($\sec \leq 0$), show that any curve $\gamma(t) = (r(t), \theta(t))$ is shorter (longer) in the metric g than in the Euclidean metric on U.

8. Around an orientable hypersurface $H \hookrightarrow (M, g)$ introduce the normal coordinates $(r, x) \in \mathbb{R} \times H$ on some neighborhood U around H. On U we have aside from the given metric g, also the radially flat metric $dt^2 + g_0$, where g_0 is the restriction of g to H. If M has sec ≥ 0 (sec ≤ 0) and $\gamma(t) = (r(t), x(t))$ is a curve, where $r \geq 0$ and the shape operator is ≤ 0 (≥ 0) at $x(t)$ for all t, show that γ is shorter (longer) with respect to g than with respect to the radially flat metric $dt^2 + g_0$.

9. An isometric immersion $(A, h) \looparrowright (M, g)$ is said to be totally geodesic iff the connection on (A, h) is the same as the restriction of the connection from M to A. Show that an immersion is totally geodesic iff all geodesics in A are mapped to geodesics in M.

10. (Frankel) Let M be an n-dimensional Riemannian manifold of positive curvature and A, B two totally geodesic submanifolds. Show that A and B must intersect if $\dim A + \dim B \geq n - 1$. Hint: assume that A and B do not intersect. Then find a segment of shortest length from A to B. Show that this segment is perpendicular to each submanifold. Then use the dimension condition to find a parallel field along this geodesic that is tangent to A and B at the endpoints to the segments. Finally use the length comparison to get a shorter curve from A to B.

11. Generalize Preissmann's theorem to show that any solvable subgroup of the fundamental group must be cyclic.

12. Let (M, g) be an oriented manifold of positive curvature and suppose we have an isometry $\varphi : M \to M$ of finite order without fixed points. Show that if $\dim M$ is even, then φ must be orientation reversing, while if $\dim M$ is odd, it must be orientation preserving.

13. Use an analog of Cartan's result on isometries of finite order in nonpositive curvature to show that any closed manifold of constant curvature $= 1$ must either be the standard sphere or have diameter $\leq \pi/2$. Generalize this to show that any closed manifold with sec ≥ 1 is either simply connected or has diameter $\leq \pi/2$. In Chapter 11 we shall show the stronger statement that a closed manifold with sec ≥ 1 and diameter $> \pi/2$ must in fact be homeomorphic to a sphere.

7
The Bochner Technique

One of the oldest and most important techniques in modern Riemannian geometry is that of the Bochner technique. In this chapter we shall prove some of the classical theorems Bochner proved about obstructions to the existence of Killing fields and harmonic 1-forms. We also explain how the Bochner technique extends to forms. This will in the next chapter lead us to a classification of compact manifolds with nonnegative curvature operator. To establish the relevant Bochner formula for forms, we have used the language of Clifford multiplication. It is, in our opinion, much easier to work consistently with Clifford multiplication, rather than trying to keep track of wedge products, interior products, Hodge star operators, exterior derivatives, and their dual counterparts. In addition, it has the effect of preparing one for the world of spinors, although we won't go into this here. In the last section we give a totally different application of the Bochner technique. In effect, we try to apply it to the curvature tensor itself. The outcome will be used in the next chapter, where manifolds with nonnegative curvature operator will be classified. The Bochner technique on spinors is only briefly mentioned in this chapter, but Appendix C is devoted to this subject.

It should be noted that we have not used a unified approach to the Bochner technique. There are many equivalent approaches and we have tried to discuss a few of them here. It is important to learn how it is used in its various guises, as one otherwise could not prove some of the results we present. We have for pedagogical reasons used Stokes' theorem throughout rather than the maximum principle. The reason is that one can then cover the material without any knowledge of geodesic geometry (Killing fields being the only exception to this rule). The maximum principle in the strongest possible form is established and used in Chapter 9. The interested reader is encouraged to learn about it there, and then go back and try it out in connection with the Bochner technique.

The Bochner technique was, as the name indicates, invented by Bochner. However, Bernstein knew about it for harmonic functions on domains in Euclidean space. Specifically, he used

$$\Delta\frac{1}{2}\left|\nabla u\right|^2 = \left|\nabla^2 u\right|^2,$$

where $u : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ and $\Delta u = 0$. It was really Bochner's genius that realized that when the same trick is attempted on Riemannian manifolds, a curvature term also appears. Namely, for $u : (M, g) \to \mathbb{R}$ with $\Delta_g u = 0$ one has

$$\Delta\frac{1}{2}\left|\nabla u\right|^2 = \left|\nabla^2 u\right|^2 + \operatorname{Ric}(\nabla u, \nabla u).$$

With this it is clear that curvature influences the behavior of harmonic functions. The next nontrivial step Bochner took was to realize that one could try to compute $\Delta\frac{1}{2}\left|\omega\right|^2$ for any harmonic form ω and then try to get information about the topology of the manifold. The key ingredient here is of course Hodge's theorem, which states that any homology class can be represented by a harmonic form. Yano further refined the Bochner technique, but it seems to be Lichnerowitz who really put things into gear, when around 1960 he presented his formulae for the Laplacian on forms and spinors. After this work, Berger, D. Meyer, Gallot, Gromov-Lawson, Witten, and others have all made significant contributions to this tremendously important subject.

7.1 Killing Fields

We shall see how Killing fields interact with curvature in various settings. But first we need to establish some general properties.

7.1.1 *Killing Fields in General*

A vector field X on a Riemannian manifold (M, g) is called a *Killing field* if the local flows generated by X act by isometries. This obviously means that X leaves g invariant. In other words, X is a Killing field iff $L_X g \equiv 0$.

Proposition 1.1 *X is a Killing field iff $v \to \nabla_v X$ is a skew symmetric* $(1, 1)$-*tensor.*

Proof. Note that

$$\begin{aligned}(L_X g)(Y, Z) &= L_X g(Y, Z) - g(L_X Y, Z) - g(Y, L_X Z) \\ &= (\nabla_X g)(Y, Z) + g(\nabla_Y X, Z) + g(Y, \nabla_Z X) \\ &= g(\nabla_Y X, Z) + g(Y, \nabla_Z X).\end{aligned}$$

Thus, $L_X g \equiv 0$ iff $g(\nabla_Y X, Z) = -g(Y, \nabla_Z X)$ for all Y, Z. □

Proposition 1.2 *For a given $p \in M$, a Killing field X is uniquely determined by $X(p)$ and $(\nabla X)(p)$.*

Proof. The equation $L_X g \equiv 0$ is linear in X, so the space of Killing fields is a vector space. It therefore suffices to show that if $X(p) = 0$ and $(\nabla X)(p) = 0$, then $X \equiv 0$ on M. Using an open-closed argument, we can reduce our considerations to a neighborhood of p.

Let φ^t be the local flow for X near p. The condition $X(p) = 0$ implies that $\varphi^t(p) = p$ for all t. Thus $D\varphi^t : T_pM \to T_pM$. We claim that also $D\varphi^t = I$. This follows from the fact that X commutes with any vector field at p:

$$\begin{aligned}[X, Y](p) &= \nabla_{X(p)} Y - \nabla_{Y(p)} X \\ &= \nabla_0 Y - 0 = 0.\end{aligned}$$

As the flow diffeomorphisms act by isometries, we can conclude that they must be the identity map, and hence $X = 0$ in a neighborhood of p. □

Proposition 1.3 *The set of Killing fields $\mathfrak{iso}(M, g)$ forms a Lie algebra of dimension $\leq ((n+1)n)/2$. Furthermore, if M is compact, then $\mathfrak{iso}(M, g)$ is the Lie algebra of* $\mathrm{Iso}(M, g)$.

Proof. Note that $L_{[X,Y]} = [L_X, L_Y]$. So if $L_X g = L_Y g = 0$, we also have that $L_{[X,Y]} g = 0$. Thus, $\mathfrak{iso}(M, g)$ does form a Lie algebra. We have just seen that the map $X \to (X(p), (\nabla X)(p))$ is linear and has trivial kernel. So

$$\begin{aligned}\dim(\mathfrak{iso}(M, g)) &\leq \dim T_pM \\ &\quad + \dim(\text{skew-symmetric transformations of } T_pM) \\ &= n + \frac{n(n-1)}{2} = \frac{(n+1)n}{2}.\end{aligned}$$

The last statement is not easy to prove. Observe, however, that since M is compact, each vector field generates a global flow on M. Each Killing field therefore generates a 1-parameter subgroup of $\mathrm{Iso}(M, g)$. If we take it for granted that $\mathrm{Iso}(M, g)$ is a Lie group, then the identity component is of course generated by the 1-parameter subgroups, and each such group by definition generates a Killing field. □

With a little more work one can prove the previous theorem for complete manifolds as well.

Recall that $\dim(\mathrm{Iso}(S^n_k)) = ((n+1)n)/2$. Thus, all space forms have maximal dimension for their isometry groups. If we consider other complete spaces with constant curvature, then we know they look like S^n_k / Γ, where $\Gamma \subset \mathrm{Iso}(S^n_k)$ acts freely and discontinuously on S_k^n. The isometries on the quotient S_k^n / Γ can now be identified with those isometries of $\mathrm{Iso}(S_k^n)$ that commute with all elements in Γ. At least when $k \neq 0$, we know that all elements in $\mathrm{Iso}(S_k^n)$ are linear maps on a vector space. Now, two linear maps commute iff they are simultaneously diagonalizable. The dimension of $\mathrm{Iso}(S_k^n / \Gamma)$ is therefore heavily reduced unless the elements of Γ are homotheties. Thus, Γ can essentially only be $\{I, -I\}$. But $-I$ acts freely only on the sphere. Thus, only one other constant-curvature space form has maximal dimension for the isometry group, namely $\mathbb{R}P^n$.

More generally, one can prove that if (M, g) is complete and $\dim \mathrm{Iso}(M, g) = (n(n+1))/2$, then (M, g) has constant curvature. To see this, we need a new construction. The frame bundle FM of (M, g) is the set $\{(p, e_1, \ldots, e_n) : p \in M$ and $e_1, \ldots, e_n$ forms an orthonormal basis for $T_pM\}$. It is not hard to see that this is a manifold of dimension $(n(n+1))/2$. Any isometry $\varphi : M \to M$ induces a map of FM by sending $(p, e_1, \ldots, e_n)$ to $(\varphi(p), D\varphi e_1, \ldots, D\varphi e_n)$. By the uniqueness theorem for isometries, we see that the induced action of $\mathrm{Iso}(M)$ on FM cannot have any fixed points. So $\mathrm{Iso}(M)$ can be thought of as a submanifold of FM. In the case where M is compact, we therefore see that $\mathrm{Iso}(M)$ acts transitively on FM if $\dim \mathrm{Iso}(M) = (n(n+1))/2$. (This is also true even if we only assume that (M, g) is complete.) Thus any two orthonormal frames on M can be mapped to each other by an isometry of M. This clearly shows that M has constant curvature.

7.1.2 Killing Fields in Negative Ricci Curvature

We shall use the language of norms of tensors on Riemannian manifolds. For a $(1, 1)$-tensor T the norm is

$$|T|^2 = \mathrm{tr}\left(T \circ T^*\right),$$

where T^* is the adjoint. In case T is skew-symmetric or skew-adjoint, we therefore have

$$|T|^2 = -\mathrm{tr}\left(T^2\right).$$

Recall also from Chapter 2 the second covariant derivative $\nabla^2_{V,W}X = \nabla_V\nabla_W X - \nabla_{\nabla_V W}X$, which is tensorial in V and W.

Proposition 1.4 *Let X be a Killing field on (M, g) and consider the function $f = \frac{1}{2}g(X, X) = \frac{1}{2}|X|^2$. If we define a skew-adjoint $(1, 1)$-tensor by $T(v) = \nabla_v X$, then*

(1) $\nabla f = -T(X) = -\nabla_X X$.
(2) $\nabla^2 f = -T^2 - \nabla_X T - R_X$.
(3) $\Delta f = -\mathrm{Ric}(X, X) + |T|^2 = -\mathrm{Ric}(X, X) + |\nabla X|^2$.

Proof. To see (1), observe that

$$g(V, \nabla f) = D_V f$$

$$
\begin{aligned}
&= g(\nabla_V X, X) \\
&= -g(V, \nabla_X X).
\end{aligned}
$$

For (2), we compute

$$
\begin{aligned}
\nabla^2 f(V) &= \nabla_V(-\nabla_X X) \\
&= -R(V, X)X - \nabla_X \nabla_V X - \nabla_{[V,X]} X \\
&= -R_X(V) - \nabla_X \nabla_V X + \nabla_{\nabla_X V} X - \nabla_{\nabla_V X} X \\
&= -R_X(V) - T \circ T(V) - \nabla_X \nabla_V X + \nabla_{\nabla_X V} X \\
&= -R_X(V) - T \circ T(V) - \nabla^2_{X,V} X.
\end{aligned}
$$

For (3) we just take traces of (2). The problem is that the last quantity should be traceless. Note, however, that it is simply $(\nabla_X T)(V)$. As T itself is skew-symmetric, the covariant derivative should of course also be skew-symmetric. A direct calculation shows this:

$$
\begin{aligned}
g\left(-\nabla_X \nabla_V X + \nabla_{\nabla_X V} X, V\right) &= -g(\nabla_X \nabla_V X, V) + g\left(\nabla_{\nabla_X V} X, V\right) \\
&= -g(\nabla_X \nabla_V X, V) - g(\nabla_V X, \nabla_X V) \\
&= -\nabla_X g(\nabla_V X, V) \\
&= -\nabla_X 0 \\
&= 0,
\end{aligned}
$$

where we used skew symmetry of ∇X for the second and penultimate equalities. □

Theorem 1.5 (Bochner, 1946) *Suppose (M, g) is compact, oriented, and has* $\mathrm{Ric} \le 0$. *Then every Killing field is parallel. Furthermore, if* $\mathrm{Ric} < 0$, *then there are no nontrivial Killing fields.*

Proof. If we define $f = \frac{1}{2}|X|^2$ for a Killing field X, then using Stokes' theorem and the condition $\mathrm{Ric} \le 0$ gives us

$$
\begin{aligned}
0 &= \int_M \Delta f \cdot d\mathrm{vol} \\
&= \int_M \left(-\mathrm{Ric}(X, X) + |\nabla X|^2\right) \cdot d\mathrm{vol} \\
&\ge \int_M |\nabla X|^2 \cdot d\mathrm{vol} \\
&\ge 0.
\end{aligned}
$$

Thus $|\nabla X| \equiv 0$ and X must be parallel.

If, in addition, $\mathrm{Ric} < 0$, then $\mathrm{Ric}(X, X) \equiv 0$ iff $X \equiv 0$. Since X is parallel, we have $0 = \Delta f \le \mathrm{Ric}(X, X) \le 0$, which shows that $X \equiv 0$. □

Corollary 1.6 *With (M, g) as in the theorem, we have* $\dim(\mathfrak{iso}(M, g)) = \dim(\mathrm{Iso}(M, g)) \leq \dim M$, *and* $\mathrm{Iso}(M, g)$ *is finite if* $\mathrm{Ric}(M, g) < 0$.

Proof. Since any Killing field is parallel, the linear map: $X \to X(p)$ from $\mathfrak{i}(M, g)$ to T_pM is injective. This gives the result. For the second part observe that $\mathrm{Iso}(M, g)$ is compact, since M is compact, and the identity component is trivial. □

Corollary 1.7 *With (M, g) as before and $p = \dim(\mathfrak{iso}(M, g))$, we have that the universal covering splits isometrically as $\widetilde{M} = \mathbb{R}^p \times N$.*

Proof. On $\widetilde{M}$ there are p linearly independent parallel vector fields, which we can assume to be orthonormal. Since $\widetilde{M}$ is simply connected, each of these vector fields is the gradient field for a distance function. Thus we have a Riemannian submersion $\widetilde{M} \to \mathbb{R}^r$ with totally geodesic fibers (Hessian $\equiv 0$ for the distance functions). This gives the desired splitting. □

The result about nonexistence of Killing fields can actually be slightly improved to yield

Theorem 1.8 *Suppose (M, g) is a compact manifold with quasi-negative Ricci curvature, i.e.,* $\mathrm{Ric} \leq 0$ *and* $\mathrm{Ric}(v, v) < 0$ *for all* $v \in T_pM - \{0\}$ *for some* $p \in M$. *Then (M, g) admits no nontrivial Killing fields.*

Proof. We already know that any Killing field is parallel. Thus a Killing field is always zero or never zero. If the latter holds, then $\mathrm{Ric}(X, X)(p) < 0$, but this contradicts $0 = \Delta f(p) = -\mathrm{Ric}(X, X)(p) > 0$. □

This theorem has been generalized by X. Rong to a more general statement, which states that a closed Riemannian manifold with negative Ricci curvature can't admit a pure F-structure of positive rank (see [74] for the definition of F structure and proof of this). Given a closed Riemannian manifold, this essentially means that we have a finite covering of open sets U_i, and on each open set U_i there is a nowhere vanishing Killing field X_i. Furthermore, these Killing fields commute whenever they are defined at the same point, i.e., $[X_i, X_j] = 0$ on $U_i \cap U_j$. The idea of the proof is then to consider the function

$$f = \det\left(g\left(X_i, X_j\right)\right)_{i,j}.$$

If only one vector field is given on all of M, then this reduces to the function

$$f = g(X, X)$$

that we considered above. For the above expression one must show that it has a similar Bochner formula, and also that it isn't too discontinuous.

7.1.3 Killing Fields in Positive Curvature

We can actually also say something about Killing fields in positive sectional curvature. Recall that any vector field on an even-dimensional sphere has a zero, since the Euler characteristic is 2 ($\neq 0$). At some point H. Hopf conjectured that in fact any even-dimensional manifold with positive sectional curvature has positive Euler characteristic. If the curvature operator is positive, this is certainly true. Thus, the conjecture holds in dimension 2. From Chapter 6 we know that $|\pi_1| < \infty$. In dim $= 4$, therefore, we have $H_1(M, \mathbb{R}) = H_3(M, \mathbb{R}) = 0$, and hence $\chi(M) = 1 + \dim H_2(M, \mathbb{R}) + 1 \geq 2$. In higher dimensions we have only the following partial justification for the Hopf conjecture:

Theorem 1.9 (Berger, 1965) *If (M, g) is a compact, even-dimensional manifold of positive sectional curvature, then every Killing field has a zero.*

Proof. Consider as before $f = \frac{1}{2}|X|^2$. If X has no zeros, f will have a positive minimum at some point $p \in M$. Then of course $\nabla^2 f(p) \geq 0$. We also know that

$$\begin{aligned} g(\nabla^2 f(V), V) &= -g\left(\nabla_{\nabla_V X} X, V\right) - g\left(R(V, X)X, V\right) - g\left(\nabla^2_{X,V} X, V\right) \\ &= g\left(\nabla_V X, \nabla_V X\right) - g\left(R(V, X)X, V\right), \end{aligned}$$

and by assumption, $g(R(V, X)X, V) > 0$ if X and V are linearly independent. Using this, we shall find V such that $g(\nabla^2 f(V), V) < 0$ near p, thus arriving at a contradiction.

Recall that the linear endomorphism $v \to \nabla_v X$ is skew-symmetric. Furthermore, $(\nabla_X X)(p) = 0$, since $\nabla f(p) = -(\nabla_X X)(p) = 0$, and f had a minimum at p. Thus, we have a skew-symmetric map $T_pM \to T_pM$ with at least one zero eigenvalue. But then, even dimensionality of T_pM ensures us that there must be at least one more zero eigenvector $v \in T_pM$ linearly independent from X. Thus,

$$\begin{aligned} g(\nabla^2 f(v), v) &= g(\nabla_v X, \nabla_v X) - g(R(v, X)X, v) \\ &= -g(R(v, X)X, v) < 0. \end{aligned}$$

□

In odd dimensions we get completely different information from the existence of a nontrivial Killing field.

Theorem 1.10 (X. Rong, 1995) *If a closed Riemannian n-manifold (M, g) admits a nontrivial Killing field, then the fundamental group has a cyclic subgroup of index $\leq c(n)$.*

The reader should consult [73] for a more general statement and the proof. Observe, however, that should such a manifold admit a free isometric action by S^1, then the quotient M/S^1 is a positively curved manifold as well (by the O'Neill formula from the exercises of Chapter 2). Thus, we have a fibration,

$$\begin{array}{ccc} S^1 & \to & M \\ & & \downarrow \\ & & M/S^1. \end{array}$$

In case M is even-dimensional, we know that $\pi_1(M)$ is either trivial or $\mathbb{Z}_2$. So in this case there is nothing to prove. Otherwise, the quotient is even-dimensional. The long exact sequence for the homotopy groups is thus

$$\pi_2\left(M/S^1\right) \to \mathbb{Z} \to \pi_1(M) \to \pi_1\left(M/S^1\right) = \begin{cases} \{1\}, \\ \mathbb{Z}_2. \end{cases}$$

Thus, $\pi_1(M)$ must either be a finite cyclic group or contain a finite cyclic subgroup of index 2.

7.2 Hodge Theory

The reader who is not familiar with de Rham cohomology might wish to consult Appendix A before proceeding.

Recall that on a manifold M we have the *de Rham complex*

$$0 \to \Omega^0(M) \overset{d^0}{\to} \Omega^1(M) \overset{d^1}{\to} \Omega^2(M) \to \cdots \overset{d^{n-1}}{\to} \Omega^n(M) \to 0,$$

where $\Omega^k(M)$ denotes the space of k-forms on M and $d^k : \Omega^k(M) \to \Omega^{k+1}(M)$ is exterior differentiation. The *de Rham cohomology* groups

$$H^k(M) = \frac{\ker(d^k)}{\operatorname{Im}(d^{k-1})}$$

compute the real cohomology of M. We know that $H^0(M) \simeq \mathbb{R}$ if M is connected, and $H^n(M) = \mathbb{R}$ if M is orientable and compact. In this case we have a pairing,

$$\begin{aligned} \Omega^k(M) \times \Omega^{n-k}(M) &\to \mathbb{R}, \\ (\omega_1, \omega_2) &\to \int_M \omega_1 \wedge \omega_2. \end{aligned}$$

This pairing induces a nondegenerate pairing $H^k(M) \times H^{n-k}(M) \to \mathbb{R}$ on the cohomology groups. The two vector spaces $H^k(M)$ and $H^{n-k}(M)$ are therefore dual to each other and in particular have the same dimension.

Now suppose M is endowed with a Riemannian metric g. Then each of the spaces $\Omega^k(M)$ is also endowed with a pointwise inner product structure: $\Omega^k(M) \times \Omega^k(M) \to \Omega^0(M)$. This structure is obtained by declaring that if $E_1, \ldots, E_n$ is an orthonormal frame, then the dual coframe $\sigma^1, \ldots, \sigma^n$ is also orthonormal, and furthermore that all the k-forms $\sigma^{i_1} \wedge \ldots \wedge \sigma^{i_k}$, $i_1 < \cdots < i_k$, are orthonormal. A different way of introducing this inner product structure is by first observing that k-forms of the form $f_0 \cdot df_1 \wedge \ldots \wedge df_k$, where $f_0, f_1, \ldots f_k \in \Omega^0(M)$, actually span $\Omega^k(M)$. The product on $\Omega^0(M)$ is obviously just the normal multiplication of functions. On $\Omega^1(M)$ we declare that

$$\begin{aligned} g\,(f_0 \cdot df_1,\ h_0 dh_1) &= f_0 \cdot h_0 \cdot g\,(df_1, dh_1) \\ &= f_0 \cdot h_0 \cdot g\,(\nabla f_1, \nabla h_1)\,, \end{aligned}$$

and on $\Omega^k(M)$ we define

$$\begin{aligned} &g\,(f_0 \cdot df_1 \wedge \ldots \wedge df_k,\ h_0 dh_1 \wedge \ldots \wedge dh_k) \\ &\qquad = f_0 \cdot h_0\, g\,(df_1 \wedge \ldots \wedge df_k,\ dh_1 \wedge \ldots \wedge dh_k) \\ &\qquad = f_0 \cdot h_0 \ \det\left(g\,(df_i, dh_j)_{1 \le i,j \le k}\right) \\ &\qquad = f_0 \cdot h_0 \ \det\left(g\,(\nabla f_i, \nabla h_j)_{1 \le i,j \le k}\right). \end{aligned}$$

By integrating this pointwise inner product we get a real inner product on $\Omega^k(M)$:

$$(\omega_1, \omega_2) = \int_M g(\omega_1, \omega_2) d\text{vol}, \quad \omega_1, \omega_2 \in \Omega^k(M).$$

Using this inner product we can define the *Hodge star operator*

$$* : \Omega^k(M) \to \Omega^{n-k}(M)$$

by the formula

$$(*\omega_1, \omega_2) = \int_M g(*\omega_1, \omega_2) d\text{vol} = \int_M \omega_1 \wedge \omega_2.$$

The Hodge operator gives us an isomorphism $* : H^k(M) \to H^{n-k}(M)$, which depends on the metric g. The fact that it is an isomorphism comes from the following result:

Lemma 2.1 *The square of the Hodge star* $*^2 : \Omega^k \to \Omega^k$ *is simply multiplication by* $(-1)^{k(n-k)}$.

Proof. The simplest way of showing this is to prove that if $\sigma^1, \ldots, \sigma^n$ is a positively oriented orthonormal coframe then

$$*\left(\sigma^1 \wedge \cdots \wedge \sigma^k\right) = \sigma^{k+1} \wedge \cdots \wedge \sigma^n.$$

This follows easily from the above definition of $*$, the fact that the forms $\sigma^{i_{k+1}} \wedge \cdots \wedge \sigma^{i_n}$, $i_{k+1} < \cdots < i_n$ are orthonormal, and that $\sigma^1 \wedge \cdots \wedge \sigma^n$ represents the volume form on (M, g). □

Using the inner product structures on $\Omega^k(M)$, we can define the adjoint $\delta^k : \Omega^{k+1}(M) \to \Omega^k(M)$ to d^k via the formula

$$(\delta^k \omega_1, \omega_2) = (\omega_1, d^k \omega_2).$$

Using the definition of the Hodge star operator, we see that δ can be computed from d as follows:

Lemma 2.2 $\delta^k = (-1)^{(n-k)(k+1)} * d^{n-k-1} *$.

Proof. If $\omega_1 \in \Omega^{k+1}(M)$, $\omega_2 \in \Omega^k(M)$, then

$$\begin{aligned}
(\delta^k \omega_1, \omega_2) &= (\omega_1, d^k \omega_2) \\
&= (-1)^{k(n-k)} (* * \omega_1, d^k \omega_2) \\
&= (-1)^{k(n-k)} \int_M *\omega_1 \wedge d^k \omega_2 \\
&= (-1)^{n-k-1+k(n-k)} \int_M d^{n-1}(*\omega_1 \wedge \omega_2) \\
&\quad - (-1)^{n-k-1+k(n-k)} \int_M (d^{n-k-1} * \omega_1) \wedge \omega_2 \\
&= (-1)^{n-k-1+k(n-k)} \int_{\partial M} *\omega_1 \wedge \omega_2 \\
&\quad + (-1)^{n-k+k(n-k)} \int_M (d^{n-k-1} * \omega_1) \wedge \omega_2 \\
&= (-1)^{(k+1)(n-k)} (*d^{n-k-1} * \omega_1, \omega_2).
\end{aligned}$$

□

Thus, we have a diagram of complexes,

$$\begin{array}{ccccccccc}
0 & \to & \Omega^0(M) & \overset{d}{\to} & \Omega^1(M) & \overset{d}{\to} \cdots \overset{d}{\to} & \Omega^n(M) & \to & 0 \\
 & & \updownarrow * & & \updownarrow * & & \updownarrow * & & \\
0 & \to & \Omega^n(M) & \overset{\delta}{\to} & \Omega^{n-1}(M) & \overset{\delta}{\to} \cdots \overset{\delta}{\to} & \Omega^0(M) & \to & 0,
\end{array}$$

where each square commutes up to some sign.

The *Laplacian on forms,* also called the *Hodge Laplacian,* can now be defined as

$$\begin{aligned}
\triangle : \Omega^k(M) &\to \Omega^k(M), \\
\triangle \omega &= (d\delta + \delta d)\omega.
\end{aligned}$$

In the next section we shall see that on functions, the Hodge Laplacian is the negative of the previously defined Laplacian, whence the slightly different symbol $\triangle$ instead of Δ.

Lemma 2.3 $\triangle \omega = 0$ *iff* $d\omega = 0$ *and* $\delta \omega = 0$.

Proof.

$$\begin{aligned}(\Delta\omega, \omega) &= (d\delta\omega, \omega) + (\delta d\omega, \omega)\\ &= (\delta\omega, \delta\omega) + (d\omega, d\omega).\end{aligned}$$

Thus, $\Delta\omega = 0$ implies $(\delta\omega, \delta\omega) = (d\omega, d\omega) = 0$, which shows that $\delta\omega = 0, d\omega = 0$. The opposite direction is obvious. □

We can now introduce the *Hodge cohomology*:

$$\mathcal{H}^k(M) = \{\omega \in \Omega^k(M) : \Delta\omega = 0\}.$$

Since all harmonic forms are closed, we obviously have a natural map:

$$\mathcal{H}^k(M) \to H^k(M).$$

Theorem 2.4 (Hodge, 1935) *This map is an isomorphism.*

Proof. The proof of this theorem actually requires a lot of work and can't really be proved here. A good reference for a rigorous treatment is [72]. We'll try to give the essential idea. The claim, in other words, is that for any closed form ω we can find a unique exact form $d\theta$ such that $\Delta(\omega + d\theta) = 0$. The uniqueness part is obviously equivalent to the statement that the harmonic exact forms are zero everywhere. Another way of getting at the result is by observing that since $H^k(M) = (\ker d^k)/(\mathrm{Im}(d^{k-1}))$, we would be done if we could only prove that $\Omega^k(M) = \mathrm{Im} d^{k-1} \oplus \ker(\delta^{k-1})$. This statement is actually quite reasonable from the point of view of linear algebra in finite dimensions. There we know that $W = \mathrm{Im}(L) \oplus \ker(L^*)$, where $L : V \to W$ is a linear map between inner product spaces and $L^* : W \to V$ is the adjoint. This theorem extends to infinite dimensions with some modifications. Notice that such a decomposition is necessarily orthogonal.

Let us see how this implies the theorem. If $\omega \in \Omega^k$, then we can write $\omega = d\theta + \tilde{\omega}$, where $\delta\tilde{\omega} = 0$. Therefore if $d\omega = 0$, we get that $d\tilde{\omega} = 0$ as well. But then $\tilde{\omega}$ must be harmonic, and we have obtained the desired decomposition. To check uniqueness, we must show that $d\theta = 0$ if $\Delta d\theta = 0$. The equation $\Delta d\theta = 0$ reduces to $\delta d\theta = 0$. This shows that $d\theta = 0$, since we have that

$$\begin{aligned}0 &= (\delta d, \theta, \theta)\\ &= (d\theta, d\theta)\\ &= \int_M g(d\theta, d\theta)\\ &\geq 0.\end{aligned}$$

□

7.3 Harmonic Forms

We shall now see how Hodge theory can be used to get information about $b_i(M) = \dim \mathcal{H}^i(M)$ given various curvature inequalities.

7.3.1 1-Forms

Suppose that ω is a harmonic 1-form on (M, g). We shall consider $f = \frac{1}{2}g(\omega, \omega)$ just as we did for Killing fields. One of the technical problems is that we don't have a really good feeling for what $g(\omega, \omega)$ is. If X is the vector field dual to ω, i.e., $\omega(v) = g(X, v)$ for all v, then, of course, $f = \frac{1}{2}g(\omega, \omega) = \frac{1}{2}g(X, X) = \frac{1}{2}\omega(X)$. We now have to figure out what the harmonicity of ω is good for.

Proposition 3.1 *If X is a vector field on (M, g) and $\omega(v) = g(X, v)$ is the dual 1-form, then*

$$\operatorname{div}X = -\delta\omega.$$

Proof. (See also Appendix A.) We shall prove this in the case where M is compact and oriented. If $f \in \Omega^0(M)$, then δ is defined by the relationship $\int_M g(df, \omega) = \int_M f \cdot \delta\omega$. So we need to show that

$$\int_M g(df, \omega) = -\int_M f \cdot \operatorname{div}X.$$

The left-hand side is by definition equal to $\int_M df(X)$. On the other hand, recall that $\operatorname{div}X$ satisfies

$$L_X d\mathrm{vol} = \operatorname{div}X \cdot d\mathrm{vol}.$$

So we have to show that

$$\int_M df(X)d\mathrm{vol} = -\int_M f \cdot L_X(d\mathrm{vol}).$$

Now,

$$\begin{aligned} L_X(f \cdot d\mathrm{vol}) &= (L_X f) \cdot d\mathrm{vol} + f \cdot L_X d\mathrm{vol} \\ &= df(X) \cdot d\mathrm{vol} + f \cdot L_X d\mathrm{vol}, \end{aligned}$$

so we are actually reduced to proving

$$\int_M L_X(f \cdot d\mathrm{vol}) = 0$$

for all vector fields X and functions f. Stokes' theorem implies that this integral is zero, provided that $L_X(f \cdot d\mathrm{vol})$ is an exact form. To see this, we must use the formula $L_X\omega = i_X d\omega + d i_X\omega$, where $i_X\omega(X_2, \ldots, X_n) = \omega(X, X_2, \ldots, X_n)$ and ω is any k-form. Now, $f \cdot d\mathrm{vol}$ is an n-form, so $d(f \cdot d\mathrm{vol}) = 0$, and hence $L_X(f \cdot d\mathrm{vol}) = d(i_X(f \cdot d\mathrm{vol}))$. □

This result also shows that up to sign, at least, the Laplacian on functions is the same as our old definition, that is

$$\operatorname{div}\nabla = -\delta d.$$

The other result we need is

Proposition 3.2 *Suppose X, ω are as in the previous proposition. Then $v \to \nabla_v X$ is symmetric iff $d\omega = 0$.*

Proof. Recall that

$$d\omega(V, W) = D_V\omega(W) - D_W\omega(V) - \omega([V, W]).$$

Using that $\omega(Z) = g(X, Z)$, we then get

$$\begin{aligned} d\omega(V, W) &= D_V g(X, W) - D_W g(X, V) - g(X, \nabla_V W) + g(X, \nabla_W V) \\ &= g(\nabla_V X, W) - g(\nabla_W X, V) + g(X, \nabla_V W) \\ &\quad - g(X, \nabla_W V) - g(X, \nabla_V W) + g(X, \nabla_W V) \\ &= g(\nabla_V X, W) - g(\nabla_W X, V). \end{aligned}$$

□

Therefore if ω is harmonic and X is the dual vector field, we have that $\text{div}X = 0$ and ∇X is a symmetric $(1, 1)$-tensor. Using this we can now prove

Proposition 3.3 *Let X be a vector field so that ∇X is symmetric (i.e. corresponding 1-form is closed) and define $f = \frac{1}{2}|X|^2$. If we define a symmetric $(1, 1)$-tensor by $S(v) = \nabla_v X$, then*

(1) $\nabla f = \nabla_X X$.
(2) $\nabla^2 f = R_X + \nabla_X S + S^2$.
(3) $\Delta f = |\nabla X|^2 + g(X, \nabla \text{div}X) + \text{Ric}(X, X)$.

Proof. For (1) just observe that

$$\begin{aligned} g(\nabla f, V) &= D_V \frac{1}{2} g(X, X) \\ &= g(\nabla_V X, X) \\ &= g(\nabla_X X, V). \end{aligned}$$

For (2) just note that

$$\begin{aligned} \nabla^2 f(V) &= \nabla_V \nabla_X X \\ &= R(V, X)X + \nabla_X \nabla_V X + \nabla_{[V,X]} X \\ &= R(V, X)X + \left(\nabla_X \nabla_V X - \nabla_{\nabla_X V} X\right) + \nabla_{\nabla_V X} X \\ &= R_X(V) + \nabla^2_{X,V} X + \nabla_{\nabla_V X} X \\ &= R_X(V) + (\nabla_X S)(V) + S^2(V). \end{aligned}$$

Thus, $\nabla^2 f$ is decomposed into three $(1, 1)$-tensors.

For (3) we now have to compute the trace of each one of these. Clearly: $\operatorname{tr}(V \to R(V, X)X) = \operatorname{Ric}(X, X)$. Next, the tensor $v \to \nabla_v X = S(v)$ is assumed to be symmetric, so it follows that

$$\begin{aligned}|\nabla X|^2 &= \operatorname{tr}(\nabla X \circ \nabla X) \\ &= \operatorname{tr}\left(S^2\right).\end{aligned}$$

Now for the last tensor. We shall prove something slightly more general. Namely, that for any $(1, 1)$-tensor S and vector field X we have

$$\operatorname{tr}(\nabla_X S) = \nabla_X \operatorname{tr} S.$$

This will establish the desired identity since

$$\begin{aligned}\nabla_X \operatorname{tr} S &= \nabla_X \operatorname{div} X \\ &= g(X, \nabla \operatorname{div} X).\end{aligned}$$

To prove the formula we just calculate

$$\begin{aligned}\operatorname{tr}(\nabla_X S) &= \sum g((\nabla_X S)(E_i), E_i) \\ &= \sum g(\nabla_X (S(E_i)), E_i) - \sum g(S(\nabla_X E_i), E_i) \\ &= \sum \nabla_X g(S(E_i), E_i) - \sum g(S(E_i), \nabla_X E_i) \\ &\quad - \sum g(S(\nabla_X E_i), E_i) \\ &= \sum \nabla_X g\left(\nabla_{E_i} X, E_i\right) \\ &= \nabla_X \operatorname{tr} S.\end{aligned}$$

The fact that

$$\sum g(S(E_i), \nabla_X E_i) + g(S(\nabla_X E_i), E_i) = 0$$

for any orthonormal frame might seem a little mysterious. There are several ways of establishing this. On one hand one could simply prove this pointwise and suppose that the frame was chosen to be normal at a point. Alternatively the fact that the frame is orthonormal means that

$$\begin{aligned}0 &= \nabla_X g\left(E_i, E_j\right) \\ &= g\left(\nabla_X E_i, E_j\right) + g\left(E_i, \nabla_X E_j\right).\end{aligned}$$

Thus, $g\left(\nabla_X E_i, E_j\right)$ is a skew-symmetric matrix. Now, if S^* denotes the adjoint to S, we have

$$\begin{aligned}&\sum g(S(E_i), \nabla_X E_i) + g(S(\nabla_X E_i), E_i) \\ &\qquad\qquad = \sum g(S(E_i), \nabla_X E_i) + g\left(\nabla_X E_i, S^*(E_i)\right)\end{aligned}$$

$$
\begin{aligned}
&= \sum g\left(\left(S+S^{*}\right)\left(E_{i}\right), \nabla_{X} E_{i}\right) \\
&= \sum_{i,j} g\left(g\left(\left(S+S^{*}\right)\left(E_{i}\right), E_{j}\right) E_{j}, \nabla_{X} E_{i}\right) \\
&= \sum_{i,j} g\left(\left(S+S^{*}\right)\left(E_{i}\right), E_{j}\right) g\left(E_{j}, \nabla_{X} E_{i}\right) \\
&= \sum_{i<j} g\left(\left(S+S^{*}\right)\left(E_{i}\right), E_{j}\right) g\left(E_{j}, \nabla_{X} E_{i}\right) \\
&\qquad + \sum_{i=j} g\left(\left(S+S^{*}\right)\left(E_{i}\right), E_{i}\right) g\left(E_{i}, \nabla_{X} E_{i}\right) \\
&\qquad + \sum_{i>j} g\left(\left(S+S^{*}\right)\left(E_{i}\right), E_{j}\right) g\left(E_{j}, \nabla_{X} E_{i}\right) \\
&= \sum_{i<j} g\left(\left(S+S^{*}\right)\left(E_{i}\right), E_{j}\right) g\left(E_{j}, \nabla_{X} E_{i}\right) \\
&\qquad + \sum_{i<j} g\left(\left(S+S^{*}\right)\left(E_{j}\right), E_{i}\right) g\left(E_{i}, \nabla_{X} E_{j}\right) \\
&= \sum_{i<j} g\left(\left(S+S^{*}\right)\left(E_{i}\right), E_{j}\right) g\left(E_{j}, \nabla_{X} E_{i}\right) \\
&\qquad - \sum_{i<j} g\left(\left(S+S^{*}\right)\left(E_{i}\right), E_{j}\right) g\left(E_{j}, \nabla_{X} E_{i}\right) \\
&= 0.
\end{aligned}
$$

We have simply observed that skew-symmetric and symmetric matrices are orthogonal to each other, and that the term we wish to show is zero is the inner product between a skew-symmetric and a symmetric matrix. □

Note that (2) and (3) in the above result generalize the radial curvature equation.

We can now easily show the other Bochner theorem.

Theorem 3.4 (Bochner, 1946) *If* (M, g) *is compact, oriented, and has* $\mathrm{Ric} \geq 0$*, then every harmonic 1-form is parallel.*

Proof. Suppose ω is a harmonic 1-form, X the dual vector field, and $f = \frac{1}{2}g(\omega, \omega) = \frac{1}{2}|X|^2$. Then $\Delta f = |\nabla X|^2 + \mathrm{Ric}(X, X)$, since $\mathrm{div} X = 0$. Thus Stokes' theorem together with the condition $\mathrm{Ric} \geq 0$ implies

$$
\begin{aligned}
0 &= \int_M \Delta f \cdot d\mathrm{vol} \\
&= \int_M \left(|\nabla X|^2 + \mathrm{Ric}(X, X)\right) \cdot d\mathrm{vol} \\
&\geq \int_M |\nabla X|^2 \cdot d\mathrm{vol} \\
&\geq 0.
\end{aligned}
$$

We can therefore conclude that $|\nabla X| = 0$. □

Corollary 3.5 *If (M, g) is as before and furthermore has positive Ricci curvature at one point, then all harmonic 1-forms vanish everywhere.*

Proof. Since we just proved $\text{Ric}(X, X) \equiv 0$, we must have that $X(p) = 0$ if the Ricci tensor is positive on T_pM. But then $X \equiv 0$, since X is parallel. □

Corollary 3.6 *If (M, g) is compact, orientable, and satisfies* $\text{Ric} \geq 0$, *then* $b_1(M) \leq n = \dim M$, *with equality holding iff (M, g) is a flat torus.*

Proof. We know from Hodge theory that $b_1(M) = \dim \mathcal{H}^1(M)$. Now, all harmonic 1-forms are parallel, so the linear map: $\mathcal{H}^1(M) \to T_pM$ that maps $\omega \to X(p)$ is injective. In particular, $\dim \mathcal{H}^1(M) \leq n$.

If equality holds, we obviously have n linearly independent parallel fields E_i, $i = 1, \ldots, n$. This clearly implies that (M, g) is flat. Thus the universal covering is $(\mathbb{R}^n, \text{can})$ with $\Gamma = \pi_1(M)$ acting by isometries. Now pull the vector fields E_i, $i = 1, \ldots, n$, back to $\tilde{E}_i$, $i = 1, \ldots, n$, on $\mathbb{R}^n$. These vector fields are again parallel and are therefore constant vector fields. In addition, they are invariant under the action of Γ, i.e., for each $\gamma \in \Gamma$ we have $D\gamma\left(\tilde{E}_i(p)\right) = \tilde{E}_i(\gamma(p))$, $i = 1, \ldots, n$. A basis of constant vector fields can, however, be invariant only under translations. Thus, Γ consists entirely of translations. This means that Γ is finitely generated, Abelian, and torsion free, and hence must be $\mathbb{Z}^q$ for some q. To see that M is a torus, we need only show that $q = n$. If $q < n$, then $\mathbb{Z}^q$ generates a subspace V of $\mathbb{R}^n$ with dimension $< n$. Let W denote the orthogonal complement to V in $\mathbb{R}^n$. Then $M = \mathbb{R}^n/\mathbb{Z}^q = (V \oplus W)/\mathbb{Z}^q = (V/\mathbb{Z}^q) \oplus W$, which is not compact. Thus, we must have that $q = n$ and that $\Gamma = \mathbb{Z}^n$ generates $\mathbb{R}^n$. □

7.3.2 *The Bochner Technique in General*

The Bochner technique actually works in a much more general setting. Suppose we have a vector bundle $E \to M$ that is endowed with an inner product structure $\langle \cdot, \cdot \rangle$ and a connection that are compatible. To be more precise, let $\Gamma(E)$ denote the sections $s : M \to E$. The connection on E is a map

$$\begin{aligned} \nabla &: \Gamma(E) \to \Gamma(\text{Hom}(TM, E)), \\ s &\to \nabla s, \end{aligned}$$

and $\nabla s : TM \to E$. We assume that it is linear in s, tensorial in X, and compatible with the metric

$$D_X\langle s_1, s_2 \rangle = \langle \nabla_X s_1, s_2 \rangle + \langle s_1, \nabla_X s_2 \rangle.$$

If we assume that (M, g) is an oriented Riemannian manifold, then using the pointwise inner product structures on $\Gamma(E)$, $\Gamma(TM)$, and integration, we get inner product structures on $\Gamma(E)$ and $\Gamma(\mathrm{Hom}(TM, E))$ via the formulae

$$\begin{aligned}(s_1, s_2) &= \int_M \langle s_1, s_2 \rangle, \\ (S_1, S_2) &= \int_M \langle S_1, S_2 \rangle \\ &= \int_M \mathrm{tr}\left(S_1^* S_2\right),\end{aligned}$$

where $S_1^* \in \Gamma(\mathrm{Hom}(E, TM))$ is the pointwise adjoint to S_1. In case M is not compact we must of course use compactly supported sections. Since the connection is a linear map $\nabla : \Gamma(E) \to \Gamma(\mathrm{Hom}(TM, E))$, we get an adjoint $\nabla^* : \Gamma(\mathrm{Hom}(TM, E)) \to \Gamma(E)$ defined implicitly by

$$\int_M \left\langle \nabla S^*, s \right\rangle = \int_M \langle S, \nabla s \rangle .$$

The *connection Laplacian* of a section is defined as $\nabla^* \nabla s$. We do not call this Δ, since even for forms it does not equal our previous choice for the Laplacian. In fact, $\int_M \langle \nabla^* \nabla s, s \rangle = \int_M |\nabla s|^2$. Thus, the only sections which are "harmonic" with respect to this Laplacian are the parallel sections.

There is a different way of defining the connection Laplacian. Namely, consider the second covariant derivative $\nabla^2_{X,Y} s$ and take the trace $\sum_{i=1}^n \nabla^2_{E_i,E_i} s$ with respect to some orthonormal frame. This is easily seen to be invariantly defined. We shall use the notation

$$\begin{aligned}\mathrm{tr}\left(\nabla^2 s\right) &= \sum_{i=1}^n \nabla^2_{E_i,E_i} s, \\ \mathrm{tr}\nabla^2 &= \sum_{i=1}^n \nabla^2_{E_i,E_i} .\end{aligned}$$

The two Laplacians are related as follows:

Proposition 3.7 *Let (M, g) be an oriented Riemannian manifold, and $E \to M$ a vector bundle with an inner product and compatible connection. Then*

$$\nabla^* \nabla s = -\mathrm{tr}\nabla^2 s$$

for all compactly supported sections of E.

Proof. Let s_1 and s_2 be two sections which are compactly supported in the domain for an orthonormal frame E_i on M. The left-hand side of the formula can be reduced as follows:

$$
\begin{aligned}
(\nabla^*\nabla s_1, s_2) &= \int_M \langle \nabla^*\nabla s_1, s_2 \rangle \\
&= \int_M \langle \nabla s_1, \nabla s_2 \rangle \\
&= \int_M \operatorname{tr}\left((\nabla s_1)^* \nabla s_2\right) \\
&= \sum_{i=1}^n \int_M g\left(E_i, \left((\nabla s_1)^* \nabla s_2\right)(E_i)\right) \\
&= \sum_{i=1}^n \int_M g\left((\nabla s_1)(E_i), (\nabla s_2)(E_i)\right) \\
&= \sum_{i=1}^n \int_M g\left(\nabla_{E_i} s_1, \nabla_{E_i} s_2\right).
\end{aligned}
$$

The right-hand side reduces to something similar

$$
\begin{aligned}
\sum_{i=1}^n \left\langle \nabla^2_{E_i,E_i} s_1, s_2 \right\rangle &= \sum_{i=1}^n \left\langle \nabla_{E_i}\nabla_{E_i} s_1, s_2 \right\rangle - \sum_{i=1}^n \left\langle \nabla_{\nabla_{E_i} E_i} s_1, s_2 \right\rangle \\
&= -\sum_{i=1}^n \left\langle \nabla_{E_i} s_1, \nabla_{E_i} s_2 \right\rangle + \sum_{i=1}^n \nabla_{E_i} \left\langle \nabla_{E_i} s_1, s_2 \right\rangle - \sum_{i=1}^n \left\langle \nabla_{\nabla_{E_i} E_i} s_1, s_2 \right\rangle \\
&= -\sum_{i=1}^n \left\langle \nabla_{E_i} s_1, \nabla_{E_i} s_2 \right\rangle + \operatorname{div} X,
\end{aligned}
$$

where X is defined by

$$g(X, v) = \langle \nabla_v s_1, s_2 \rangle .$$

We can then integrate and use Stokes' theorem to conclude

$$\int_M \left\langle \nabla^*\nabla s_1, s_2 \right\rangle = -\int_M \left\langle \operatorname{tr}\nabla^2 s_1, s_2 \right\rangle .$$

Thus, we must have that $\nabla^*\nabla s_1 = -\operatorname{tr}\nabla^2 s_1$ for all such sections. It is now easy to establish the result for all compactly supported sections. □

With this in mind we can, as above, try to compute $\Delta \frac{1}{2} |s|^2$. Initially this works as follows:

$$
\begin{aligned}
\Delta \frac{1}{2} |s|^2 &= \sum_{i=1}^n \nabla^2_{E_i,E_i} \frac{1}{2} \langle s, s \rangle \\
&= \sum_{i=1}^n \nabla_{E_i}\nabla_{E_i} \frac{1}{2} \langle s, s \rangle - \sum_{i=1}^n \left\langle \nabla_{\nabla_{E_i} E_i} s, s \right\rangle
\end{aligned}
$$

$$= \sum_{i=1}^{n} \left(\left\langle \nabla_{E_i} s, \nabla_{E_i} s \right\rangle + \left\langle \nabla_{E_i} \nabla_{E_i} s, s \right\rangle - \left\langle \nabla_{\nabla_{E_i} E_i} s, s \right\rangle \right)$$

$$= \langle \nabla s, \nabla s \rangle + \left\langle \sum_{i=1}^{n} \nabla^2_{E_i, E_i} s, s \right\rangle$$

$$= \langle \nabla s, \nabla s \rangle - \left\langle \nabla^* \nabla s, s \right\rangle .$$

The problem now lies in getting to understand the new Laplacian $\nabla^*\nabla$. This is not always possible and can only be handled on a case-by-case basis. Later, we shall try this out in the situation where s is the curvature tensor. In the exercises, various situations where s is a $(1, 1)$-tensor are discussed.

A general procedure for handling this term comes from understanding certain differential operators. Suppose we have a second-order operator $D^2 : \Gamma(E) \to \Gamma(E)$, such as the Hodge Laplacian. Then we can often get identities of the form

$$D^2 = \nabla^*\nabla + C(R_\nabla),$$

where $R_\nabla : \Gamma(TM) \otimes \Gamma(TM) \otimes \Gamma(E) \to \Gamma(E)$ is the curvature of ∇ defined by

$$\begin{aligned} R_\nabla(X, Y)s &= \nabla^2_{X,Y} s - \nabla^2_{Y,X} s \\ &= \nabla_X \nabla_Y s - \nabla_Y \nabla_X s - \nabla_{[X,Y]} s. \end{aligned}$$

Such formulae are called *Weitzenböck formulae*. Construct $H_{D^2}(E \to M)$ as the sections with $D^2 s = 0$. If we are lucky enough to have an operator D^2 with a Weitzenböck formula, then this space will probably be some sort of topological invariant of $E \to M$, or at least be related to topological invariants of M. Therefore, if $C(R_\nabla) \geq 0$, then $D^2 s = 0$ implies $\nabla s = 0$, which means that s is parallel. Thus we can conclude that $\dim H_{D^2}(E \to M) \leq \dim E_p =$ dimension of fiber of $E \to M$.

In general, the problem is to identify $C(R_\nabla)$. Obviously, the X, Y variables in R_∇ have to be contracted in such a way that $C(R_\nabla) : \Gamma(E) \to \Gamma(E)$.

7.3.3 p-Forms

The first obvious case to try this philosophy on is that of the Hodge Laplacian on k-forms, as we already know that harmonic forms compute the topology of the underlying manifold. Thus we consider $E = \Lambda^k T^* M$ with the usual inner product and Riemannian covariant derivative. It is not hard to see that we have a Weitzenböck formula for k-forms of the form

$$\Delta = \nabla^*\nabla + C(R_\nabla).$$

Using this one can get a Bochner formula for harmonic forms. Actually, the term $C(R_\nabla)$ is a generalized version of the curvature operator. This was apparently observed by I. Singer long before D. Meyer published his different version of this fact in 1971. The general construction of both I. Singer and D. Meyer can be found in [84, Theorem 3.3]. When $p = 1$ we know that $C(R_\nabla)$ is essentially the Ricci tensor after type change. We shall find a formula for $C(R_\nabla)$ that will establish the theorem below. The reason for delaying the proof is that it is a little long, as we need to introduce some new language. Also, the Bochner technique has the wonderful property that it can be used without an understanding of how one proves the formulae that are used. From the theorem we get some nice topological restrictions:

Theorem 3.8 *If the curvature operator $\mathfrak{R} \geq 0$, then $C(R_\nabla) \geq 0$ on k-forms; and if $\mathfrak{R} > 0$, then $C(R_\nabla) > 0$.*

Corollary 3.9 *Suppose M is orientable. If $\mathfrak{R} \geq 0$, then $b_k(M) \leq \binom{n}{k}$, with equality holding only for the flat torus; and if $\mathfrak{R} > 0$ somewhere, then $b_k(M) = 0$ for $k = 1, 2, \ldots, n-1$.*

Proof. Evidently we have that harmonic forms must be parallel. In the case of positive curvature no such forms can exist, and if the curvature is nonnegative, then the Betti number estimate follows from the fact that a parallel form is completely determined by its value at a point. Thus

$$\begin{aligned} b_k &= \dim H^k \\ &= \dim \mathcal{H}^k \\ &\leq \dim \Lambda^k \left(T_p^* M\right) \\ &= \frac{n!}{k!\,(n-k)!}. \end{aligned}$$

□

We now have a pretty good understanding of manifolds with nonnegative (or positive) curvature operator. From the generalized Gauss-Bonnet theorem we know that the Euler characteristic is $\geq 0\,(=2)$. Thus, one of the Hopf problems is settled for this class of manifolds.

H. Hopf is famous for another problem: Does $S^2 \times S^2$ admit a metric with positive sectional curvature? We already know that this space has positive Ricci curvature and also that it doesn't admit a positive curvature operator, as $\chi\left(S^2 \times S^2\right) = 4$. It is also interesting to observe that $\mathbb{C}P^2$ has positive sectional curvature but doesn't admit a metric with positive curvature operator either, as $\chi\left(\mathbb{C}P^2\right) = 3$. Thus, even among 4-manifolds, there seems to be a big difference between simply connected manifolds that admit $\operatorname{Ric} > 0$, $\sec > 0$, and $\mathfrak{R} > 0$. We shall in Chapter 11 describe a simply connected manifold that has $\operatorname{Ric} > 0$ but doesn't even admit a metric with $\sec \geq 0$.

Actually, manifolds with nonnegative curvature operator can be classified (see Chapter 8). From this classification it follows that there are many manifolds that have positive or nonnegative sectional curvature but admit no metric with nonnegative curvature operator.

Example 3.10 We can exhibit a metric with nonnegative sectional curvature on $\mathbb{C}P^2 \sharp \overline{\mathbb{C}}P^2$ by observing that it is an S^1 quotient of $S^2 \times S^3$. Namely, let S^1 act on the three-sphere by the Hopf action and on the two-sphere by rotations. If the total rotation on the two-sphere is $2\pi k$, then the quotient is $S^2 \times S^2$ if k is even, and $\mathbb{C}P^2 \sharp \overline{\mathbb{C}}P^2$ if k is odd. From the above-mentioned classification it follows, however, that the only simply connected spaces with nonnegative curvature operator are topologically equivalent to $S^2 \times S^2$, S^4, or $\mathbb{C}P^2$.

The Bochner technique has found many generalizations. It has, for instance, proven very successful in the study of manifolds with nonnegative scalar curvature. This is explained in more detail in Appendix C. Briefly, what happens is that *spin manifolds* (this a condition similar to saying that a manifold is orientable) admit certain *spinor bundles*. These bundles come with a natural first-order operator called the *Dirac operator,* often denoted by $\partial\!\!\!/$ or $D\!\!\!\!/$. The square of this operator has a Weitzenböck formula of the form

$$D\!\!\!\!/^2 = \nabla^*\nabla + \frac{1}{4}\text{scal}.$$

This formula was discovered (again this was done earlier by I. Singer in unpublished work, as pointed out in [84]) and used by Lichnerowicz to show that a sophisticated invariant called the $\hat{A}$-genus vanishes for spin manifolds with positive scalar curvature. Using some simple generalizations of this formula, Gromov-Lawson showed that any metric on a torus with scal ≥ 0 is in fact flat. We just proved this for metrics with Ric ≥ 0. Dirac operators and their Weitzenböck formulae have also been of extreme importance in physics and 4-manifolds theory. Much of Witten's work (e.g., the positive mass conjecture) uses these ideas. Also, some of the recent work of Seiberg-Witten, which has had a revolutionary impact on 4-manifold geometry, is related to these ideas.

In relation to our discussion above on positively curved manifolds, we should note that there are still no known examples of simply connected manifolds that admit positive scalar curvature but not positive Ricci curvature, this despite the fact that if (M, g) is any closed Riemannian manifold, then for small enough ε the product $\left(M \times S^2, g + \varepsilon^2 ds_2^2\right)$ clearly has positive scalar curvature. This example shows that there are manifolds with positive scalar curvature that don't admit even nonnegative Ricci curvature. To see this, select your favorite surface M^2 with $b_1 > 4$. Then $b_1\left(M^2 \times S^2\right) > 4$ and therefore by Bochner's theorem can't support a metric with nonnegative Ricci curvature.

7.4 Clifford Multiplication on Forms

In order to give a little perspective on the proof of the Weitzenböck formula for p-forms and also to give an indication of some of the basic ideas in spin geometry, we shall develop some new structures on forms. Instead of first developing Clifford algebras in the linear algebra setting, we just go ahead and define the desired structure on a manifold. Appendix C covers some of the basic spin geometry constructions and also mentions briefly how Clifford algebras are defined.

This section, however, is a prerequisite for Appendix C. Throughout, we fix a Riemannian manifold (M, g) of dimension n.

We shall use the *musical* isomorphisms, $\sharp$ (sharp) and $\flat$ (flat), between 1-forms and vector fields. Thus, if X is a vector field, the dual 1-form is defined as $X^\flat(v) = g(X, v)$, and conversely, if ω is a 1-form, then the vector field $\omega^\sharp$ is defined by $\omega(v) = g(\omega^\sharp, v)$.

Recall that $\Omega^*(M)$ denotes the space of all forms on M, while $\Omega^p(M)$ is the space of p-forms. On $\Omega^*(M)$ we can define a product structure that is different from the wedge product. This product is called *Clifford multiplication,* and for $\theta, \omega \in \Omega^*$ it is denoted $\theta \cdot \omega$. If $\theta \in \Omega^1(M)$ and $\omega \in \Omega^p(M)$, then

$$\theta \cdot \omega = \theta \wedge \omega - i_{\theta^\sharp}\omega,$$
$$\omega \cdot \theta = (-1)^p (\theta \wedge \omega + i_{\theta^\sharp}\omega).$$

By declaring the product to be bilinear and associative, we can use these properties to define the product of any two forms. Note that even when ω is a p-form, the Clifford product with a 1-form gives a mixed form. The important property of this new product structure is that for 1-forms we have

$$\theta \cdot \theta = -|\theta|^2.$$

We can polarize this formula to get

$$\theta_1 \cdot \theta_2 + \theta_2 \cdot \theta_1 = -2g(\theta_1, \theta_2).$$

Thus, orthogonal 1-forms anticommute. Also, we see that orthogonal forms satisfy

$$\omega_1 \cdot \omega_2 = \omega_1 \wedge \omega_2.$$

Hence, we see that Clifford multiplication not only depends on the inner product, wedge product, and interior product, but actually reproduces these three items. This is the tremendous advantage of this new structure. Namely, after one gets used to Clifford multiplication, it becomes unnecessary to work with wedge products and interior products.

There are a few more important properties, which are easily established.

Proposition 4.1 *For $\omega_1, \omega_2 \in \Omega^*(M)$ we have*

$$g(\theta \cdot \omega_1, \omega_2) = -g(\omega_1, \theta \cdot \omega_2) \quad \textit{for any} \quad \textit{1-form } \theta,$$
$$g([\psi, \omega_1], \omega_2) = -g(\omega_1, [\psi, \omega_2]) \quad \textit{for any} \quad \textit{2-form } \psi.$$

Here, the commutator is defined by $[\omega_1, \omega_2] = \omega_1 \cdot \omega_2 - \omega_1 \cdot \omega_2$.

Proof. Evidently both formulae refer to the fact that the linear maps

$$\omega \to \theta \cdot \omega,$$
$$\omega \to [\psi, \omega]$$

are skew-symmetric. To prove the identities, one therefore only needs to prove that for any p-form,

$$g(\theta \cdot \omega, \omega) = 0,$$
$$g([\psi, \omega], \omega) = 0.$$

Both of these identities follow directly from the definition of Clifford multiplication, and the fact that the two maps

$$\Omega^p \to \Omega^{p+1},$$
$$\omega \to \theta \wedge \omega,$$

$$\Omega^{p+1} \to \Omega^p,$$
$$\omega \to i_{\theta^\#}\omega,$$

are adjoint to each other. Namely, Clifford multiplication is the difference between these two operations, and since they are adjoint to each other this must be a skew-symmetric operation as desired. □

Proposition 4.2 *For* $\omega_1, \omega_2 \in \Omega^*(M)$ *and vector fields* X, Y *we have*

$$\nabla_X(\omega_1 \cdot \omega_2) = (\nabla_X\omega_1) \cdot \omega_2 + \omega_1 \cdot (\nabla_X\omega_2),$$
$$R(X,Y)(\omega_1 \cdot \omega_2) = (R(X,Y)\omega_1) \cdot \omega_2 + \omega_1 \cdot R(X,Y)\omega_2.$$

Proof. In case $\omega_1 = \omega_2 = \theta$ is a 1-form, we have

$$\begin{aligned}\nabla_X(\theta \cdot \theta) &= -\nabla_X|\theta|^2\\ &= -2g(\nabla_X\theta, \theta)\\ &= (\nabla_X\theta) \cdot \theta + \theta \cdot (\nabla_X\theta).\end{aligned}$$

More generally, we must use the easily established Leibniz rules for interior and exterior products. In case $\omega_1 = \theta$ is a 1-form and $\omega_2 = \omega$ is a general form, we have that

$$\nabla_X(\theta \wedge \omega) = (\nabla_X\theta) \wedge \omega + \theta \wedge (\nabla_X\omega),$$
$$\nabla_X(i_{\theta^\#}\omega) = i_{\nabla_X\theta^\#}\omega + i_{\theta^\#}(\nabla_X\omega),$$

from which we conclude,

$$\begin{aligned}\nabla_X(\theta \cdot \omega) &= \nabla_X(\theta \wedge \omega - i_{\theta^\#}\omega)\\ &= (\nabla_X\theta) \wedge \omega + \theta \wedge (\nabla_X\omega)\\ &\quad - i_{\nabla_X\theta^\#}\omega - i_{\theta^\#}(\nabla_X\omega)\\ &= (\nabla_X\theta) \wedge \omega - i_{\nabla_X\theta^\#}\omega\\ &\quad + \theta \wedge (\nabla_X\omega) - i_{\theta^\#}(\nabla_X\omega)\\ &= (\nabla_X\theta) \cdot \omega + \theta \cdot (\nabla_X\omega).\end{aligned}$$

One can then easily extend this to all forms. The second formula is almost immediate from the first formula. □

We can now define the *Dirac operator* on forms:

$$D : \Omega^*(M) \to \Omega^*(M),$$
$$D(\omega) = \sum_{i=1}^{n} \theta^i \cdot \nabla_{E_i}\omega,$$

where E_i is any frame and θ^i the dual coframe. The definition is clearly independent of the frame field. We can now relate this Dirac operator to the standard exterior derivative and its adjoint. But first we need to know how to relate these two derivatives to covariant differentiation. This is done as follows:

Proposition 4.3 *Given a frame E_i and its dual coframe θ^i, then we have the formulae*

$$d\omega = \theta^i \wedge \nabla_{E_i}\omega,$$
$$\delta\omega = -i_{(\theta^i)^\sharp}\nabla_{E_i}\omega.$$

Proof. First one sees, as usual, that the right-hand sides are invariantly defined and give operators with the usual properties. (Note, in particular, that $d = \theta^i \wedge \nabla_{E_i}$ on functions and that $\delta = -i_{(\theta^i)^\sharp}\nabla_{E_i}\omega$ on 1-forms.) Thus, one can compute, say, $\theta^i \wedge \nabla_{E_i}\omega$ from knowing how to compute this when $\omega = \theta^j$. Then we take an orthonormal frame such that $(\theta^i)^\sharp = E_i$, and finally we assume that the frame is normal at $p \in M$ and establish the formulae at that point. However, the assumption that the frame is normal insures us that all the quantities vanish when we use $\omega = \theta^j$. Thus, the formulae follow. □

From this proposition it is now immediate that

$$D = d + \delta.$$

Thus, we have

$$D^2 = (d+\delta)^2 = d\delta + \delta d = \Delta.$$

With this we can now prove the following nice formula:

Proposition 4.4 *Given a frame E_i and its dual coframe θ^i, we have:*

$$D^2\omega = \sum_{i,j=1}^{n} \theta^i \cdot \theta^j \cdot \nabla^2_{E_i,E_j}\omega$$
$$= \sum_{i,j=1}^{n} \left(\nabla^2_{E_i,E_j}\omega\right) \cdot \theta^j \cdot \theta^i.$$

Proof. First, recall that

$$\nabla^2_{E_i,E_j} = \nabla_{E_i}\nabla_{E_j} - \nabla_{\nabla_{E_i}E_j}$$

is tensorial in both E_i and E_j, and thus the two expressions on the right-hand side are invariantly defined. Using invariance, we need only prove the formula at a point $p \in M$, where the frame is assumed to be normal, i.e., $(\nabla E_i)(p) = 0$ and consequently also $(\nabla\theta^i)(p) = 0$. We can then compute at p,

$$\begin{aligned} D^2\omega &= \theta^i \cdot \left(\nabla_{E_i}\left(\theta^j \cdot \nabla_{E_j}\omega\right)\right) \\ &= \theta^i \cdot \left(\nabla_{E_i}\theta^j\right) \cdot \nabla_{E_j}\omega + \theta^i \cdot \theta^j \cdot \nabla_{E_i}\nabla_{E_j}\omega \\ &= \theta^i \cdot \theta^j \cdot \nabla_{E_i}\nabla_{E_j}\omega + \theta^i \cdot \theta^j \nabla_{\nabla_{E_i}E_j}\omega \\ &= \sum_{i,j=1}^{n} \theta^i \cdot \theta^j \cdot \nabla^2_{E_i,E_j}\omega. \end{aligned}$$

For the second formula the easiest thing to do is to observe that for a p-form ω we have

$$\hat{D}\omega = \left(\nabla_{E_i}\omega\right) \cdot \theta^i = (-1)^p\,(d - \delta)\,.$$

Thus also,

$$\hat{D}^2 = \Delta = D^2.$$

This finishes the proof. □

We can now establish the relevant Weitzenböck formula.

Theorem 4.5 *Given a frame E_i and its dual coframe θ^i, we have*

$$\begin{aligned} D^2\omega &= \nabla^*\nabla\omega + \frac{1}{2}\sum_{i,j=1}^{n} \theta^i \cdot \theta^j \cdot R\left(E_i, E_j\right)\omega \\ &= \nabla^*\nabla\omega + \frac{1}{2}\sum_{i,j=1}^{n} R\left(E_i, E_j\right)\omega \cdot \theta^j \cdot \theta^i. \end{aligned}$$

Proof. Using the above identities for D^2, it clearly suffices to check

$$\nabla^*\nabla\omega + \frac{1}{2}\sum_{i,j=1}^{n} \theta^i \cdot \theta^j \cdot R\left(E_i, E_j\right)\omega = \sum_{i,j=1}^{n} \theta^i \cdot \theta^j \cdot \nabla^2_{E_i,E_j}\omega,$$

$$\nabla^*\nabla\omega + \frac{1}{2}\sum_{i,j=1}^{n} R\left(E_i, E_j\right)\omega \cdot \theta^j \cdot \theta^i = \sum_{i,j=1}^{n} \left(\nabla^2_{E_i,E_j}\omega\right) \cdot \theta^j \cdot \theta^i.$$

These formulae are of course established in the same way, so let us concentrate on the first. As usual, note that everything is invariant. We can therefore pick a frame that is orthonormal and normal at $p \in M$ and compute at $p \in M$,

$$\begin{aligned}\sum_{i,j=1}^{n} \theta^i \cdot \theta^j \cdot \nabla^2_{E_i,E_j}\omega &= -\sum_{i=1}^{n} \nabla^2_{E_i,E_i}\omega + \sum_{i\neq j} \theta^i \cdot \theta^j \cdot \nabla^2_{E_i,E_j}\omega \\ &= -\sum_{i=1}^{n} \nabla^2_{E_i,E_i}\omega + \sum_{i<j} \theta^i \cdot \theta^j \cdot \left(\nabla^2_{E_i,E_j}\omega - \nabla^2_{E_j,E_i}\omega\right) \\ &= -\sum_{i=1}^{n} \nabla^2_{E_i,E_i}\omega + \sum_{i<j} \theta^i \cdot \theta^j \cdot R\left(E_i, E_j\right)\omega \\ &= -\sum_{i=1}^{n} \nabla^2_{E_i,E_i}\omega + \frac{1}{2}\sum_{i,j=1}^{n} \theta^i \cdot \theta^j \cdot R\left(E_i, E_j\right)\omega,\end{aligned}$$

where we used the relations

$$\begin{aligned}\theta^i \cdot \theta^i &= -1, \\ \theta^i \cdot \theta^j &= -\theta^j \cdot \theta^i,\end{aligned}$$

Now use that we know

$$\nabla^*\nabla = -\sum_{i=1}^{n} \nabla^2_{E_i,E_i}$$

to finish the proof. □

We can now establish the desired Bochner formula for forms.

Corollary 4.6 *Given an orthonormal frame E_i and its dual coframe θ^i, we have for any harmonic form ω, i.e., $D\omega = 0$, that*

$$0 = \nabla^*\nabla\omega + \frac{1}{4}\sum_{i,j=1}^{n} g\left[\theta^i \cdot \theta^j, R\left(E_i, E_j\right)\omega\right].$$

Proof. First, we use that the frame is orthonormal to conclude that

$$\sum_{i,j=1}^{n} R\left(E_i, E_j\right)\omega \cdot \theta^j \cdot \theta^i = -\sum_{i,j=1}^{n} R\left(E_i, E_j\right)\omega \cdot \theta^i \cdot \theta^j.$$

Thus, we have

$$\begin{aligned}D^2\omega &= \nabla^*\nabla + \frac{1}{2}\sum_{i,j=1}^{n} \theta^i \cdot \theta^j \cdot R\left(E_i, E_j\right)\omega, \\ D^2\omega &= \nabla^*\nabla - \frac{1}{2}\sum_{i,j=1}^{n} R\left(E_i, E_j\right)\omega \cdot \theta^i \cdot \theta^j.\end{aligned}$$

Adding these equations and dividing by 2 then yields

$$D^2\omega = \nabla^*\nabla + \frac{1}{4}\sum_{i,j=1}^{n}\left[\theta^i \cdot \theta^j, R\left(E_i, E_j\right)\omega\right].$$

Therefore, if $D\omega = 0$ we get

$$0 = \nabla^*\nabla\omega + \frac{1}{4}\sum_{i,j=1}^{n}\left[\theta^i \cdot \theta^j, R\left(E_i, E_j\right)\omega\right],$$

which yields the desired equation. □

Having identified the curvature terms in the Weitzenböck and Bochner formulae, it now remains to be seen that this term is nonnegative when the curvature operator is nonnegative. Before doing this, let us deconstruct the curvature terms in the following way:

Lemma 4.7 *For an orthonormal frame E_i and dual coframe θ^i we have*

$$\begin{aligned} R(X,Y)\omega &= \frac{1}{4}\sum_{i,j=1}^{n} g\left(R(X,Y)E_i, E_j\right)\left(\theta^i \cdot \theta^j \cdot \omega - \omega \cdot \theta^i \cdot \theta^j\right) \\ &= \frac{1}{4}\sum_{i,j=1}^{n} g\left(R(X,Y)E_i, E_j\right)\left[\theta^i \cdot \theta^j, \omega\right]. \end{aligned}$$

Proof. Needless to say, as the right-hand side is invariant, we can assume that the frame is orthonormal and normal at $p \in M$. Moreover, both sides are derivations in ω, so it suffices to check the identities for 1-forms. Finally, we can restrict attention to 1-forms of the type $\omega = \theta^k$ and then compute $\theta^i \cdot \theta^j \cdot \theta^k - \theta^k \cdot \theta^i \cdot \theta^j$. This term depends on whether $k = i$ or $k = j$ or $k \neq i, j$. We can also assume that $i \neq j$, as those terms are zero in the above expression. We then get

$$\theta^i \cdot \theta^j \cdot \theta^k - \theta^k \cdot \theta^i \cdot \theta^j = \begin{cases} 0, & k \neq i, j, \\ -2\theta^i, & k = j, \\ 2\theta^j, & k = i. \end{cases}$$

Using this we can now compute

$$\begin{aligned} \sum_{i,j=1}^{n} g\left(R(X,Y)E_i, E_j\right)\left[\theta^i \cdot \theta^j, \theta^k\right] &= -2\sum_{i=1}^{n} g\left(R(X,Y)E_i, E_k\right)\theta^i \\ &\quad + 2\sum_{j=1}^{n} g\left(R(X,Y)E_k, E_j\right)\theta^j \\ &= 4\sum_{i=1}^{n} g\left(R(X,Y)E_k, E_i\right)\theta^i. \end{aligned}$$

It is now easy to see that the last term is equivalent to the 1-form $4R(X,Y)\theta^k$. □

With this last formula we can now relate the curvature term in the Bochner formula to the curvature operator.

Lemma 4.8 *For an orthonormal frame E_i and its dual coframe θ^i we have that*

$$\sum_{i,j=1}^{n} g\left(\left[\theta^i \cdot \theta^j, R\left(E_i, E_j\right)\omega\right], \omega\right) = \sum_{\alpha} \lambda_\alpha \left|\left[\Theta_\alpha, \omega\right]\right|^2,$$

where λ_α are the eigenvalues for the curvature operator and Θ_α the duals of eigenvectors for the curvature operator.

Proof. Using the skew symmetry of $\omega \to \left[\theta^i \cdot \theta^j, \omega\right]$ and the definition of the curvature operator, we can compute

$$\begin{aligned}
&\sum_{i,j=1}^{n} g\left(\left[\theta^i \cdot \theta^j, R\left(E_i, E_j\right)\omega\right], \omega\right) \\
&= -\sum_{i,j=1}^{n} g\left(R\left(E_i, E_j\right)\omega, \left[\theta^i \cdot \theta^j, \omega\right]\right) \\
&= -\frac{1}{4}\sum_{i,j=1}^{n}\sum_{k,l=1}^{n} g\left(R\left(E_i, E_j\right)E_k, E_l\right) g\left(\left[\theta^k \cdot \theta^l, \omega\right], \left[\theta^i \cdot \theta^j, \omega\right]\right) \\
&= \frac{1}{4}\sum_{i,j,k,l=1}^{n} g\left(\mathfrak{R}\left(E_i \wedge E_j\right), E_k \wedge E_l\right) g\left(\left[\theta^k \cdot \theta^l, \omega\right], \left[\theta^i \cdot \theta^j, \omega\right]\right) \\
&= \sum_{i<j,k<l}^{n} g\left(\mathfrak{R}\left(E_i \wedge E_j\right), E_k \wedge E_l\right) g\left(\left[\theta^k \cdot \theta^l, \omega\right], \left[\theta^i \cdot \theta^j, \omega\right]\right).
\end{aligned}$$

Now observe that the $E_i \wedge E_j$ form an orthonormal basis for $\Lambda^2 TM$, and the $\theta^i \cdot \theta^j$ are the dual basis for $\Omega^2(M)$. The expression we have arrived at is obviously invariant under change of bases in $\Lambda^2 TM$. So select an orthonormal basis Ξ_α for $\Lambda^2 TM$ such that $\mathfrak{R}(\Xi_\alpha) = \lambda_\alpha \Xi_\alpha$. With Θ_α denoting the dual basis for $\Omega^2(M)$, we then get

$$\sum_{i,j=1}^{n} g\left(\left[\theta^i \cdot \theta^j, R\left(E_i, E_j\right)\omega\right], \omega\right) = \sum_{\alpha} \lambda_\alpha \left|\left[\Theta_\alpha, \omega\right]\right|^2$$

as desired. □

Theorem 4.9 *On a compact oriented Riemannian n-manifold with nonnegative curvature operator every harmonic form is parallel. Moreover, if the curvature operator is positive, then harmonic p-forms vanish, $p = 1, \ldots, n-1$.*

Proof. If ω is harmonic, then we have from the previous section that

$$\begin{aligned} 0 &= \int_M \langle \nabla^* \nabla \omega, \omega \rangle + \frac{1}{4} \int_M \sum_\alpha \lambda_\alpha \left| [\Theta_\alpha, \omega] \right|^2 \\ &= \int_M |\nabla \omega|^2 + \frac{1}{4} \int_M \sum_\alpha \lambda_\alpha \left| [\Theta_\alpha, \omega] \right|^2 . \end{aligned}$$

As both terms are nonnegative, they both vanish. In particular, $\nabla \omega = 0$.

We also have that $\lambda_\alpha \left| [\Theta_\alpha, \omega] \right|^2 = 0$. The only way this can happen, if all $\lambda_\alpha > 0$, is if $[\Theta_\alpha, \omega] = 0$ for all α. Since the Θ_α form a basis for the 2-forms, this means that $[\psi, \omega] = 0$ for all 2-forms. To see that this makes $\omega = 0$, just pick $\psi = \theta^i \cdot \theta^j$, $\omega = \theta^{i_1} \cdot \cdots \cdot \theta^{i_p}$, and compute:

$$\left[\theta^i \cdot \theta^j, \theta^{i_1} \cdot \cdots \cdot \theta^{i_p} \right] = \begin{cases} 0, & i, j \notin \left\{ i_1, \ldots, i_p \right\}, \\ 0, & i, j \in \left\{ i_1, \ldots, i_p \right\}, \\ 2\theta^i \cdot \theta^j \cdot \theta^{i_1} \cdot \cdots \cdot \theta^{i_p}, & \text{otherwise.} \end{cases}$$

In general, we can write

$$\omega = \sum_{i_1 < \cdots < i_p} a_{i_1 \cdots i_p} \theta^{i_1} \cdot \cdots \cdot \theta^{i_p}.$$

Therefore, $\left[\theta^i \cdot \theta^j, \omega \right]$ can only vanish if $a_{i_1 \ldots i_p} = 0$ whenever $i \in \left\{ i_1, \ldots, i_p \right\}$ or $j \in \left\{ i_1, \ldots, i_p \right\}$ but not both i and j belong to $\left\{ i_1, \cdots, i_p \right\}$. Using this in the situation where $i < j$ shows that ω must be zero unless p is 0 or n. □

7.5 The Curvature Tensor

It is now time to apply the Bochner technique to the most natural tensor, the curvature tensor. It is by no means clear that this will yield anything. It seems to us both miraculous and profound that something comes out of this. We shall present results by Lichnerowicz (see [56, Chapter 1] and also [57] for an in-depth discussion on the meaning of these matters in physics), Berger, and Tachibana (see [78]) that combine to show that a compact Riemannian manifold with $\text{div} R = 0$ and nonnegative sectional curvature (or nonnegative curvature operator) has parallel Ricci tensor (parallel curvature tensor).

Recall that if we consider the $(1, 3)$ version of the curvature tensor R, then we can construct two $(0, 4)$-tensors: $\text{div} \nabla R$ and $\nabla \text{div} R$. If for our present purposes we use the notation $R^\flat (X, Y, Z, W) = g(X, R(Y, Z) W)$, then we can take inner products of the three tensors $R^\flat$, $\text{div} \nabla R$, and $\nabla \text{div} R$. Note that $R^\flat$ is not the usual $(0, 4)$-tensor. This will be very important in the proof below.

Theorem 5.1 (Lichnerowicz, 1958) *The curvature tensor R on a compact oriented Riemannian manifold satisfies*

$$2\int_M |\mathrm{div}R|^2 - 2\int_M K = \int_M |\nabla R|^2,$$

where $K = g\left(R^\flat, \mathrm{div}\nabla R - \nabla\mathrm{div}R\right)$.

Proof. By far the most important ingredient in the proof is that we have the second Bianchi identity at our disposal. To establish the formula, we compute at a point p where we have an orthonormal frame E_i with $(\nabla E_i)(p) = 0$:

$$\begin{aligned}
\Delta\frac{1}{2}|R|^2 &= \frac{1}{2}\sum_{i=1}^n \nabla_{E_i}\nabla_{E_i}|R|^2 \\
&= \frac{1}{2}\sum_{i=1}^n \nabla_{E_i}\nabla_{E_i} g(R,R) \\
&= \sum_{i=1}^n \nabla_{E_i} g\left(\left(\nabla_{E_i}R\right), R\right) \\
&= \sum_{i=1}^n g\left(\left(\nabla_{E_i}\left(\nabla_{E_i}R\right)\right), R\right) \\
&\quad + \sum_{i=1}^n g\left(\left(\nabla_{E_i}R\right), \left(\nabla_{E_i}R\right)\right) \\
&= \sum_{i=1}^n g\left(\left(\nabla_{E_i}\left(\nabla_{E_i}R\right)\right), R\right) \\
&\quad + |\nabla R|^2.
\end{aligned}$$

We now claim that

$$\sum_{i=1}^n g\left(\left(\nabla_{E_i}\left(\nabla_{E_i}R\right)\right), R\right) = 2g\left(R^\flat, \mathrm{div}\nabla R\right).$$

Using that ∇R has the same symmetry properties as R, we first compute

$$\begin{aligned}
2\mathrm{div}\nabla R\left(E_j, E_k, E_l, E_m\right) &= 2\sum_{i=1}^n g\left(\left(\nabla_{E_i}(\nabla R)\right)\left(E_j, E_k, E_l, E_m\right), E_i\right) \\
&= 2\sum_{i=1}^n g\left(\nabla_{E_i}\left((\nabla R)\left(E_j, E_k, E_l, E_m\right)\right), E_i\right) \\
&= 2\sum_{i=1}^n g\left(\nabla_{E_i}\left(\left(\nabla_{E_j}R\right)(E_k, E_l)\,E_m\right), E_i\right)
\end{aligned}$$

$$= 2\sum_{i=1}^{n} \nabla_{E_i} g\left(\left(\nabla_{E_j} R\right)(E_k, E_l) E_m, E_i\right)$$
$$= 2\sum_{i=1}^{n} \nabla_{E_i} g\left(\left(\nabla_{E_j} R\right)(E_m, E_i) E_k, E_l\right)$$
$$= 2\sum_{i=1}^{n} g\left(\nabla_{E_i}\left(\nabla_{E_j} R\right)(E_m, E_i) E_k, E_l\right)$$

and then observe that

$$2g\left(R^{\flat}, \operatorname{div}\nabla R\right)$$
$$= 2\sum_{i,j,k,l,m=1}^{n} g\left(\nabla_{E_i}\left(\nabla_{E_j} R\right)(E_k, E_l) E_m, E_i\right) g\left(E_j, R(E_k, E_l) E_m\right)$$
$$= 2\sum_{i,j,k,l,m=1}^{n} g\left(\nabla_{E_i}\left(\nabla_{E_j} R\right)(E_k, E_l) E_m, E_i\right) g\left(R(E_k, E_l) E_m, E_j\right)$$
$$= 2\sum_{i,j,k,l,m=1}^{n} g\left(\nabla_{E_i}\left(\nabla_{E_j} R\right)(E_k, E_l) E_m, E_i\right) g\left(R\left(E_j, E_m\right) E_l, E_k\right).$$

On the other hand, we have using the second Bianchi identity,

$$\sum_{i=1}^{n} g\left(\nabla_{E_i}\left(\nabla_{E_i} R\right)\left(E_j, E_k\right) E_l, E_m\right) = \sum_{i=1}^{n} \nabla_{E_i} g\left(\left(\nabla_{E_i} R\right)\left(E_j, E_k\right) E_l, E_m\right)$$
$$= -\sum_{i=1}^{n} \nabla_{E_i} g\left(\left(\nabla_{E_j} R\right)(E_k, E_i) E_l, E_m\right) - \sum_{i=1}^{n} \nabla_{E_i} g\left(\left(\nabla_{E_k} R\right)\left(E_i, E_j\right) E_l, E_m\right)$$
$$= -\sum_{i=1}^{n} \nabla_{E_i} g\left(\left(\nabla_{E_j} R\right)(E_k, E_i) E_l, E_m\right) + \sum_{i=1}^{n} \nabla_{E_i} g\left(\left(\nabla_{E_k} R\right)\left(E_j, E_i\right) E_l, E_m\right),$$

and so,

$$\sum_{i=1}^{n} g\left(\left(\nabla_{E_i}\left(\nabla_{E_i} R\right)\right), R\right)$$
$$= \sum_{i,j,k,l,m=1}^{n} g\left(\nabla_{E_i}\left(\nabla_{E_i} R\right)\left(E_j, E_k\right) E_l, E_m\right) g\left(R\left(E_j, E_k\right) E_l, E_m\right)$$
$$= -\sum_{i,j,k,l,m=1}^{n} \nabla_{E_i} g\left(\left(\nabla_{E_j} R\right)(E_k, E_i) E_l, E_m\right) g\left(R\left(E_j, E_k\right) E_l, E_m\right)$$

$$\begin{aligned}
&+\sum_{i,j,k,l,m=1}^{n} \nabla_{E_i} g\left(\left(\nabla_{E_k} R\right)\left(E_j, E_i\right) E_l, E_m\right) g\left(R\left(E_j, E_k\right) E_l, E_m\right) \\
&= \sum_{i,j,k,l,m=1}^{n} \nabla_{E_i} g\left(\left(\nabla_{E_j} R\right)\left(E_k, E_i\right) E_l, E_m\right) g\left(R\left(E_k, E_j\right) E_l, E_m\right) \\
&\quad+\sum_{i,j,k,l,m=1}^{n} \nabla_{E_i} g\left(\left(\nabla_{E_k} R\right)\left(E_j, E_i\right) E_l, E_m\right) g\left(R\left(E_j, E_k\right) E_l, E_m\right) \\
&= 2\sum_{i,j,k,l,m=1}^{n} \nabla_{E_i} g\left(\left(\nabla_{E_j} R\right)\left(E_k, E_i\right) E_l, E_m\right) g\left(R\left(E_k, E_j\right) E_l, E_m\right) \\
&= 2\sum_{i,j,k,l,m=1}^{n} \nabla_{E_i} g\left(\left(\nabla_{E_j} R\right)\left(E_m, E_l\right) E_k, E_i\right) g\left(R\left(E_j, E_k\right) E_l, E_m\right) \\
&= 2\sum_{i,j,k,l,m=1}^{n} \nabla_{E_i} g\left(\left(\nabla_{E_j} R\right)\left(E_k, E_l\right) E_m, E_i\right) g\left(R\left(E_j, E_m\right) E_l, E_k\right) \\
&= 2g\left(R^{\flat}, \operatorname{div}\nabla R\right).
\end{aligned}$$

Using the definition of K, we then arrive at

$$\Delta\frac{1}{2}|R|^2 = |\nabla R|^2 + 2g\left(R^{\flat}, \nabla\operatorname{div}R\right) + 2K.$$

From Stokes' theorem (see also Appendix A) it follows that

$$\int_M \Delta\frac{1}{2}|R|^2 = 0,$$

$$\int_M g\left(R^{\flat}, \nabla\operatorname{div}R\right) = -\int_M |\operatorname{div}R|^2.$$

This clearly gives us the desired formula. □

We are now interested in understanding when K is nonnegative. In order to analyze this better we shall go through some generalities.

For any tensor T we can consider the curvature

$$\begin{aligned}
R(X,Y)T &= (\nabla_X(\nabla_Y T)) - (\nabla_Y(\nabla_X T)) - \left(\nabla_{[X,Y]}T\right) \\
&= \nabla^2_{X,Y}T - \nabla^2_{Y,X}T
\end{aligned}$$

as a new tensor of the same type. This new tensor is tensorial in X and Y. Moreover, it is also tensorial in T, so we have for any function f

$$R(X,Y)(fT) = fR(X,Y)T.$$

More importantly, one can easily show that

$$(R(X,Y)T)(X_1,\ldots,X_k) = R(X,Y)(T(X_1,\ldots,X_k))$$

$$
\begin{aligned}
&- T(R(X,Y)X_1,\ldots,X_k)\\
&\vdots\\
&- T(X_1,\ldots,R(X,Y)X_k).
\end{aligned}
$$

To understand this new curvature, we can simply break in down to the point where we need to worry only about how it acts on vector fields and 1-forms. And this we already know how to deal with.

We are particularly interested in the case where T is of type $(1,k)$. In that case we can make a special contraction. Namely, if we choose an orthonormal frame E_i, then

$$((\mathrm{div}\nabla - \nabla\mathrm{div})T)(Y,X_1,\ldots,X_k) = \sum_{i=1}^{n} g((R(E_i,Y)T)(X_1,\ldots,X_k),E_i).$$

It therefore appears that $(\mathrm{div}\nabla - \nabla\mathrm{div})T$ is something like the Ricci curvature of T. This is in line with our Weitzenböck formulae, where the curvature term is some sort of contraction in the curvature. If we make the type change $T^\flat(Y,X_1,\ldots,X_k) = g(Y,T(X_1,\ldots,X_k))$, then we get the quadratic expression for this Ricci curvature

$$K = g\left(T^\flat,(\mathrm{div}\nabla - \nabla\mathrm{div})T\right).$$

The claim is that this quantity is nonnegative whenever the curvature operator is nonnegative and $T = R$. In order to make our argument a little more transparent, let us first show a similar but easier result.

Lemma 5.2 (Berger) *Suppose T is a symmetric $(1,1)$-tensor on a Riemannian manifold (M,g) with* $\sec \geq 0$, *then*

$$K = g\left(T^\flat,(\mathrm{div}\nabla - \nabla\mathrm{div})T\right) \geq 0.$$

Proof. We shall calculate at a point p, where an orthonormal frame has been chosen such that $T(E_i) = \lambda_i E_i$:

$$
\begin{aligned}
g\left(T^\flat,(\mathrm{div}\nabla - \nabla\mathrm{div})T\right) &= \sum_{i,j,k=1}^{n} g(E_j,T(E_k))\, g\left(\left(R\left(E_i,E_j\right)T\right)(E_k),E_i\right)\\
&= \sum_{i,j,k=1}^{n} g(E_j,T(E_k))\, g\left(R\left(E_i,E_j\right)T(E_k),E_i\right)\\
&\quad - \sum_{i,j,k=1}^{n} g(E_j,T(E_k))\, g\left(T\left(R\left(E_i,E_j\right)E_k\right),E_i\right)\\
&= \sum_{i,j,k=1}^{n} g\left(R\left(E_i,g(E_j,T(E_k))E_j\right)T(E_k),E_i\right)
\end{aligned}
$$

$$
\begin{aligned}
&- \sum_{i,j,k=1}^{n} g\left(T\left(R\left(E_i, g\left(E_j, T\left(E_k\right)\right) E_j\right) E_k\right), E_i\right) \\
&= \sum_{i,k=1}^{n} g\left(R\left(E_i, T\left(E_k\right)\right) T\left(E_k\right), E_i\right) \\
&\quad - \sum_{i,k=1}^{n} g\left(R\left(E_i, T\left(E_k\right)\right) E_k, T\left(E_i\right)\right) \\
&= \sum_{i,k=1}^{n} \lambda_k^2 \cdot g\left(R\left(E_i, E_k\right) E_k, E_i\right) \\
&\quad - \sum_{i,k=1}^{n} \lambda_k \lambda_i \cdot g\left(R\left(E_i, E_k\right) E_k, E_i\right) \\
&= \sum_{i,k=1}^{n} \left(\lambda_k^2 - \lambda_k \lambda_i\right) \sec\left(E_i, E_k\right) \\
&= \sum_{i<k} \left(\lambda_k^2 - \lambda_k \lambda_i\right) \sec\left(E_i, E_k\right) \\
&\quad + \sum_{i>k} \left(\lambda_k^2 - \lambda_k \lambda_i\right) \sec\left(E_i, E_k\right) \\
&= \sum_{i<k} \left(\lambda_k^2 - \lambda_k \lambda_i\right) \sec\left(E_i, E_k\right) \\
&\quad + \sum_{i<k} \left(\lambda_i^2 - \lambda_k \lambda_i\right) \sec\left(E_i, E_k\right) \\
&= \sum_{i<k} \left(\lambda_k - \lambda_i\right)^2 \sec\left(E_i, E_k\right) \\
&\geq 0.
\end{aligned}
$$

This finishes the proof. □

Given this, one might suspect that we should be able to do something for the Ricci tensor, given that the sectional curvature is nonnegative. This is only partially true, as we don't have a Bochner formula for the Ricci tensor. Given that the manifold has divergence-free curvature tensor, one can find a Bochner formula and then get that the Ricci tensor must be parallel. The proofs are not hard and are deferred to the exercises. Note that we can't more generally hope that the Ricci tensor is parallel if it is divergence free, as all of the Berger spheres have divergence-free Ricci tensor, but only the standard sphere has parallel Ricci tensor.

We can now go over to the more complicated result we are interested in. It was first established in [78], and then a "new" proof appeared in [38]. After that, the result seems to have fallen into oblivion. We shall present a more general version that is analogous to the above lemma, but the proof is essentially the one proposed by Tachibana.

Theorem 5.3 (Tachibana, 1974) *If* $\mathfrak{R} \geq 0$, *then* $g\left(T^{\flat}, (\mathrm{div}\nabla - \nabla\mathrm{div})\, T\right) \geq 0$ *for any* $(1,3)$*-tensor* T *that induces a self-adjoint map* $\mathfrak{T} : \Lambda^2 TM \to \Lambda^2 TM$.

Proof. The fact that $\mathfrak{T} : \Lambda^2 TM \to \Lambda^2 TM$ is self-adjoint means that T enjoys the properties

$$\begin{aligned} &g\left(T\left(X,Y,Z\right),W\right) = -g\left(T\left(X,Y,W\right),Z\right) = g\left(T\left(Y,X,W\right),Z\right),\\ &g\left(T\left(X,Y,Z\right),W\right) = g\left(T\left(Z,W,X\right),Y\right).\end{aligned}$$

Thus, we have a tensor with the same properties as the curvature tensor, with the exception of Bianchi's identities. Let us first divide K into four terms:

$$\begin{aligned} K &= g\left(T^{\flat}, (\mathrm{div}\nabla - \nabla\mathrm{div})\, T\right)\\ &= \sum_{i,j,k,l,m=1}^{n} g\left(E_j, T\left(E_k, E_l, E_m\right)\right) g\left(\left(R\left(E_i, E_j\right) T\right)\left(E_k, E_l, E_m\right), E_i\right)\\ &= \sum_{i,j,k,l,m=1}^{n} g\left(E_j, T\left(E_k, E_l, E_m\right)\right) g\left(R\left(E_i, E_j\right)\left(T\left(E_k, E_l, E_m\right)\right), E_i\right)\\ &\quad + \sum_{i,j,k,l,m=1}^{n} -g\left(E_j, T\left(E_k, E_l, E_m\right)\right) g\left(T\left(R\left(E_i, E_j\right) E_k, E_l, E_m\right), E_i\right)\\ &\quad + \sum_{i,j,k,l,m=1}^{n} -g\left(E_j, T\left(E_k, E_l, E_m\right)\right) g\left(T\left(E_k, R\left(E_i, E_j\right) E_l, E_m\right), E_i\right)\\ &\quad + \sum_{i,j,k,l,m=1}^{n} -g\left(E_j, T\left(E_k, E_l, E_m\right)\right) g\left(T\left(E_k, E_l, R\left(E_i, E_j\right) E_m\right), E_i\right)\\ &= A + B + C + D.\end{aligned}$$

We now compute each of the terms A, B, C, and D:

$$\begin{aligned} A &= \sum_{i,j,k,l,m=1}^{n} g\left(E_j, T\left(E_k, E_l, E_m\right)\right) g\left(R\left(E_i, E_j\right)\left(T\left(E_k, E_l, E_m\right)\right), E_i\right)\\ &= \sum_{i,j,k,l,m=1}^{n} g\left(R\left(E_i, g\left(E_j, T\left(E_k, E_l, E_m\right)\right) E_j\right)\left(T\left(E_k, E_l, E_m\right)\right), E_i\right)\\ &= \sum_{i,k,l,m=1}^{n} g\left(R\left(E_i, T\left(E_k, E_l, E_m\right)\right)\left(T\left(E_k, E_l, E_m\right)\right), E_i\right)\\ &= \sum_{i,k,l,m=1}^{n} g\left(\mathfrak{R}\left(E_i \wedge T\left(E_k, E_l, E_m\right)\right), E_i \wedge T\left(E_k, E_l, E_m\right)\right);\\ B &= \sum_{i,j,k,l,m=1}^{n} -g\left(E_j, T\left(E_k, E_l, E_m\right)\right) g\left(T\left(R\left(E_i, E_j\right) E_k, E_l, E_m\right), E_i\right)\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i,j,k,l,m=1}^{n} -g\left(E_j, T\left(E_k, E_l, E_m\right)\right) g\left(T\left(E_m, E_i, R\left(E_i, E_j\right) E_k\right), E_l\right) \\
&= \sum_{i,j,k,l,m=1}^{n} g\left(E_j, T\left(E_k, E_l, E_m\right)\right) g\left(T\left(E_m, E_i, E_l\right), R\left(E_i, E_j\right) E_k\right) \\
&= \sum_{i,k,l,m=1}^{n} g\left(T\left(E_m, E_i, E_l\right), R\left(E_i, T\left(E_k, E_l, E_m\right)\right) E_k\right) \\
&= \sum_{i,k,l,m=1}^{n} g\left(R\left(E_i, T\left(E_k, E_l, E_m\right)\right) E_k, T\left(E_m, E_i, E_l\right)\right) \\
&= -\sum_{i,k,l,m=1}^{n} g\left(R\left(E_i, T\left(E_k, E_l, E_m\right)\right) T\left(E_m, E_i, E_l\right), E_k\right) \\
&= -\sum_{i,k,l,m=1}^{n} g\left(\mathfrak{R}\left(E_i \wedge T\left(E_k, E_l, E_m\right)\right), E_k \wedge T\left(E_m, E_i, E_l\right)\right).
\end{aligned}
$$

Similarly,

$$
C = \sum_{i,k,l,m=1}^{n} g\left(\mathfrak{R}\left(E_i \wedge T\left(E_k, E_l, E_m\right)\right), E_l \wedge T\left(E_m, E_i, E_k\right)\right).
$$

Finally, we have

$$
\begin{aligned}
D &= \sum_{i,j,k,l,m=1}^{n} -g\left(E_j, T\left(E_k, E_l, E_m\right)\right) g\left(T\left(E_k, E_l, R\left(E_i, E_j\right) E_m\right), E_i\right) \\
&= \sum_{i,j,k,l,m=1}^{n} g\left(E_j, T\left(E_k, E_l, E_m\right)\right) g\left(R\left(E_i, E_j\right) E_m, T\left(E_k, E_l, E_i\right)\right) \\
&= -\sum_{i,k,l,m=1}^{n} g\left(R\left(E_i, T\left(E_k, E_l, E_m\right)\right) T\left(E_k, E_l, E_i\right), E_m\right) \\
&= -\sum_{i,k,l,m=1}^{n} g\left(\mathfrak{R}\left(E_i \wedge T\left(E_k, E_l, E_m\right)\right), E_m \wedge T\left(E_k, E_l, E_i\right)\right) \\
&= \sum_{i,k,l,m=1}^{n} g\left(\mathfrak{R}\left(E_i \wedge T\left(E_k, E_l, E_m\right)\right), E_m \wedge T\left(E_l, E_k, E_i\right)\right).
\end{aligned}
$$

Therefore, if we define elements $\theta_{iklm} \in \Lambda^2 TM$ by

$$
\begin{aligned}
\theta_{iklm} = {} & E_i \wedge T\left(E_k, E_l, E_m\right) \\
& + E_k \wedge T\left(E_i, E_m, E_l\right) \\
& + E_l \wedge T\left(E_m, E_i, E_k\right) \\
& + E_m \wedge T\left(E_l, E_k, E_i\right),
\end{aligned}
$$

then one easily checks that

$$\sum_{i,k,l,m=1}^{n} g\left(\mathfrak{R}\left(\theta_{iklm}\right), \theta_{iklm}\right) = 4K$$

by observing that after multiplying out, there are 16 terms on the left-hand side, which can be collected in groups of four. After reindexing some of the sums, each of these groups consists of four equal terms that correspond to one of A, B, C, or D. Since the left-hand side is assumed to be nonnegative, we have proven the desired result. □

Corollary 5.4 (Tachibana, 1974) *If (M, g) is a compact oriented Riemannian manifold with* $\operatorname{div} R = 0$ *and* $\mathfrak{R} \geq 0$, *then* $\nabla R = 0$. *If in addition,* $\mathfrak{R} > 0$, *then* (M, g) *has constant curvature.*

Proof. The first part is immediate from the above theorems. For the second part we have again that $K = 0$. Since $\mathfrak{R}$ is assumed to be positive, we must therefore have that

$$\begin{aligned}\theta_{iklm} &= E_i \wedge R(E_k, E_l) E_m \\ &\quad + E_k \wedge R(E_i, E_m) E_l \\ &\quad + E_l \wedge R(E_m, E_i) E_k \\ &\quad + E_m \wedge R(E_l, E_k) E_i \\ &= 0.\end{aligned}$$

From this one can see that the curvature must be constant. A different proof of this can be found using the material from Chapter 8. □

7.6 Further Study

For more general and complete accounts of the Bochner technique and spin geometry we recommend the two texts [84] and [54]. The latter book also has a complete proof of the Hodge theorem. Other sources for this particular result are [50], [72], and [82]. For more information about Killing fields and related matters we refer the reader to [52, Chapter II]. There is also a good elementary account of Killing fields in O'Neill's book [65, Chapter 9].

For other generalizations to manifolds with integral curvature bounds the reader should consult [36]. In there the reader will find a complete discussion on generalizations of the above mentioned results about Betti numbers.

7.7 Exercises

1. Let $f : (M, g) \to (\mathbb{R}^k, \text{can})$ be a Riemannian submersion and let (M, g) be complete. If $\nabla^2 f = 0$ (each of the components has zero Hessian), then $(M, g) = (N, h) \times (\mathbb{R}^k, \text{can})$.

2. Suppose we have a frame of Killing fields $X_1, \ldots, X_n$ on a Riemannian manifold (M, g). Show that the structure constants c_{ij}^k defined by

$$[X_i, X_j] = c_{ij}^k X_k$$

are constant. Thus, a Killing frame is always a finite-dimensional Lie algebra. Recall that we used a Killing frame to compute the curvatures of the Berger spheres. What can you say about manifolds with globally defined Killing frames?

3. Given two Killing fields X and Y on a Riemannian manifold, develop a formula for $\Delta g(X, Y)$. Use this to give a formula for the Ricci curvature in a Killing frame.

4. For a vector field X define the Lie derivative of the connection as follows:

$$\begin{aligned}(L_X \nabla)(U, V) &= L_X(\nabla_U V) - \nabla_{L_X U} V - \nabla_U L_X V \\ &= [X, \nabla_U V] - \nabla_{[X,U]} V - \nabla_U [X, V].\end{aligned}$$

   (a) Show that $L_X \nabla$ is a $(1, 2)$-tensor.

   (b) We say that X is an affine vector field if $L_X \nabla = 0$. Show that for such a field we have
$$\nabla^2_{U,V} X = -R(X, U) V.$$
   (Hint: Show that: $R(W, U) V + \nabla^2_{U,V} W = (L_W \nabla)(U, V)$.)

   (c) Show that Killing fields are affine. Give an example of an affine field on $\mathbb{R}^n$ which is not a Killing field.

5. Let K be a Killing field on a Riemannian manifold.

   (a) Show that $dK^\flat(X, Y) = 2g(\nabla_X K, Y)$.

   (b) For any vector field X, show more generally that
$$g(\nabla_v X, w) = \frac{1}{2}\left((L_X g)(v, w) + \left(dX^\flat\right)(v, w)\right).$$
   Use this to conclude that
$$|L_X g|^2 = 2|\nabla X|^2 + 2\text{tr}(\nabla X)^2.$$

(c) Establish the following integral formulae on a closed oriented Riemannian manifold:

$$\int_M \left(\operatorname{Ric}(X, X) + \operatorname{tr}(\nabla X)^2 - (\operatorname{div} X)^2\right) = 0,$$

$$\int_M \left(\operatorname{Ric}(X, X) + g\left(\operatorname{tr}\nabla^2 X, X\right) + \frac{1}{2}\left|L_X g\right|^2 - (\operatorname{div} X)^2\right) = 0.$$

(d) Finally, show that X is a Killing field iff

$$\begin{aligned} \operatorname{div} X &= 0, \\ \operatorname{tr}\nabla^2 X &= -\operatorname{Ric}(X). \end{aligned}$$

6. (Yano) If X is an affine vector field show that $\operatorname{tr}\nabla^2 X = -\operatorname{Ric}(X)$ and that $\operatorname{div} X$ is constant. Use this together with the above characterizations of Killing fields to show that on closed manifolds affine fields are Killing fields.

7. If K is a Killing field show that L_K and Δ commute as operators on forms. Conversely show that X is a Killing field if L_X and Δ commute on functions.

8. Suppose (M, g) is compact and has $b_1 = k$. If $\operatorname{Ric} \geq 0$, then the universal covering splits: $\left(\tilde{M}, g\right) = (N, h) \times \left(\mathbb{R}^k, \mathrm{can}\right)$.

9. Let (M, g) be an n-dimensional Riemannian manifold that is isometric to Euclidean space outside some compact subset $K \subset M$, i.e., $M - K$ is isometric to $\mathbb{R}^n - C$ for some compact set $C \subset \mathbb{R}^n$. If $\operatorname{Ric}_g \geq 0$, show that $M = \mathbb{R}^n$. (Hint: (1) Find a metric on the n-torus that is isometric to a neighborhood of $K \subset M$ somewhere and otherwise flat. (2) Alternatively, show that any parallel 1-form on $\mathbb{R}^n - C$ extends to a harmonic 1-form on M. Then apply Bochner's formula to show that it must in fact be parallel when $\operatorname{Ric}_g \geq 0$, and use this to conclude that the manifold is flat.)

10. Given two vector fields X and Y on (M, g) such that ∇X and ∇Y are symmetric, develop Bochner formulae for $\nabla^2 \frac{1}{2} g(X, Y)$ and $\Delta \frac{1}{2} g(X, Y)$.

11. For general sections s_1 and s_2 of an appropriate bundle show in analogy with the formula $\Delta \frac{1}{2} |s|^2 = |\nabla s|^2 + \left(\operatorname{tr}\nabla^2 s, s\right)$ that:

$$\Delta \langle s_1, s_2 \rangle = 2 \langle \nabla s_1, \nabla s_2 \rangle + \left\langle \operatorname{tr}\nabla^2 s_1, s_2 \right\rangle + \left\langle s_1, \operatorname{tr}\nabla^2 s_2 \right\rangle.$$

Use this on forms and spinors to develop Bochner formulae from the Weitzenböck formulae for inner products of such sections.

More generally we can consider the 1-form defined by $\omega(v) = \langle \nabla_v s_1, s_2 \rangle$ which represents half of the differential of $\langle s_1, s_2 \rangle$. Show that

$$\begin{aligned} -\delta\omega &= \langle \nabla s_1, \nabla s_2 \rangle + \left\langle \operatorname{tr}\nabla^2 s_1, s_2 \right\rangle \\ &= \left\langle \left(\nabla^*\nabla + \operatorname{tr}\nabla^2\right) s_1, s_2 \right\rangle, \\ d\omega(X, Y) &= \langle R(X, Y) s_1, s_2 \rangle - \langle \nabla_X s_1, \nabla_Y s_2 \rangle + \langle \nabla_Y s_1, \nabla_X s_2 \rangle. \end{aligned}$$

12. Show that in dimension 2,

$$K = g\left(R^{\flat}, (\mathrm{div}\nabla - \nabla\mathrm{div})\, R\right) = 0.$$

13. (Berger) Let (M, g) be a Riemannian manifold with a $(1, 1)$-tensor field T that is symmetric and whose covariant derivative is symmetric $(\nabla_X T)(Y) = (\nabla_Y T)(X)$. Show that

$$\Delta\frac{1}{2}|T|^2 = |\nabla T|^2 + g\left(T^{\flat}, \nabla\mathrm{div}T\right) + g\left(T^{\flat}, (\mathrm{div}\nabla - \nabla\mathrm{div})\, T\right).$$

And when M is compact and oriented conclude that if $\sec \geq 0$ and $\mathrm{div}T = 0$, then $\nabla T = 0$. Moreover, if $\sec > 0$, then $T = c \cdot I$ for some constant c. In case S is not symmetric establish a Bochner formula that can be used to arrive at the above results.

14. (Berger) On a closed Riemannian manifold (M, g) show that if $\mathrm{div}R = 0$ and $\sec \geq 0$, then $\nabla\mathrm{Ric} = 0$. (Hint: use an exercise from Chapter 2 to get the symmetry for $\nabla\mathrm{Ric}$ and also the formula $2\mathrm{div}\mathrm{Ric} = d\,(\mathrm{scal})$ to conclude that $\mathrm{div}\mathrm{Ric} = 0$.)

15. Let $(M, g) \looparrowright \mathbb{R}^{n+1}$ be an isometric immersion of an oriented manifold.

   (a) Using the Codazzi equations, show that

   $$\Delta\frac{1}{2}|S|^2 = |\nabla S|^2 + g\left(S^{\flat}, \nabla\mathrm{div}S\right) + K,$$

   where S is the shape operator and K is as usual defined by

   $$K = g\left(S^{\flat}, (\mathrm{div}\nabla - \nabla\mathrm{div})\, S\right).$$

   (b) Assuming that M is compact, show that

   $$\int |\nabla S|^2 = \int |d\,(\mathrm{tr}S)|^2 - \int K.$$

   (Recall that we proved in the exercises to Chapter 4 that $\mathrm{div}S = d\,(\mathrm{tr}S)$.)

   (c) Show Liebmann's theorem: If (M, g) has constant mean curvature ($\mathrm{tr}S = \text{constant}$) and nonnegative shape operator, then (M, g) is a constant-curvature sphere. (Hint: using Chapter 4, find out something about the curvature from the positivity of S; then use $K = \sum_{i<j}\left(\lambda_j - \lambda_i\right)^2 \cdot \sec\left(E_i, E_j\right)$.)

In case $M = S^2$, H. Hopf showed that one can prove this theorem without using the nonnegativity of the shape operator. This is not too hard to believe, as we know that

$$K(p) = (\lambda_2 - \lambda_1)^2 \cdot \sec(p),$$

$$\int \sec(p)\, d\mathrm{vol} = 4\pi,$$

indicating that $\int K$ should be nonnegative. On the other hand, Wente has exhibited immersed tori with constant mean curvature (see Wente's article in [41]).

16. Show that if one defines the divergence of a p-form by

$$\begin{aligned} \mathrm{div}\omega\left(X_2,\ldots,X_p\right) &= \sum_{i=1}^{n}\left(\nabla_{E_i}\omega\right)\left(E_i,X_2,\ldots,X_p\right) \\ &= \sum_{i=1}^{n} i_{E_i}\left(\nabla_{E_i}\omega\right)\left(X_2,\ldots,X_p\right), \end{aligned}$$

where E_i is an orthonormal frame, then $\delta = -\mathrm{div}$.

17. Suppose we have a Killing field K on a closed oriented Riemannian manifold (M, g). Assume that ω is a harmonic form.

    (a) Show that $L_K\omega = 0$. (Hint: Show that $L_K\omega$ is also harmonic.)

    (b) Show that $i_K\omega$ is closed, but not necessarily harmonic.

18. Let (M, g) be a Kähler manifold with Kähler form ω. Show using the exercises from Chapter 2 that

$$\omega^k = \underbrace{\omega\wedge\cdots\wedge\omega}_{\text{ktimes}}$$

is closed but not exact by showing that $\omega^{(\dim M)/2}$ is proportional to the volume form. Conclude that none of the even homology groups vanish.

19. Let $E \to M$ be a vector bundle with connection ∇.

    (a) Show that ∇ induces a natural connection on $\mathrm{Hom}(E, E)$ that we also denote by ∇.

    (b) Let $\Omega^p(M, E)$ denote the alternating p-linear maps from TM to E (note that $\Omega^0(M, E) = \Gamma(E)$.) Show that $\Omega^*(M)$ acts in a natural way from both left and right on $\Omega^*(M, E)$ by wedge product. Show also that there is a natural wedge product $\Omega^p(M, \mathrm{Hom}(E, E)) \times \Omega^q(M, E) \to \Omega^{p+q}(M, E)$.

    (c) Show that there is a connection-dependent exterior derivative $d^\nabla : \Omega^p(M, E) \to \Omega^{p+1}(M, E)$ with the property that it satisfies the exterior derivative version of Leibniz's rule with respect to the above defined wedge products, and such that for $s \in \Gamma(E)$ we have: $d^\nabla s = \nabla s$.

    (d) If we think of the curvature $R(X, Y)s$ as an element of $\Omega^2(M, \mathrm{Hom}(E, E))$, show that: $\left(d^\nabla \circ d^\nabla\right)(s) = R \wedge s$ for any $s \in \Omega^p(M, E)$ and that Bianchi's second identity can be stated as $d^\nabla R = 0$.

20. If we let $E = TM$ in the previous exercise, then $\Omega^1(M, TM) = \text{Hom}(TM, TM)$ will just consist of all $(1, 1)$-tensors.

    (a) Show that in this case $d^\nabla s = 0$ iff s is a Codazzi tensor.

    (b) The entire chapter seems to indicate that whenever we have a tensor bundle E $(= \mathbb{R}, TM, \Lambda^2 M, \text{etc.})$ and an element $s \in \Omega^p(M, E)$ with $d^\nabla s = 0$, then there is a Bochner-type formula for s. Moreover, when in addition s is "divergence free" and some sort of curvature is nonnegative, then s should be parallel. Can you develop a theory in this generality?

    (c) Show that if X is a vector field, then ∇X is a Codazzi tensor iff $R(\cdot, \cdot) X = 0$. Give an example of a vector field such that ∇X is Codazzi but X itself is not parallel. Is it possible to establish a Bochner type formula for exact tensors like $\nabla X = d^\nabla X$ even if they are not closed?

21. (Thomas) Show that in dimensions $n > 3$ the Gauss equations ($\mathfrak{R} = S \wedge S$) imply the Codazzi equations $(d^\nabla S = 0)$ provided $\det S \neq 0$. (Hint: use the second Bianchi identity and be very careful with how things are defined. It will also be useful to study the linear map $\text{Hom}(\Lambda^2 V, V) \to \text{Hom}(\Lambda^3 V, \Lambda^2 V)$ defined by $T \to T \wedge S$ for a linear map $S : V \to V$. In particular, one can see that this map is injective only when the rank of S is ≥ 4.)

22. Aside from the Euler characteristic, there are other topological invariants. In dimensions $4n$ we have that the Hodge $* : H^{2n}(M) \to H^{2n}(M)$ satisfies $** = I$. The difference in the dimensions of the eigenspaces for ± 1 is called the *signature* of M :

$$\tau(M) = \sigma(M) = \dim(\ker(* - I) - \ker(* + I)).$$

One can show that this does not depend on the metric used to define $*$, by observing that it is the index of the symmetric bilinear map

$$H^{2n}(M) \times H^{2n}(M) \to \mathbb{R},$$
$$(\omega_1, \omega_2) \to \int \omega_1 \wedge \omega_2.$$

Recall that the index of a symmetric bilinear map is the difference between positive and negative diagonal elements when it has been put into diagonal form. In dimension 4 one can show that

$$\sigma(M) = \frac{1}{12\pi^2} \int_M \left(|W^+|^2 - |W^-|^2 \right).$$

Using the exercises from Chapter 4, show that for an Einstein metric in dimension 4 we have

$$\chi(M) \geq \frac{3}{2} \sigma(M),$$

with equality holding iff the metric is Ricci flat and $W^- = 0$. Conclude that not all four manifolds admit Einstein metrics. (Hint: consider connected sums of $\mathbb{C}P^2$ with itself k times.) In higher dimensions there are no known obstructions to the existence of Einstein metrics.

23. Recall the curvature forms defined using an orthonormal frame E_i:

$$\Omega_i^j (X, Y) E_j = R(X, Y) E_i.$$

They yield a skew-symmetric matrix of 2-forms:

$$\Omega = \left(\Omega_i^j\right).$$

From linear algebra we know that there are various invariant polynomials that depend on the entries of matrices, e.g., the trace and determinant. We can define similar objects in this case as follows:

$$p_{2l}(\Omega) = \sum \Omega_{i_1}^{i_2} \wedge \Omega_{i_2}^{i_3} \wedge \cdots \wedge \Omega_{i_l}^{i_1}.$$

These are known as the *Pontryagin forms*. Show that they yield globally defined forms that are closed (you need to look at the exercises in Chapter 2 and also understand what the second Bianchi identity has to do with $d\Omega$). Show that they are zero when l is odd. Thus, they generate homology classes $p_l \in H^{4l}$, which are known as the Pontryagin classes of the manifold. It can be shown that these classes do not depend on the metric.

Show that the Pontryagin classes are zero on a manifold with constant curvature. (Hint: use that we know what the curvature tensor looks like.) Thus even in the case where $4l = n = \dim M$, we do not necessarily have that p_l is the Euler class.

Try to compute $p_1 \in H^4$ for some of the standard 4-manifolds.

24. In case the manifold has even dimension $n = 2m$, we can construct the *Euler form*:

$$e(\Omega) = \varepsilon^{i_1 \cdots i_n} \cdot \Omega_{i_2}^{i_1} \wedge \cdots \wedge \Omega_{i_n}^{i_{n-1}},$$
$$\varepsilon^{i_1 \cdots i_n} = \text{sign of the permutation } (i_1 \cdots i_n),$$

which modulo a factor generates the Euler class, or characteristic, of the manifold. Show that this form also yields a globally defined closed form. Note that this is essentially the square root of the determinant of Ω. However, as this determinant is a $2n$ form, it is always zero and therefore doesn't yield anything interesting. The cohomology class of $e(\Omega)$ can also be seen to be independent of the metric. Moreover, as discussed in Chapter 4, it is proportional to the Euler characteristic.

8

Symmetric Spaces and Holonomy

In this chapter we shall give a brief overview of (locally) symmetric spaces and holonomy. Only the simplest proofs will be presented. Thus, we will have to be sketchy in places. Still, most of the standard results are proved or at least mentioned. We give some explicit examples, including the complex projective space, in order to show how one can compute curvatures on symmetric spaces relatively easily. There is a brief introduction to holonomy and the de Rham decomposition theorem. We give a few interesting consequences of this theorem and then proceed to discuss how holonomy and symmetric spaces are related. Finally, we classify all compact manifolds with nonnegative curvature operator. We shall in a few places use results from Chapter 9. They will therefore have to be taken for granted at this point.

As we have already seen, Riemann showed that locally there is only one constant-curvature geometry. After Lie's work on "continuous" groups it became clear that one had many more interesting models for geometries. Next to constant curvature spaces, the most natural type of geometry to try to understand is that of (locally) symmetric spaces. At the end of the nineteenth century it became a well-defined problem to classify all such geometries. The history of symmetric spaces parallels that of the Lebesgue integral. Namely, one person managed to take all the glory, Elie Cartan. He started out in his thesis with classifying all simple complex Lie algebras. Using this he later classified all the simple real Lie algebras. With the help of this and many of his different characterizations of symmetric spaces, Cartan, by the mid 1920s had managed to give a complete (local) classification of all symmetric spaces. This was an astonishing achievement even by today's deconstructionist standards, not least because Cartan also had to classify the simple Lie algebras. This in itself takes so much work that most courses on Lie algebras these days give up after having settled the complex case.

After Cartan's work, a few people worked on getting a better understanding of some of these new geometries and also on giving a more global classification. Still, not much happened until the 1950s, when people realized a serious connection between symmetric spaces and holonomy. Here we are thinking of the de Rham decomposition theorem and Berger's classification of holonomy groups. With this work it became clear that almost all holonomy groups occurred for symmetric spaces and therefore gave good approximating geometries to most holonomy groups. An even more interesting question also came out of this, namely, What about those few holonomy groups that do not occur for symmetric spaces? This is related to the study of Kähler manifolds and then some exotic geometries in dimensions 7 and 8. The Kähler case seems to be quite well understood by now, not least because of Yau's work on the Calabi conjecture. The exotic geometries have only very recently become better understood with D. Joyce's work.

8.1 Symmetric Spaces

There are many ways of representing symmetric spaces. Below we shall see how they can be described via homogeneous spaces, Lie algebras, and finally, by their curvature tensor.

8.1.1 *The Homogeneous Description*

We say that a Riemannian manifold (M, g) is a *symmetric space* if for each $p \in M$ the isotropy group Iso_p contains an isometry I_p such that $DI_p : T_pM \to T_pM$ is the antipodal map $-I$. Since isometries preserve geodesics, we immediately see that for any geodesic $\gamma(t)$ such that $\gamma(0) = p$ we have that $I_p \circ \gamma(t) = \gamma(-t)$. Using this, it is easy to show that symmetric spaces are homogeneous and complete. Namely, if two points are joined by a geodesic, then the symmetry in the midpoint between these points on the geodesic is an isometry that maps these points to each other. Thus, any two points that can be joined by a broken sequence of geodesics can be mapped to each other by an isometry. This shows that the space is homogeneous. It is then easy to show that the space is complete. In conclusion, we see that any symmetric space looks like

$$G/H = \mathrm{Iso}/\mathrm{Iso}_p.$$

Given a homogeneous space $G/H = \mathrm{Iso}/\mathrm{Iso}_p$, we see that it is symmetric provided that the symmetry I_p exists for just one p. The symmetry I_q can then be constructed by selecting an isometry g that takes p to q and then observing that

$$g \circ I_p \circ g^{-1}$$

has the correct differential at q. This means, in particular, that any Lie group G with bi-invariant metric is a symmetric space, since $g \to g^{-1}$ is the desired symmetry around the identity element. Let us list some of the important families of homogeneous spaces that are symmetric. They come in pairs of compact and noncompact spaces. Below we list just a few families of examples. There are many more families and several exceptional examples as well.

Lie groups with bi-invariant metrics

group	rank	dim
$SU(n+1)$	n	$n(n+2)$
$SO(2n+1)$	n	$n(2n+1)$
$Sp(n)$	n	$n(2n+1)$
$SO(2n)$	n	$n(2n-1)$

Noncompact analogues of bi-invariant metrics

(complexified group)/group	rank	dim
$SL(n+1,\mathbb{C})/SU(n+1)$	n	$n(n+2)$
$SO(2n+1,\mathbb{C})/SO(2n+1)$	n	$n(2n+1)$
$Sp(n,\mathbb{C})/Sp(n)$	n	$n(2n+1)$
$SO(2n,\mathbb{C})/SO(2n)$	n	$n(2n-1)$

Compact homogeneous examples

Iso	Iso_p	dim	rank	description
$SO(n+1)$	$SO(n)$	n	1	Sphere
$O(n+1)$	$O(n)\times\{1,-1\}$	n	1	$\mathbb{R}P^n$
$U(n+1)$	$U(n)\times U(1)$	$2n$	1	$\mathbb{C}P^n$
$Sp(n+1)$	$Sp(n)\times Sp(1)$	$4n$	1	$\mathbb{H}P^n$
F_4	$Spin(9)$	16	1	Cayley plane
$SO(p+q)$	$SO(p)\times SO(q)$	pq	$\min(p,q)$	real Grassmannian
$SU(p+q)$	$S(U(p)\times U(q))$	$2pq$	$\min(p,q)$	complex Grassmannian

Noncompact homogeneous examples

Iso	Iso_p	dim	rank	description
$SO(n,1)$	$SO(n)$	n	1	Hyperbolic space
$O(n,1)$	$O(n)\times\{1,-1\}$	n	1	Hyperbolic $\mathbb{R}P^n$
$U(n,1)$	$U(n)\times U(1)$	$2n$	1	Hyperbolic $\mathbb{C}P^n$
$Sp(n,1)$	$Sp(n)\times Sp(1)$	$4n$	1	Hyperbolic $\mathbb{H}P^n$
F_4^{-20}	$Spin(9)$	16	1	Hyperbolic Cayley plane
$SO(p,q)$	$SO(p)\times SO(q)$	pq	$\min(p,q)$	Hyperbolic Grassmannian
$SU(p,q)$	$S(U(p)\times U(q))$	$2pq$	$\min(p,q)$	Complex hyperbolic Grassmannian

Recall that $Spin(n)$ is the universal double covering of $SO(n)$ for $n>1$. We also have the following special identities for low dimensions:

$$Spin(3)=SU(2)=Sp(1),$$

$$Spin(4) = Spin(3) \times Spin(3).$$

Note that all of the compact examples have $\sec \geq 0$, by O'Neill's formula. It also follows from this formula that all the projective spaces (compact and noncompact) have quarter pinched metrics, i.e., the ratio between the smallest and largest curvatures is $\frac{1}{4}$. Below we shall do some concrete calculations to justify these remarks.

In the above list of examples we have a column called *rank*. The *rank* of a geodesic $\gamma : \mathbb{R} \to M$ is simply the dimension of parallel fields E along γ such that $g(R(E(t), \dot{\gamma}(t))\dot{\gamma}(t), E(t)) = 0$ for all t. The rank of a geodesic is therefore always ≥ 1. The rank of a Riemannian manifold is now defined as the minimum rank over all of the geodesics in M. The concept of rank for symmetric spaces has to do with maximal tori in Lie groups and is therefore more or less algebraic. For a general manifold there might of course be metrics with different ranks, but this is actually not so obvious. Is it, for example, possible to find a metric on the sphere of rank > 1? A general remark is that of course any Cartesian product has rank ≥ 2, and also many symmetric spaces have rank ≥ 2. In general it is unclear to what extent other manifolds can also have rank ≥ 2. However, see below for the case of nonpositive curvature and nonnegative curvature operators. Note that there are five compact rank one symmetric spaces (CROSS) in the above lists. These are the only simply connected compact rank 1 symmetric spaces.

To get back to the rank for symmetric spaces, let us start with a Lie group G of dimension $n \geq 3$ endowed with a bi-invariant metric. Now, rank has to do with having lots of zero curvature. On G we know from the exercises from Chapter 2 that $\sec(x, y) = 0$ iff $[x, y] = 0$, where $x, y \in T_eG = \mathfrak{g}$. It is therefore not surprising that the rank can be computed knowing only the group structure. Namely, it must be the dimension of the largest Abelian subalgebra $\mathfrak{a} \subset \mathfrak{g}$. Recall the linear map $L : \Lambda^2\mathfrak{g} \to \mathfrak{g}$ defined by $x \wedge y \to [x, y]$. The kernel of this map is exactly $\Lambda^2\mathfrak{a}$, and thus $\mathfrak{a}$ is uniquely defined and satisfies

$$(\dim \mathfrak{a})(\dim \mathfrak{a} - 1) \geq n(n-3).$$

Therefore, the rank is always > 1 when the dimension is > 3. Using this, one can see that most quotients of Lie groups also have rank > 1.

8.1.2 Isometries and Parallel Curvature

Another interesting property for symmetric spaces is that they have parallel curvature tensor. This is because the symmetries I_p leave the curvature tensor and its covariant derivative invariant. In particular, we have

$$DI_p((\nabla_X R)(Y, Z, W)) = \left(\nabla_{DI_pX} R\right)\left(DI_pY, DI_pZ, DI_pW\right),$$

which at p implies

$$\begin{aligned} -(\nabla_X R)(Y, Z, W) &= (\nabla_{-X} R)(-Y, -Z, -W) \\ &= (\nabla_X R)(Y, Z, W). \end{aligned}$$

Thus, $\nabla R = 0$. This almost characterizes symmetric spaces.

Theorem 1.1 (E. Cartan) *If (M, g) is a Riemannian manifold with parallel curvature tensor, then for each $p \in M$ there is an isometry I_p defined in a neighborhood of p with $DI_p = -I$ on T_pM. Moreover, if (M, g) is simply connected and complete, then the symmetry is defined on all of M, and in particular, the space is symmetric.*

Proof. The global statement follows from the local one using an analytic continuation argument and Theorem 1.2, which proved below. Note that for the local statement we already have a candidate for a map. Namely, if ε is so small that $\exp_p : B(0, \varepsilon) \to B(p, \varepsilon)$ is a diffeomorphism, then we can just define $I_p(x) = -x$ in these coordinates. It now remains to see why this is an isometry when we have parallel curvature tensor. To this effect, we must show that in these coordinates the metric is the same at x and $-x$. Switching to polar coordinates, we have the usual equations relating curvature and the metric. So the claim follows if we can prove that the curvature tensor is the same when we go in opposite directions. To check this, first observe

$$R(\cdot, v)v = R(\cdot, -v)(-v).$$

So the curvatures start out being the same. If ∂_r is the radial field, we also have

$$\left(\nabla_{\partial_r} R\right) = 0.$$

Thus, the curvature tensors not only start out being equal, but also satisfy the same simple first-order equation. Thus, they must remain the same as we go in opposite directions. □

A Riemannian manifold with parallel curvature tensor is called a *locally symmetric space*.

We should also mention at this point that there are left-invariant metrics that are not even locally symmetric. Namely, in the exercises to Chapter 3 it is shown that the Berger spheres ($\varepsilon \neq 1$) and the Heisenberg group do not have parallel curvature tensor. In fact, they don't even have parallel Ricci tensor, but this comes as no surprise, as they are 3-dimensional.

With very little extra work we can generalize the above theorem on the existence of local symmetries. Recall that in our discussion about existence of isometries with a given differential in Chapter 6 we decided that they could exist only when the spaces had the same constant curvature. However, there is a generalization to symmetric spaces. Namely, we know that any isometry preserves the curvature tensor. Thus, if we start with a linear isometry that preserves the curvatures at a point, then we should be able to extend this map in the situation where curvatures are everywhere the same. This is the content of the next theorem.

Theorem 1.2 (E. Cartan) *Suppose we have a simply connected symmetric space (M, g) and a complete locally symmetric space (N, h) of the same dimension. Given a linear isometry $L : T_pM \to T_qN$ such that*

$$L\left(R^g(x, y)z\right) = R^h(Lx, Ly)Lz$$

for all $x, y, z \in T_pM$, there is a unique Riemannian isometry $M \to N$ such that $D_p\varphi = L$.

Proof. The proof of this is, as usual, by analytic continuation, given that we can find these isometries locally. Given that there is an isometry defined locally, we know that it must look like

$$\varphi = \exp_q \circ L \circ \exp_p^{-1}.$$

To see that this indeed defines an isometry, we have to show that the metrics in exponential coordinates are the same via the identification of the tangent spaces by L. As usual the radial curvatures from the centers determine the metrics. In addition, the curvatures are parallel and therefore satisfy the same first-order equation. Now initially, we assume that the curvatures are the same at p and q via the linear isometry. But then they must be the same in frames that are radially parallel around these points. Consequently, the spaces are locally isometric. □

This result shows that the curvature tensor completely characterizes the symmetric space. We shall study this further below.

8.1.3 Algebraic Descriptions of Symmetric Spaces

It is worthwhile to get a more algebraic description of symmetric spaces. Note that there are many ways of writing homogeneous spaces as quotients G/H, e.g.,

$$S^3 = SU(2) = SO(4)/SO(3) = U(2)/(U(1) \times U(1)).$$

But only one of these, $SO(4)/SO(3)$, tells us that S^3 is a symmetric space. However, there are also several ways of writing it as a symmetric space:

$$SO(4)/SO(3) = O(4)/O(3) = Spin(4)/Spin(3).$$

Note, however, that at the Lie algebra level those three descriptions look the same. To get a more complete picture, we also have to understand how the involution acts, not just on the space M, but as a map in $\mathrm{Iso}(M, g)$, and then in the Lie algebra $\mathfrak{iso}(M, g)$ of Killing fields. We shall here give the *isometry description* of a symmetric space.

Let us fix a symmetric space (M, g) and a point $p \in M$. Recall from Chapter 7 that the map

$$\begin{aligned} \mathfrak{iso} &\to T_pM \times \mathfrak{so}\left(T_pM\right), \\ X &\to (X(p), (\nabla X)(p)) \end{aligned}$$

is a linear isomorphism. Since (M, g) is homogeneous, this linear map will be onto the first factor. Thus, $\mathfrak{iso}$ can be identified with $T_pM \times \mathfrak{iso}_p$. This then induces a Lie algebra structure on $T_pM \times \mathfrak{iso}_p$ from that on $\mathfrak{iso}$. To understand this structure a little better, let us first observe that the decomposition $T_pM \times \mathfrak{iso}_p$ at the level of Killing fields looks like

$$\begin{aligned} X \in T_pM \quad &\text{iff} \quad (\nabla X)(p) = 0, \\ X \in \mathfrak{iso}_p \quad &\text{iff} \quad X(p) = 0. \end{aligned}$$

So as not to confuse Killing fields with vectors, let us introduce the terminology

$$\mathfrak{t}_p = \{X \in \mathfrak{iso} : (\nabla X)(p) = 0\}.$$

Let us check where the Lie brackets of various combinations of Killing fields X, Y lie.

(a) If $X, Y \in \mathfrak{t}_p$ or $X, Y \in \mathfrak{iso}_p$, then

$$[X, Y](p) = \nabla_{X(p)}Y - \nabla_{Y(p)}X = 0.$$

So we conclude that $[X, Y] \in \mathfrak{iso}_p$ in these cases. In the case where $X, Y \in \mathfrak{iso}_p$, we even have that the Lie bracket coincides, up to sign, with the Lie bracket coming from $\mathfrak{so}(T_pM)$. Namely, the map

$$\nabla^2_{V,W}Y = \nabla_V\nabla_W Y - \nabla_{\nabla_V W}Y$$

is tensorial in V and W. Therefore, if $v \in T_pM$ and $X(p) = 0$, we see that

$$\nabla_v\nabla_X Y = \nabla_{\nabla_v X}Y.$$

But this implies, in the case where $X, Y \in \mathfrak{iso}_p$ and $v \in T_pM$, that

$$\begin{aligned} [\nabla X(p), \nabla Y(p)](v) &= (\nabla X \circ \nabla Y - \nabla Y \circ \nabla X)(v) \\ &= \nabla_{\nabla_v Y}X - \nabla_{\nabla_v X}Y \\ &= \nabla_v(\nabla_Y X - \nabla_X Y) \\ &= -\nabla_v[X, Y]. \end{aligned}$$

Hence, the element $[X, Y] \in \mathfrak{iso}$ is identified with $-[\nabla X(p), \nabla Y(p)]$ inside $\mathfrak{so}(T_pM)$.

(b) If $X \in \mathfrak{t}_p$ and $Y \in \mathfrak{iso}_p$, then

$$[X, Y](p) = \nabla_{X(p)}Y = (\nabla Y)(X(p)),$$

which is simply the way the elements $Y \in \mathfrak{so}(T_pM)$ act on T_pM.

In conclusion, we see that the Lie algebra $\mathfrak{iso}$ can be represented as a direct sum: $\mathfrak{iso} = \mathfrak{t}_p \oplus \mathfrak{iso}_p$, where $\mathfrak{t}_p$ is a vector space with a Euclidean metric, and $\mathfrak{iso}_p$

is a subalgebra of the skew-symmetric transformations on $\mathfrak{t}_p$. Moreover, the Lie algebra structure on $\mathfrak{iso} = \mathfrak{t}_p \oplus \mathfrak{iso}_p$ is given by

$$\begin{aligned} [h_1, h_2] &= -(h_1 \circ h_2 - h_2 \circ h_1) \quad \text{if} \quad h_i \in \mathfrak{iso}_p, \\ [h, x] &= -[x, h] = h(x) \quad \text{if} \quad h \in \mathfrak{iso}_p \quad \text{and} \quad x \in \mathfrak{t}_p, \\ [x, y] &\in \mathfrak{iso}_p \quad \text{for} \quad x, y \in \mathfrak{t}_p. \end{aligned}$$

Thus, the only Lie brackets that are not given canonically are $[x, y]$, where $x, y \in \mathfrak{t}_p$.

All of this, of course, works for homogeneous spaces, so what about the involution? It can actually be seen as acting on both Iso and $\mathfrak{iso}$. In the latter case, we can guess that it should be the identity on $\mathfrak{iso}_p$, but multiplication by -1 on $\mathfrak{t}_p$. At the Lie group level, it is defined as the map $\sigma : \text{Iso} \to \text{Iso}$ such that

$$\sigma(g) = I_p \circ g \circ I_p.$$

Then we have found an automorphism that is characterized by

$$\begin{aligned} \sigma(g) &= g \quad \text{iff} \quad g \in \text{Iso}_p, \\ \sigma \circ \sigma &= id. \end{aligned}$$

If we take the differential of σ, then we get a linear map $D\sigma : \mathfrak{iso} \to \mathfrak{iso}$, which is a Lie algebra automorphism such that

$$\begin{aligned} D\sigma(h) &= h \quad \text{for all} \quad h \in \mathfrak{iso}_p, \\ D\sigma(x) &= -x \quad \text{for all} \quad x \in \mathfrak{t}_p. \end{aligned}$$

Thus, $D\sigma$ acts in the desired way on T_pM. Since σ also fixes Iso_p, it clearly induces a map on Iso/Iso_p whose differential on $T_{\text{Iso}_p}(\text{Iso}/\text{Iso}_p)$ is $-id$. But then we have found a completely algebraic description of a symmetric space.

Conversely, suppose we have a Lie algebra $\mathfrak{g}$ and a Lie algebra automorphism $L : \mathfrak{g} \to \mathfrak{g}$ that is an involution. Then we can construct a symmetric space as follows: First decompose $\mathfrak{g} = \mathfrak{t} \oplus \mathfrak{k}$ where $\mathfrak{t}$ is the -1 eigenspace for L and $\mathfrak{k}$ is the 1 eigenspace for L. Then observe that $\mathfrak{k}$ is a Lie subalgebra, since

$$\begin{aligned} L[h_1, h_2] &= [Lh_1, Lh_2] \\ &= [h_1, h_2]. \end{aligned}$$

Note also that for similar reasons,

$$\begin{aligned} [\mathfrak{k}, \mathfrak{t}] &\subset \mathfrak{t}, \\ [\mathfrak{t}, \mathfrak{t}] &\subset \mathfrak{k}. \end{aligned}$$

Suppose now that there is a compact Lie group K such that its Lie algebra is $\mathfrak{k}$. Then we can choose a Euclidean metric on $\mathfrak{t}$ making the adjoint action of K on $\mathfrak{t}$ isometric. Then we see that the decomposition $\mathfrak{g} = \mathfrak{t} \oplus \mathfrak{k}$ is exactly of the type described for $\mathfrak{iso}$. Now pick a bi-invariant metric on K such that $\mathfrak{g}$ gets a Euclidean metric. Therefore, if we can also choose a Lie group $G \supset K$ whose Lie algebra is $\mathfrak{g}$, then we have constructed a Riemannian manifold G/K. To make it symmetric we need to be able to find an involution σ on G such that $D\sigma = L$. Note that if K is simply connected then we can take G to be the unique simply connected Lie group whose Lie algebra is $\mathfrak{g}$. Then σ can also easily be constructed.

Note that the isometry approach and curvature approach give different ways of representing a given symmetric space in an algebraic way. For Euclidean space we can, aside from the standard way using $\mathfrak{g} = \mathfrak{iso}$, also simply use $\mathfrak{g} = \mathbb{R}^n$ and let the involution be multiplication by -1 on all of $\mathfrak{g}$.

It is important to realize that a Lie algebra $\mathfrak{g}$, in itself, does not give rise to a symmetric space. The involution is really an integral part of the construction and does not necessarily exist on a given Lie algebra. The map $-id$ can, for instance, not be used, as it does not preserve the bracket. Rather, it is an *antiautomorphism*. This is particularly interesting if $\mathfrak{g}$ comes from a Lie group G with bi-invariant metric. Then the involution on G, which shows that G is symmetric, is the anti-automorphism: $g \to g^{-1}$, whose differential at e is $-id$. This just shows that the Lie algebra involution and the differential of the symmetry are not related in this simple fashion on Lie groups.

8.1.4 Curvature Description of Symmetric Spaces

Given the algebraic nature of symmetric spaces, there must of course be a purely algebraic way of computing the curvatures. This is the content of our next lemma. Note that the formula is similar to the one that was developed for bi-invariant metrics in the exercises to Chapter 2.

Lemma 1.3 *On a symmetric space we have that if $X, Y, Z \in \mathfrak{t}_p$, then*

$$R(X, Y)Z = [Z, [X, Y]]$$

at p.

Proof. By assumption, we suppose that the Killing fields are globally defined and satisfy $\nabla X = \nabla Y = \nabla Z = 0$ at p. The right-hand side does lie in T_pM rather than $\mathfrak{iso}_p$, so we are on the right track. The proof follows from the fact, proved below, that if K is a Killing field on a Riemannian manifold, then

$$\nabla^2_{X,Y} K = -R(K, X)Y.$$

Using this and $\nabla X = \nabla Y = \nabla Z = 0$ at p, we can calculate at p

$$\begin{aligned} R(X, Y)Z &= R(X, Z)Y - R(Y, Z)X \\ &= -\nabla_Z \nabla_Y X + \nabla_Z \nabla_X Y \\ &= \nabla_Z [X, Y] \\ &= [Z, [X, Y]], \end{aligned}$$

which is what we wanted to prove. □

Lemma 1.4 *If K is a Killing field on a Riemannian manifold (M, g), then*

$$\nabla^2_{X,Y} K = -R(K, X) Y.$$

Proof. The fact that K is a Killing field is used in the sense that $Y \to \nabla^2_{X,Y} K$ is skew adjoint. This fact was established in the proof of Proposition 1.4 in Chapter 7. For any vector field Z we can now compute

$$\begin{aligned} g\left(\nabla^2_{X,Y} K, Z\right) &= -g\left(\nabla^2_{X,Z} K, Y\right) \\ &= -g\left(\nabla^2_{Z,X} K, Y\right) - g\left(R(X, Z) K, Y\right) \\ &= g\left(\nabla^2_{Z,Y} K, X\right) - g\left(R(X, Z) K, Y\right) \\ &= g\left(\nabla^2_{Y,Z} K, X\right) + g\left(R(Z, Y) K, X\right) - g\left(R(X, Z) K, Y\right) \\ &= -g\left(\nabla^2_{Y,X} K, Z\right) + g\left(R(Z, Y) K, X\right) - g\left(R(X, Z) K, Y\right) \\ &= -g\left(\nabla^2_{X,Y} K, Z\right) - g\left(R(Y, X) K, Z\right) + g\left(R(Z, Y) K, X\right) \\ &\quad - g\left(R(X, Z) K, Y\right). \end{aligned}$$

Thus,

$$2g\left(\nabla^2_{X,Y} K, Z\right) = -g\left(R(Y, X) K, Z\right) + g\left(R(Z, Y) K, X\right) - g\left(R(X, Z) K, Y\right).$$

Bianchi's first identity, together with the other symmetry properties of the curvature tensor, now tells us that

$$\begin{aligned} g&\left(R(Z, Y) K, X\right) - g\left(R(X, Z) K, Y\right) \\ &= -g\left(R(K, X) Y, Z\right) + g\left(R(Y, K) X, Z\right) + g\left(R(X, Y) K, Z\right) \\ &= -2g\left(R(K, X) Y, Z\right). \end{aligned}$$

Hence

$$2g\left(\nabla^2_{X,Y} K, Z\right) = -2g\left(R(K, X) Y, Z\right),$$

which yields the desired property. □

Note that the curvatures therefore contain all the information about the Lie algebra structure that is needed for defining the brackets of vectors in $\mathfrak{t}_p$. This can be used to give a more efficient description of a symmetric space than the one using Iso. This description is called the *curvature description*. Suppose (M, g) is a symmetric space and fix $p \in M$. Suppose $\mathfrak{r}_p \subset \mathfrak{so}\left(T_p M\right)$ is the Lie algebra generated by the skew-symmetric endomorphisms $R(x, y) : T_p M \to T_p M$. Then we get a bracket operation on $\mathfrak{c}_p = T_p M \oplus \mathfrak{r}_p$ by defining

$$\begin{aligned} &[x, y] = R(x, y) \in \mathfrak{r}_p \quad \text{for} \quad x, y \in T_p M, \\ &[r, x] = -[x, r] = r(x) \in T_p M \quad \text{for} \quad x \in T_p M \quad \text{and} \quad r \in \mathfrak{r}_p, \\ &[r, s] = -(r \circ s - s \circ r) \in \mathfrak{r}_p \quad \text{for} \quad r, s \in \mathfrak{r}_p. \end{aligned}$$

Using Bianchi's first identity for the curvature tensor, one can show that the Jacobi identity holds. Thus, this bracket operation defines a Lie algebra. Also, the linear involution L, which is the identity on $\mathfrak{r}_p$ and multiplication by -1 on T_pM, is a Lie algebra automorphism. Since this construction works on any manifold, we still have to worry about why it reconstructs the symmetric space we started with. We shall show below that $R(x, y) \in \mathfrak{iso}_p$, using a holonomy argument. From this it follows that $(\mathfrak{c}_p, \mathfrak{r}_p) \subset (\mathfrak{iso}, \mathfrak{iso}_p)$, $\mathfrak{iso} \cap \mathfrak{c}_p = \mathfrak{r}_p$, and that L is merely the restriction of $D\sigma$ onto $\mathfrak{c}_p$. From this it is easy to see that this new description gives a possibly different way of representing the symmetric space.

Note also that given any Lie algebra description $(\mathfrak{g}, L)$ for a symmetric space, we can use this description to compute the curvature tensor. This is because any such description yields a homomorphism $h : \mathfrak{g} \to \mathfrak{iso}$ such that $h \circ L = D\sigma \circ h$. Thus, we can divide out by the kernel and get a smaller description that actually is contained in $(\mathfrak{iso}, D\sigma)$.

8.2 Examples of Symmetric Spaces

We shall here try to explain how some of the above constructions work in the concrete case of the Grassmannian manifold and its hyperbolic counterpart. We shall also look at complex Grassmannians, but there we restrict attention to the complex projective space. After these examples we give a formula for the curvature tensor on a compact Lie group with bi-invariant metric. Finally, we briefly discuss the symmetric space structure of $Sl(n)/SO(n)$. The moral of all of these examples and the above Lie algebra descriptions is that one can compute the curvature tensor algebraically without knowing the connection. Based on some general features of these examples, we shall see in the next section that the simplest symmetric spaces have either nonnegative or nonpositive curvature operator.

8.2.1 The Compact Grassmannian

First consider the Grassmannian of oriented k-planes in $\mathbb{R}^{k+l}$, denoted by $M = \tilde{G}_k(\mathbb{R}^{k+l})$. Thus, each element in M is a k-dimensional subspace of $\mathbb{R}^{k+l}$ together with an orientation; in particular, $\tilde{G}_1(\mathbb{R}^{n+1}) = S^n$. We shall assume that we have the orthogonal splitting $\mathbb{R}^{k+l} = \mathbb{R}^k \oplus \mathbb{R}^l$, where the distinguished element $p = \mathbb{R}^k$ takes up the first k coordinates in $\mathbb{R}^{k+l}$ and is endowed with its natural positive orientation.

Let us first identify M as a homogeneous space. Observe that $O(k+l)$ acts on $\mathbb{R}^{k+l}$. As such, it maps k-dimensional subspaces to k-dimensional subspaces, and does something uncertain to the orientations of these subspaces. We therefore get that $O(k+l)$ acts transitively on M. This is, however, not the isometry group, which is really what we wish to find, as the matrix $-I \in SO(k+l)$ acts trivially if k and l are even.

The isotropy group consists of those elements that keep $\mathbb{R}^k$ fixed as well as preserving the orientation. Clearly, the correct isotropy group is then $SO(k) \times O(l) \subset O(k+l)$.

The tangent space at $p = \mathbb{R}^k$ is naturally identified with the space of $k \times l$ matrices $Mat_{k\times l}$, or equivalently, with $\mathbb{R}^k \otimes \mathbb{R}^l$. To see this, just observe that any k-dimensional subspace of $\mathbb{R}^{k+l}$ can be represented as a linear graph over $\mathbb{R}^k$ with values in the orthogonal complement $\mathbb{R}^l$. The isotropy action of $SO(k) \times O(l)$ on $Mat_{k\times l}$ now acts as follows:

$$\begin{aligned} SO(k) \times O(l) \times Mat_{k\times l} &\to Mat_{k\times l}, \\ (A, B, X) &\to AXB^{-1} = AXB^t. \end{aligned}$$

If we define X to be the matrix that is 1 in the $(1, 1)$ entry and otherwise zero, then $AXB^t = A_1 (B_1)^t$, where A_1 is the first column of A and B_1 is the first column of B. Thus, the orbit of X, under the isotropy action, generates a basis for $Mat_{k\times l}$ but does not cover all of the space. Thus, we have an example of an irreducible action on Euclidean space that is not transitive on the unit sphere. This representation, when seen as acting on $\mathbb{R}^k \otimes \mathbb{R}^l$, is denoted by $SO(k) \otimes O(l)$.

To see that M is a symmetric space, we have to show that the isotropy group contains the required involution. On the tangent space $T_pM = Mat_{k\times l}$ it is supposed to act by multiplication by -1. Thus, we have to find $(A, B) \in SO(k) \times O(l)$ such that for all X,

$$AXB^t = -X.$$

Clearly, we can just set

$$\begin{aligned} A &= I_k, \\ B &= -I_l. \end{aligned}$$

Depending on k and l, other choices are possible, but they will act in the same way.

We have now exhibited M as a symmetric space, although we didn't use the isometry group of the space. Instead, we used a finite covering of the isometry group and then had some extra elements that acted trivially.

Let us now give the Lie algebra description and compute the curvature tensor. Since we actually found the isometry group modulo a finite covering, we see that

$$\begin{aligned} \mathfrak{iso} &= \mathfrak{so}(k+l), \\ \mathfrak{iso}_p &= \mathfrak{so}(k) \times \mathfrak{so}(l). \end{aligned}$$

We shall use the block decomposition of matrices in $\mathfrak{so}(k+l)$:

$$X = \begin{pmatrix} X_1 & B \\ -B^t & X_2 \end{pmatrix},$$
$$X_1 \in \mathfrak{so}(k), X_2 \in \mathfrak{so}(l), B \in Mat_{k\times l}.$$

If we set

$$\mathfrak{t}_p = \left\{ \begin{pmatrix} 0 & B \\ -B^t & 0 \end{pmatrix} : B \in Mat_{k\times l} \right\},$$

then we have an orthogonal decomposition:

$$\mathfrak{so}(k+l) = \mathfrak{t}_p \oplus \mathfrak{so}(k) \oplus \mathfrak{so}(l),$$

where we can identify $\mathfrak{t}_p = T_pM$. The inner product on $\mathfrak{t}_p$ is the standard Euclidean metric defined by

$$\begin{aligned}
\left\langle \begin{pmatrix} 0 & A \\ -A^t & 0 \end{pmatrix}, \begin{pmatrix} 0 & B \\ -B^t & 0 \end{pmatrix} \right\rangle &= \operatorname{tr}\left(\begin{pmatrix} 0 & A \\ -A^t & 0 \end{pmatrix} \begin{pmatrix} 0 & B \\ -B^t & 0 \end{pmatrix}^t \right) \\
&= \operatorname{tr}\left(\begin{pmatrix} 0 & A \\ -A^t & 0 \end{pmatrix} \begin{pmatrix} 0 & -B \\ B^t & 0 \end{pmatrix} \right) \\
&= \operatorname{tr} \begin{pmatrix} AB^t & 0 \\ 0 & A^tB \end{pmatrix} \\
&= \operatorname{tr}\left(AB^t\right) + \operatorname{tr}\left(A^tB\right) \\
&= 2\operatorname{tr}\left(AB^t\right).
\end{aligned}$$

Thus, it is twice the usual Euclidean metric on $\mathbb{R}^{k\cdot l}$ that we used above. But that, of course, does not change matters much.

We now have to compute Lie brackets of elements in $\mathfrak{t}_p$ and then see how $\mathfrak{so}(k) \oplus \mathfrak{so}(l)$ acts on $\mathfrak{t}_p$, in order to find the curvature tensor. For $A, B \in \mathfrak{t}_p$ we have

$$\begin{aligned}
[A,B] &= \begin{pmatrix} 0 & A \\ -A^t & 0 \end{pmatrix} \begin{pmatrix} 0 & B \\ -B^t & 0 \end{pmatrix} - \begin{pmatrix} 0 & B \\ -B^t & 0 \end{pmatrix} \begin{pmatrix} 0 & A \\ -A^t & 0 \end{pmatrix} \\
&= \begin{pmatrix} -AB^t & 0 \\ 0 & -A^tB \end{pmatrix} - \begin{pmatrix} -BA^t & 0 \\ 0 & -B^tA \end{pmatrix} \\
&= \begin{pmatrix} BA^t - AB^t & 0 \\ 0 & B^tA - A^tB \end{pmatrix} \in \mathfrak{so}(k) \oplus \mathfrak{so}(l).
\end{aligned}$$

Observe that there is a basis for $\mathfrak{so}(k) \oplus \mathfrak{so}(l)$ that can be written in this way, so there will be no difference between the curvature and isometry descriptions. Now take $C \in \mathfrak{t}_p$ and compute

$$\begin{aligned}
R(A,B)C &= [C,[A,B]] \\
&= \left[\begin{pmatrix} 0 & C \\ -C^t & 0 \end{pmatrix}, \begin{pmatrix} BA^t - AB^t & 0 \\ 0 & B^tA - A^tB \end{pmatrix} \right] \\
&= \begin{pmatrix} 0 & \begin{matrix} C\left(B^tA - A^tB\right) \\ -\left(BA^t - AB^t\right)C \end{matrix} \\ \begin{matrix} -\left(A^tB - B^tA\right)C^t \\ +C^t\left(AB^t - BA^t\right) \end{matrix} & 0 \end{pmatrix}.
\end{aligned}$$

This does not seem very illuminating, so let us find the sectional curvatures by considering the directional curvature transformation

$$R(A,B)B = \begin{pmatrix} 0 & BB^tA - 2BA^tB + AB^tB \\ -A^tBB^t + 2B^tAB^t - B^tBA^t & 0 \end{pmatrix}.$$

We now have to take the inner product with A, which gives us

$$\begin{aligned}
\langle R(A,B)B, A\rangle &= \operatorname{tr}\left(BB^tAA^t - 2BA^tBA^t + AB^tBA^t\right) \\
&\quad + \operatorname{tr}\left(A^tBB^tA - 2B^tAB^tA + B^tBA^tA\right) \\
&= \operatorname{tr}\left(BB^tAA^t\right) - 2\operatorname{tr}\left(BA^tBA^t\right) + \operatorname{tr}\left(AB^tBA^t\right) \\
&\quad + \operatorname{tr}\left(A^tBB^tA\right) - 2\operatorname{tr}\left(B^tAB^tA\right) + \operatorname{tr}\left(B^tBA^tA\right) \\
&= \operatorname{tr}\left(BA^tAB^t\right) - 2\operatorname{tr}\left(BA^tBA^t\right) + \operatorname{tr}\left(AB^tBA^t\right) \\
&\quad + \operatorname{tr}\left(A^tBB^tA\right) + \operatorname{tr}\left(B^tAA^tB\right) - 2\operatorname{tr}\left(B^tAB^tA\right) \\
&= \langle BA^t, BA^t\rangle - 2\langle BA^t, AB^t\rangle + \langle AB^t, AB^t\rangle \\
&\quad + \langle A^tB, A^tB\rangle - 2\langle A^tB, B^tA\rangle + \langle B^tA, B^tA\rangle \\
&= \left|BA^t - AB^t\right|^2 + \left|A^tB - B^tA\right|^2 \geq 0.
\end{aligned}$$

Here we recklessly used Euclidean norms for matrices in various different spaces. The conclusion is that the sectional curvatures are all ≥ 0.

When $k = 1$ or $l = 1$, it is easy to see that one gets a metric of constant positive curvature. Otherwise, the metric will have some zero sectional curvatures.

8.2.2 *The Hyperbolic Grassmannian*

Let us now turn to the hyperbolic analogue. In the Euclidean space $\mathbb{R}^{k+l}$ we use, instead of the positive definite inner product $v^t \cdot w$, the quadratic form:

$$\begin{aligned}
v^t I_{k,l} w &= v^t \begin{pmatrix} I_k & 0 \\ 0 & -I_l \end{pmatrix} w \\
&= \sum_{i=1}^{k} v_i w_i - \sum_{i=k+1}^{k+l} v_i w_i.
\end{aligned}$$

With this form we usually write $\mathbb{R}^{k,l}$ to indicate both the dimension $k+l$ and the type of quadratic form used on the Euclidean space. The group linear transformations that preserve this form is denoted by $O(k,l)$. These transformations are defined by the relation

$$X \cdot I_{k,l} \cdot X^t = I_{k,l}.$$

Note that if $k, l > 0$, then $O(k,l)$ is not compact. But it clearly contains the (maximal) compact subgroup $O(k) \times O(l)$.

The Lie algebra $\mathfrak{so}(k,l)$ of $O(k,l)$ consists of the matrices satisfying

$$Y \cdot I_{k,l} + I_{k,l} \cdot Y^t = 0.$$

If we use the same block decomposition for Y as we did for $I_{k,l}$ above, then we have that it looks like

$$\begin{aligned} Y &= \begin{pmatrix} A & B \\ B^t & C \end{pmatrix}, \\ A &\in \mathfrak{so}(k), \\ C &\in \mathfrak{so}(l), \\ B &\in Mat_{k\times l}. \end{aligned}$$

We now consider only those (oriented) k-dimensional subspaces of $\mathbb{R}^{k,l}$ on which this quadratic form generates a positive definite inner product. This space is the hyperbolic Grassmannian $M = \tilde{G}_k\left(\mathbb{R}^{k,l}\right)$. Our selected point is as before $p = \mathbb{R}^k$. One can easily see that topologically: $\tilde{G}_k\left(\mathbb{R}^{k,l}\right)$ is an open subset of $\tilde{G}_k\left(\mathbb{R}^{k+l}\right)$. The metric on this space is another story, however. Clearly, $O(k,l)$ acts transitively on M, and those elements that fix p are of the form $SO(k)\times O(l)$. One can, as before, find the desired involution, and thus exhibit M as a symmetric space. Again some of these elements act trivially, but at the Lie algebra level this makes no difference. Thus, we have

$$\begin{aligned} \mathfrak{iso} &= \mathfrak{so}(k,l), \\ \mathfrak{iso}_p &= \mathfrak{so}(k)\times\mathfrak{so}(l), \\ \mathfrak{t}_p &= \left\{\begin{pmatrix} 0 & A \\ A^t & 0 \end{pmatrix} : B \in Mat_{k\times l}\right\}. \end{aligned}$$

On $\mathfrak{t}_p$ we use the Euclidean metric

$$\begin{aligned} \left\langle \begin{pmatrix} 0 & A \\ A^t & 0 \end{pmatrix}, \begin{pmatrix} 0 & B \\ B^t & 0 \end{pmatrix} \right\rangle &= \operatorname{tr}\left(\begin{pmatrix} 0 & A \\ A^t & 0 \end{pmatrix}\cdot\begin{pmatrix} 0 & B \\ B^t & 0 \end{pmatrix}^t\right) \\ &= \operatorname{tr}\left(\begin{pmatrix} 0 & A \\ A^t & 0 \end{pmatrix}\cdot\begin{pmatrix} 0 & B \\ B^t & 0 \end{pmatrix}\right) \\ &= \operatorname{tr}\begin{pmatrix} AB^t & 0 \\ 0 & A^tB \end{pmatrix} \\ &= \operatorname{tr}\left(AB^t\right) + \operatorname{tr}\left(A^tB\right) \\ &= 2\operatorname{tr}\left(AB^t\right). \end{aligned}$$

So while $\mathfrak{t}_p$ looks different, we seem to use the same metric.

On $\mathfrak{t}_p$ we have the Lie bracket

$$\begin{aligned} \left[\begin{pmatrix} 0 & A \\ A^t & 0 \end{pmatrix}, \begin{pmatrix} 0 & B \\ B^t & 0 \end{pmatrix}\right] &= \begin{pmatrix} 0 & A \\ A^t & 0 \end{pmatrix}\begin{pmatrix} 0 & B \\ B^t & 0 \end{pmatrix} \\ &\quad - \begin{pmatrix} 0 & B \\ B^t & 0 \end{pmatrix}\begin{pmatrix} 0 & A \\ A^t & 0 \end{pmatrix} \\ &= \begin{pmatrix} AB^t & 0 \\ 0 & A^tB \end{pmatrix} - \begin{pmatrix} BA^t & 0 \\ 0 & B^tA \end{pmatrix} \end{aligned}$$

$$= \begin{pmatrix} AB^t - BA^t & 0 \\ 0 & A^tB - B^tA \end{pmatrix}$$
$$\in \mathfrak{so}(k) \oplus \mathfrak{so}(l).$$

This is the negative of what we had before. We can now compute the curvature tensor:

$$\begin{aligned} R(A,B)C &= [C,[A,B]] \\ &= \left[\begin{pmatrix} 0 & C \\ C^t & 0 \end{pmatrix}, \begin{pmatrix} AB^t - BA^t & 0 \\ 0 & A^tB - B^tA \end{pmatrix}\right] \\ &= \begin{pmatrix} 0 & \begin{matrix} C(A^tB - B^tA) \\ -(AB^t - BA^t)C \end{matrix} \\ \begin{matrix} C^t(AB^t - BA^t) \\ -(A^tB - B^tA)C^t \end{matrix} & 0 \end{pmatrix}. \end{aligned}$$

If we let $C = B$ and compute the sectional curvature as before, we arrive at

$$\begin{aligned} &\langle R(A,B)B, A\rangle \\ &= \operatorname{tr}\left(\begin{pmatrix} 0 & \begin{matrix} 2BA^tB \\ -BB^tA - AB^tB \end{matrix} \\ \begin{matrix} 2B^tAB^t \\ -B^tBA^t - A^tBB^t \end{matrix} & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & A \\ A^t & 0 \end{pmatrix}\right) \\ &= \operatorname{tr}\begin{pmatrix} \begin{matrix} 2BA^tBA^t \\ -BB^tAA^t - AB^tBA^t \end{matrix} & 0 \\ 0 & \begin{matrix} 2B^tAB^tA \\ -B^tBA^tA - A^tBB^tA \end{matrix} \end{pmatrix}. \end{aligned}$$

This is exactly the negative of the expression we got in the compact case. Hence, the hyperbolic Grassmannians have nonpositive curvature. When $k = 1$, we have reconstructed the hyperbolic space together with its isometry group.

8.2.3 *Complex Projective Space Revisited*

We shall view complex projective space as a complex Grassmannian. Namely, let $M = \mathbb{C}P^n = G_1(\mathbb{C}^{n+1})$, i.e., the complex lines in $\mathbb{C}^{n+1}$. More generally we can consider $G_k(\mathbb{C}^{k+l})$ and of course the hyperbolic counterparts $G_k(\mathbb{C}^{k,l})$, but we leave this to the reader.

The group $U(n+1) \subset SO(2n+2)$ consists of those orthogonal transformations that also preserve the complex structure. If we use complex coordinates, then the Hermitian metric on $\mathbb{C}^{n+1}$ can be written as

$$z^*w = \sum \bar{z}_i w_i,$$

where as usual, $A^* = \bar{A}^t$ is the conjugate transpose. Thus, the elements of $U(n+1)$ satisfy

$$A^{-1} = A^*.$$

As with the Grassmannian, $U(n+1)$ acts on M, but this time, all of the transformations of the form aI, where $a\bar{a} = 1$, act trivially. Thus, we restrict attention to $SU(n+1)$, which still acts transitively, but now also almost effectively.

If we let $p = \mathbb{C}$ be the first coordinate axis, then the isotropy group is $S(U(1) \times U(n))$, i.e., those matrices in $U(1) \times U(n)$ of determinant 1. This group is, of course, naturally isomorphic to $U(n)$ via the map

$$A \to \begin{pmatrix} \det A^{-1} & 0 \\ 0 & A \end{pmatrix}.$$

The involution that makes M symmetric is then given by

$$\begin{pmatrix} (-1)^n & 0 \\ 0 & -I_n \end{pmatrix}.$$

Let us now pass to the Lie algebra level in order to compute the curvature tensor. From the above, we have

$$\begin{aligned} \mathfrak{iso} &= \mathfrak{su}(n+1) = \{A : A = -A^*, \operatorname{tr} A = 0\}, \\ \mathfrak{iso}_p &= \mathfrak{u}(n) = \{B : B = -B^*\}. \end{aligned}$$

The inclusion looks like

$$B \to \begin{pmatrix} -\operatorname{tr} B & 0 \\ 0 & B \end{pmatrix}.$$

Thus we should write elements of $\mathfrak{su}(n+1)$ in the form

$$\begin{pmatrix} -\operatorname{tr} B & -z^* \\ z & B \end{pmatrix},$$

and then identify $\mathfrak{k}_p = \left\{ \begin{pmatrix} 0 & -z^* \\ z & 0 \end{pmatrix} : z \in \mathbb{C}^n \right\}$ and use the inner product

$$\begin{aligned} \left\langle \begin{pmatrix} 0 & -z^* \\ z & 0 \end{pmatrix}, \begin{pmatrix} 0 & -w^* \\ w & 0 \end{pmatrix} \right\rangle &= \frac{1}{2} \operatorname{tr} \left(\begin{pmatrix} 0 & -z^* \\ z & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & -w^* \\ w & 0 \end{pmatrix}^* \right) \\ &= \frac{1}{2} \operatorname{tr} \begin{pmatrix} 0 & -z^* \\ z & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & w^* \\ -w & 0 \end{pmatrix} \\ &= \frac{1}{2} \operatorname{tr} \begin{pmatrix} z^* w & 0 \\ 0 & z w^* \end{pmatrix} \\ &= \frac{1}{2} \left(z^* w + \operatorname{tr} z w^* \right) \\ &= \frac{1}{2} \left(z^* w + w^* z \right) \\ &= \operatorname{Re} \langle z, w \rangle . \end{aligned}$$

Here $\langle z, w \rangle$ is the usual Hermitian inner product on $\mathbb{C}^n$, which is conjugate linear in the w variable.

For the curvature tensor we first compute the Lie bracket on $\mathfrak{t}_p$:

$$\begin{aligned}
\left[\begin{pmatrix} 0 & -z^* \\ z & 0 \end{pmatrix}, \begin{pmatrix} 0 & -w^* \\ w & 0 \end{pmatrix}\right] &= \begin{pmatrix} 0 & -z^* \\ z & 0 \end{pmatrix}\begin{pmatrix} 0 & -w^* \\ w & 0 \end{pmatrix} \\
&\quad - \begin{pmatrix} 0 & -w^* \\ w & 0 \end{pmatrix}\begin{pmatrix} 0 & -z^* \\ z & 0 \end{pmatrix} \\
&= \begin{pmatrix} -z^*w & 0 \\ 0 & -zw^* \end{pmatrix} - \begin{pmatrix} -w^*z & 0 \\ 0 & -wz^* \end{pmatrix} \\
&= \begin{pmatrix} w^*z - z^*w & 0 \\ 0 & wz^* - zw^* \end{pmatrix}.
\end{aligned}$$

Then, we get

$$\begin{aligned}
R(z, w)\, w &= \left[\begin{pmatrix} 0 & -w^* \\ w & 0 \end{pmatrix}, \begin{pmatrix} w^*z - z^*w & 0 \\ 0 & wz^* - zw^* \end{pmatrix}\right] \\
&= \begin{pmatrix} 0 & -w^* \\ w & 0 \end{pmatrix}\begin{pmatrix} w^*z - z^*w & 0 \\ 0 & wz^* - zw^* \end{pmatrix} \\
&\quad - \begin{pmatrix} w^*z - z^*w & 0 \\ 0 & wz^* - zw^* \end{pmatrix}\begin{pmatrix} 0 & -w^* \\ w & 0 \end{pmatrix} \\
&= \begin{pmatrix} 0 & \begin{matrix} w^*(zw^* - wz^*) \\ +(w^*z - z^*w)\, w^* \end{matrix} \\ \begin{matrix} w\,(w^*z - z^*w) \\ +(zw^* - wz^*)\, w \end{matrix} & 0 \end{pmatrix}.
\end{aligned}$$

Now identify $\mathfrak{t}_p$ with $\mathbb{C}^n$ and observe that

$$R(z, w)\, w = w\left(w^*z - z^*w\right) + \left(zw^* - wz^*\right) w.$$

To compute the sectional curvatures we need to pick an orthonormal basis z, w for a plane. This means that $|z|^2 = |w|^2 = 1$ and $\operatorname{Re}\langle z, w\rangle = 0$. The sectional curvature of the plane spanned by z, w is therefore

$$\begin{aligned}
\sec(z, w) &= \operatorname{Re}\left\langle w\left(w^*z - z^*w\right) + \left(zw^* - wz^*\right) w, z\right\rangle \\
&= \operatorname{Re} z^*w\left(w^*z - z^*w\right) + \operatorname{Re} z^*\left(zw^* - wz^*\right) w \\
&= |\langle w, z\rangle|^2 - 2\operatorname{Re}\left(\langle w, z\rangle^2\right) + 1 \\
&= 1 + 3\,|\operatorname{Im}\langle w, z\rangle|^2.
\end{aligned}$$

Thus, if z, w are orthogonal with respect to the Hermitian metric, i.e., $\langle z, w\rangle = 0$, then $\sec(z, w) = 1$, while if, e.g., $w = iz$, then we get that the sectional curvature of a complex line is $\sec(z, iz) = 4$. Since $|\operatorname{Im}\langle w, z\rangle| \leq |z|\,|w| = 1$, all other curvatures lie between these two values. Thus we have shown that the complex projective space is quarter pinched.

8.2.4 Lie Groups with Bi-Invariant Metrics

In a more abstract vein, let us see how Lie groups with bi-invariant metrics behave when considered as symmetric spaces. To this end, suppose we have a compact Lie group G with a bi-invariant metric. As usual, the Lie algebra $\mathfrak{g}$ of G is identified with T_pM and is also the set of left-invariant vector fields on G. The object is then to find an appropriate Lie algebra description.

The claim is that a Lie algebra description is $(\mathfrak{g} \oplus \mathfrak{g}, L)$, where $L(X, Y) = (Y, X)$. Clearly, the diagonal $\mathfrak{g}^{\Delta} = \{(X, X) : X \in \mathfrak{g}\}$ is the 1-eigenspace, while the complement $\mathfrak{g}^{\perp} = \{(X, -X) : X \in \mathfrak{g}\}$ is the -1-eigenspace. Thus, we should identify

$$\begin{aligned} \mathfrak{k} &= \mathfrak{g}^{\Delta} \cong \mathfrak{g}, \\ \mathfrak{t} &= \mathfrak{g}^{\perp}. \end{aligned}$$

We already know that $\mathfrak{g}$ corresponds to the compact Lie group G, so we are simply saying that

$$G = (G \times G)/G^{\Delta}.$$

On $\mathfrak{t}$, the Lie bracket looks like

$$\begin{aligned} [(X, -X), (Y, -Y)] &= ([X, Y], [-X, -Y]) \\ &= ([X, Y], [X, Y]) \in \mathfrak{k}. \end{aligned}$$

Thus, the curvature tensor can be computed as follows:

$$\begin{aligned} R(X, Y)Z &= R((X, -X), (Y, -Y))(Z, -Z) \\ &= [(Z, -Z), ([X, Y], [X, Y])] \\ &= ([Z, [X, Y]], -[Z, [X, Y]]) \in \mathfrak{t}. \end{aligned}$$

Hence, we arrive at that the formula

$$R(X, Y)Z = [Z, [X, Y]]$$

for the curvature tensor on a compact Lie group with bi-invariant metric. This formula looks exactly like the one for the curvature of a symmetric space, but it is interpreted differently. Another curious feature is that if one computes the curvature tensor in the standard way using a bi-invariant metric, then the formula has a factor $\frac{1}{4}$ on it (see the exercises to Chapter 2). The reason for this discrepancy is that left- or right-invariant vector fields do not lie in $\mathfrak{t}$ unless they are parallel. And conversely, a Killing field from $\mathfrak{t}$ is left- or right-invariant only when it is parallel.

8.2.5 $Sl(n)/SO(n)$

The manifold is the quotient space of the $n \times n$ matrices with determinant 1 by the orthogonal matrices. The Lie algebra of $Sl(n)$ is $\mathfrak{sl}(n) = \{X \in Mat_{n\times n} : \mathrm{tr}X = 0\}$.

This Lie algebra is naturally divided up into symmetric and skew-symmetric matrices

$$\mathfrak{sl}(n) = \mathfrak{k} \oplus \mathfrak{so}(n),$$

where $\mathfrak{k}$ consists of the symmetric matrices. On $\mathfrak{k}$ we can use the usual Euclidean metric. The involution is obviously given by $-id$ on $\mathfrak{k}$ and id on $\mathfrak{so}(n)$. Holistically, this is the map

$$L(X) = -X^t.$$

Let us now grind out the curvature tensor:

$$\begin{aligned} R(X,Y)Z &= [Z,[X,Y]] \\ &= Z[X,Y] - [X,Y]Z \\ &= ZXY - ZYX - XYZ + YXZ. \end{aligned}$$

This yields

$$\begin{aligned} g(R(X,Y)Z,W) &= \operatorname{tr}\left([Z,[X,Y]]\,W^t\right) \\ &= \operatorname{tr}([Z,[X,Y]]\,W) \\ &= \operatorname{tr}(Z[X,Y]W - [X,Y]ZW) \\ &= \operatorname{tr}(WZ[X,Y] - [X,Y]ZW) \\ &= \operatorname{tr}([X,Y][W,Z]) \\ &= -\operatorname{tr}\left([X,Y][W,Z]^t\right) \\ &= -g([X,Y],[W,Z]). \end{aligned}$$

In particular, the sectional curvatures must be nonpositive.

8.3 Holonomy

First we discuss holonomy for general manifolds and the de Rham decomposition theorem. We then use holonomy to give a brief discussion of how symmetric spaces can be classified according to whether they are compact or not.

8.3.1 The Holonomy Group

Let (M,g) be a Riemannian n-manifold. If $c:[a,b]\to M$ is a smooth curve, then

$$P_{c(a)}^{c(b)} : T_{c(a)}M \to T_{c(b)}M$$

denotes the effect of parallel translating a vector in $T_{c(a)}M$ along c to $T_{c(b)}M$. This property will in general depend not only on the endpoints of the curve, but also on the actual curve. We can generalize this to work for piecewise smooth curves by breaking up the process at the breakpoints in the curve.

Suppose now the curve is a loop, i.e., $c(a) = c(b) = p$. Then parallel translation gives an isometry on T_pM. The set of all such isometries is called the *holonomy group* at p and is denoted by $\mathrm{Hol}_p = \mathrm{Hol}_p(M, g)$. One can easily see that this forms a subgroup of $O(T_pM) = O(n)$. Moreover, it is actually a Lie group, which is usually a closed subgroup of $O(n)$. We also have the *restricted holonomy group* $\mathrm{Hol}^0_p = \mathrm{Hol}^0_p(M, g)$, which is the connected normal subgroup that comes from using only contractible loops. It can be shown that the restricted holonomy group is always compact. Here are some elementary properties that are easy to establish:

(a) $\mathrm{Hol}_p(\mathbb{R}^n) = \{1\}$.

(b) $\mathrm{Hol}_p(S^n(r)) = SO(n)$.

(c) $\mathrm{Hol}_p(H^n) = SO(n)$.

(d) $\mathrm{Hol}_p(M, g) \subset SO(n)$ iff M is orientable.

(e) $\mathrm{Hol}_p\left(\tilde{M}, \tilde{g}\right) = \mathrm{Hol}^0_p\left(\tilde{M}, \tilde{g}\right) = \mathrm{Hol}^0_p(M, g)$, where $\tilde{M}$ is the universal covering of M.

(f) $\mathrm{Hol}_{(p,q)}(M_1 \times M_2, g_1 + g_2) = \mathrm{Hol}_p(M_1, g_1) \times \mathrm{Hol}_q(M_2, g_2)$.

(g) $\mathrm{Hol}_p(M, g)$ is conjugate to $\mathrm{Hol}_q(M, g)$ via parallel translation along any curve from p to q.

(h) A tensor on (M, g) is parallel iff it is invariant under the (restricted) holonomy group; e.g., if ω is a 2-form, then $\nabla\omega = 0$ iff $\omega(Pv, Pw) = \omega(v, w)$ for all $P \in \mathrm{Hol}^0_p(M, g)$ and $v, w \in T_pM$.

We are now ready to study how the Riemannian manifold decomposes according to the holonomy. Guided by (e) we see that Cartesian products are reflected in a product structure at the level of the holonomy. Furthermore, (g) shows that if the holonomy decomposes at just one point, then it decomposes everywhere.

To make things more precise, let us consider the action of Hol^0_p on T_pM. If $E \subset T_pM$ is an invariant subspace, i.e., $\mathrm{Hol}^0_p(E) \subset E$, then the orthogonal complement is also preserved, i.e., $\mathrm{Hol}^0_p\left(E^\perp\right) \subset E^\perp$. Thus, T_pM decomposes into irreducible invariant subspaces:

$$T_pM = E_1 \oplus \cdots \oplus E_k.$$

Here, irreducible means that there are no invariant subspaces. Since parallel translation around loops at p preserves this decomposition, we see that parallel translation along any curve from p to q preserves this decomposition. Thus, we get a global decomposition of the tangent bundle into distributions each of which is invariant under parallel translation:

$$TM = \eta_1 \oplus \cdots \oplus \eta_k.$$

With this we can prove de Rham's decomposition theorem.

Theorem 3.1 (de Rham, 1952) *If we decompose the tangent bundle of a Riemannian manifold (M, g) into irreducible components according to the holonomy*

$$TM = \eta_1 \oplus \cdots \oplus \eta_k,$$

then around each point $p \in M$ there is a neighborhood U that has a product structure of the form

$$\begin{aligned}(U, g) &= (U_1 \times \cdots \times U_k, g_1 + \cdots + g_k), \\ TU_i &= \eta_{i|U_i}.\end{aligned}$$

Moreover, if (M, g) is simply connected and complete, then there is a global splitting

$$\begin{aligned}(M, g) &= (M_1 \times \cdots \times M_k, g_1 + \cdots + g_k), \\ TM_i &= \eta_i.\end{aligned}$$

Proof. Given the decomposition into parallel distributions, we first observe that each of the distributions must be integrable. Thus, we do get a local splitting into submanifolds at the manifold level. To see that the metric splits as well, just observe that the submanifolds are totally geodesic, as their tangent spaces are invariant under parallel translation. This gives the local splitting. The global result is not just a trivial analytic continuation argument. Apparently, one must understand what simple connectivity has to do with the maximal integral submanifolds being embedded submanifolds. Instead of doing that, let M_i be the maximal integral submanifolds, and define abstractly the Riemannian manifold $(M_1 \times \cdots \times M_k, g_1 + \cdots + g_k)$. Locally, (M, g) and $(M_1 \times \cdots \times M_k, g_1 + \cdots + g_k)$ are isometric to each other. Given that (M, g) is complete, it is not hard to see that also $(M_1 \times \cdots \times M_k, g_1 + \cdots + g_k)$ is complete. Therefore, if M is also simply connected, we can find an isometric embedding $(M, g) \to (M_1 \times \cdots \times M_k, g_1 + \cdots + g_k)$. Completeness insures us that the map is onto. To see that it is also one to one requires a little more work. □

It is therefore reasonable when studying classification problems for Riemannian manifolds to study only those Riemannian manifolds that are *irreducible*, i.e., those where the holonomy has no invariant subspaces. Guided by this we have some nice characterizations of Einstein manifolds.

Theorem 3.2 *If (M, g) is an irreducible Riemannian manifold with parallel Ricci tensor, then (M, g) is Einstein. In particular, irreducible symmetric spaces are Einstein.*

Proof. The fact that $\nabla \mathrm{Ric} = 0$ means that the Ricci tensor is invariant under parallel translation. Now decompose

$$T_pM = E_1 \oplus \cdots \oplus E_k$$

into the eigenspaces for $\mathrm{Ric} : T_pM \to T_pM$ with respect to distinct eigenvalues $\lambda_1 < \cdots < \lambda_k$. As above, we can now parallel translate these eigenspaces to get a global decomposition

$$TM = \eta_1 \oplus \cdots \oplus \eta_k$$

into parallel distributions, with the property that

$$\mathrm{Ric}_{|\eta_i} = \lambda_i \cdot I.$$

But then the decomposition theorem tells us that (M, g) is reducible unless there is only one eigenvalue, which means that the metric is Einstein. □

8.3.2 *A Different Curvature Characterization of Symmetric Spaces*

There is a very interesting result by Tricerri and Vanhecke in [81] that shows that only symmetric spaces have curvature tensors that look like the curvature tensor for a symmetric space. Recall that a locally symmetric space is completely determined by its curvature tensor at a point. If we identify the tangent space with $\mathbb{R}^n$, then a curvature tensor is simply a special type of 4-linear map on $\mathbb{R}^n$. Suppose we have specified such a potential curvature tensor $\bar{R} : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$. Throughout this short subsection, suppose we have a Riemannian n-manifold (M, g) with algebraically constant curvature tensor $\bar{R}$, in the sense that for each $p \in M$, there is an isometry $L_p : T_pM \to \mathbb{R}^n$ and a number $\lambda(p)$ such that

$$g(R(x, y)z, w) = \lambda(p) \cdot \bar{R}\left(L_px, L_py, L_pz, L_pw\right).$$

Such a condition is purely algebraic and certainly doesn't necessarily imply that the space is locally symmetric. Still, one does get a certain amount of information. First, we note that since L_p is an isometry, it doesn't matter whether we are using the Euclidean metric or the Riemannian metric. Next, observe that we have

$$|R|^2 = |\lambda|^2 \left|\bar{R}\right|^2,$$

which implies that λ is a smooth function on M, as $\left|\bar{R}\right|^2$ is constant.

Suppose now $\bar{R}$ has the form

$$\bar{R}(x, y, z, w) = k \cdot (\langle y, z\rangle \langle x, w\rangle - \langle x, z\rangle \langle y, w\rangle)$$

for some number k and where $\langle\cdot,\cdot\rangle$ is the usual Euclidean inner product. Then it must follow that (M, g) has constant sectional curvature $\lambda(p) \cdot k$ at each $p \in M$. It will then follow from Schur's lemma (Chapter 2) that the space has constant curvature provided that $n > 2$. Similarly, if $\bar{R}$ is Einstein, in the sense that for some orthonormal basis e_i for $\mathbb{R}^n$ we have

$$\sum_{i=1}^{n} \bar{R}(x, e_i, e_i, w) = k \cdot \langle x, w\rangle,$$

then it will also be true that (M, g) is Einstein, again provided that $n > 2$.

From the holonomy section we know that any irreducible symmetric space is Einstein. So what happens if $\bar{R}$ is the curvature tensor for such a space? Certainly it is Einstein, but in fact, it is locally symmetric. To see this we must use Lichnerowicz's Bochner formula for the curvature tensor from Chapter 7 in a different way. From the proof of the Lichnerowicz identity we get the relationship

$$\Delta \frac{1}{2} |R|^2 = |\nabla R|^2 + 2g\left(R^\flat, \nabla \mathrm{div} R\right) + 2K(R),$$

where K is some quantity that depends on R in a purely algebraic way. We already know that the metric is Einstein, so it must follow that λ (and consequently $|R|^2$) is constant and $\mathrm{div} R = 0$. The formula therefore reduces to

$$0 = |\nabla R|^2 + 2K(R).$$

For a symmetric space, we know that $\nabla R = 0$, so also $K = 0$ for such a space. As K is purely algebraic in the curvature tensor, we can therefore compute $K\left(\bar{R}\right)$ and conclude that it is zero. Since $R = \lambda \bar{R}$, it will then also follow that $K(R) = 0$, which finally implies that $|\nabla R|^2 = 0$. Thus, (M, g) is locally symmetric.

There are other possibilities for applying this technique. Using Tachibana's result that K is nonnegative when the curvature operator is nonnegative, we see that (M, g) must be locally symmetric, provided that $\bar{R}$ is Einstein and has nonnegative curvature operator.

8.3.3 Rough Classification of Symmetric Spaces

Guided by our examples and the results on holonomy, we can now try to classify irreducible symmetric spaces. They seem to come in three groups.

Compact Type: If the Einstein constant is positive, then it follows from Myers' diameter bound (Chapter 9) that the space is compact. In this case one can show that the curvature operator is nonnegative.

Flat Type: If the space is Ricci flat, then it follows that it must be flat. In case the space is compact, this is immediate from Bochner's theorem, while if the space is noncompact and complete a little more work is needed. Thus, the only Ricci flat irreducible examples are S^1 and $\mathbb{R}^1$.

Noncompact Type: If the Einstein constant is negative, then it follows from Bochner's theorem that the space is noncompact. In this case, one can show that the curvature operator is nonpositive.

We won't give a complete list of all irreducible symmetric spaces, but one interesting feature is that they come in compact/noncompact dual pairs, as described in the above lists. Also, there is a further subdivision. Among the compact types there are Lie groups with bi-invariant metrics and then all the others. Similarly, in the noncompact regime there are the duals to the bi-invariant metrics and then the rest. This gives us the following division:

Type I: Compact irreducible symmetric spaces of the form G/K where G is a compact simple real Lie group and K a maximal compact subgroup. Example: $SO(k+l)/(SO(k)\times SO(l))$.

Type II: Compact irreducible symmetric spaces G, where G is a compact simple real Lie group with a bi-invariant metric. Example: $SO(n)$.

Type III: Noncompact symmetric spaces G/K, where G is a noncompact simple real Lie group and K a maximal compact subgroup. Example: $SO(k,l)/(SO(k)\times SO(l))$ or $Sl(n)/SO(n)$.

Type IV: Noncompact symmetric spaces G/K, where K is a compact simple real Lie group and G its complexification. Example: $SO(n,\mathbb{C})/SO(n)$.

The difference algebraically between compact and noncompact can be seen by looking at the examples above. There we saw that in the compact case $\mathfrak{t}$ consists of skew-symmetric matrices, while in the noncompact case $\mathfrak{t}$ consists of symmetric matrices. Thus, the metric looks like

$$g(X,Y)=\mp\operatorname{tr}(XY),$$

where the minus is for the compact case and the plus for the noncompact case. It is this difference that ultimately gave us the different sign for the curvatures. But even before this, we see that for $X,Y\in\mathfrak{t}$ and $K\in\mathfrak{k}$,

$$\begin{aligned}
g([X,K],Y) &= \mp\operatorname{tr}((XK-KX)Y)\\
&= \mp(\operatorname{tr}XKY-\operatorname{tr}KXY)\\
&= \mp(\operatorname{tr}KYX-\operatorname{tr}KXY)\\
&= \pm\operatorname{tr}(K(XY-YX))\\
&= \pm\operatorname{tr}(K[X,Y])\\
&= \mp\langle K,[X,Y]\rangle,
\end{aligned}$$

where for elements of $\mathfrak{k}$ we use that they are always skew-symmetric, and therefore their inner product is given by

$$\langle K_1,K_2\rangle=-\operatorname{tr}(K_1K_2).$$

Using this, one can see that

$$\begin{aligned}
g(R(X,Y)Z,W) &= g([Z,[X,Y]],W)\\
&= \mp\langle [X,Y],[Z,W]\rangle.
\end{aligned}$$

With this information we can compute the diagonal terms for the curvature operator

$$\begin{aligned} g\left(\mathfrak{R}\left(\sum X_i \wedge Y_i\right), \left(\sum X_i \wedge Y_i\right)\right) &= \sum g\left(R\left(X_i, Y_i\right) Y_j, X_j\right) \\ &= \mp \sum \left\langle \left[X_i, Y_i\right], \left[Y_j, X_j\right] \right\rangle \\ &= \pm \left| \sum \left[X_i, Y_i\right] \right|^2 \end{aligned}$$

and conclude that it is either nonnegative or nonpositive according to type. This seems to have been noticed for the first time in the literature in [38]. This also means that Tachibana's results from the last section are not vacuous. In fact, they give a characterization of symmetric spaces of compact type.

Note that as compact type symmetric spaces have nonnegative curvature operator, it becomes relatively easy to compute their cohomology. The Bochner technique tells us that all harmonic forms are parallel. Now, a parallel form is necessarily invariant under the holonomy. Thus, we are left with a classical invariance problem. Namely, determine all forms on a Euclidean space that are invariant under a given group action on the space. It is particularly important to know the cohomology of the real and complex Grassmannians, as one can use that information to define Pontryagin and Chern classes for vector bundles. We refer the reader to [76, vol. 5] and [61] for more on this.

8.3.4 The Golden Triangle

We take here a short detour to explain what some call the Golden Triangle of Riemannian geometry. It looks like this:

Parallel Transport	$\rightarrow$	Covariant Derivatives
$\nwarrow$		$\swarrow$
	Curvature	

The meaning is that any of these concepts can de derived from any of the others. Curvature was defined using the connection and therefore covariant differentiation. We saw above that curvature generates $\mathfrak{hol}$, thus parallel transport can be recaptured from knowledge of curvature. Finally, the connection, and therefore all covariant differentiation operations, can be recaptured from knowledge of parallel transport. This can be seen as follows: Suppose we wish to compute $\nabla_X Y$ at $p \in M$. Select a curve $c : I \to M$ with $\dot{c}(0) = X(p)$. Now along c we select a parallel frame $E_i(t)$ and write $Y \circ c(t) = \alpha^i(t) E_i(t)$. Then we have

$$\nabla_{X(p)} Y = \dot{\alpha}^i(0) E_i(0).$$

The philosophy of this triangle seems to have been noticed first by Levi-Civita and Ricci.

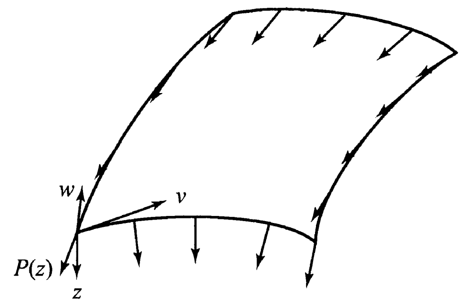


FIGURE 8.1.

8.4 Curvature and Holonomy

To get a better understanding of holonomy and how it relates to symmetric spaces, we need to figure out how it can be computed from the curvature tensor.

We let the Lie algebra of $\mathrm{Hol}_p^0 \subset SO(n)$ be denoted by $\mathfrak{hol}_p \subset \mathfrak{so}(n)$. This Lie algebra is therefore an algebra of skew-symmetric transformations of T_pM. We have on T_pM several other skew-symmetric transformations. Namely, for each pair of vectors $v, w \in T_pM$ there is the curvature tensor $R(v, w) : T_pM \to T_pM$ that maps x to $R(v, w)x$. In fact, if we let V, W be two commuting vector fields such that $V(p) = v$, $W(p) = w$, then we can for each $t > 0$ consider the loop c_t at p obtained by first following the flow of V for time t, then the flow of W for time t, then the flow of $-V$ for time t, and finally the flow of $-W$ for time t. Since the vector fields commute, this is indeed a loop. Now let P_t be parallel translation along this loop (see Figure 8.1). Then one can easily prove (first done by Cartan)

$$R(v, w) = \lim_{t \to 0} \frac{P_t - I}{t}.$$

To completely determine $\mathfrak{hol}_p$, it is of course necessary to look at all contractible loops, not just the short ones. However, each contractible loop can be decomposed into lassos, that is, loops that consist of a curve emanating from p and ending at some q, and then at q we have a very small loop (see Figure 8.2). Thus, any element of $\mathfrak{hol}_p$ is the composition of elements of the form

$$P^{-1} \circ R(P(v), P(w)) \circ P : T_pM \to T_pM,$$

where $P : T_pM \to T_qM$ denotes parallel translation along some curve from p to q. This characterization of holonomy was first proved by Amrose and Singer; the proof we have indicated is due to Nijenhuis. For the complete proofs the reader is referred to [11].

It is therefore possible in principle to compute holonomy from knowledge of the curvature tensor at all points. In reality this is not so useful, but for locally symmetric spaces we know that the curvature tensor is invariant under parallel translation, so we have

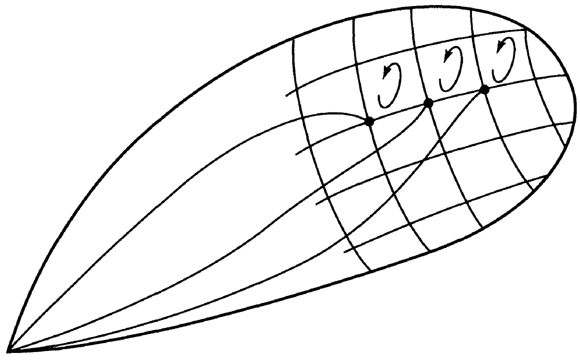


FIGURE 8.2.

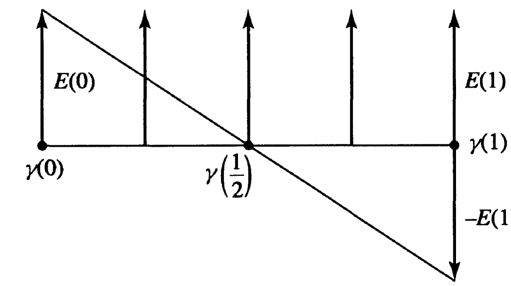


FIGURE 8.3.

Theorem 4.1 *For a locally symmetric space the holonomy Lie algebra* $\mathfrak{hol}_p$ *is generated by curvature transformations* $R(v, w)$, *where* $v, w \in T_pM$. *Moreover,* $\mathfrak{hol}_p \subset \mathfrak{iso}_p$.

Proof. We have already indicated the first part. For the second we could use our curvature description of the symmetric space. Instead, we give a more geometric proof, which also establishes that $R(x, y) \in \mathfrak{iso}_p$ as a by-product.

Observe that not only do isometries map geodesics to geodesics, but also parallel fields to parallel fields. Therefore, if we have a geodesic $\gamma : [0, 1] \to M$ and a parallel field E along γ, then we could apply the involution $I_{\gamma(1/2)}$ to E (this involution exists if the geodesic is sufficiently short). This involution reverses γ and at the same time changes the sign of E. Thus we have $DI_{\gamma(\frac{1}{2})}(E(0)) = -E(1)$, or in other words (see also Figure 8.3)

$$P_{\gamma(0)}^{\gamma(1)} = -DI_{\gamma(\frac{1}{2})}.$$

Now use that any curve can be approximated by a broken geodesic to conclude that parallel translation along any curve can be approximated by a successive composition of differentials of isometries. For a loop that is also a broken geodesic, we see that the composition of these isometries must belong to Iso_p. Hence, we have shown the stronger statement that

$$\mathrm{Hol}_p \subset \mathrm{Iso}_p.$$

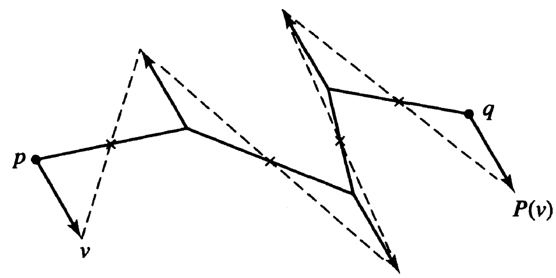


FIGURE 8.4.

In Figure 8.4 we have sketched how one can parallel translate along a broken geodesic from p to q in a symmetric space. This finishes the proof. □

Armed with this information, it is possible for us to determine the holonomy of irreducible symmetric spaces.

Corollary 4.2 *For an irreducible symmetric space*

$$\mathfrak{hol}_p = \mathfrak{iso}_p.$$

Proof. We know that $\mathfrak{hol}_p \subset \mathfrak{iso}_p$ and that $\mathfrak{iso}_p$ acts effectively on T_pM. By assumption we have that $\mathfrak{hol}_p$ acts irreducibly on T_pM. It is now a question of using that an irreducible symetric space has a unique Lie algebra description to finish this proof. Proving this uniqueness result is a little beyond what we wish to do here. □

Note that irreducibility is important for this theorem since Euclidean space has trivial holonomy. Also, Iso_p might contain orientation-reversing elements, so we cannot show equality at the Lie group level.

We shall now mention, without any indication of proof whatsoever, the classification of connected irreducible holonomy groups.

Theorem 4.3 (Berger, 1955) *Given a simply connected irreducible Riemannian n-manifold* (M, g), *then either the holonomy* Hol_p *acts transitively on the unit sphere in* T_pM *or* (M, g) *is a symmetric space of rank* ≥ 2. *Moreover, in the first case the holonomy is one of the following groups:*

$\dim = n$	Hol_p	Properties
n	$SO(n)$	Generic case
$n = 2m$	$U(m)$	Kähler metric
$n = 2m$	$SU(m)$	Kähler metric and $\mathrm{Ric} = 0$
$n = 4m$	$Sp(1) \cdot Sp(m)$	Quaternionic-Kähler and Einstein
$n = 4m$	$Sp(m)$	Hyper-Kähler and $\mathrm{Ric} = 0$
$n = 16$	$Spin(9)$	Symmetric and Einstein
$n = 8$	$Spin(7)$	$\mathrm{Ric} = 0$
$n = 7$	G_2	$\mathrm{Ric} = 0$

It is curious that all but the two largest irreducible holonomy groups, $SO(n)$ and $U(m)$, force the metric to be Einstein and in some cases even Ricci flat. Looking at the relationship between curvature and holonomy, it is clear that having small holonomy forces the curvature tensor to have some special properties. One can then in a case-by-case check see that various traces of the curvature tensor must be zero, thus forcing the metric to be either Einstein or Ricci flat (see [11] for details.) Note that Kähler metrics do not have to be Einstein (see the exercises to Chapter 3), and quaternionic Kähler manifolds are not necessarily Kähler, as $Sp(1) \cdot Sp(m)$ is not contained in $U(2m)$. Using a little bit of the theory of Kähler manifolds, it is not hard to see that metrics with holonomy $SU(n)$ are Ricci flat. Since $Sp(m) \subset SU(2m)$, we then get that also hyper-Kähler manifolds are Ricci flat. One can see that the last two holonomies occur only for Ricci flat manifolds. In particular, they never occur as the holonomy of a symmetric space. That all other holonomies do occur for symmetric spaces follows from Berger's result and the fact that the rank one symmetric spaces have holonomy $SO(n)$, $U(m)$, $Sp(1) \cdot Sp(m)$, or $Spin(9)$.

This leads to another profound question. Are there compact simply connected Ricci flat spaces with holonomy $SU(m)$, $Sp(m)$, G_2, or $Spin(7)$? The answer is yes. But it is a highly nontrivial yes. Yau got the Fields medal for establishing the $SU(m)$ case. Actually, he solved the Calabi conjecture, and the holonomy question was a by-product (see, e.g., [11] for more information on the Calabi conjecture). Note that we have the Eguchi-Hanson metric which is a complete Ricci flat Kähler metric and therefore has $SU(2)$ as holonomy group. Recently, D. Joyce solved the cases of $Spin(7)$ and G_2 by methods similar to those employed by Yau. An even more interesting question is whether there are compact simply connected manifolds that are Ricci flat but have $SO(n)$ as a holonomy group. Note that the Schwarzschild metric is complete Ricci flat and has $SO(4)$ as holonomy group. For more in-depth information on these issues we refer the reader to [11].

A general remark about how special ($\neq SO(n)$) holonomies occur: It seems that they are all related to the existence of parallel forms. In the Kähler case, for example, the Kähler form is a parallel nondegenerate 2-form. Correspondingly, one has a parallel 4-form for quaternionic-Kähler manifolds and a parallel 8-form for Cayley-Kähler manifolds (which are all known to be locally symmetric). This is studied in more detail in the proof of the classification of manifolds with nonnegative curvature operator below. For the last two exceptional holonomies there are also some special 4-forms that do the job.

From the classification of holonomy groups we immediately get an interesting corollary.

Corollary 4.4 *If a Riemannian manifold has the property that the holonomy doesn't act transitively on the unit sphere, then it is either reducible or a locally symmetric space of rank* ≥ 2. *In particular, the rank must be* ≥ 2.

It is unclear to what extent the converse fails for general manifolds. For nonpositive curvature, however, there is the famous higher-rank rigidity result proved independently by W. Ballmann and K. Burns-R. Spatzier (see [5] and [16]).

Theorem 4.5 *A compact Riemannian manifold of nonpositive curvature of rank* > 1 *does not have transitive holonomy. In particular it must be either reducible or locally symmetric.*

It is worthwhile mentioning that in [7] it was shown that the rank of a compact non-positively curved manifold can be computed from the fundamental group. Thus, a good deal of geometric information is automatically encoded into the topology. The rank rigidity theorem is proved by dynamical systems methods. The idea is to look at the geodesic flow on the unit sphere bundle, i.e., the flow that takes a unit vector and moves it time t along the unit speed geodesic in the direction of the unit vector. This flow has particularly nice properties on non-positively curved manifolds, which we won't go into. The idea is to use the flat parallel fields to show that the holonomy can't be transitive. Berger's result then gives us that the manifold has to be locally symmetric. Nice as this method of proof is, it would be very pleasing to have a proof that goes more along the lines of the Bochner technique. In nonpositive curvature this method is different. It usually centers on studying harmonic maps into the space rather than harmonic forms on the space (for more on this see [84]).

In nonnegative curvature, however, it is possible to find irreducible spaces that are not symmetric and have rank ≥ 2. On $S^2 \times S^2$ we have a product metric that is reducible and has rank 3. But if we take another metric on this space that comes as a quotient of $S^2 \times S^3$ by an action of S^1 (acting by rotations on the first factor and the Hopf action on the second), then we get a metric which has rank 2. The only way in which a rank 2 metric can split off a de Rham factor is if it splits off something 1-dimensional, but that is topologically impossible in this case. So in conclusion, the holonomy must be transitive and irreducible.

By assuming the stronger condition that the curvature operator is nonnegative, one can almost classify all such manifolds. This was first done in [38] and in more generality in Chen's article in [41]. This classification allows us to conclude that higher rank gives rigidity. The theorem and proof are a nice synthesis of everything we have learned in this and the previous chapter. In particular, the proof uses the Bochner technique in the two most nontrivial cases we have covered: for forms and the curvature tensor.

Theorem 4.6 (S. Gallot and D. Meyer, 1975) *If* (M, g) *is a compact Riemannian n-manifold with nonnegative curvature operator, then one of the following cases must occur:*

(1) (M, g) *has rank* > 1 *and is either reducible or locally symmetric.*

(2) $\mathrm{Hol}^0(M, g) = SO(n)$ *and the universal covering is homeomorphic to a sphere.*

(3) $\mathrm{Hol}^0(M,g) = U\left(\frac{n}{2}\right)$ *and the universal covering is biholomorphic to* $\mathbb{C}P^{\frac{n}{2}}$.

(4) $\mathrm{Hol}^0(M,g) = Sp(1)\,Sp\left(\frac{n}{4}\right)$ *and the universal covering is up to scaling isometric to* $\mathbb{H}P^{\frac{n}{4}}$.

(5) $\mathrm{Hol}^0(M,g) = Spin(9)$ *and the universal covering is up to scaling isometric to* CaP^2.

Proof. First observe that by the splitting theorem (see Chapter 9), a finite covering of M is isometric to a product $N \times T^n$. So if we assume that (M,g) is irreducible, then the fundamental group is finite, and we can therefore assume that we work with a simply connected manifold M. Now we observe that either all of the homology groups $H^p(M,\mathbb{R}) = 0$ for $p = 1,\ldots,n-1$, in which case the space is a homology sphere, or some homology group $H^p(M,\mathbb{R}) \neq 0$ for some $p \neq 0, n$. In the latter case, we then have a harmonic p-form by the Hodge theorem. The Bochner technique now tells us that this form must be parallel, since the curvature operator is nonnegative.

The proof is now based on the following observation: A Riemannian n-manifold with holonomy $SO(n)$ cannot admit any nontrivial p-forms for $0 < p < n$. Note that the volume form is always parallel, so it is clearly necessary to use the condition $p \neq 0, n$. We are also allowed to assume that $p \leq \frac{n}{2}$, since the Hodge star $*\omega$ of ω is parallel iff ω is parallel. The observation is proved by contradiction, so suppose that ω is a parallel p-form, where $0 < p < n$.

First suppose $p = 1$. Then the dual of the 1-form is a parallel vector field. This means that the manifold splits locally. In particular, it must be reducible and have special holonomy.

More generally, when $p \leq n-2$ is odd, we can for $v_1,\ldots,v_p \in T_pM$ find an element of $P \in SO(n)$ such that $P(v_i) = -v_i$. Therefore, if the holonomy is $SO(n)$ and ω is invariant under parallel translation, we must have

$$\begin{aligned}
\omega\left(v_1,\ldots,v_p\right) &= \omega\left(Pv_1,\ldots,Pv_p\right)\\
&= \omega\left(-v_1,\ldots,-v_p\right)\\
&= -\omega\left(v_1,\ldots,v_p\right).
\end{aligned}$$

This shows that $\omega = 0$. In case n is odd, we can then conclude using the Hodge star operator that no parallel forms exist when the holonomy is $SO(n)$.

We can then assume that we have an even dimensional manifold and that ω is a parallel p-form with $p \leq \frac{n}{2}$ even. We claim again that if the holonomy is $SO(n)$, then $\omega = 0$. First select vectors $v_1,\ldots,v_p \in T_pM$; then find orthonormal vectors $e_1,\ldots,e_p \in T_pM$ such that $\mathrm{span}\{v_1,\ldots,v_p\} = \mathrm{span}\{e_1,\ldots,e_p\}$. Then we know that $\omega\left(e_1,\ldots,e_p\right)$ is zero iff $\omega\left(v_1,\ldots,v_p\right)$ is zero. Now use that $p \leq \frac{n}{2}$ to find $P \in SO(n)$ such that

$$\begin{aligned}
P(e_1) &= e_2,\\
P(e_2) &= e_1,\\
P(e_i) &= e_i \quad \text{for} \quad i = 3,\ldots,p.
\end{aligned}$$

Using invariance of ω under P then yields

$$\begin{aligned}\omega\left(e_1,\ldots,e_p\right) &= \omega\left(Pe_1,\ldots,Pe_p\right)\\ &= \omega\left(e_2,e_1,e_3,\ldots,e_p\right)\\ &= -\omega\left(e_1,\ldots,e_p\right).\end{aligned}$$

In summary, we have thus shown that any Riemannian manifold with nonnegative curvature operator has holonomy $SO(n)$ only if all homology groups vanish. Supposing that the manifold is irreducible and has transitive holonomy, we can then use the above classification to see what holonomy groups are potentially allowed. The Ricci flat cases are, however, not allowed, as the nonnegative curvature would then make the manifold flat. Thus, we have only the three possibilities $U\left(\frac{n}{2}\right)$, $Sp(1)\,Sp\left(\frac{n}{4}\right)$, or $Spin(9)$. In the latter two cases one can show from holonomy considerations that the manifold must be Einstein. Thus, Tachibana's result from Chapter 7 shows that the metric must be locally symmetric. From the classification of symmetric spaces it then follows that the space is isometric to either $\mathbb{H}P^{\frac{n}{4}}$ or CaP^2. This leaves us with the Kähler case. In this situation we can show that the cohomology ring must be the same as that of $\mathbb{C}P^{\frac{n}{2}}$, i.e., there is a homology class $\omega \in H^2(M,\mathbb{R})$ such that any homology class is proportional to some power of $\omega : \omega^k = \omega \wedge \ldots \wedge \omega$. This can be seen as follows. If the holonomy is $U\left(\frac{n}{2}\right)$, first observe that there must be an almost complex structure on the tangent spaces that is invariant under parallel translation. Then we get a Kähler form ω from this structure by type change. This 2-form is necessarily parallel and cannot be exact, as the $\frac{n}{2}$th power $\omega^{\frac{n}{2}}$ must be proportional to the volume form. Thus, we have $H^2(M,\mathbb{R}) \neq 0$. If ω^k doesn't generate H^{2k}, then each form not proportional to ω^k will by the above arguments reduce the holonomy to a proper subgroup of $U\left(\frac{n}{2}\right)$.

To get the stronger conclusions on the topological type one must use results from [58] and [63]. □

There are two questions left over in this classification. Namely, for the sphere and complex projective space we get only topological rigidity. For the sphere one can clearly perturb the standard metric and still have positive curvature operator, so one couldn't expect more there. On $\mathbb{C}P^2$, say, we know that the curvature operator has exactly two zero eigenvalues. These two zero eigenvalues and eigenvectors are actually forced on us by the fact that the metric is Kähler. Therefore, if we perturb the standard metric, while keeping the same Kähler structure, then these two zero eigenvalues will persist and the positive eigenvalues will stay positive. Thus, the curvature operator stays nonnegative.

Given that there is such a big difference between the classes of manifolds with nonnegative curvature operator and nonnegative sectional curvature, one might think the same is true for nonpositive curvature. However, the above rank rigidity theorem tells us that in fact nonpositive sectional curvature is much more rigid than nonnegative sectional curvature. So the question is, Are there any compact rank 1 manifolds of nonpositive sectional curvature that do not admit a metric with nonpositive curvature operator?

8.5 Further Study

We have eliminated many important topics about symmetric spaces. For more in-depth information we recommend the texts by Besse, Helgason, and Jost (see [11, Chapters 7,10], [12, Chapter 3], [49], and [50, Chapter 6]). O'Neill's book [65, Chapter 8] also has a nice elementary account of symmetric spaces.

8.6 Exercises

1. Here are two problems on the connection between holonomy and Kähler manifolds.
   (a) Show that the holonomy of $\mathbb{C}P^n$ is $U(n)$.
   (b) Show that the holonomy of a Riemannian manifold is $U(m)$ iff it has a Kähler structure.
2. Show that $SO(n, \mathbb{C})/SO(n)$ and $Sl(n, \mathbb{C})/SU(n)$ are symmetric spaces with nonpositive curvature operator.
3. The *quaternionic projective space* is defined as being the quaternionic lines in $\mathbb{H}^{n+1}$. Here the quaternions $\mathbb{H}$ are the complex matrices
$$\begin{pmatrix} z & \bar{w} \\ -w & \bar{z} \end{pmatrix}.$$
If we identify $\mathbb{H}$ with $\mathbb{R}^4$, then we usually write elements as $x_1 + ix_2 + jx_3 + kx_4$. Multiplication is done using
$$\begin{aligned} i^2 = j^2 = k^2 &= -1, \\ ij = k &= -ji, \\ jk = i &= -kj, \\ ki = j &= -ik. \end{aligned}$$
   (a) Show that if we define
$$\begin{aligned} i &= \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \\ j &= \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \\ k &= \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \end{aligned}$$

then these two descriptions are equivalent.

(b) The *symplectic group* $Sp(n) \subset SU(2n) \subset SO(4n)$ consists of those orthogonal matrices that commute with the three complex structures generated by i, j, k on $\mathbb{R}^{4n}$. A better way of looking at this group is by considering $n \times n$ matrices A with quaternionic entries such that

$$A^{-1} = A^*.$$

Here the conjugate of a quaternion is

$$\overline{x_1 + ix_2 + jx_3 + kx_4} = x_1 - ix_2 - jx_3 - kx_4,$$

so we have as usual that

$$|q|^2 = q\bar{q}.$$

Now show that

$$\mathbb{H}P^n = Sp(n+1)/(Sp(1) \times Sp(n)).$$

Use this to exhibit $\mathbb{H}P^n$ as a symmetric space. Show that the holonomy is $Sp(1) \times Sp(n)$ and that the space is quarter pinched.

4. Construct the hyperbolic analogues of the complex and quaternionic projective spaces. Show that they have negative curvature and are quarter pinched.

5. Show that any locally symmetric space (not necessarily complete) is locally isometric to a symmetric space. Conclude that a simply connected locally symmetric space admits a monodromy map into a unique symmetric space. Show that if the locally symmetric space is complete, then the monodromy map is bijective.

6. Let (M, g) be a compact Riemannian n-manifold that is irreducible and with $\mathfrak{R} \geq 0$. Show that the following are equivalent:

   (a) $\chi(M) > 0$.
   (b) The odd Betti numbers are zero.
   (c) $-I \in \mathrm{Hol}_p^0$.
   (d) The dimension n is even.

   Use this to show that any compact manifold with $\mathfrak{R} \geq 0$ has $\chi(M) \geq 0$.

7. Show that if an irreducible symmetric space has strictly positive or negative curvature operator, then it has constant curvature.

8. Using the skew-symmetric linear maps

$$x \wedge y : T_pM \to T_pM,$$
$$x \wedge y(v) = g(x, v)\, y - g(y, v)\, x,$$

show that $\Lambda^2 T_p M = \mathfrak{so}(T_p M)$. Using this identification, show that the image of the curvature operator $\mathfrak{R}(\Lambda^2 T_p M) \subset \mathfrak{hol}_p$, with equality for symmetric spaces. Use this to conclude that the holonomy is $SO(n)$ if the curvature operator is positive or negative.

9
Ricci Curvature Comparison

In this chapter we shall introduce some of the fundamental theorems for manifolds with lower Ricci curvature bounds. Two important techniques will be developed: relative volume comparison and weak upper bounds for the Laplacian of distance functions. With these techniques we shall show numerous results on restrictions of fundamental groups of such spaces and also present a different proof of the estimate for the first Betti number by Bochner. The proof of the splitting theorem is self-contained. It uses the generalized maximum principle, but we show how one can get around the regularity issue for harmonic distance functions using some of our previous work on distance functions.

Until around 1970, when Cheeger and Gromoll proved their splitting theorem, one had only the Bochner technique and Myers' diameter estimate as part of the theory of Ricci curvature. In the mid 1970s, Cheng proved his maximal diameter result, which shows that only the sphere has maximal diameter. Around 1980, Gromov exposed the world to his view of how volume comparison can be used. The relative volume comparison theorem was actually first proved by Bishop in [13]. At the time, however, one only considered balls of radius less than the injectivity radius. Later, Gromov observed that the result holds for all balls and immediately put it to use in many situations. In particular, he showed how one could generalize the Betti number estimate from Bochner's theorem using only topological methods and volume comparison. Anderson then refined this to get information about fundamental groups. One's intuition about Ricci curvature has generally been borrowed from experience with sectional curvature. This has led to many naive conjectures that haven proven to be false by constructing several interesting examples of manifolds with nonnegative Ricci curvature. On the other hand, much good work has also come out of this, as we shall see. The reason for treating Ricci curvature before the more advanced results on sectional curvature is that we want to break the link between the two. The techniques for dealing with these two subjects, while similar, are not the same. In a way, it is even easier to get into the theory of Ricci curvature, as one doesn't have to struggle through the proof of Toponogov's theorem.

9.1 Volume Comparison

9.1.1 *The Basic Comparison Estimates*

Throughout this section, assume that we have a complete Riemannian manifold (M, g) of dimension n. Furthermore, we are given a point $p \in M$ and with that the distance function $f(x) = d(x, p)$. We know that this distance function is smooth on the image of the segment domain. Using polar coordinates, we can write the metric as $\left(g_{\alpha\beta}(r, \theta)\right)$. Moreover, if we let $m = \Delta f$ and $\lambda = \sqrt{\det\left(g_{\alpha\beta}(r, \theta)\right)}$, then we have three fundamental relationships between these quantities:

(tr1) $\partial_r m + m^2/(n-1) \leq \partial_r m + \left|\nabla^2 f\right|^2 = -\mathrm{Ric}(\partial_r, \partial_r)$,
(tr2) $\partial_r \lambda = m \cdot \lambda$,
(tr3) $\partial_r^2 \sqrt[n-1]{\lambda} \leq -(\mathrm{Ric}(\partial_r, \partial_r)/(n-1)) \cdot \sqrt[n-1]{\lambda}$,

where the initial conditions for λ are

$$\sqrt[n-1]{\lambda}(r, \theta) = r + O\left(r^2\right),$$
$$\partial_r \sqrt[n-1]{\lambda}(r, \theta) = 1 + O(r).$$

Consequently, (tr2) gives us that

$$m(r, \theta) \sim \frac{n-1}{r} \quad \text{as} \quad r \to 0.$$

With this information we can prove

Lemma 1.1 (Ricci Comparison Result) *Suppose that* (M, g) *has* Ric $\geq (n-1) \cdot k$ *for some* $k \in \mathbb{R}$. *Then*

$$\sqrt[n-1]{\lambda}(r, \theta) \leq \mathrm{sn}_k(r)$$

and

$$m(r, \theta) \leq (n-1)\frac{\mathrm{sn}_k'(r)}{\mathrm{sn}_k(r)}.$$

Proof. Notice that the right-hand sides of the inequalities correspond exactly to what one would get in constant curvature k.

The first inequality follows from the fact that

$$\begin{aligned}\partial_r^2 \sqrt[n-1]{\lambda} &\leq -k \cdot \sqrt[n-1]{\lambda}, \\ \sqrt[n-1]{\lambda}(r,\theta) &= r + O\left(r^2\right), \\ \partial_r \sqrt[n-1]{\lambda}(r,\theta) &= 1 + O(r).\end{aligned}$$

For the second, we use that

$$\partial_r m + \frac{m^2}{n-1} \leq -(n-1)\cdot k.$$

This can be solved by separation of variables:

$$\int_a^b \frac{\partial_r \frac{m}{n-1}}{k + \left(\frac{m}{n-1}\right)^2} \leq (a-b).$$

The integral can be computed explicitly, and depending on the sign of k, we get something with either arccot, $1/m$, or arccoth. Inverting this and using the initial condition $m(r,\theta) \sim (n-1)/r$ then gives the desired inequality. One could also use the substitution $u = (n-1)/m$, together with the initial condition $u(0) = 0$, to avoid infinities. □

9.1.2 The Diameter Estimate

With these simple preliminary observations we can now generalize Bonnet's diameter estimate from Chapter 6.

Theorem 1.2 (Myers, 1941) *Suppose (M, g) is a complete Riemannian manifold with* $\mathrm{Ric} \geq (n-1)k > 0$. *Then* $\mathrm{diam}(M, g) \leq \pi/\sqrt{k}$. *Furthermore, (M, g) has finite fundamental group.*

Proof. To see that $\mathrm{diam}(M,g) \leq \pi/\sqrt{k}$, we show that the segment domain for any point p is contained in the closed ball $B\left(0, \pi/\sqrt{k}\right)$. This shows that no distance function can take values that are larger than $\pi/\sqrt{k}$. It was just shown that the volume density in polar coordinates satisfies

$$\sqrt[n-1]{\lambda}(r,\theta) \leq \mathrm{sn}_k(r) = \frac{\sin\left(\sqrt{k}\cdot r\right)}{\sqrt{k}}.$$

This quantity goes to zero as $r \to \pi/\sqrt{k}$. Thus, no matter what direction we go in from a point, we must develop a conjugate before we reach $\pi/\sqrt{k}$.

This implies, in particular, that M is compact. Thus, the universal covering (which also has $\mathrm{Ric} \geq (n-1)k$) must also be compact, showing that $\pi_1(M)$ is finite. □

Example 1.3 The universal covering of the incomplete Riemannian manifold $S^2 - \{p, -p\}$ clearly has constant curvature 1 but is not compact.

Example 1.4 $S^1 \times \mathbb{R}^3$ admits a complete doubly warped product metric $dr^2 + \varphi^2(r)d\theta^2 + \psi^2(r)ds_2^2$, which has $\mathrm{Ric} > 0$ everywhere. For $t \geq 1$ just let $\varphi(t) = t^{-1/4}$ and $\psi(t) = t^{3/4}$ and then adjust φ and ψ near $t = 0$ to make things work out.

9.1.3 Volume Estimation

Our next applications are to volume comparison. To define volume, first recall that if $\sigma^1, \ldots, \sigma^n$ is an orthonormal coframing on $U \subset M$, then $\omega = \sigma^1 \wedge \ldots \wedge \sigma^n$ is the Riemannian volume form on (U, g). Therefore, if $A \subset U$, then $\mathrm{vol}A = |\int_A \sigma^1 \wedge \ldots \wedge \sigma^n|$. Strictly speaking, (M, g) only has a volume form if it is orientable. To compute volumes of subsets, however, we always take absolute values afterwards, so we don't need to worry about signs. Another convenient observation is that any Riemannian manifold admits $U \subset M$ such that $M - U$ has measure zero (zero measure is independent of the Riemannian metric) and a coordinate chart $\varphi : U \to \mathbb{R}^n$. Thus, any open set $O \subset M$ satisfies $\mathrm{vol}(O) = \mathrm{vol}(O \cap U)$, where the latter quantity can be computed. The volume form on U must, of course, be of the form $\lambda(x) \cdot dx^1 \wedge \ldots \wedge dx^n$, and it is not hard to check that $\lambda(x) = \sqrt{\det g(\partial_i, \partial_j)} = \sqrt{\det g_{ij}}$.

Our first volume comparison gives the obvious upper volume bound, which comes from our upper bound on the volume density.

Lemma 1.5 *If (M, g) has $\mathrm{Ric} \geq (n-1) \cdot k$, then*

$$\mathrm{vol}B(p, r) \leq v(n, k, r),$$

where $v(n, k, r)$ denotes the volume of a ball of radius r in the constant-curvature space form S_k^n.

Proof. Above, we showed that in polar coordinates around p we have

$$\lambda(r, \theta) \leq \mathrm{sn}_k^{n-1}(r).$$

In S_k^n the volume density in polar coordinates is exactly $\mathrm{sn}_k^{n-1}(r)$. If the volume form on $S^{n-1}(1)$ is denoted by $d\theta$, then we have

$$\begin{aligned} \mathrm{vol}B(p, r) &= \int_{\mathrm{seg}_p \cap B(0,r)} \lambda(r, \theta)\, dr \wedge d\theta \\ &\leq \int_{B(0,r)} \mathrm{sn}_k^{n-1}(r)\, dr \wedge d\theta \\ &= v(k, n, r). \end{aligned}$$

□

With a little more technical work, the above absolute volume comparison result can be improved in a rather interesting direction. The result one obtains is referred to as the relative volume comparison estimate. It will prove invaluable in many situations throughout the rest of the text.

Lemma 1.6 (Relative Volume Comparison, Bishop-Cheeger-Gromov, 1964–1980) *Suppose (M, g) is a complete Riemannian manifold with* $\mathrm{Ric} \geq (n-1)\cdot k$. *Then*

$$r \to \frac{\mathrm{vol} B(p, r)}{v(n, k, r)}$$

is a nonincreasing function whose limit as $r \to 0$ is 1.

Proof. We will use exponential polar coordinates. The volume form $\lambda(r, \theta)dr \wedge d\theta$ for (M, g) is initially defined only on some star-shaped subset of $T_pM = \mathbb{R}^n = (0, \infty) \times S^{n-1}$, but we can just set $\lambda(x) = 0$ outside this set. The volume form $\lambda_k(r, \theta)dr \wedge d\theta$ is defined on all of $\mathbb{R}^n$ when $k \leq 0$, but only on $B\left(0, \pi/\sqrt{k}\right)$ for $k > 0$. We can likewise extend $\lambda_k = 0$ outside $B\left(0, \pi/\sqrt{k}\right)$. Myers' theorem says that $\lambda = 0$ on $\mathbb{R}^n - B\left(0, \pi/\sqrt{k}\right)$ in this case. So we might as well just consider $r < \pi/\sqrt{k}$ when $k > 0$.

We now have

$$\frac{\mathrm{vol} B(p, R)}{v(n, k, R)} = \frac{\int_0^R \int_{S^{n-1}} \lambda dr \wedge d\theta}{\int_0^R \int_{S^{n-1}} \lambda_k dr \wedge d\theta},$$

where $\omega_{n-1} = \int_{S^{n-1}} d\theta = \mathrm{vol}\left(S^{n-1}(1)\right)$, and we know that $0 \leq \lambda(r, \theta) \leq \lambda_k(r, \theta) = \mathrm{sn}_k^{n-1}(r)$ everywhere.

Differentiation of this quotient with respect to R yields

$$\begin{aligned}
&\frac{d}{dR}\left(\frac{\mathrm{vol} B(p, R)}{v(n, k, R)}\right) \\
&\quad = \frac{\left(\int_{S^{n-1}} \lambda(R, \theta)\, d\theta\right)\left(\int_0^R \int_{S^{n-1}} \lambda_k(r, \theta)\, dr \wedge d\theta\right)}{(v(n, k, R))^2} \\
&\qquad - \frac{\left(\int_{S^{n-1}} \lambda_k(R, \theta)\, d\theta\right)\left(\int_0^R \int_{S^{n-1}} \lambda(r, \theta)\, dr \wedge d\theta\right)}{(v(n, k, R))^2} \\
&\quad = (v(n, k, R))^{-2} \cdot \int_0^R \left[\left(\int_{S^{n-1}} \lambda(R, \theta)\, d\theta\right) \cdot \left(\int_{S^{n-1}} \lambda_k(r, \theta)\, d\theta\right)\right. \\
&\qquad \left. - \left(\int_{S^{n-1}} \lambda_k(R, \theta)\, d\theta\right)\left(\int_{S^{n-1}} \lambda(r, \theta)\, d\theta\right)\right] dr.
\end{aligned}$$

So to see that

$$R \to \frac{\mathrm{vol} B(p, R)}{v(n, k, R)}$$

is nonincreasing, it therefore suffices to check that

$$\frac{\int_{S^{n-1}} \lambda(r,\theta)\, d\theta}{\int_{S^{n-1}} \lambda_k(r,\theta)\, d\theta} = \frac{1}{\omega_{n-1}} \int_{S^{n-1}} \frac{\lambda(r,\theta)}{\lambda_k(r,\theta)} d\theta$$

is nonincreasing. This follows from

$$\begin{aligned}\partial_r\left(\log\frac{\lambda(r,\theta)}{\lambda_k(r,\theta)}\right) &= \partial_r \log\lambda - \partial_r \log\lambda_k \\ &= \frac{\partial_r\lambda}{\lambda} - \frac{\partial_r\lambda_k}{\lambda_k} \\ &= m - m_k \\ &\leq 0. \end{aligned}$$

□

9.1.4 Maximal Diameter Rigidity

Given Myers' diameter estimate, it is natural to ask what happens if the diameter obtains its maximal value. The next result shows that only the sphere has this property.

Theorem 1.7 (S.Y. Cheng, 1975) *If (M,g) is a complete Riemannian manifold with* $\mathrm{Ric} \geq (n-1)k > 0$ *and* $\mathrm{diam} = \pi/\sqrt{k}$, *then (M,g) is isometric to S^n_k.*

Proof. Fix $p, q \in M$ such that $d(p,q) = \pi/\sqrt{k}$. Define $f(x) = d(x,p)$, $h(x) = d(x,q)$. We will show that

(1) $f + h = d(p,x) + d(x,q) = d(p,q) = \pi/\sqrt{k},\ x \in M.$
(2) f, h are smooth on $M - \{p,q\}$.
(3) $\nabla^2 f = (\mathrm{sn}_k'/\mathrm{sn}_k)(I - \partial_r \cdot dr)$ on $M - \{p,q\}$.
(4) $g = dr^2 + \mathrm{sn}_k^2 ds_{n-1}^2$.

We know that (3) implies (4) and that (4) implies M must be S^n_k.

Proof of (1): The triangle inequality shows that $d(p,x) + d(x,q) \geq \pi/\sqrt{k}$, so if (1) does not hold, we can find $\varepsilon > 0$ such that (see Figure 9.1)

$$d(p,x) + d(x,q) = 2\cdot\varepsilon + \frac{\pi}{\sqrt{k}} = 2\cdot\varepsilon + d(p,q).$$

Then the metric balls $B(p,r_1)$, $B(q,r_2)$, and $B(x,\varepsilon)$ are pairwise disjoint, when $r_1 = d(p,x) - \varepsilon$, $r_2 = d(q,x) - \varepsilon$ (note that $r_1 + r_2 = \pi/\sqrt{k}$). Thus,

$$\begin{aligned} 1 = \frac{\mathrm{vol}M}{\mathrm{vol}M} &\geq \frac{\mathrm{vol}B(x,\varepsilon) + \mathrm{vol}B(p,r_1) + \mathrm{vol}B(q,r_2)}{\mathrm{vol}M} \\ &\geq \frac{v(n,k,\varepsilon)}{v\left(n,k,\frac{\pi}{\sqrt{k}}\right)} + \frac{v(n,k,r_1)}{v\left(n,k,\frac{\pi}{\sqrt{k}}\right)} + \frac{v(n,k,r_2)}{v\left(n,k,\frac{\pi}{\sqrt{k}}\right)} \end{aligned}$$

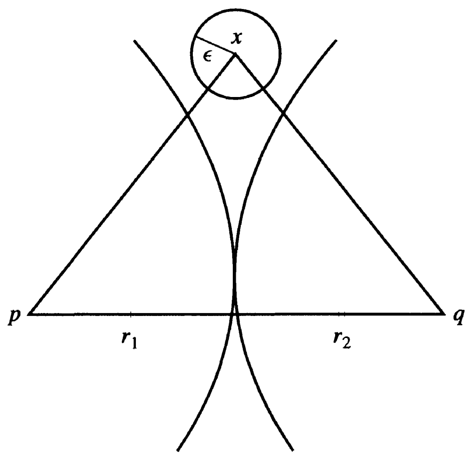


FIGURE 9.1.

$$= \frac{v(n,k,\varepsilon)}{v\left(n,k,\frac{\pi}{\sqrt{k}}\right)} + 1 > 1,$$

which is a contradiction.

Proof of (2): If $x \in M - \{q, p\}$, then x can be joined to both p and q by segments σ_1, σ_2. The previous statement says that if we put these two segments together, then we get a segment from p to q through x. Such a segment must be smooth, and thus σ_1 and σ_2 are both subsegments of a larger segment. This implies from our characterization of when distance functions are smooth that both f and h are smooth at $x \in M - \{p, q\}$.

Proof of (3): We have $f(x) + h(x) = \pi/\sqrt{k}$, thus $\Delta f = -\Delta h$. On the other hand,

$$\begin{aligned}
(n-1)\frac{\mathrm{sn}_k'(f(x))}{\mathrm{sn}_k(f(x))} &\geq \Delta f(x) \\
&= -\Delta h(x) \\
&\geq -(n-1)\frac{\mathrm{sn}_k'(h(x))}{\mathrm{sn}_k(h(x))} \\
&= -(n-1)\frac{\mathrm{sn}_k'\left(\frac{\pi}{\sqrt{k}} - f(x)\right)}{\mathrm{sn}_k\left(\frac{\pi}{\sqrt{k}} - f(x)\right)} \\
&= (n-1)\frac{\mathrm{sn}_k'(f(x))}{\mathrm{sn}_k(f(x))}.
\end{aligned}$$

Thus,

$$\Delta f = (n-1)\frac{\mathrm{sn}_k'}{\mathrm{sn}_k}$$

and

$$\begin{aligned}-(n-1)\cdot k &= \partial_r(\Delta f) + \frac{(\Delta f)^2}{n-1}\\ &\leq \partial_r(\Delta f) + |\nabla^2 f|^2\\ &\leq -\mathrm{Ric}(\partial_r, \partial_r)\\ &\leq -(n-1)k.\end{aligned}$$

Hence, all inequalities are equalities, and in particular $(\Delta f)^2 = (n-1)|\nabla^2 f|^2$. This can, however, happen only when

$$\nabla^2 f = \frac{\Delta f}{n-1}(I - \partial_r \cdot d_r) = \frac{\mathrm{sn}_k'}{\mathrm{sn}_k}(I - \partial_r \cdot d_r).$$

To see this, just think of $\nabla^2 f$ as being diagonalized with eigenvalues $0 = \lambda_1, \lambda_2, \ldots, \lambda_n$ then the equality says

$$(\lambda_2 + \cdots + \lambda_n)^2 = (n-1)\left(\lambda_2^2 + \cdots + \lambda_n^2\right).$$

This can be true only provided that $\lambda_2 = \cdots = \lambda_n$. □

We have now proved that any complete manifold with $\mathrm{Ric} \geq (n-1)\cdot k > 0$ has diameter $\leq \pi/\sqrt{k}$, where equality holds only when the space is S_k^n. A natural perturbation question is therefore, If (M, g) has $\mathrm{Ric} \geq (n-1)\cdot k > 0$ and $\mathrm{diam} \approx \pi/\sqrt{k}$, must M be homeomorphic or diffeomorphic to a sphere?

For $n = 2, 3$ this is true when $n \geq 4$, however, there are counterexamples. The case $n = 2$ will be settled later, while $n = 3$ goes beyond the scope of this book (see [75]). The examples for $n \geq 4$ are divided into two cases: $n = 4$ and $n \geq 5$.

Example 1.8 (Anderson, see [3]) For $n = 4$ consider metrics on $I \times S^3$ of the form $dr^2 + \varphi^2\sigma_1^2 + \psi^2(\sigma_2^2 + \sigma_3^2)$. If we define

$$\varphi(r) = \begin{cases} \frac{\sin(ar)}{a} & r \leq r_0, \\ c_1 \sin(r + \delta) & r \geq r_0, \end{cases}$$

$$\psi = \begin{cases} br^2 + c & r \leq r_0, \\ c_2 \sin(r + \delta) & r \geq r_0, \end{cases}$$

and then reflect these function in $r = \pi/2 - \delta$, we get a metric on $\mathbb{C}P^2\sharp\mathbb{C}P^2$. For any small $r_0 > 0$ we can now adjust the parameters so that φ and ψ become C^1 and generate a metric with $\mathrm{Ric} \geq (n-1)$. For smaller and smaller choices of r_0 we see that $\delta \to 0$, so the interval $I \to [0, \pi]$ as $r_0 \to 0$. This means that the diameters converge to π.

Example 1.9 (Otsu, see [66]) For $n \geq 5$ we only need to consider standard doubly warped products: $dr^2 + \varphi^2 \cdot ds_2^2 + \psi^2 ds_{n-3}^2$ on $I \times S^2 \times S^{n-3}$. Similar choices for φ and ψ will yield metrics on $S^2 \times S^{n-2}$ with $\mathrm{Ric} \geq n-1$ and diameter $\to \pi$.

In both of the above examples, we actually only constructed C^1 functions φ, ψ and therefore only C^1 metrics. So one also needs to smooth out these functions to make them C^∞. This is not hard to accomplish but is still a fairly technical and nasty task to perform.

9.2 Fundamental Groups and Ricci Curvature

We shall now attempt to generalize the estimate on the first Betti number we obtained using the Bochner technique to the situation where one has more general Ricci curvature bounds. This requires some knowledge about how fundamental groups are tied in with the geometry.

9.2.1 *The First Betti Number*

We now wish to use the relative volume comparison estimate to re-prove and extend Bochner's Betti number bound.

Suppose M is a compact Riemannian manifold of dimension n and $\tilde{M}$ its universal covering space. The fundamental group $\pi_1(M)$ acts by isometries on $\tilde{M}$. Recall from algebraic topology that $H_1(M,\mathbb{Z}) = \pi_1(M)/[\pi_1(M),\pi_1(M)]$, where $[\pi_1(M),\pi_1(M)]$ is the commutator subgroup. Thus, $H_1(M,\mathbb{Z})$ acts by deck transformations on the covering space $\tilde{M}/[\pi_1(M),\pi_1(M)]$ with quotient M. Since $H_1(M,\mathbb{Z})$ is a finitely generated Abelian group, we know that the set of torsion elements T is a finite normal subgroup. We can then consider $\Gamma = H_1(M,\mathbb{Z})/T$ as acting by deck transformations on $\bar{M} = \tilde{M}/[\pi_1(M),\pi_1(M)]/T$. Thus, we have a covering $\pi : \bar{M} \to M$ with torsion-free and Abelian Galois group of deck transformations. The rank of the torsion-free group Γ is clearly equal to $b_1(M) = \dim H_1(M,\mathbb{R})$. Now recall that any finite-index subgroup of Γ has the same rank as Γ. So if we can find a finite-index subgroup that is generated by elements that can be geometrically controlled, then we might be able to bound b_1. To this end we have a very interesting result.

Lemma 2.1 (M. Gromov, 1980) *For fixed $x \in \bar{M}$ there exists a finite-index subgroup $\Gamma' \subset \Gamma$ that is generated by elements $\gamma_1,\ldots,\gamma_m$ such that*

$$d(x,\gamma_i(x)) \leq 2 \cdot \operatorname{diam}(M).$$

Furthermore, for all $\gamma \in \Gamma' - \{1\}$ we have

$$d(x,\gamma(x)) > \operatorname{diam}(M).$$

Proof. First we find a finite-index subgroup that can be generated by elements satisfying the first condition. Then we modify this group so that it also satisfies the second condition.

For each $\varepsilon > 0$ let Γ_ε be the group generated by

$$\{\gamma \in \Gamma : d(x, \gamma(x)) < 2\mathrm{diam}(M) + \varepsilon\},$$

and let $\pi_\varepsilon : \bar{M} \to \bar{M}/\Gamma_\varepsilon$ denote the covering projection. We claim that for each $z \in \bar{M}$ we have $d(\pi_\varepsilon(z), \pi_\varepsilon(x)) < \mathrm{diam}(M) + \varepsilon$; in particular, Γ_ε will then have finite index, as $\bar{M}/\Gamma_\varepsilon$ is compact. Otherwise, we could find $z \in \bar{M}$ such that $d(x, z) = d(\pi_\varepsilon(z), \pi_\varepsilon(x)) = \mathrm{diam}(M) + \varepsilon$. Now, we can find $\gamma \in \Gamma$ such that $d(\gamma(x), z) \leq \mathrm{diam}(M)$, but then we would have

$$\begin{aligned} d(\pi_\varepsilon(\gamma(x)), \pi_\varepsilon(x)) &\geq d(\pi_\varepsilon(z), \pi_\varepsilon(x)) - d(\pi_\varepsilon(z), \pi_\varepsilon(\gamma(x))) \geq \varepsilon, \\ d(x, \gamma(x)) &\leq d(x, z) + d(z, \gamma(x)) \leq 2\mathrm{diam}(M) + \varepsilon. \end{aligned}$$

Here we have a contradiction, as the first line says that $\gamma \notin \Gamma_\varepsilon$, while the second line says $\gamma \in \Gamma_\varepsilon$. Now observe that there are at most finitely many elements in the set $\{\gamma \in \Gamma : d(x, \gamma(x)) < 3\mathrm{diam}(M)\}$, as Γ acts discretely on $\bar{M}$. Hence, there must be a sufficiently small $\varepsilon > 0$ such that $\{\gamma \in \Gamma : d(x, \gamma(x)) < 2\mathrm{diam}(M) + \varepsilon\} = \{\gamma \in \Gamma : d(x, \gamma(x)) \leq 2\mathrm{diam}(M)\}$. Then we have a finite-index subgroup Γ_ε of Γ generated by $\{\gamma \in \Gamma : d(x, \gamma(x)) \leq 2\mathrm{diam}(M)\} = \{\gamma_1, \ldots, \gamma_m\}$. We shall now modify these generators until we get the desired group Γ'.

First, observe that as the rank of Γ_ε is b_1, we can assume that $\{\gamma_1, \ldots, \gamma_{b_1}\}$ are linearly independent and generate a subgroup $\Gamma'' \subset \Gamma_\varepsilon$ of finite index. Next, we recall that only finitely many elements γ in Γ'' lie in $\{\gamma \in \Gamma : d(x, \gamma(x)) \leq 2\mathrm{diam}(M)\}$. We can therefore choose $\{\tilde{\gamma}_1, \ldots, \tilde{\gamma}_{b_1}\} \subset \{\gamma \in \Gamma : d(x, \gamma(x)) \leq 2\mathrm{diam}(M)\}$ with the following properties (we use additive notation here, as it is easier to read):

(1) $\mathrm{span}\{\tilde{\gamma}_1, \ldots, \tilde{\gamma}_k\} \subset \mathrm{span}\{\gamma_1, \ldots, \gamma_k\}$ has finite index for all $k = 1, \ldots, b_1$.

(2) $\tilde{\gamma}_k = l_{1k} \cdot \gamma_1 + \cdots + l_{kk} \cdot \gamma_k$ is chosen such that l_{kk} is maximal in absolute value among all elements in $\Gamma'' \cap \{\gamma \in \Gamma : d(x, \gamma(x)) \leq 2\mathrm{diam}(M)\}$.

The group Γ' generated by $\{\tilde{\gamma}_1, \ldots, \tilde{\gamma}_{b_1}\}$ clearly has finite index in Γ'' and hence also in Γ. The generators lie in $\{\gamma \in \Gamma : d(x, \gamma(x)) \leq 2\mathrm{diam}(M)\}$, as demanded by the first property. It only remains to show that the second property is also satisfied. The see this, let $\gamma = m_1 \cdot \tilde{\gamma}_1 + \cdots + m_k \cdot \tilde{\gamma}_k$ be chosen such that $m_k \neq 0$. If $d(x, \gamma(x)) \leq \mathrm{diam}(M)$, then we also have that

$$\begin{aligned} d\left(x, \gamma^2(x)\right) &\leq d(x, \gamma(x)) + d\left(\gamma(x), \gamma^2(x)\right) \\ &= 2d(x, \gamma(x)) \\ &\leq 2\mathrm{diam}(M). \end{aligned}$$

Thus, $\gamma^2 \in \Gamma'' \cap \{\gamma \in \Gamma : d(x, \gamma(x)) \leq 2\mathrm{diam}(M)\}$, and also,

$$\begin{aligned} \gamma^2 &= 2m_1 \cdot \tilde{\gamma}_1 + \cdots + 2m_k \cdot \tilde{\gamma}_k \\ &= \sum_{i=1}^{k-1} n_i \cdot \gamma_i + 2m_k \cdot l_{kk} \cdot \gamma_k. \end{aligned}$$

But this violates the maximality of l_{kk}, as we assumed $m_k \neq 0$. □

With this lemma we can now give Gromov's proof of

Theorem 2.2 (S. Gallot and M. Gromov, 1980) *If M is a Riemannian manifold of dimension n such that* $\mathrm{Ric} \geq (n-1)k$ *and* $\mathrm{diam}(M) \leq D$, *then there is a function $C(n, k \cdot D^2)$, with the property that $\lim_{x\to 0} C(n,x) = n$, such that*

$$b_1(M) \leq C\left(n, k \cdot D^2\right).$$

In particular, there is $\varepsilon(n) > 0$ such that if $k \cdot D^2 \geq -\varepsilon(n)$, then $b_1(M) \leq n$.

Proof. First observe that for $k > 0$ there is nothing to prove, as we know that $b_1 = 0$ from Myers' theorem.

Suppose we have chosen a covering $\bar{M}$ of M with torsion-free Abelian Galois group of deck transformations $\Gamma = \langle \gamma_1, \ldots, \gamma_{b_1} \rangle$ such that for some $x \in \bar{M}$ we have

$$d(x, \gamma_i(x)) \leq 2\mathrm{diam}(M),$$
$$d(x, \gamma(x)) > \mathrm{diam}(M), \gamma \neq 1.$$

Then we clearly have that all of the balls $B(\gamma(x), (\mathrm{diam}(M))/2)$ are disjoint. Now set

$$I_r = \left\{\gamma \in \Gamma : \gamma = l_1 \cdot \gamma_1 + \cdots + l_{b_1} \cdot \gamma_{b_1}, |l_1| + \cdots + \left|l_{b_1}\right| \leq r\right\}.$$

If $b_1(M) \geq b$, then we must have that the cardinality of I_r is larger than $\left(\frac{r}{b}\right)^b$. On the other hand, for $\gamma \in I_r$ we have $B(\gamma(x), (\mathrm{diam}(M))/2) \subset B(x, r \cdot 2\mathrm{diam}(M) + (\mathrm{diam}(M))/2)$. All of these balls are disjoint and have the same volume, as γ acts isometrically. We can therefore use the relative volume comparison theorem to conclude that the cardinality of I_r is bounded from above by

$$\frac{\mathrm{vol}B\left(x, r \cdot 2\mathrm{diam}(M) + \frac{\mathrm{diam}(M)}{2}\right)}{\mathrm{vol}B\left(x, \frac{\mathrm{diam}(M)}{2}\right)} \leq \frac{v\left(n, k, r \cdot 2\mathrm{diam}(M) + \frac{\mathrm{diam}(M)}{2}\right)}{v\left(n, k, \frac{\mathrm{diam}(M)}{2}\right)}.$$

Hence, we have

$$\left(\frac{r}{b}\right)^b \leq \frac{v\left(n, k, r \cdot 2\mathrm{diam}(M) + \frac{\mathrm{diam}(M)}{2}\right)}{v\left(n, k, \frac{\mathrm{diam}(M)}{2}\right)}$$

for all $r > 0$. If we set $r = 2b$, we clearly get a bound for b. Thus, b_1 is bounded from above in terms of k and $\mathrm{diam}(M)$. To get better information, observe that

$$\frac{v\left(n, k, r \cdot 2\mathrm{diam}(M) + \frac{\mathrm{diam}(M)}{2}\right)}{v\left(n, k, \frac{\mathrm{diam}(M)}{2}\right)} \leq \frac{v\left(n, k, \left(r \cdot 2 + \frac{1}{2}\right) D\right)}{v\left(n, k, \frac{D}{2}\right)}$$

$$= \frac{\int_0^{(r\cdot 2+\frac{1}{2})D} \left(\frac{\sinh(\sqrt{-k}t)}{\sqrt{-k}}\right)^{n-1} dt}{\int_0^{\frac{1}{2}D} \left(\frac{\sinh(\sqrt{-k}t)}{\sqrt{-k}}\right)^{n-1} dt}$$

$$= \frac{\int_0^{(r\cdot 2+\frac{1}{2})D\sqrt{-k}} \sinh^{n-1}(t)\,dt}{\int_0^{\frac{1}{2}D\sqrt{-k}} \sinh^{n-1}(t)\,dt}$$

$$= 2^n\left(r\cdot 2+\frac{1}{2}\right)^n + \cdots \leq 5^n \cdot r^n,$$

where in the last step we assume that $D\sqrt{-k}$ is very small relative to r. We now get the desired bound for b_1 by using $r = b^b$ and $b = b_1$. If we assume that $b_1 \geq b = n+1$ and we fix $r = 1 + 5^n \cdot (n+1)^{n+1}$, then for small $D\sqrt{-k}$, depending on this choice for r, we get

$$\left(\frac{r}{n+1}\right)^{n+1} \leq 5^n \cdot r^n,$$

which is impossible. □

Gallot's proof of the above theorem uses techniques that are sophisticated generalizations of the Bochner technique.

9.2.2 *Finiteness of Fundamental Groups*

One can get even more information from these volume comparison techniques. Instead of considering just the first homology group, we can actually get some information about fundamental groups as well.

For our next result we need a different kind of understanding of how fundamental groups can be represented.

Lemma 2.3 (M. Gromov, 1980) *For a Riemannian manifold M and $\tilde{x} \in \tilde{M}$, we can always find generators $\{\gamma_1, \ldots \gamma_m\}$ for the fundamental group $\Gamma = \pi_1(M)$ such that $d(\tilde{x}, \gamma_i(\tilde{x})) \leq 2\operatorname{diam}(M)$ and such that all relations for Γ in these generators are of the form $\gamma_i \cdot \gamma_j \cdot \gamma_k^{-1} = 1$.*

Proof. For any $\varepsilon \in (0, \operatorname{inj}(M))$ choose a triangulation of M such that adjacent vertices in this triangulation are joined by a curve of length less that ε. Let $\{x_1, \ldots, x_k\}$ denote the set of vertices and $\{e_{ij}\}$ the edges joining adjacent vertices (thus, e_{ij} is not necessarily defined for all i, j). If x is the projection of $\tilde{x} \in \tilde{M}$, then join x and x_i by a segment σ_i for all $i = 1, \ldots, k$ and construct the loops

$$\sigma_{ij} = \sigma_i e_{ij} \sigma_j^{-1}$$

for adjacent vertices. Now, any loop in M based at x is homotopic to a loop in the 1-skeleton of the triangulation. Furthermore, any loop in this triangulation is homotopic to a product of loops of the form σ_{ij}. Thus, these loops generate Γ.

Now observe that if three vertices x_i, x_j, x_k are adjacent to each other, then they span a 2-simplex Δ_{ijk}. Thus, we have that the loop $\sigma_{ij}\sigma_{jk}\sigma_{ki} = \sigma_{ij}\sigma_{jk}\sigma_{ik}^{-1}$ is homotopically trivial. We claim that these are the only relations needed to describe Γ. To see this, let σ be any loop in the 1-skeleton that is homotopically trivial. Now use that σ in fact contracts in the 2-skeleton. Thus, a homotopy corresponds to a collection of 2-simplices Δ_{ijk}. In this way we can represent the relation $\sigma = 1$ as a product of elementary relations of the form $\sigma_{ij}\sigma_{jk}\sigma_{ik}^{-1} = 1$.

Finally, use discreteness of Γ to get rid of ε as in the above case. □

A simple example might be instructive here.

Example 2.4 Consider $M_k = S^3/\mathbb{Z}_k =$ the constant-curvature 3-sphere divided out by the cyclic group of order k. As $k \to \infty$ the volume of these manifolds goes to zero, while the curvature is 1 and the diameter $\frac{\pi}{2}$. Thus, the fundamental groups can only get bigger at the expense of having small volume. If we insist on writing the cyclic group $\mathbb{Z}_k$ in the above manner, then the number of generators needed goes to infinity as $k \to \infty$.

For numbers $n \in \mathbb{N}, k \in \mathbb{R}, v, D \in (0,\infty)$, let $\mathfrak{M}(n,k,v,D)$ denote the class of compact Riemannian n-manifolds with

$$\begin{aligned} \text{Ric} &\geq (n-1)k, \\ \text{vol} &\geq v, \\ \text{diam} &\leq D. \end{aligned}$$

We can now prove:

Theorem 2.5 (M. Anderson, 1990) *There are only finitely many fundamental groups among the manifolds in $\mathfrak{M}(n,k,v,D)$ for fixed n, k, v, D.*

Proof. Choose generators $\{\gamma_1, \ldots, \gamma_m\}$ as in the lemma. Since the number of possible relations is bounded by 2^{m^3}, we have reduced the problem to showing that m is bounded. We have that $d(x, \gamma_i(x)) \leq 2D$. Fix a fundamental domain $F \subset \tilde{M}$ that contains x, i.e., a closed set such that $\pi : F \to M$ is onto and $\text{vol}F = \text{vol}M$. One could, for example, choose $F = \left\{z \in \tilde{M} : d(x,z) \leq d(\gamma(x), z) \text{ for all } \gamma \in \pi_1(M)\right\}$. Then we have that the sets $\gamma_i(F)$ are disjoint up to sets of measure 0, all have the same volume, and all lie in the ball $B(x, 4D)$. Thus,

$$m \leq \frac{\text{vol}B(x,4D)}{\text{vol}F} \leq \frac{v(n,k,4D)}{v}.$$

In other words, we have bounded the number of generators in terms of n, D, v, k alone. □

Another related result shows that groups generated by short loops must in fact be finite.

Lemma 2.6 (M. Anderson, 1990) *For fixed numbers $n \in \mathbb{N}$, $k \in \mathbb{R}$, $v, D \in (0, \infty)$ we can find $L = L(n, k, v, D)$ and $N = N(n, k, v, D)$ such that if $M \in \mathfrak{M}(n, k, v, D)$, then any subgroup of $\pi_1(M)$ that is generated by loops of length $\leq L$ must have order $\leq N$.*

Proof. Let $\Gamma \subset \pi_1(M)$ be a group generated by loops $\{\gamma_1, \dots, \gamma_k\}$ of length $\leq L$. Consider the universal covering $\pi : \tilde{M} \to M$ and let $x \in \tilde{M}$ be chosen such that the loops are based at $\pi(x)$. Then select a fundamental domain $F \subset \tilde{M}$ such that $x \in F$. One could, for example, choose $F = \left\{z \in \tilde{M} : d(x, z) \leq d(\gamma(x), z) \text{ for all } \gamma \in \pi_1(M)\right\}$. We have that for any $\gamma_1, \gamma_2 \in \pi_1(M)$, either $\gamma_1 = \gamma_2$ or $\gamma_1(F) \cap \gamma_2(F)$ has measure 0.

Now define $U(r)$ as the set of $\gamma \in \Gamma$ such that γ can be written as a product of at most r elements from $\{\gamma_1, \dots, \gamma_k\}$. We assumed that $d(x, \gamma_i(x)) \leq L$ for all i, and thus $d(x, \gamma(x)) \leq r \cdot L$ for all $\gamma \in U(r)$. This means that $\gamma(F) \subset B(x, r \cdot L + D)$. As the sets $\gamma(F)$ are disjoint up to sets of measure zero, we then obtain

$$\begin{aligned} |U(r)| &\leq \frac{\mathrm{vol} B(x, r \cdot L + D)}{\mathrm{vol} F} \\ &\leq \frac{v(n, k, r \cdot L + D)}{v}. \end{aligned}$$

If we assume that Γ has order larger than N and we suppose that $L < D/N$, then for $r = N$,

$$N < \frac{v(n, k, 2D)}{v}.$$

In other words, if we set

$$\begin{aligned} N &= \frac{\mathrm{vol} B(x, 2D)}{v} + 1, \\ L &= \frac{D}{2N}, \end{aligned}$$

then we get the desired conclusion. □

9.3 Manifolds of Nonnegative Ricci Curvature

In this section we shall prove the splitting theorem of Cheeger-Gromoll. This theorem is analogous to the maximal diameter theorem in many ways. It also has far-reaching consequences for compact manifolds with nonnegative Ricci curvature. For instance, we shall see that $S^3 \times \mathbb{R}$ does not admit any complete metrics with zero Ricci curvature. One of the critical ingredients in the proof of the splitting theorem is the maximum principle for continuous functions. These analytical matters will be taken care of in the first subsection.

9.3.1 The Maximum Principle

We shall now try to understand how one can assign second derivatives to (distance) functions at points where the function is not smooth. Later, we shall also discuss generalized gradients, but this theory is completely different and works only for Lipschitz functions.

The key observation for our development of generalized Hessians and Laplacians is

Lemma 3.1 *If $f, h : (M, g) \to \mathbb{R}$ are C^2 functions such that $f(p) = h(p)$ and $f(x) \geq h(x)$ for all x near p, then*

$$\begin{aligned} \nabla f(p) &= \nabla h(p), \\ \nabla^2 f(p) &\geq \nabla^2 h(p), \\ \Delta f(p) &\geq \Delta h(p). \end{aligned}$$

Proof. If $(M, g) \subset (\mathbb{R}, \text{can})$, then the theorem is simple calculus. In general, We can take $\gamma : (-\varepsilon, \varepsilon) \to M$ to be a geodesic with $\gamma(0) = p$, then use this observation on $f \circ \gamma, h \circ \gamma$ to see that

$$\begin{aligned} df(\dot{\gamma}(0)) &= dh(\dot{\gamma}(0)), \\ g\left(\nabla^2 f(\dot{\gamma}(0)), \dot{\gamma}(0)\right) &\geq g\left(\nabla^2 h(\dot{\gamma}(0)), \dot{\gamma}(0)\right). \end{aligned}$$

This clearly implies the lemma if we let $v = \dot{\gamma}(0)$ run over all $v \in T_pM$. □

This lemma implies that a C^2 function $f : M \to \mathbb{R}$ has $\nabla^2 f(p) \geq S$, where S is a symmetric linear map on T_pM (or $\Delta f(p) \geq a \in \mathbb{R}$), iff for every $\varepsilon > 0$ there exists a function $f_\varepsilon(x)$ defined in a neighborhood of p such that

(1) $f_\varepsilon(p) = f(p)$.
(2) $f(x) \geq f_\varepsilon(x)$ in some neighborhood of p.
(3) $\nabla^2 f_\varepsilon(p) \geq S - \varepsilon \cdot I$ (or $\Delta f_\varepsilon(p) \geq a - \varepsilon$).

Such functions f_ε are called *support functions from below*. One can analogously use *support functions from above* to find upper bounds for $\nabla^2 f$ or Δf.

For a continuous function $f : (M, g) \to \mathbb{R}$ we say that: $\nabla^2 f(p) \geq S$ (or $\Delta f(p) \geq a$) iff there exist smooth support functions f_ε satisfying (1) to (3).

One can easily check that if $(M, g) \subset (\mathbb{R}, \text{can})$, then f has $\nabla^2 f \geq 0$ everywhere iff f is convex. Thus, $f : (M, g) \to \mathbb{R}$ has $\nabla^2 f \geq 0$ everywhere iff $f \circ \gamma$ is convex for all geodesics γ. Using this, one can easily prove

Theorem 3.2 *If $f : (M, g) \to \mathbb{R}$ is continuous with $\nabla^2 f \geq 0$ everywhere, then f is constant near any local maximum. In particular, f cannot have a global maximum unless f is constant.*

We shall need a more general version of this theorem called the maximum principle. As stated below, it was first proved for smooth functions by E. Hopf in 1927 and then later for continuous functions by Calabi in 1958.

Theorem 3.3 (The Strong Maximum Principle) *If $f : (M, g) \to \mathbb{R}$ is continuous and has $\Delta f \geq 0$ everywhere, then f is constant in a neighborhood of every local maximum. In particular, f can have a global maximum only if f is constant.*

Proof. First, suppose that $\Delta f > 0$ everywhere. Then f can't have any local maxima at all. For if f has a local maximum at $p \in M$, then there would exist a smooth support function $f_\varepsilon(x)$ with

(1) $f_\varepsilon(p) = f(p)$,
(2) $f_\varepsilon(x) \leq f(x)$ for all x near p,
(3) $\Delta f_\varepsilon(p) > 0$.

Here (1) and (2) imply that f_ε must also have a local maximum at p. But this implies that $\nabla^2 f_\varepsilon(p) \leq 0$, which contradicts (3).

Next just assume that $\Delta f \geq 0$ and let $p \in M$ be a local maximum for f. For sufficiently small $r < \text{inj}(p)$ we therefore have a function $f : (B(p, r), g) \to \mathbb{R}$ with $\Delta f \geq 0$ and a global maximum at p. If f is constant on $B(p, r)$, then we are done; otherwise, we can assume (by possibly decreasing r) that f is not equal to $f(p)$ on $S(p, r) = \{x \in M : d(p, x) = r\}$. Then define $V = \{x \in S(p, r) : f(x) = f(p)\}$. Now construct a smooth function $h = e^{\alpha\varphi} - 1$ such that

$$
\begin{aligned}
h &< 0 \quad \text{on} \quad V, \\
h(p) &= 0, \\
\Delta h &> 0 \quad \text{on} \quad \bar{B}(p, r).
\end{aligned}
$$

This function is found by first selecting an open disc $U \subset S(p, r)$ that contains V. We can then find φ such that

$$
\begin{aligned}
\varphi(p) &= 0, \\
\varphi &< 0 \quad \text{on} \quad U, \\
\nabla\varphi &\neq 0 \quad \text{on} \quad \bar{B}(p, r).
\end{aligned}
$$

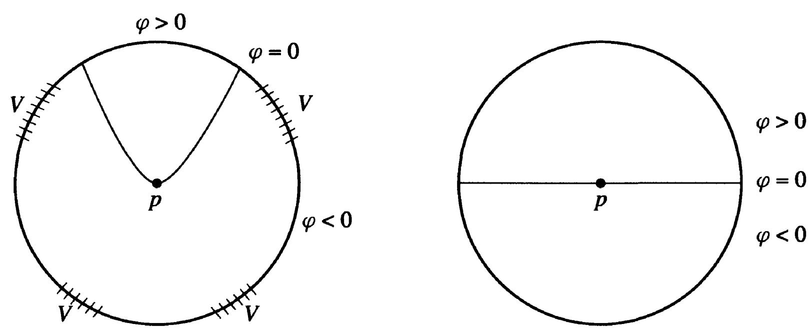


FIGURE 9.2.

In an appropriate coordinate system $\left(x^1, \ldots, x^n\right)$ we can simply assume that U looks like the lower half-plane: $x^1 < 0$ and then define $\varphi = x^1$ (see also Figure 9.2). Then choose α so big that $\Delta h = \alpha e^{\alpha\varphi}(\alpha|\nabla\varphi|^2 + \Delta\varphi) > 0$ on $\overline{B}(p, r)$.

Now consider the function $f_\delta = f + \delta h$ on $\overline{B}(p, r)$. Provided that δ is very small, this function has a local maximum in the interior $B(p, r)$, since

$$\begin{aligned} f_\delta(p) &= f(p) \\ &> \max\{f(x) + \delta h(x) = f_\delta(x) : x \in \partial B(p, r)\}. \end{aligned}$$

On the other hand, we can also show that f_δ has positive Laplacian, thus giving a contradiction with the first part of the proof. To see that the Laplacian is positive, select f_ε a support function from below for f at $q \in B(p, r)$. Then $f_\varepsilon + \delta h$ is a support function from below for f_δ at q. The Laplacian of this support function is estimated by

$$\Delta(f_\varepsilon + \delta h)(q) \geq -\varepsilon + \delta\Delta h(q),$$

which for given δ must become positive as $\varepsilon \to 0$. □

A continuous function $f : (M, g) \to \mathbb{R}$ is said to be *linear* if $\nabla^2 f \equiv 0$ (i.e., both of the inequalities $\nabla^2 f \geq 0$, $\nabla^2 f \leq 0$ hold everywhere). One can easily prove that this implies that $f \circ \gamma$ really is linear for each geodesic γ. This implies that $f \circ \exp_p(x) = f(p) + g(v_p, x)$ for each $p \in M$ and some $v_p \in T_pM$. This shows in particular that f is C^∞ with $\nabla f_p = v_p$.

More generally, we have the concept of a harmonic function. This is a continuous function $f : (M, g) \to \mathbb{R}$ with $\Delta f = 0$. The maximum principle shows that if M is closed, then all harmonic functions are constant. On incomplete or complete open manifolds, however, there are often many harmonic functions. This is in contrast to the existence of linear functions, where ∇f is necessary parallel and therefore splits the manifold locally into a product where one factor is an interval. It is a fairly subtle fact that any harmonic function is C^∞ if the metric is C^∞.

We finish this subsection with a new piece of notation. A continuous function $f : (M, g) \to \mathbb{R}$ with $\Delta f \geq 0$ everywhere is said to be *subharmonic*. If instead, $\Delta f \leq 0$, then f is *superharmonic*.

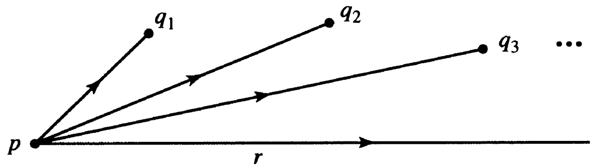


FIGURE 9.3.

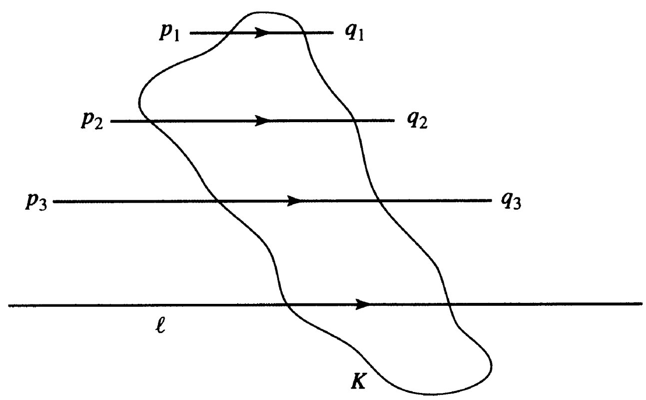


FIGURE 9.4.

9.3.2 *Rays and Lines*

We will work only with complete manifolds in this section. A *ray* $r(t) : [0, \infty) \to (M, g)$ is a unit speed geodesic such that $d(r(t), r(s)) = |t - s|$ for all $t, s \geq 0$. One can think of a ray as a semi-infinite segment or as a segment from $r(0)$ to infinity. A *line* $\ell(t) : \mathbb{R} \to (M, g)$ is a unit speed geodesic such that $d(\gamma(t), \gamma(s)) = |t - s|$ for all $t, s \in \mathbb{R}$.

Lemma 3.4 *If $p \in (M, g)$, then there is always a ray emanating from p. If M is disconnected at infinity then (M, g) contains a line.*

Proof. Let $p \in M$ and consider a sequence $q_i \to \infty$. Find unit vectors $v_i \in T_pM$ such that: $\sigma_i(t) = \exp_p(tv_i)$, $t \in [0, d(p, q_i)]$ is a segment from p to q_i. By possibly passing to a subsequence, we can assume that $v_i \to v \in T_pM$ (see Figure 9.3). Now $\sigma(t) = \exp_p(tv)$, $t \in [0, \infty)$, becomes a segment. This is because σ_i converges pointwise to σ by continuity of $\exp_p$, and thus

$$d(\sigma(s), \sigma(t)) = \lim d(\sigma_i(s), \sigma_i(t)) = |s - t|.$$

A complete manifold is *connected at infinity* if for every compact set $K \subset M$ there is a compact set $C \supset K$ such that any two points in $M - C$ can be joined by a curve in $M - K$. If M is not connected at infinity, we say that M is *disconnected at infinity*.

If M is disconnected at infinity, we can obviously find a compact set K and sequences of points $p_i \to \infty$, $q_i \to \infty$ such that any curve from p_i to q_i must pass through K. If we join these points by segments $\sigma_i : (-a_i, b_i) \to M$ such that $a_i, b_i \to \infty$, $\sigma_i(0) \in K$, then the sequence will subconverge to a line (see Figure 9.4). $\square$

Example 3.5 Surfaces of revolution $dr^2 + \varphi^2(r)ds_{n-1}^2$, where $\varphi : [0, \infty) \to [0, \infty)$ and $\dot\varphi(t) < 1, \ddot\varphi(t) < 0, t > 0$, cannot contain any lines. These manifolds look like paraboloids.

Example 3.6 Any complete metric on $S^{n-1} \times \mathbb{R}$ must contain a line, since the manifold is disconnected at infinity.

Example 3.7 The Schwarzschild metric on $S^2 \times \mathbb{R}^2$ does not contain any lines. This will also follow from our main result in this section.

Theorem 3.8 (The Splitting Theorem, Cheeger-Gromoll, 1971) *If (M, g) contains a line and has* Ric ≥ 0, *then (M, g) is isometric to a product $(H \times \mathbb{R}, g_0 + dt^2)$.*

The proof is quite involved and will require several constructions. The main idea is to find a distance function $f : M \to \mathbb{R}$ (i.e., $|\nabla f| \equiv 1$) that is linear (i.e., $\nabla^2 f \equiv 0$). Having found such a function, one can easily see that $M = U_0 \times \mathbb{R}$, where $U_0 = f^{-1}(0)$ and $g = dt^2 + g_0$. The maximum principle will play a key role in showing that f, when it has been constructed, is both smooth and linear. Recall that in the proof of the maximal diameter theorem we used two distance functions f, h placed at maximal distance from each other and then proceeded to show that $f + h =$ constant. This implied that f, h were smooth, except at the two chosen points, and that Δf is exactly what it is in constant curvature. We then used the rigidity part of the Cauchy-Schwartz inequality to compute $\nabla^2 f$. In the construction of our linear distance function we shall do something similar. In this situation the two ends of the line play the role of the points at maximal distance. We will then construct two distance functions $b_\pm$ from infinity, using this line, that are continuous and satisfy $b_+ + b_- \geq 0$ (from the triangle inequality), $\Delta b_\pm \leq 0$, and $b_+ + b_- = 0$ on the line. Thus, $b_+ + b_-$ is superharmonic and has a global minimum. The minimum principle will therefore show that $b_+ + b_- \equiv 0$. Thus, $b_+ = -b_-$ and $0 \geq \Delta b_+ = -\Delta b_- \geq 0$, which shows that both of $b_\pm$ are harmonic and therefore C^∞. We then show that they are actually distance functions (i.e., $|\nabla b_\pm| \equiv 1$). Then we can conclude that

$$\begin{aligned} 0 &= \nabla b_\pm(\Delta b_\pm) + \frac{(\Delta b_\pm)^2}{n-1} \\ &\leq \nabla b_\pm(\Delta b_\pm) + |\nabla^2 b_\pm|^2 \\ &= |\nabla^2 b_\pm|^2 \\ &\leq -\mathrm{Ric}(\nabla b_\pm, \nabla b_\pm) \leq 0. \end{aligned}$$

This establishes that $|\nabla^2 b_\pm|^2 = 0$, so that we have two linear distance functions $b_\pm$ as desired.

The proof proceeds through several results, which we will need later and which are also of some interest in their own right.

9.3.3 Laplacian Comparison

Lemma 3.9 (E. Calabi, 1958) *Let* $f(x) = d(x, p)$, $p \in (M, g)$. *If* $\mathrm{Ric}(M, g) \geq 0$, *then*

$$\Delta f(x) \leq \frac{n-1}{f(x)} \quad \text{for all} \quad x \in M.$$

Proof. We know that the result is true whenever f is smooth. For any other $q \in M$, choose a unit speed segment $\sigma : [0, \ell] \to M$ with $\sigma(0) = p$, $\sigma(\ell) = q$. Then the triangle inequality implies that $f_\varepsilon(x) = \varepsilon + d(\sigma(\varepsilon), x)$ is a support function from above for f at q. If all these functions are smooth at q, then

$$\begin{aligned}\Delta f_\varepsilon(q) &\leq \frac{n-1}{f_\varepsilon(q) - \varepsilon} \\ &= \frac{n-1}{f(q) - \varepsilon} \\ &\leq \frac{n-1}{f(q)} + \varepsilon \cdot \frac{2(n-1)}{(f(q))^2}\end{aligned}$$

for small ε, and hence $\Delta f(q) \leq \frac{n-1}{f(q)}$ in the support sense.

Now for the smoothness. Fix $\varepsilon > 0$ and suppose f_ε is not smooth at q. Then we know that either

(1) there are two segments from $\sigma(\varepsilon)$ to q, or
(2) q is a critical value for $\exp_{\sigma(\varepsilon)} : \mathrm{seg}(\sigma(\varepsilon)) \to M$.

Case (1) would gives us a nonsmooth curve of length ℓ from p to q, which we know is impossible. Thus, case (2) must hold. To get a contradiction out of this, we show that this implies that $h(x) = d(x, q)$ can't be smooth at $\sigma(\varepsilon)$, which we know to be an incorrect conclusion. For convenience, assume that σ is reparametrized so that $|\dot\sigma| = d(p, q) - \varepsilon = \ell - \varepsilon$ and $\sigma(b) = q$.

Using that q is critical for $\exp_{\sigma(\varepsilon)}$, we find $w \in T\mathrm{seg}(\sigma(\varepsilon))$ such that $D\exp_{\sigma(\varepsilon)} w = 0$. If we consider the curve $s \to sw + \dot\sigma(\varepsilon)$, then we have that

$$\frac{d}{ds}\left(\exp_{\sigma(\varepsilon)}(sw + \dot\sigma(\varepsilon))\right)(0) = 0.$$

Now consider the triangle $(s, t) \to t \cdot (sw + \dot\sigma(\varepsilon))$ and with it the mapping: $(s, t) \to \exp_{\sigma(\varepsilon)}(t \cdot (sw + \dot\sigma(\varepsilon)))$. Then we get a vector field $J(t)$ along σ by

$$J(t) = \frac{d}{ds}\left[\exp_{\sigma(\varepsilon)}(t \cdot (sw + \dot\sigma(\varepsilon)))\right](0).$$

This vector field satisfies

$$
\begin{aligned}
J(0) &= 0, \\
J(1) &= 0, \\
\dot{J}(0) &= w, \\
\ddot{J}(t) &= -R_{\dot{\sigma}}(J(t)).
\end{aligned}
$$

The equation $\ddot{J}(t) = -R_{\dot{\sigma}}(J(t))$ is very easy to establish, as J has constant coefficients in polar coordinates around $\sigma(\varepsilon)$ and therefore satisfies $\dot{J} = \left(\nabla^2 f_\varepsilon\right)(J)$. Thus, J is a solution to a second-order linear ODE. In particular, J is uniquely determined by $J(0) = 0$ and $\dot{J}(0)$. Since $\dot{J}(0) \neq 0$, it must follow that $\dot{J}(1) \neq 0$, for otherwise J must be everywhere zero. Now consider the triangle $(s,t) \to (1-t)\left(s\dot{J}(1) - \dot{\sigma}(\ell)\right)$. This generates a vector field K along σ that satisfies

$$
\begin{aligned}
K(t) &= \frac{d}{ds}\left[\exp_q\left((1-t)\left(s\dot{J}(1) - \dot{\sigma}(\ell)\right)\right)\right](0), \\
K(1) &= 0 = J(1), \\
\dot{K}(1) &= \dot{J}(1), \\
\ddot{K} &= -R_{\dot{\sigma}}(K).
\end{aligned}
$$

Thus, we conclude that $J = K$ and in particular, $J(0) = K(0) = 0$. This, however, implies that $\exp_q$ is critical at $\sigma(\varepsilon)$. □

By a similar analysis, we can prove

Lemma 3.10 *If (M, g) is complete and $\mathrm{Ric}(M, g) \geq (n-1)\cdot k$, then any distance function $f(x) = d(x, p)$ satisfies:*

$$
\Delta f(x) \leq (n-1)\frac{\mathrm{sn}_k'(f(x))}{\mathrm{sn}_k(f(x))}.
$$

This lemma can be used to give a different proof of Cheng's diameter theorem that does not use relative volume comparison.

As before, consider $h(x) = d(x, q)$, $f(x) = d(x, p)$, where $d(p, q) = \pi/\sqrt{k}$. Then we have $f + h \geq \pi/\sqrt{k}$, and equality will hold for any $x \in M - \{p, q\}$ that lies on a segment joining p and q. On the other hand, the above lemma tells us that

$$
\begin{aligned}
\Delta(f+h) &\leq \Delta f + \Delta h \\
&\leq (n-1)\sqrt{k}\cot(\sqrt{k}f(x)) + (n-1)\sqrt{k}\cdot\cot(\sqrt{k}h(x)) \\
&\leq (n-1)\sqrt{k}\cot(\sqrt{k}f(x)) + (n-1)\sqrt{k}\cot\left(\sqrt{k}\left(\frac{\pi}{\sqrt{k}} - f(x)\right)\right) \\
&= (n-1)\sqrt{k}(\cot(\sqrt{k}f(x)) + \cot(\pi - \sqrt{k}f(x))) = 0.
\end{aligned}
$$

So $f + h$ is superharmonic on $M - \{p, q\}$ and has a global minimum on this set. Thus, the minimum principle tells us that $f + h = \pi/\sqrt{k}$ on M. The proof can now be completed as before.

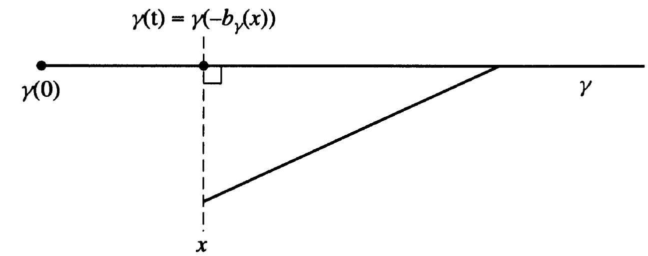


FIGURE 9.5.

9.3.4 *Busemann Functions*

Now let $\gamma : [0, \infty) \to (M, g)$ be a unit speed ray, and define $b_t(x) = d(x, \gamma(t)) - t$.

Proposition 3.11 (1) *For fixed x, the function $t \to b_t(x)$ is decreasing and bounded in absolute value by $d(x, \gamma(0))$.*
(2) $|b_t(x) - b_t(y)| \leq d(x, y)$.
(3) $\Delta b_t(x) \leq \frac{n-1}{b_t + t}$ *everywhere.*

Proof. (2) and (3) are obvious, since $b_t(x) + t$ is a distance function from $\gamma(t)$. For (1), first observe that the triangle inequality implies

$$|b_t(x)| = |d(x, \gamma(t))| - t| = |d(x, \gamma(t)) - d(\gamma(0), \gamma(t))| \leq d(x, \gamma(0)).$$

Second, if $s < t$ then

$$\begin{aligned} b_t(x) - b_s(x) &= d(x, \gamma(t)) - t - d(x, \gamma(s)) + s \\ &= d(x, \gamma(t)) - d(x, \gamma(s)) - d(\gamma(t), \gamma(s)) \\ &\leq d(\gamma(t), \gamma(s)) - d(\gamma(t), \gamma(s)) = 0. \end{aligned}$$

□

Thus, the family of functions $\{b_t\}_{t \geq 0}$ forms a pointwise bounded equicontinuous family that is also pointwise decreasing. Thus, b_t must converge to a distance-decreasing function b_γ satisfying

$$\begin{aligned} |b_\gamma(x) - b_\gamma(y)| &\leq d(x, y), \\ |b_\gamma(x)| &\leq d(x, \gamma(0)), \end{aligned}$$

and

$$b_\gamma(\gamma(r)) = \lim b_t(\gamma(r)) = \lim(d(\gamma(r), \gamma(t)) - t) = -r.$$

This function b_γ is called the *Busemann function* for γ and should be interpreted as a distance function from "$\gamma(\infty)$."

Example 3.12 If $M = (\mathbb{R}^n, \text{can})$, then all Busemann functions are of the form $b_\gamma(x) = \dot{\gamma}(0) \cdot \gamma(0) \cdot x$ (see Figure 9.5).

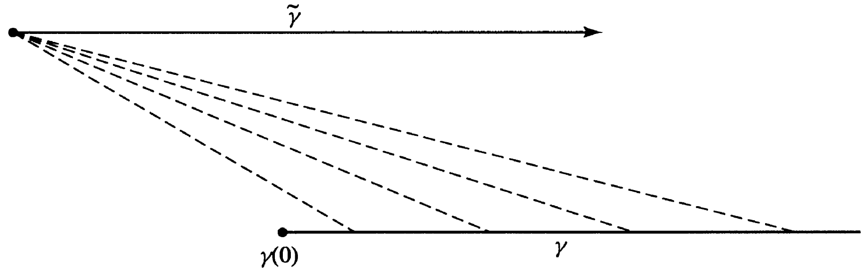


FIGURE 9.6.

The level sets $b_\gamma^{-1}(t)$ are called *horospheres*. In $(\mathbb{R}^n, \mathrm{can})$ these are obviously hyperplanes.

Given our ray γ, as before, and $p \in M$, consider a family of unit speed segments $\sigma_t : [0, \ell_t] \to (M, g)$ from p to $\gamma(t)$. As when we constructed rays, this family must subconverge to some ray $\tilde{\gamma} : [0, \infty) \to M$, with $\tilde{\gamma}(0) = p$. A ray coming from such a construction is called an *asymptote* for γ from p (see Figure 9.6). Such asymptotes from p need not be unique.

Proposition 3.13 (1) $b_\gamma(x) \le b_\gamma(p) + b_{\tilde{\gamma}}(x)$.
(2) $b_\gamma(\tilde{\gamma}(t)) = b_\gamma(p) + \tilde{b}_{\tilde{\gamma}}(\tilde{\gamma}(t)) = b_\gamma(p) - t$.

Proof. Let $\sigma_i : [0, \ell_i] \to (M, g)$ be the segments converging to $\tilde{\gamma}$. To check (1), observe that

$$\begin{aligned} d(x, \gamma(s)) - s &\le d(x, \tilde{\gamma}(t)) + d(\tilde{\gamma}(t), \gamma(s)) - s \\ &= d(x, \tilde{\gamma}(t)) - t + d(p, \tilde{\gamma}(t)) + d(\tilde{\gamma}(t), \gamma(s)) - s \\ &\to d(x, \tilde{\gamma}(t)) - t + d(p, \tilde{\gamma}(t)) + b_\gamma(\tilde{\gamma}(t)) \quad \text{as} \quad s \to \infty. \end{aligned}$$

Thus, we see that (1) is true provided that (2) is true. To establish (2), we notice that

$$d(p, \gamma(t_i)) = d(p, \sigma_i(s)) + d(\sigma_i(s), \gamma(ti))$$

for some sequence $t_i \to \infty$. Now, $\sigma_i(s) \to \tilde{\gamma}(s)$, so we obtain

$$\begin{aligned} b_\gamma(p) &= \lim(d(p, \gamma(ti)) - ti) \\ &= \lim(d(p, \tilde{\gamma}(s)) + d(\tilde{\gamma}(s), \gamma(ti)) - ti) \\ &= d(p, \tilde{\gamma}(s)) + \lim(d(\tilde{\gamma}(s), \gamma(ti)) - ti) \\ &= s + b_\gamma(\tilde{\gamma}(s)) \\ &= -b_{\tilde{\gamma}}(\tilde{\gamma}(s)) + b_\gamma(\tilde{\gamma}(s)). \end{aligned}$$

□

Thus, b_γ has $b_\gamma(p) + b_{\tilde{\gamma}}$ as support function from above at $p \in M$.

Lemma 3.14 *If* $\mathrm{Ric}(M, g) \ge 0$, *then* $\Delta b_\gamma \le 0$ *everywhere.*

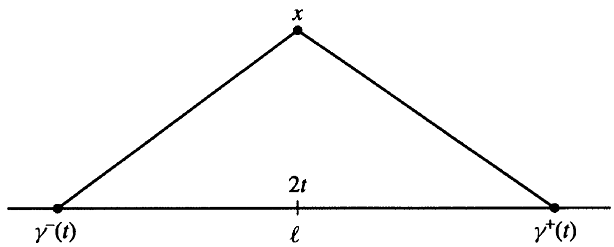


FIGURE 9.7.

Proof. Since $b_\gamma(p) + b_{\tilde{\gamma}}$ is a support function from above at p, we only need to check that $\Delta b_{\tilde{\gamma}}(p) \leq 0$. To see this, observe that the functions $b_t(x) = d(x, \tilde{\gamma}(t)) - t$ are actually support functions from above for $b_{\tilde{\gamma}}$ at p. Furthermore, these functions are smooth at p with $\Delta b_t(p) \leq (n-1)/t \to 0$ as $t \to \infty$. □

Now suppose (M, g) has Ric ≥ 0 and contains a line $\gamma(t) : \mathbb{R} \to M$. Let b^+ be the Busemann function for $\gamma : [0, \infty) \to M$, and b^- the Busemann function for $\gamma : (-\infty, 0] \to M$. Thus,

$$b^+(x) = \lim_{t \to +\infty} (d(x, \gamma(t)) - t),$$
$$b^-(x) = \lim_{t \to +\infty} (d(x, \gamma(-t)) - t).$$

Clearly,

$$b^+(x) + b^-(x) = \lim_{t \to +\infty} (d(x, \gamma(t)) + d(x, \gamma(-t)) - 2t),$$

so $(b^+ + b^-)(x) \geq 0$ for all x, by the triangle inequality, and $(b^+ + b^-)(\gamma(t)) = 0$, since γ is a line (see Figure 9.7).

Thus we have a function $b^+ + b^-$ with $\Delta(b^+ + b^-) \leq 0$ and a global minimum at $\gamma(t)$. The minimum principle then shows that $b^+ + b^- = 0$ everywhere. In particular, $b^+ = -b^-$, and thus $\Delta b^+ = \Delta b^- = 0$ everywhere.

To finish the proof of the splitting theorem, we still need to show that $b^\pm$ are distance functions, i.e., $|\nabla b^\pm| \equiv 1$. To see this, let $p \in M$ and construct asymptotes $\tilde{\gamma}^\pm$ for $\gamma^\pm$ from p. Then consider $b_t^\pm(x) = d(x, \tilde{\gamma}^\pm(t)) - t$, and observe:

$$b_t^+(x) \geq b^+(x) = -b^-(x) \geq -b_t^-(x)$$

with equality holding for $x = p$. Since both $b_t^\pm$ are smooth at p with unit gradient, we must therefore have that $\nabla b_t^+(p) = -\nabla b_t^-(p)$. Then also, $b^\pm$ must be differentiable at p with unit gradient. We have therefore shown (without using that $b^\pm$ are smooth from $\Delta b^\pm = 0$) that $b^\pm$ are everywhere differentiable with unit gradient. We are therefore finished if we believe that $\Delta b^\pm = 0 \Rightarrow b^\pm$ is C^∞. Even without this regularity result we can still prove that $\nabla^2 b^\pm = 0$ in the support sense. The argument is due to Eschenburg and Heintze.

First we prove that the asymptotes $\tilde{\gamma}^\pm(t)$ fit together to form a line. We already know that

$$\nabla b^\pm \left(\tilde{\gamma}^\pm(0)\right) = \nabla b_t^\pm \left(\tilde{\gamma}^\pm(0)\right) = -\frac{d}{dt}\tilde{\gamma}^\pm(0).$$

As $b^+ = -b^-$, we therefore have that the two asymptotes fit together to form a smooth geodesic $\tilde{\gamma}$. This means in particular that the asymptotes are uniquely defined at p. Now fix some $q = \tilde{\gamma}^{\pm}(s)$, say $q = \tilde{\gamma}^+(s)$, and form the asymptote $\hat{\gamma}$ for γ^+ at q. We know that the support functions $\hat{b}_t(x) = d\left(x, \hat{\gamma}(t)\right) - t$ are smooth at q with $\nabla \hat{b}_t(q) = -d/dt\hat{\gamma}(0)$. However, b^+ is also differentiable at q, and since $b_t^+(x)$ ($t > s$) is also a support function at $q = \tilde{\gamma}(s)$, we have that $-d/dt\hat{\gamma}(0) = \nabla \hat{b}_t(q) = \nabla b^+(q) = \nabla b_t^+(\tilde{\gamma}^+(s)) = -d/dt\tilde{\gamma}^+(s)$. Thus $\hat{\gamma}$ and $\tilde{\gamma}^+$ must coincide up to a shift in parameter. This, however, means that the asymptote for γ^- at q must also coincide with the geodesic formed by $\tilde{\gamma}$. In particular, we have that the asymptote is a ray as seen form any point on it and must therefore be a line.

Putting this together, we have that the Busemann functions for this line are equal up to sign. But this implies that the support functions $b_t^{\pm}(x) = d(x, \tilde{\gamma}^{\pm}(t)) - t$ are in fact equal to each other for all $x = \tilde{\gamma}(s)$, $-t < s < t$. To show that $\nabla^2 b^{\pm} = 0$, it therefore suffices to show that $\nabla^2 b_t^{\pm}(\tilde{\gamma}(s)) \to 0$ as $t \to \infty$. As the functions $b_t^{\pm}(x)$ are equal to each other at $x = \tilde{\gamma}(s)$ and otherwise decreasing in t, we have that $\nabla^2 b_T^{\pm}(\tilde{\gamma}(s)) \geq \nabla^2 b_t^{\pm}(\tilde{\gamma}(s))$ for $T < t$. Thus these Hessians are uniformly bounded as $t \to \infty$ and converge to some operators $S^{\pm}(s)$. In fact they converge uniformly to $S^{\pm}(s)$, since we have the equation

$$\nabla_{\frac{d}{ds}\tilde{\gamma}(s)} \nabla^2 b_t^{\pm}(\tilde{\gamma}(s)) + \left(\nabla^2 b_t^{\pm}(\tilde{\gamma}(s))\right)^2 = -R_{\frac{d}{ds}\tilde{\gamma}(s)}.$$

Here the two terms $\left(\nabla^2 b_t^{\pm}(\tilde{\gamma}(s))\right)^2$ and $-R_{d/ds\tilde{\gamma}(s)}$ are bounded operators on any fixed compact interval $[-\varepsilon, \varepsilon]$, and thus the derivatives $\nabla_{d/ds\tilde{\gamma}(s)} \nabla^2 b_t^{\pm}(\tilde{\gamma}(s))$ remain bounded as $t \to \infty$. This insures us that the convergence is uniform on compact intervals and that the limit operators $S^{\pm}(s)$ are continuous in s.

We now need to show that $S^{\pm}(s) = 0$. To see this, first use that

$$\frac{d}{ds}\Delta b_t^{\pm}(\tilde{\gamma}(s)) + \left|\nabla^2 b_t^{\pm}(\tilde{\gamma}(s))\right|^2 \leq 0,$$

which implies

$$\Delta b_t^{\pm}(\tilde{\gamma}(\varepsilon)) - \Delta b_t^{\pm}(\tilde{\gamma}(-\varepsilon)) + \int_{-\varepsilon}^{\varepsilon} \left|\nabla^2 b_t^{\pm}(\tilde{\gamma}(s))\right|^2 ds \leq 0.$$

The first two terms go to zero as $t \to \infty$, so we must have that $\int_{-\varepsilon}^{\varepsilon} \left|S^{\pm}(s)\right|^2 ds = 0$ for all $\varepsilon > 0$. As $\left|S^{\pm}(s)\right|^2$ is continuous, we therefore conclude that $S^{\pm}(s) = 0$ on $[-\varepsilon, \varepsilon]$ for all $\varepsilon > 0$.

9.3.5 Structure Results in Nonnegative Ricci Curvature

Corollary 3.15 $S^3 \times \mathbb{R}$ *does not admit any Ricci flat metrics.*

Proof. Any complete metric on $S^p \times \mathbb{R}$ with nonnegative Ricci curvature must split, since $S^p \times \mathbb{R}$ has two ends. If the original metric is Ricci flat, then after the splitting, we will get a Ricci flat metric on S^p. If $p \leq 3$, such a metric must also be flat. But we know that S^p, $p = 2, 3$ does not admit any flat metrics. □

When $p \geq 4$ it is not known whether S^p admits a Ricci flat metric.

Theorem 3.16 (Structure Theorem for Nonnegative Ricci Curvature, Cheeger-Gromoll, 1971) *Suppose (M, g) is a compact Riemannian manifold with* $\text{Ric} \geq 0$. *Then the universal cover $(\tilde{M}, \tilde{g})$ splits isometrically as a product $N \times \mathbb{R}^p$, where N is a compact manifold.*

Proof. By the splitting theorem, we can write $\tilde{M} = N \times \mathbb{R}^p$, where N does not contain any lines. Observe that if $\gamma(t) = (\gamma_1(t), \gamma_2(t)) \in N \times \mathbb{R}^p$ is a geodesic, then both $\gamma_{1,2}$ are geodesics, and if γ is a line, then both $\gamma_{1,2}$ are also lines. Thus, all lines in $\tilde{M}$ must be of the form $\gamma(t) = (x, \sigma(t))$, where $x \in N$ and σ is a line in $\mathbb{R}^p$.

If N is not compact, then it must contain a ray $\gamma(t) : [0, \infty) \to N$. If $\pi : \tilde{M} \to M$ is the covering map, then we can consider $c(t) = \pi \circ (\gamma(t), 0)$ in M. This is of course a geodesic in M, and since M is compact, there must be a sequence $t_i \to \infty$ such that $\dot{c}(t_i) \to v \in T_xM$ for some $x \in M$, $v \in T_xM$. Choose $\tilde{x} \in \tilde{M}$ such that $\pi(\tilde{x}) = x$, and consider lifts $\gamma_i(t) : [-t_i, \infty) \to \tilde{M}$ of $c(t + t_i)$, where $D\pi(\dot{\gamma}_i(0)) = \dot{c}(t_i)$ and $\gamma_i(0) \to \tilde{x}$. On the one hand, these geodesics converge to a geodesic $\hat{\gamma} : (-\infty, \infty) \to \tilde{M}$ with $\hat{\gamma}(0) = \tilde{x}$. On the other hand, since $D\pi(\dot{\gamma}(t_i)) = \dot{c}(t_i)$, there must be deck transformations $g_i \in \pi_1(M)$ such that $g_i \circ \gamma(t + t_i) = \gamma_i(t)$. Thus, the γ_i are rays and must converge to a line. From our earlier observations, this line must be in $\mathbb{R}^p$. The deck transformations g_i therefore map $\dot{\gamma}(t + t_i)$, which are tangent to N, to vectors that are almost perpendicular to N.

Now let $\varphi : \tilde{M} \to \tilde{M}$ be an isometry, e.g., $\varphi = g_i$. If $\ell(t)$ is a line in $\tilde{M}$, then $\varphi \circ \ell$ must also be a line in $\tilde{M}$. Since all lines in $\tilde{M}$ lie in $\mathbb{R}^p$ and every vector tangent to $\mathbb{R}^p$ is the velocity of some line, we see that $D\varphi$ must preserve $T\mathbb{R}^p \subset T\tilde{M}$. Since TN is the orthogonal complement to $T\mathbb{R}^p$ in $T\tilde{M}$ and $D\varphi$ is a linear isometry, we see that $D\varphi$ also maps TN to TN. This, of course, implies that φ must be of the form $\varphi = (\varphi_1, \varphi_2)$, where φ_1 is an isometry of $\mathbb{R}^p$ and φ_2 an isometry of N. □

This theorem also shows that $\pi_1(M)$ looks almost like the fundamental group of a flat manifold, i.e., it contains a finite normal subgroup of $\pi_1(M)$ such that the quotient group is the fundamental group of a flat manifold.

Another interesting piece of information is the following:

Corollary 3.17 *Suppose (M, g) is a complete, compact Riemannian manifold with* $\text{Ric} \geq 0$. *If M is $K(\pi, 1)$, i.e., the universal cover is contractible, then the universal covering is Euclidean space and (M, g) is a flat manifold.*

Proof. We know that $\widetilde{M} = \mathbb{R}^p \times C$, where C is compact. The only way in which this space can be contractible is if C is contractible. But the only compact manifold that is contractible is the one-point space. □

We can also use the splitting theorem to get different proofs of some of the theorems we proved using the Bochner technique.

Corollary 3.18 *If (M, g) is compact with* Ric ≥ 0 *and has* Ric > 0 *on some tangent space T_pM, then $\pi_1(M)$ is finite.*

Proof. Since Ric > 0 on an entire tangent space, the universal cover cannot split into a product $\mathbb{R}^p \times C$, where $p \geq 1$. Thus, the universal covering is compact. □

Corollary 3.19 *If (M, g) is compact and has* Ric ≥ 0*, then $b_1(M) \leq \dim M = n$, with equality holding iff (M, g) is a flat torus.*

Proof. If $\widehat{M} \to M$ is a finite cover, then we have a map $H^1(M) \to H^1(\widehat{M})$ between the de Rham cohomologies that is injective. (You should prove this.) Thus, $b_1(M) \leq b_1(\widehat{M})$.

Now, $\widetilde{M} = \mathbb{R}^p \times C$. If we let H be the subgroup of $\pi_1(M)$ that acts by translations on $\mathbb{R}^p$, then one can check that $H \simeq \mathbb{Z}^p$. Thus $\widehat{M} = \widetilde{M}/H$ has $b_1 = p \leq n$ and is a finite covering of M. In the case where $b_1(M) = n$, we must then have $p = n$, so M is clearly flat. To see that M is a torus, we must prove that $H = \pi_1(M)$.

First observe that all elements of $\pi_1(M)$ are torsion free, since any isometry of $\mathbb{R}^n$ of finite order must have a fixed point. Second, H is normal in $\pi_1(M)$ with finite index. Thus, the sequence $1 \to H \to \pi_1(M) \to G \to 1$ cannot split, as $\pi_1(M)$ is torsion free. Finally, $H^1(M) = Ab(\pi_1(M)) \otimes \mathbb{R}$, where $Ab(\Gamma) = \Gamma/[\Gamma : \Gamma]$ is the Abelianized group of Γ. Now, the previous remarks show that $Ab(\pi_1(M))$ has rank $\leq n$ and that the rank can be n only if $\pi_1(M)$ is Abelian. Hence, $\pi_1(M) = H$. □

9.4 Further Study

The adventurous reader could consult [43] for further discussions. While this is probably one of the most interesting books in geometry, it has the defect of having many mistakes. An English translation is forthcoming, and it is hoped that it might be more user-friendly. Anderson's article [2] contains the finiteness results for fundamental groups mentioned here and also some interesting examples of manifolds with nonnegative Ricci curvature. For the examples with almost maximal diameter we refer the reader to [3] and [66]. It is also worthwhile to consult the original paper on the splitting theorem [26] and the elementary proof of it in [33]. We already mentioned in Chapter 7 Gallot's contributions to Betti number bounds, and reference [36] works here as well. The reader should also consult the articles by Colding, Perel'man, and Zhu in [46] to get an idea of how rapidly this subject has grown in the past few years.

9.5 Exercises

1. Assume the distance function $f = d(\cdot, p)$ is smooth on $B(p, r)$. Suppose that on vectors w perpendicular to the gradient, the Hessian is being given by
$$\nabla^2 f(w) = \frac{\mathrm{sn}_k'(f)}{\mathrm{sn}_k(f)} w,$$
then all sectional curvatures on $B(p, r)$ are equal to k.

2. Show that if (M, g) has $\mathrm{Ric} \geq (n-1)k$ and for some $p \in M$ we have $\mathrm{vol}B(p, r) = v(n, k, r)$, then the metric has constant curvature k on $B(p, r)$.

3. Show that a complete manifold (M, g) with the property that
$$\mathrm{Ric} \geq 0,$$
$$\lim_{r \to \infty} \frac{\mathrm{vol}B(p, r)}{\omega_n r^n} = 1,$$
for some $p \in M$, must be isometric to Euclidean space.

4. (Cheeger) The relative volume comparison estimate can be generalized as follows: Suppose (M, g) has $\mathrm{Ric} \geq (n-1)k$ and dimension n. Select points $p_1, \ldots, p_k \in M$. Then the function
$$r \to \frac{\mathrm{vol}\left(\bigcup_{i=1}^{k} B(p_i, r)\right)}{v(n, k, r)}$$
is nonincreasing and converges to k as $r \to 0$.
If $A \subset M$, then
$$r \to \frac{\mathrm{vol}\left(\bigcup_{p \in A} B(p, r)\right)}{v(n, k, r)}$$
is nonincreasing. To prove this, use the above with the finite collection of points taken to be very dense in A.

5. The absolute volume comparison can also be slightly generalized. Namely, for $p \in M$ and a subset $\Gamma \subset T_pM$ of unit vectors, consider the cones defined in polar coordinates:
$$B^{\Gamma}(p, r) = \{(t, \theta) \in M : t \leq r \quad \text{and} \quad \theta \in \Gamma\}.$$

If $\text{Ric}M \geq (n-1)k$, show that

$$\text{vol}B^{\Gamma}(p,r) \leq \text{vol}\Gamma \cdot \int_0^r (\text{sn}_k(t))^{n-1}\,dt.$$

6. Let G be a compact Lie group with a bi-invariant metric. Use the previous exercises on Lie groups from Chapters 1, 2, and 5 and the results from this chapter to prove

   (a) If G has finite center, then G has finite fundamental group.

   (b) A finite covering of G looks like $G' \times T^k$, where G' is compact simply connected, and T^k is a torus.

   (c) If G has finite fundamental group, then the center is finite.

7. Show that a compact Riemannian manifold with irreducible holonomy and $\text{Ric} \geq 0$ has finite fundamental group.

8. Let (M, g) be an n-dimensional Riemannian manifold that is isometric to Euclidean space outside some compact subset $K \subset M$, i.e., $M - K$ is isometric to $\mathbb{R}^n - C$ for some compact set $C \subset \mathbb{R}^n$. If $\text{Ric}_g \geq 0$, show that $M = \mathbb{R}^n$. (In Chapter 7 we gave two different hints for this problem; here is a third. Use the splitting theorem.)

10
Convergence

In this chapter we will give an introduction to several of the convergence ideas for Riemannian manifolds. The goal is to understand what it means for a sequence of Riemannian manifolds, or more generally metric spaces, to converge to a space. In the first section we develop the weakest convergence concept: Gromov-Hausdorff convergence. We then go on to explain some of the elliptic regularity theory we need for some of the later developments. We have confined ourselves to the simpler Hölder and Schauder theories. In Section 3 we develop the idea of norms of Riemannian manifolds. This is a concept developed by the author in the hope that it will make it easier to understand convergence theory as a parallel to the easier Hölder theory for functions (as is explained in Section 2). At the same time, we also feel that it has made some parts of the theory more concise. In this section we examine some stronger convergence ideas that were developed by Cheeger and Gromov and study their relation to the norms of manifolds. These preliminary discussions will enable us in subsequent sections to establish the convergence theorem of Riemannian geometry and its generalizations by Anderson and others. These convergence theorems contain the Cheeger finiteness theorem, which states that certain classes of Riemannian manifolds contain only finitely many diffeomorphism types.

The idea of measuring the distance between subspaces of a given space goes back to Hausdorff and was extensively studied in the Polish and Russian schools of topology. The more abstract versions we use here seem to begin with Shikata's proof of the differentiable sphere theorem. In Cheeger's thesis, the idea that abstract manifolds can converge to each other is also evident. In fact, as we shall see below, he proved his finiteness theorem by showing that certain classes of manifolds are precompact in various topologies. After these two early forays into convergence theory, not much appears until Gromov bombarded the mathematical community with his highly original approaches. He introduced a very weak kind of convergence that is simply an abstract version of Hausdorff's early work. The first use of this new idea was to prove a group-theoretic question about the nilpotency of groups with polynomial growth. Soon after the introduction of this weak convergence, the earlier ideas on strong convergence by Cheeger resurfaced. There are various conflicting accounts on who did what and when. Certainly, the Russian school, notably Nikolaev and Berestovskii, deserve a lot of credit for their work on synthetic geometry, which could have been used in the convergence context. However, it appears that they were concerned mostly with studying generalized metrics rather than convergence. By contrast, the western school studied convergence and thereby developed an appreciation for studying Riemannian manifolds with little regularity, and even metric spaces.

10.1 Gromov-Hausdorff Convergence

10.1.1 Hausdorff Versus Gromov Convergence

At the beginning of the twentieth century, Hausdorff introduced what we call the *Hausdorff distance* between subsets of a metric space. If (X, d) is the metric space and $A, B \subset X$, then we define

$$\begin{aligned} d(A, B) &= \inf\{d(a, b) : a \in A, b \in B\}, \\ B(A, \varepsilon) &= \{x \in X : d(x, A) < \varepsilon\}, \\ d_H(A, B) &= \inf\{\varepsilon : A \subset B(B, \varepsilon), B \subset B(A, \varepsilon)\}. \end{aligned}$$

Thus, $d(A, B)$ is small if some points in these sets are close, while the Hausdorff distance $d_H(A, B)$ is small iff every point of A is close to a point in B and vice versa. One can easily see that the Hausdorff distance defines a metric on the closed subsets of X and that this collection is compact when X is compact.

We shall concern ourselves only with compact metric spaces and *proper* metric spaces. The latter have by definition proper distance functions, i.e., all closed balls are compact. This implies, in particular, that the spaces are separable, complete, and locally compact.

Around 1980, Gromov extended this concept to that of a distance between two abstract metric spaces. If X and Y are metric spaces, then an *admissible* metric on the disjoint union $X \amalg Y$ is a metric that extends the given metrics on X and Y. With this we can define the *Gromov-Hausdorff distance* as

$$d_{G-H}(X, Y) = \inf\{d_H(X, Y) : \text{admissible metrics on } X \amalg Y\}.$$

Thus, we try to put a metric on $X \amalg Y$ such that X and Y are as close as possible in the Hausdorff distance, with the constraint that the extended metric equals the given metrics on X and Y. In other words, we are trying to define distances between points in X and Y without violating the triangle inequality.

Example 1.1 If Y is the one-point space, then

$$\begin{aligned} d_{G-H}(X, Y) &= \operatorname{rad} X \\ &= \inf_{y \in X} \sup_{x \in X} d(x, y) \\ &= \text{radius of smallest ball covering } X. \end{aligned}$$

Example 1.2 By defining $d(x, y) = D/2$, where $\operatorname{diam} X, \operatorname{diam} Y \leq D$, we see that

$$d_{G-H}(X, Y) \leq D/2.$$

Let $(\mathcal{M}, d_{G-H})$ denote the collection of compact metric spaces. We shall study this class as a metric space in its own right. To justify this we must show that only isometric spaces are within distance zero of each other.

Lemma 1.3 *Suppose X and Y are complete metric spaces with $d_{G-H}(X, Y) = 0$. Then X and Y are isometric.*

Proof. Choose a sequence of metrics d_i on $X \amalg Y$ such that the Hausdorff distance between X and Y in this metric is $< i^{-1}$. Then we can find (possibly discontinuous) maps

$$\begin{aligned} I_i &: X \to Y, \quad \text{where} \quad d_i(x, I_i(x)) \leq i^{-1}, \\ J_i &: Y \to X, \quad \text{where} \quad d_i(y, J_i(y)) \leq i^{-1}. \end{aligned}$$

Using the triangle inequality and that d_i restricted to either X or Y is the given metric d on these spaces yields

$$\begin{aligned} d\left(I_i(x), I_j(x)\right) &\leq 2i^{-1}, \\ d\left(J_i(y), J_j(y)\right) &\leq 2i^{-1}, \\ d(I_i(x_1), I_i(x_2)) &\leq 2i^{-1} + d(x_1, x_2), \\ d(J_i(y_1), J_i(y_2)) &\leq 2i^{-1} + d(y_1, y_2), \\ d(x, J_i \circ I_i(x)) &\leq 2i^{-1}, \\ d(y, I_i \circ J_i(y)) &\leq 2i^{-1}. \end{aligned}$$

The first two inequalities show that $I_i : X \to Y$ and $J_i : Y \to X$ converge to maps $I : X \to Y$ and $J : Y \to X$. The next two inequalities state that these two maps are distance nonincreasing. The last two inequalities imply that the two maps are inverses to each other. It then easily follows that both maps are isometries that are inverses of each other. □

Both symmetry and the triangle inequality are easily established for d_{G-H}. Thus, $(\mathcal{M}, d_{G-H})$ is a pseudometric space, and if we divide out by isometric spaces, we get a metric space. We shall show that this metric space is both complete and separable. But first we show how spaces can be approximated by finite metric spaces.

Example 1.4 Let X be compact and suppose we have a finite subset $A \subset X$ such that every point in X is within distance ε of some element in A, i.e., $d_H(A, X) \leq \varepsilon$. Such sets A are called *ε-dense* in X. It is then clear that if we use the metric on A induced by X, then also $d_{G-H}(X, A) \leq \varepsilon$. The importance of this remark is that for any $\varepsilon > 0$ we can in fact find such finite subsets of X, since X is compact.

Example 1.5 Suppose we have ε-dense subsets

$$\begin{aligned} A &= \{x_1, \dots, x_k\} \subset X, \\ B &= \{y_1, \dots, y_k\} \subset Y, \end{aligned}$$

with the further property that

$$\left| d\left(x_i, x_j\right) - d\left(y_i, y_j\right) \right| \leq \varepsilon, \ 1 \leq i, j \leq k.$$

Then $d_{G-H}(X, Y) \leq 3\varepsilon$. We already have that the finite subsets are ε-close to the spaces, so by the triangle inequality it suffices to show that $d_{G-H}(A, B) \leq \varepsilon$. For this we must exhibit a metric d on $A \amalg B$ that makes A and B ε-Hausdorff close. Define

$$\begin{aligned} d(x_i, y_i) &= \varepsilon, \\ d\left(x_i, y_j\right) &= \min_k \left\{ d(x_i, x_k) + \varepsilon + d\left(y_j, y_k\right) \right\}. \end{aligned}$$

Thus, we have extended the given metrics on A and B in such a way that no points from A and B get identified, and in addition the potential metric is symmetric. It then remains to check the triangle inequality. Here we must show

$$\begin{aligned} d\left(x_i, y_j\right) &\leq d(x_i, z) + d\left(y_j, z\right), \\ d\left(x_i, x_j\right) &\leq d(y_k, x_i) + d\left(y_k, x_j\right), \\ d\left(y_i, y_j\right) &\leq d\left(x_k, y_j\right) + d\left(x_k, y_j\right). \end{aligned}$$

It suffices to check the first two cases. In the first one we can assume that $z = x_k$. Then we can find l such that

$$d\left(y_j, x_k\right) = \varepsilon + d\left(y_j, y_l\right) + d(x_l, x_k).$$

Hence,

$$\begin{aligned} d(x_i, x_k) + d\left(y_j, x_k\right) &= d(x_i, x_k) + \varepsilon + d\left(y_j, y_l\right) + d(x_l, x_k) \\ &\geq d(x_i, x_l) + \varepsilon + d\left(y_j, y_l\right) \\ &\geq d\left(x_i, y_j\right). \end{aligned}$$

For the second case select l, m with

$$d(y_k, x_i) = d(y_k, y_l) + \varepsilon + d(x_l, x_i),$$
$$d(y_k, x_j) = d(y_k, y_m) + \varepsilon + d(x_m, x_j).$$

Then, using our assumption about the comparability of the metrics on A and B, we have

$$\begin{aligned} d(y_k, x_i) + d(y_k, x_j) &= d(y_k, y_l) + \varepsilon + d(x_l, x_i) + d(y_k, y_m) + \varepsilon + d(x_m, x_j) \\ &\geq d(x_k, x_l) + d(x_l, x_i) + d(x_k, x_m) + d(x_m, x_j) \\ &\geq d(x_i, x_j). \end{aligned}$$

Example 1.6 Suppose $M_k = S^3/\mathbb{Z}_k$ with the usual metric induced from $S^3(1)$. Then we have a Riemannian submersion $M_k \to S^2\left(\frac{1}{2}\right)$ whose fibers have diameter $\frac{2\pi}{k} \to 0$ as $k \to \infty$. Using the previous example, we can therefore easily check that $M_k \to S^2\left(\frac{1}{2}\right)$ in the Gromov-Hausdorff topology.

One can similarly see that the Berger metrics $\left(S^3, g_\varepsilon\right) \to S^2\left(\frac{1}{2}\right)$ as $\varepsilon \to 0$. Notice that in both cases the volume goes to zero, but the curvatures and diameters are uniformly bounded. In the second case the manifolds are even simply connected. It should also be noted that the topology changes rather drastically from the sequence to the limit, and in the first case the elements of the sequence even have mutually different fundamental groups.

Proposition 1.7 *The "metric space" $(\mathcal{M}, d_{G-H})$ is separable and complete.*

Proof. To see that it is separable, first observe that the collection of all finite metric spaces is dense in this collection. Now take the "countable" collection of all finite metric spaces that in addition have the property that all distances are rational. Clearly, this collection is dense as well.

To show completeness, select a Cauchy sequence $\{X_n\}$. To show convergence of this sequence, it suffices to check that some subsequence is convergent. Select a subsequence $\{X_i\}$ such that $d_{G-H}(X_i, X_{i+1}) < 2^{-i}$ for all i. Then select metrics $d_{i,i+1}$ on $X_i \amalg X_{i+1}$ making these spaces 2^{-i}-Hausdorff close. Now define a metric $d_{i,i+j}$on $X_i \amalg X_{i+j}$ by

$$d_{i,i+j}\left(x_i, x_{i+j}\right) = \min_{\{x_{i+k} \in X_{i+k}\}} \left\{\sum_{k=0}^{j-1} d\left(x_{i+k}, x_{i+k+1}\right)\right\}.$$

We have then defined a metric d on $Y = \amalg_i X_i$ with the property that in this metric $d_H\left(X_i, X_{i+j}\right) \leq 2^{-i+1}$. This metric space is not complete, but the "boundary" of the completion is exactly our desired limit space. To define it, first consider

$$\hat{X} = \left\{\{x_i\} : x_i \in X_i \text{ and } d\left(x_i, x_j\right) \to 0 \text{ as } i, j \to \infty\right\}.$$

This space has a pseudometric defined by

$$d(\{x_i\}, \{y_i\}) = \lim_{i \to \infty} d(x_i, y_i).$$

Given that we are only considering Cauchy sequences $\{x_i\}$, this must yield a metric on the quotient space X, obtained by the equivalence relation

$$\{x_i\} \sim \{y_i\} \quad \text{iff} \quad d(\{x_i\}, \{y_i\}) = 0.$$

Now we can extend the metric on Y to one on $X \amalg Y$ by declaring

$$d(x_k, \{x_i\}) = \lim_{i\to\infty} d(x_k, x_i).$$

Using that $d_H(X_j, X_{j+1}) \leq 2^{-j}$, we can for any $x_i \in X_i$ find a sequence $\{x_{i+j}\}$ such that $x_{i+0} = x_i$ and $d(x_{i+j}, x_{i+j+1}) \leq 2^{-j}$. Then we must have $d(x_i, \{x_{i+j}\}) \leq 2^{-i+1}$. Thus, every X_i is 2^{-i+1}-close to the limit space X. Conversely, for any given sequence $\{x_i\}$ we can find an equivalent sequence $\{y_i\}$ with the property that $d(y_k, \{y_i\}) \leq 2^{-k+1}$ for all k. Thus, X is 2^{-i+1}-close to X_i. □

From the proof of this theorem we get the useful information that Gromov-Hausdorff convergence can always be thought of as Hausdorff convergence. In other words, if we know that $X_i \to X$ in the Gromov-Hausdorff sense, then after possibly passing to a subsequence, we can assume that there is a metric on $X \amalg (\amalg_i X_i)$ in which the X_i Hausdorff converge to X. With such a selection of a metric, it then makes sense to say that $x_i \to x$, where $x_i \in X_i$ and $x \in X$. We shall often use this without explicitly mentioning the ambient metric on $X \amalg (\amalg_i X_i)$.

Before going any further, we should mention an equivalent way of picturing convergence. For a metric space X, let $C(X)$ denote the continuous functions on X, and $L^\infty(X)$ the bounded measurable functions with the sup-norm (not the essential sup-norm). We know that $L^\infty(X)$ is a Banach space. When X is bounded, we construct a map $X \to L^\infty(X)$, by sending x to the continuous function $d(x, \cdot)$. This is usually called the *Kuratowski embedding* when we consider it as a map into $C(X)$. From the triangle inequality, we can easily see that this is in fact a distance-preserving map. Thus, any compact metric space is isometric to a subset of some Banach space $L^\infty(X)$. The important observation now is that two such spaces $L^\infty(X)$ and $L^\infty(Y)$ are isometric if the spaces X and Y are Borel equivalent (there exists a measurable bijection). Also, if $X \subset Y$, then $L^\infty(X)$ sits isometrically as a linear subspace of $L^\infty(Y)$. Now recall that any compact space is Borel equivalent to some subset of $[0, 1]$. Thus all compact metric spaces X are isometric to some subset of $L^\infty([0, 1])$. We can then define

$$d_{G-H}(X, Y) = \inf d_H(i(X), j(Y)),$$

where $i : X \to L^\infty([0, 1])$ and $j : Y \to L^\infty([0, 1])$ are distance-preserving maps. The completeness issue then becomes a little less abstract to deal with.

10.1.2 Pointed Convergence

So far, we haven't really dealt with noncompact spaces in a serious way. There is, of course, nothing wrong with defining the Gromov-Hausdorff distance between unbounded spaces, but it will almost never be finite. In order to change this, we should have in mind what is done for convergence of functions on unbounded domains. There, one usually speaks about convergence on compact subsets. To do something similar, we first define the pointed Gromov-Hausdorff distance

$$d_{G-H}((X,x),(Y,y)) = \inf\{d_H(X,Y) + d(x,y)\}.$$

Here we take as usual the infimum over all Hausdorff distances and in addition require the selected points to be close. The above results are still true for this modified distance. We can then introduce the Gromov-Hausdorff topology on the collection of proper pointed metric spaces $\mathcal{M}_* = \{(X,x,d)\}$ in the following way: We say that $(X_i, x_i, d_i) \to (X, x, d)$ in the *pointed Gromov-Hausdorff topology* if for all R, the closed metric balls $\left(\bar{B}(x_i, R), x_i, d_i\right) \to \left(\bar{B}(x, R), x, d\right)$ converge with respect to the pointed Gromov-Hausdorff metric.

10.1.3 Convergence of Maps

We shall also have recourse to speak about *convergence of maps*. Suppose we have

$$\begin{aligned} f_k &: X_k \to Y_k, \\ X_k &\to X, \\ Y_k &\to Y. \end{aligned}$$

Then we say that f_k converges to $f : X \to Y$ if for every sequence $x_k \in X_x$ converging to $x \in X$ we have that $f_k(x_k) \to f(x)$. This definition obviously depends in some sort of way on having the spaces converge in the Hausdorff sense, but we shall ignore this. It is really a very strong kind of convergence for if we assume that $X_k = X$, $Y_k = Y$, and $f_k = f$, then f can converge to itself only if it is continuous.

Note also that convergence of functions preserves such properties as being distance preserving or submetries.

Another useful observation is that we can regard the sequence of maps f_k as one continuous map $F : (\amalg_i X_i) \to Y \amalg (\amalg_i Y_i)$. The sequence converges iff this map has an extension $X \amalg (\amalg_i X_i) \to Y \amalg (\amalg_i Y_i)$, in which case the limit map is the restriction to X. Thus, a sequence is convergent iff the map $F : (\amalg_i X_i) \to Y \amalg (\amalg_i Y_i)$ is uniformly continuous.

A sequence of functions as above is called *equicontinuous,* if for every $\varepsilon > 0$ there is an $\delta > 0$ such that $f_k(B(x_k, \delta)) \subset B(f_k(x_k), \varepsilon)$ for all k and $x_k \in X_k$. A sequence is therefore equicontinuous if, for example, all the functions are Lipschitz continuous with the same constant. As for standard equicontinuous sequences, we have the Arzela-Ascoli lemma:

Lemma 1.8 *An equicontinuous family $f_k : X_k \to Y_k$, where $X_k \to X$, and $Y_k \to Y$ in the (pointed) Gromov-Hausdorff topology, has a convergent subsequence. When the spaces are not compact, we also assume that f_k preserves the base point.*

Proof. The standard proof carries over without much change. Namely, first choose dense subsets $A_i = \{a_1^i, a_2^i, \ldots\} \subset X_i$ such that the sequences $\{a_j^i\} \to a_j \in X$. Then also, $A = \{a_j\} \subset X$ is dense. Next, use a diagonal argument to find a subsequence of functions that converge on the above sequences. Finally, show that this sequence converges as promised. □

10.1.4 Compactness of Classes of Metric Spaces

We now turn our attention to convergence of spaces. Namely, we need some good criteria for when a collection of (pointed) spaces is precompact (i.e., closure is compact).

For a compact metric space X, define

$$\begin{aligned} \mathrm{Cap}(\varepsilon) = \mathrm{Cap}_X(\varepsilon) &= \text{ maximum number of disjoint } \frac{\varepsilon}{2} - \text{balls in } X, \\ \mathrm{Cov}(\varepsilon) = \mathrm{Cov}_X(\varepsilon) &= \text{ minimum number of } \varepsilon - \text{balls it takes to cover } X. \end{aligned}$$

First, we should observe that $\mathrm{Cov}(\varepsilon) \le \mathrm{Cap}(\varepsilon)$. For if the balls $B(x_i, \varepsilon/2)$ are disjoint, then the collection $B(x_i, \varepsilon)$ must cover. Otherwise, there would be some $x \in X - \cup B(x_i, \varepsilon)$, but this would imply that $B(x, \varepsilon/2)$ is disjoint from all of the balls $B(x_i, \varepsilon/2)$, thus violating maximality.

Another important observation is that if two compact metric spaces X and Y satisfy $d_{G-H}(X, Y) < \delta$, then it follows from the triangle inequality that:

$$\begin{aligned} \mathrm{Cov}_X(\varepsilon + 2\delta) &\le \mathrm{Cov}_Y(\varepsilon), \\ \mathrm{Cap}_X(\varepsilon) &\ge \mathrm{Cap}_Y(\varepsilon + 2\delta). \end{aligned}$$

With this information we can now characterize precompact classes of compact metric spaces.

Lemma 1.9 (M. Gromov, 1980) *For a class $\mathcal{C} \subset (\mathcal{M}, d_{G-H})$, the following statements are equivalent:*

(1) *$\mathcal{C}$ is precompact, i.e., every sequence in $\mathcal{C}$ has a subsequence that is convergent in $(\mathcal{M}, d_{G-H})$.*

(2) *There is a function $N(\varepsilon) : (0, \alpha) \to (0, \infty)$ such that $\mathrm{Cap}_X(\varepsilon) \le N(\varepsilon)$ for all $X \in \mathcal{C}$.*

(3) *There is a function $N(\varepsilon) : (0, \alpha) \to (0, \infty)$ such that $\mathrm{Cov}_X(\varepsilon) \le N(\varepsilon)$ for all $X \in \mathcal{C}$.*

Proof. (1) $\Rightarrow$ (2): If $\mathcal{C}$ is precompact, then for every $\varepsilon > 0$ we can find $X_1, \ldots, X_k \in \mathcal{C}$ such that for any $X \in \mathcal{C}$ we have that $d_{G-H}(X, X_i) < \varepsilon/4$ for some i. Then, of course, $\mathrm{Cap}_X(\varepsilon) \le \mathrm{Cap}_{X_i}(\varepsilon/2) \le \max_i \mathrm{Cap}_{X_i}(\varepsilon/2)$. Thus we find a bound for $\mathrm{Cap}_X(\varepsilon)$ for each $\varepsilon > 0$.

(2) $\Rightarrow$ (3) is obvious.

(3) $\Rightarrow$ (1): It suffices to show that for each $\varepsilon > 0$ we can find finitely many metric spaces $X_1, \ldots, X_k \in \mathcal{M}$ such that any metric space in $\mathcal{C}$ is within ε of some X_i in the Gromov-Hausdorff metric. Since $\mathrm{Cov}_X(\varepsilon/2) \leq N(\varepsilon/2)$, we know that any $X \in \mathcal{C}$ is within $\varepsilon/2$ of some finite metric space, with at most $N(\varepsilon/2)$ elements in it. Next, observe that $\mathrm{diam}X \leq 2\delta \mathrm{Cov}_X(\delta)$ for any fixed δ, so we can assume that these finite metric spaces have no distances that are bigger than $\varepsilon N(\varepsilon/2)$. The metric on such a finite metric space then consists of a matrix (d_{ij}), $1 \leq i, j \leq N(\varepsilon/2)$, where each entry satisfies $d_{ij} \in [0, \varepsilon N(\varepsilon/2)]$. From among all such finite metric spaces it is then possible to select a finite number of them such that any of the matrices (d_{ij}) is within $\varepsilon/2$ of one matrix from the finite selection of matrices. This means that the spaces are within $\frac{\varepsilon}{2}$ of each other. We have then found the desired finite collection of metric spaces. □

As a corollary we can also get a precompactness theorem in the pointed category.

Corollary 1.10 *A collection $\mathcal{C} \subset \mathcal{M}_*$ is precompact iff for each $R > 0$ the collection $\{B(x, R) : B(x, R) \subset (X, x) \in \mathcal{C}\} \subset (\mathcal{M}, d_{G-H})$ is precompact.*

Using the relative volume comparison theorem we can then show

Corollary 1.11 *For any integer $n \geq 2$, $k \in \mathbb{R}$, and $D > 0$ we have that the following classes are precompact:*

(1) *The collection of closed Riemannian n-manifolds with* $\mathrm{Ric} \geq (n-1)k$ *and* $\mathrm{diam} \leq D$.

(2) *The collection of pointed complete Riemannian n-manifolds with* $\mathrm{Ric} \geq (n-1)k$.

Proof. It suffices to prove (2). So fix $R > 0$. Then we have to show that there can't be too many disjoint balls inside $B(x, R) \subset M$. To see this, suppose $B(x_1, \varepsilon), \ldots, B(x_\ell, \varepsilon) \subset B(x, R)$ are disjoint. If $B(x_i, \varepsilon)$ is the ball with the smallest volume, we have

$$\ell \leq \frac{\mathrm{vol}B(x, R)}{\mathrm{vol}B(x_i, \varepsilon)} \leq \frac{\mathrm{vol}B(x_i, 2R)}{\mathrm{vol}(x_i, \varepsilon)} \leq \frac{v(n, k, 2R)}{v(n, k, \varepsilon)}.$$

This gives the desired bound. □

It seems intuitively clear that an n-dimensional space should have $\mathrm{Cov}(\varepsilon) \sim \varepsilon^{-n}$ as $\varepsilon \to 0$. In fact, one could define the Hausdorff dimension of a metric space as

$$\dim X = \limsup_{\varepsilon \to 0} \frac{\log \mathrm{Cov}(\varepsilon)}{-\log \varepsilon}.$$

This definition will then give the right answer for Riemannian manifolds. Some fractal spaces might, however, have nonintegral dimension. Now observe that

$$\frac{v(n,k,2R)}{v(n,k,\varepsilon)} \sim \varepsilon^{-n}.$$

Therefore, if we can show that covering functions carry over to limit spaces, then we will have shown that manifolds with lower curvature bounds can only collapse in dimension. To this end we have

Lemma 1.12 *Let $C(N(\varepsilon))$ be the collection of metric spaces with* $\mathrm{Cov}(\varepsilon) \le N(\varepsilon)$. *Suppose N is continuous. Then $C(N(\varepsilon))$ is compact.*

Proof. We already know that this class is precompact. So we only have to show that if $X_i \to X$ and $\mathrm{Cov}_{X_i}(\varepsilon) \le N(\varepsilon)$, then also $\mathrm{Cov}_X(\varepsilon) \le N(\varepsilon)$. This follows easily from

$$\mathrm{Cov}_X(\varepsilon) \le \mathrm{Cov}_{X_i}(\varepsilon - 2d_{G-H}(X, X_i)) \le N(\varepsilon - 2d_{G-H}(X, X_i)),$$

since $N(\varepsilon - 2d_{G-H}(X, X_i)) \to N(\varepsilon)$ as $i \to \infty$. □

10.2 Hölder Spaces and Schauder Estimates

First, we shall define the Hölder norms and Hölder spaces. We will then briefly discuss the necessary estimates we need for elliptic operators for later applications. The standard reference for all the material here is the classic by Courant and Hilbert [28], especially Chapter IV, and the thorough text [40], especially Chapters 1 through 6. A more modern text that also explains how PDE's are used in geometry, including some of the facts we need, is [79], especially vol. III.

10.2.1 Hölder Spaces

Let us fix a bounded domain $\Omega \subset \mathbb{R}^n$. The continuous functions from Ω to $\mathbb{R}^k$ are denoted by $C^0(\Omega, \mathbb{R}^k)$, and we use the sup-norm, denoted by

$$\|u\|_{C^0} = \sup_{x\in\Omega} |u(x)|,$$

on this space. This makes $C^0(\Omega, \mathbb{R}^k)$ into a Banach space. We wish to generalize this so that we still have a Banach space, but in addition also take into account derivatives of the functions. The first natural thing to do is to define $C^m(\Omega, \mathbb{R}^k)$ as the functions with m continuous partial derivatives. Using multi-index notation, we define

$$\partial^i u = \partial_1^{i_1} \cdots \partial_n^{i_n} u = \frac{\partial^l u}{\partial (x^1)^{i_1} \cdots \partial (x^n)^{i_n}},$$

where $i = (i_1, \ldots, i_n)$ and $l = |i| = i_1 + \cdots + i_n$. Then the C^m-norm is

$$\|u\|_{C^m} = \sup_{x\in\Omega} |u(x)| + \sum_{|i|\le m} \sup_{\Omega} \left|\partial^i u\right|.$$

This norm does result in a Banach space, but the inclusions $C^m\left(\Omega, \mathbb{R}^k\right) \subset C^{m-1}\left(\Omega, \mathbb{R}^k\right)$ do not yield closed subspaces. For instance, $f(x) = |x|$ is in the closure of $C^1([-1,1], \mathbb{R}) \subset C^0([-1,1], \mathbb{R})$.

To accommodate this problem, we define for each $\alpha \in (0, 1]$ the C^α-pseudonorm of $u : \Omega \to \mathbb{R}^k$ as

$$\|u\|_\alpha = \sup_{x,y\in\Omega} \frac{|u(x) - u(y)|}{|x - y|^\alpha}.$$

When $\alpha = 1$, this gives the best Lipschitz constant for u. The important information about this norm is that any family of functions with $\|u\|_\alpha \le K$ is equicontinuous and therefore precompact by the Arzela-Ascoli lemma.

Define the *Hölder space* $C^{m,\alpha}\left(\Omega, \mathbb{R}^k\right)$ as being the functions in $C^m\left(\Omega, \mathbb{R}^k\right)$ such that all mth-order partial derivatives have finite C^α-pseudonorm. On this space we use the norm

$$\|u\|_{C^{m,\alpha}} = \|u\|_{C^m} + \sum_{|i|=m} \left\|\partial^i u\right\|_\alpha.$$

If we wish to be specific about the domain on which we take these norms, then we write

$$\|u\|_{C^{m,\alpha},\Omega}.$$

We can now show

Lemma 2.1 *$C^{m,\alpha}\left(\Omega, \mathbb{R}^k\right)$ is a Banach space with the $C^{m,\alpha}$-norm. Furthermore, the inclusion $C^{m,\alpha}\left(\Omega, \mathbb{R}^k\right) \subset C^{m,\beta}\left(\Omega, \mathbb{R}^k\right)$, where $\beta < \alpha$ is always compact, i.e., it maps closed bounded sets to compact sets.*

Proof. We only need to show this in the case where $m = 0$; the more general case is then a fairly immediate consequence.

First, we must show that any Cauchy sequence $\{u_i\}$ in $C^\alpha\left(\Omega, \mathbb{R}^k\right)$ converges. From the Arzela-Ascoli lemma we get that some sequence is convergent in the C^0-norm. Thus, we have that $u_i \to u \in C^0$ in the C^0-norm. Now observe that

$$\frac{|u_i(x) - u_i(y)|}{|x - y|^\alpha} \to \frac{|u(x) - u(y)|}{|x - y|^\alpha}.$$

As the left-hand side is uniformly bounded, we also get that the right-hand side is bounded, thus showing that $u \in C^\alpha$.

Now for the last statement. We know that the inclusion $C^\alpha\left(\Omega, \mathbb{R}^k\right) \subset C\left(\Omega, \mathbb{R}^k\right)$ is compact. We then use

$$\frac{|u(x) - u(y)|}{|x - y|^\beta} = \left(\frac{|u(x) - u(y)|}{|x - y|^\alpha}\right)^{\beta/\alpha} \cdot |u(x) - u(y)|^{1-\beta/\alpha}$$

to conclude that

$$\|u\|_\beta \leq (\|u\|_\alpha)^{\beta/\alpha} \cdot (2 \cdot \|u\|_{C^0})^{1-\beta/\alpha}.$$

If a sequence converges in C^0 and is bounded in C^α, we therefore also have that it converges in C^β, as long as $\beta < \alpha \leq 1$. □

10.2.2 Elliptic Estimates

We now turn our attention to *elliptic operators*. We shall consider equations of the form

$$Lu = a^{ij}\partial_i\partial_j u + b^i\partial_i u = f,$$

where $a^{ij} = a^{ji}$. The operator is called *elliptic* if the matrix (a^{ij}) is positive definite. We shall throughout assume that all eigenvalues for (a^{ij}) lie in $[\lambda, \lambda^{-1}]$, $\lambda > 0$,

$$\left\|a^{ij}\right\|_\alpha \leq \lambda^{-1},$$
$$\left\|b^i\right\|_\alpha \leq \lambda^{-1}.$$

Let us state without proof the a priori estimates, usually called the *Schauder estimates,* or *elliptic estimates*, that we shall need.

Theorem 2.2 *Let $\Omega \subset \mathbb{R}^n$ be an open domain of diameter $\leq D$ and $K \subset \Omega$ a subdomain such that $d(K, \partial\Omega) \geq \delta$. Moreover assume $\alpha \in (0, 1)$; then there is a constant $C = C(n, \alpha, \lambda, \delta, D)$ such that*

$$\|u\|_{C^{2,\alpha},K} \leq C\left(\|Lu\|_{C^\alpha,\Omega} + \|u\|_{C^\alpha,\Omega}\right),$$
$$\|u\|_{C^{1,\alpha},K} \leq C\left(\|Lu\|_{C^0,\Omega} + \|u\|_{C^\alpha,\Omega}\right).$$

Furthermore, if Ω has smooth boundary and $u = \varphi$ on $\partial\Omega$, then there is a constant $C = C(n, \alpha, \lambda, D)$ such that on all of Ω we have

$$\|u\|_{C^{2,\alpha},\Omega} \leq C\left(\|Lu\|_{C^\alpha,\Omega} + \|\varphi\|_{C^{2,\alpha},\partial\Omega}\right).$$

One way of proving these results is to establish them first for the simplest operator: $Lu = \Delta u = \delta^{ij}\partial_i\partial_j u$. Then observe that a linear change of coordinates shows that we can handle operators with constant coefficients: $Lu = \Delta u = a^{ij}\partial_i\partial_j u$. Finally, Schauder's trick is that the assumptions about the functions a^{ij} imply that they are almost constant locally. A partition of unity argument then finishes the analysis.

The first-order term doesn't cause much trouble and can even be swept under the rug in the case where the operator is in divergence form:

$$Lu = a^{ij}\partial_i\partial_j u + b^i\partial_i u = \partial_i\left(a^{ij}\partial_j u\right).$$

Such operators are particularly nice when one wishes to use integration by parts, as we have

$$\int_\Omega \left(\partial_i\left(a^{ij}\partial_j u\right)\right) h = -\int_\Omega a^{ij}\partial_j u \partial_i h$$

when $h = 0$ on $\partial\Omega$. This is interesting in the context of geometric operators, for if we use the Laplacian on manifolds, then in local coordinates it will look like

$$\begin{aligned} Lu &= \Delta_g u \\ &= \frac{1}{\sqrt{\det g_{ij}}} \partial_i \left(\sqrt{\det g_{ij}} \cdot g^{ij} \cdot \partial_j u \right). \end{aligned}$$

The above theorem has an almost immediate corollary.

Corollary 2.3 *If in addition we assume that* $\|a^{ij}\|_{C^{m,\alpha}}, \|b^i\|_{C^{m,\alpha}} \le \lambda^{-1}$, *then there is a constant* $C = C(n, m, \alpha, \lambda, \delta, D)$ *such that*

$$\|u\|_{C^{m+2,\alpha},K} \le C\left(\|Lu\|_{C^{m,\alpha},\Omega} + \|u\|_{C^\alpha,\Omega}\right).$$

And on a domain with smooth boundary,

$$\|u\|_{C^{m+2,\alpha},\Omega} \le C\left(\|Lu\|_{C^{m,\alpha},\Omega} + \|\varphi\|_{C^{m+2,\alpha},\Omega}\right).$$

The Schauder estimates can be used to show that the Dirichlet problem always has a unique solution.

Theorem 2.4 *Suppose* $\Omega \subset \mathbb{R}^n$ *is a bounded domain with smooth boundary; then the Dirichlet problem*

$$\begin{aligned} Lu &= f, \\ u &= \varphi \quad \text{on} \quad \partial\Omega \end{aligned}$$

always has a unique solution $u \in C^{2,\alpha}(\Omega)$ *if* $f \in C^\alpha(\Omega)$ *and* $\varphi \in C^{2,\alpha}(\partial\Omega)$.

Observe that uniqueness is an immediate consequence of the maximum principle. Existence can be established using the *Perron method.*

10.2.3 Harmonic Coordinates

The above theorem makes it possible to introduce *harmonic coordinates* on Riemannian manifolds.

Lemma 2.5 *If* (M, g) *is an n-dimensional Riemannian manifold and* $p \in M$, *then there is a neighborhood* $U \ni p$ *on which we can find a harmonic coordinate system* $x = (x^1, \ldots, x^n) : U \to \mathbb{R}^n$, *i.e., a coordinate system such that the functions* x^i *are harmonic with respect to the Laplacian on* (M, g).

Proof. First select a coordinate system $y = (y^1, \ldots, y^n)$ on a neighborhood around p such that $y(p) = 0$. We can then think of M as being an open subset of $\mathbb{R}^n$ and $p = 0$. The metric g is then written

$$g = g_{ij} = g(\partial_i, \partial_j) = g\left(\frac{\partial}{\partial y^i}, \frac{\partial}{\partial y^j}\right)$$

in the standard Cartesian coordinates $(y^1, \ldots, y^n)$. We must then find a coordinate transformation $y \to x$ such that

$$\Delta x^k = \frac{1}{\sqrt{\det g_{ij}}} \partial_i \left(\sqrt{\det g_{ij}} \cdot g^{ij} \cdot \partial_j u \right) = 0.$$

To find these coordinates, fix a small ball $B(0, \varepsilon)$ and solve the Dirichlet problem

$$\begin{aligned} \Delta x^k &= 0, \\ x^k &= y^k \quad \text{on} \quad \partial B(0, \varepsilon). \end{aligned}$$

We have then found n harmonic functions that should be close to the original coordinates. The only problem is that we don't know if they actually are coordinates. The Schauder estimates tell us that

$$\begin{aligned} \|x - y\|_{C^{2,\alpha}, B(0,\varepsilon)} &\le C \left(\|\Delta (x - y)\|_{C^\alpha, B(0,\varepsilon)} + \left\|(x - y)_{|\partial B(0,\varepsilon)}\right\|_{C^{2,\alpha}, \partial B(0,\varepsilon)} \right) \\ &= C \, \|\Delta y\|_{C^\alpha, B(0,\varepsilon)}. \end{aligned}$$

If matters were arranged such that $\|\Delta y\|_{C^\alpha, B(0,\varepsilon)} \to 0$ as $\varepsilon \to 0$, then we could conclude that Dx and Dy are close for small ε. Since y does form a coordinates system, we would then also be able to conclude that x formed a coordinate system.

Now we just observe that if y were chosen as exponential Cartesian coordinates, then we would have that $\partial_k g_{ij} = 0$ at p. The formula for Δy then shows that $\Delta y = 0$ at p. Hence, we have $\|\Delta y\|_{C^\alpha, B(0,\varepsilon)} \to 0$ as $\varepsilon \to 0$. Finally recall that the constant C depends only on an upper bound for the diameter of the domain aside from α, n, λ. Thus, $\|x - y\|_{C^{2,\alpha}, B(0,\varepsilon)} \to 0$ as $\varepsilon \to 0$. □

One reason for using harmonic coordinates on Riemannian manifolds is that both the Laplacian and Ricci curvature tensor have particularly nice formulae in such coordinates.

Lemma 2.6 *Let (M, g) be an n-dimensional Riemannian manifold and suppose we have a harmonic coordinate system $x : U \to \mathbb{R}^n$. Then*

(1) $\Delta u = (\sqrt{\det g_{ij}})^{-1} \partial_i \left(\sqrt{\det g_{ij}} \cdot g^{ij} \cdot \partial_j u \right) = g^{ij} \partial_i \partial_j u.$

(2) $\frac{1}{2} \Delta g_{ij} + Q(g, \partial g) = -\text{Ric}_{ij} = -g\left(\text{Ric}(\partial_i), \partial_j\right)$. *Here Q is some universal analytic expression that is polynomial in the matrix g, quadratic in ∂g, and has a denominator term depending on $\sqrt{\det g_{ij}}$.*

Proof. (1) By definition, we have that

$$\begin{aligned} 0 &= \Delta x^k \\ &= \frac{1}{\sqrt{\det g_{ij}}} \partial_i \left(\sqrt{\det g_{ij}} \cdot g^{ij} \cdot \partial_j x^k \right) \end{aligned}$$

$$\begin{aligned}
&= g^{ij}\partial_i\partial_j x^k + \frac{1}{\sqrt{\det g_{ij}}}\partial_i\left(\sqrt{\det g_{ij}}\cdot g^{ij}\right)\cdot\partial_j x^k \\
&= g^{ij}\partial_i\delta_j^k + \frac{1}{\sqrt{\det g_{ij}}}\partial_i\left(\sqrt{\det g_{ij}}\cdot g^{ij}\right)\cdot\delta_j^k \\
&= 0 + \frac{1}{\sqrt{\det g_{ik}}}\partial_i\left(\sqrt{\det g_{ij}}\cdot g^{ik}\right) \\
&= \frac{1}{\sqrt{\det g_{ik}}}\partial_i\left(\sqrt{\det g_{ij}}\cdot g^{ik}\right).
\end{aligned}$$

Thus, it follows that

$$\begin{aligned}
\Delta u &= \frac{1}{\sqrt{\det g_{ij}}}\partial_i\left(\sqrt{\det g_{ij}}\cdot g^{ij}\cdot\partial_j u\right) \\
&= g^{ij}\partial_i\partial_j u + \frac{1}{\sqrt{\det g_{ij}}}\partial_i\left(\sqrt{\det g_{ij}}\cdot g^{ij}\right)\cdot\partial_j u \\
&= g^{ij}\partial_i\partial_j u.
\end{aligned}$$

(2) Recall that if u is harmonic, then the Bochner formula for ∇u is

$$\begin{aligned}
\Delta\left(\frac{1}{2}|\nabla u|^2\right) &= \frac{1}{2}\Delta g(\nabla u, \nabla u) \\
&= \left|\nabla^2 u\right|^2 + \mathrm{Ric}(\nabla u, \nabla u) \\
&= \mathrm{tr}\left(\nabla^2 u\circ\nabla^2 u\right) + \mathrm{Ric}(\nabla u, \nabla u).
\end{aligned}$$

Here the term $\left|\nabla^2 u\right|^2$ can be computed explicitly and depends only on the metric and its first derivatives. Namely, $\nabla u = g^{ij}\partial_j u\partial_i$, and consequently $\nabla(\nabla u)$ depends in the desired way upon the metric.

Thus, we have the formula

$$\frac{1}{2}\Delta g\left(\nabla x^k, \nabla x^k\right) - \mathrm{tr}\left(\nabla^2 x^k\circ\nabla^2 x^k\right) = \mathrm{Ric}\left(\nabla x^k, \nabla x^k\right).$$

Polarizing this quadratic expression gives us an identity of the form

$$\frac{1}{2}\Delta g\left(\nabla x^i, \nabla x^j\right) - \mathrm{tr}\left(\nabla^2 x^i\circ\nabla^2 x^j\right) = \mathrm{Ric}\left(\nabla x^i, \nabla x^j\right).$$

Now use that $\nabla x^k = g^{ij}\partial_j x^k\partial_i = g^{ik}\partial_i$ to see that $g\left(\nabla x^i, \nabla x^j\right) = g^{ij}$. We then have

$$\frac{1}{2}\Delta g^{ij} - \mathrm{tr}\left(\nabla^2 x^i\circ\nabla^2 x^j\right) = \mathrm{Ric}\left(\nabla x^i, \nabla x^j\right),$$

which in matrix form looks like

$$\frac{1}{2}\Delta\left(g^{ij}\right) - \left(\mathrm{tr}\left(\nabla^2 x^i\circ\nabla^2 x^j\right)\right) = \left(g^{ij}\right)\cdot\left(g^{ij}\right)\cdot\left(\mathrm{Ric}\left(\partial_i, \partial_j\right)\right).$$

This is, of course, not the promised formula. Instead, it is a similar formula for the inverse of g. One can now use the matrix equation $(g_{ij}) \cdot (g^{ij}) = I$ to conclude that

$$\begin{aligned} 0 &= \Delta\left((g_{ij}) \cdot (g^{ij})\right) \\ &= (\Delta(g_{ij})) \cdot (g^{ij}) + 2g\left(\nabla(g_{ij}), \nabla(g^{ij})\right) + (g_{ij}) \cdot (\Delta(g^{ij})). \end{aligned}$$

Inserting this in the above equation yields

$$\begin{aligned} \Delta(g_{ij}) &= -(g_{ij}) \cdot 2g\left(\nabla(g_{ij}), \nabla(g^{ij})\right) - (g_{ij}) \cdot (g_{ij}) \cdot (\Delta(g^{ij})) \\ &= -2(g_{ij}) \cdot g\left(\nabla(g_{ij}), \nabla(g^{ij})\right) \\ &\quad - 2(g_{ij}) \cdot (g_{ij}) \cdot \left(\mathrm{tr}\left(\nabla^2 x^i \circ \nabla^2 x^j\right)\right) \\ &\quad - 2(g_{ij}) \cdot (g_{ij}) \cdot (g^{ij}) \cdot (g^{ij}) \cdot \left(\mathrm{Ric}(\partial_i, \partial_j)\right) \\ &= -2(g_{ij}) \cdot g\left(\nabla(g_{ij}), \nabla(g^{ij})\right) \\ &\quad - 2(g_{ij}) \cdot (g_{ij}) \cdot \left(\mathrm{tr}\left(\nabla^2 x^i \circ \nabla^2 x^j\right)\right) - 2\left(\mathrm{Ric}(\partial_i, \partial_j)\right). \end{aligned}$$

At the entry level, we therefore have an equation of the form

$$\frac{1}{2}\Delta g_{ij} + Q(g, \partial g) = -\mathrm{Ric}_{ij}.$$

Here, the Q term is computed from the matrix product

$$-(g_{ij}) \cdot g\left(\nabla(g_{ij}), \nabla(g^{ij})\right) - (g_{ij}) \cdot (g_{ij}) \cdot \left(\mathrm{tr}\left(\nabla^2 x^i \circ \nabla^2 x^j\right)\right).$$

One can easily see that this is a polynomial in the two matrices (g_{ij}),(g^{ij}), and their first derivatives. Also, all the derivative terms are quadratic. □

It is interesting to apply this formula to the case of an Einstein metric, where $\mathrm{Ric}_{ij} = (n-1)kg_{ij}$. In this case, it reads

$$\frac{1}{2}\Delta g_{ij} = -(n-1)kg_{ij} - Q(g, \partial g).$$

This formula makes sense even when g_{ij} is only $C^{1,\alpha}$. Namely, multiply by some test function, integrate, and use integration by parts to obtain a formula that uses only first derivatives of g_{ij}. If now g_{ij} is $C^{1,\alpha}$, then the left-hand side lies in C^{α}; but then our elliptic estimates show that g_{ij} must be in $C^{2,\alpha}$. This can be continued until we have that the metric is C^{∞}. In fact, one can even show that it is analytic. We can therefore conclude that any metric that in harmonic coordinates is a weak solution to the Einstein equation must in fact be smooth. We have obviously left out a few details about weak solutions, but they can easily be filled in if you consult [79, vol. III].

10.3 Norms and Convergence of Manifolds

We shall now explain how the $C^{m,\alpha}$ norm and convergence concepts for functions generalize to Riemannian manifolds. We shall also see how these ideas can be used to prove various compactness and finiteness theorems for classes of Riemannian manifolds.

10.3.1 Norms of Riemannian Manifolds

Before defining norms for manifolds, let us discuss which spaces should have norm zero. Clearly Euclidean space is a candidate. But what about open subsets of Euclidean space and other flat manifolds? If we agree that all open subsets of Euclidean space also have norm zero, then any flat manifold becomes a union of manifolds with norm zero and should therefore also have norm zero. In order to create a useful theory, it is of course best to have only one space with zero norm. Thus we must agree that subsets of Euclidean space cannot have norm zero. To accommodate this problem, we define the norm of a Riemannian manifold as a function $N : (0, \infty) \to (0, \infty)$. The number $N(r)$ then measures the degree of flatness on the scale of r, where the standard measure of flatness on the scale of r is the Euclidean ball $B(0, r)$. For small r, all flat manifolds then have norm zero; but as r increases we see that the space looks less and less like $B(0, r)$, and therefore the norm will become positive unless the space is Euclidean space.

For the precise definition, suppose A is a subset of a Riemannian n-manifold (M, g). We say that the $C^{m,\alpha}$*-norm on the scale of* r *of* $A \subset (M, g)$,

$$\|A \subset (M, g)\|_{C^{m,\alpha},r} \quad \text{is less than} \quad Q,$$

if we can find charts $\varphi_s : B(0, r) \subset \mathbb{R}^n \longleftrightarrow U_s \subset M$ such that

(n1) Every ball $B\left(p, \frac{1}{10}e^{-Q}r\right)$, $p \in A$ is contained on some U_s.

(n2) $|D\varphi_s| \leq e^Q$ on $B(0, r)$ and $\left|D\varphi_s^{-1}\right| \leq e^Q$ on U_s.

(n3) $r^{|j|+\alpha} \left\|D^j g_{s\cdot\cdot}\right\|_\alpha \leq Q$ for all multi indices j with $0 \leq |j| \leq m$. Here $g_{s\cdot\cdot}$ is the matrix of functions of metric coefficients in the φ_s coordinates regarded as a matrix on $B(0, r)$.

(n4) $\left\|\varphi_s^{-1} \circ \varphi_t\right\|_{C^{m+1,\alpha}} \leq (10 + r)\, e^Q$.

First, observe that we think of the charts as maps from the fixed space $B(0, r)$ into the manifold. This is in order to have domains for the functions which do not refer to M itself. This simplifies some technical issues and makes it more clear that we are trying to measure how different the manifolds are from the standard objects, namely, Euclidean balls. The first condition says that we have a Lebesgue number for the covering of A. The second condition tells us that in the chosen coordinates the metric coefficients are bounded from below and above (in particular, we have uniform ellipticity). The third condition, then, in addition gives us bounds on the derivatives of the metric. The fourth condition is just there to ensure that the bounds for the metric in individual coordinates don't vary drastically in places where coordinates overlap. This last condition can be eliminated in many cases. We shall give another norm concept below that does this.

One of the first simple properties one should note is that if we scale the metric, then we have

$$\begin{aligned}\min\{|\log\lambda|,\lambda\}\|A\subset(M,g)\|_{C^{m,\alpha},r} &\leq \left\|A\subset\left(M,\lambda^2 g\right)\right\|_{C^{m,\alpha},r}\\ &\leq \max\{|\log\lambda|,\lambda\}\|A\subset(M,g)\|_{C^{m,\alpha},r}.\end{aligned}$$

At a given scale r, it is therefore always possible to make norms large or small by scaling the metric.

It will be necessary on occasion to work with Riemannian manifolds that are not smooth. The above definition clearly only requires that the metric be $C^{m,\alpha}$ in the coordinates we use, and so there is no reason to assume more about the metric. Some of the basic constructions, like exponential maps, then come into question, and indeed, if $m \leq 1$ these items might not be well-defined. We shall therefore have to be a little careful in some situations.

When it is clear from the context where A is, we shall merely write $\|A\|_{C^{m,\alpha},r}$, or for the whole space, $\|(M,g)\|_{C^{m,\alpha},r}$ or $\|M\|_{C^{m,\alpha},r}$. If A is precompact in M, then it is clear that the norm is bounded for all r. For unbounded domains or manifolds it might, however, not be finite.

Example 3.1 Suppose (M,g) is a complete flat manifold. Then $\|(M,g)\|_{C^{m,\alpha},r} = 0$ for all $r \leq \operatorname{inj}(M,g)$. In particular, $\|(\mathbb{R}^n,\mathrm{can})\|_{C^{m,\alpha},r} = 0$ for all r. We shall later see that these properties characterize flat manifolds and Euclidean space.

10.3.2 Convergence of Riemannian Manifolds

Now for the convergence concept that relates to this new norm. As we can't subtract manifolds, we have to resort to a different method for defining this. If we fix a closed manifold M, or more generally a precompact subset $A \subset M$, then we say that a sequence of functions converges in $C^{m,\alpha}$, on A, if they converge in the charts for some fixed finite covering of coordinate patches. This definition is clearly independent of the finite covering we choose. We can then more generally say that a sequence of tensors converges in $C^{m,\alpha}$ if the components of the tensors converge in these patches. This then makes it possible to speak about convergence of Riemannian metrics on compact subsets of a fixed manifold.

A sequence of pointed complete Riemannian manifolds is said to *converge in the pointed* $C^{m,\alpha}$ *topology* $(M_i,p_i,g_i)\to(M,p,g)$ if for every $R>0$ we can find a domain $\Omega \supset B(p,R) \subset M$ and embeddings $\varphi_i:\Omega\to M_i$ for large i such that $\varphi_i(\Omega)\supset B(p_i,R)$ and $\varphi_i^* g_i \to g$ on Ω in the $C^{m,\alpha}$ topology. When all manifolds in question are closed, then we have that the maps φ_i are diffeomorphisms. This means that for closed manifolds we can speak about unpointed convergence. In this case, convergence can therefore only happen if all the manifolds in the tail end of the sequence are diffeomorphic. In particular, we have that classes of closed Riemannian manifolds that are precompact in some $C^{m,\alpha}$ topology contain at most finitely many diffeomorphism types.

A warning about this kind of convergence is in order here. Suppose we have a sequence of metrics g_i on a fixed manifold M. It is possible that these metrics might converge in the sense just defined, without converging in the traditional sense of converging in some fixed coordinate systems. To be more specific, let g be the standard metric on $M = S^2$. Now define diffeomorphisms φ_t as being the dynamical system corresponding to the vector field that is 0 at the north and south poles and otherwise points in the direction of the south pole. As t increases, the diffeomorphisms will try to map the whole sphere down to a small neighborhood of the south pole. The metrics $\varphi_t^* g$ will therefore in some fixed coordinates converge to 0 (except at the poles). They can therefore not converge in the classical sense. If, however, we pull these metrics back by the diffeomorphisms φ_{-t}, then we just get back to g. Thus the sequence (M, g_t), from the new point of view we are considering, is a constant sequence. This is really the right way to think about this, for the spaces $\left(S^2, \varphi_t^* g\right)$ are all isometric as abstract metric spaces.

10.3.3 Properties of the Norm

Let us now consider some of the elementary properties of norms and their relation to convergence.

Proposition 3.2 *If $A \subset (M, g)$ is precompact, then*

(1) $\|A \subset (M, g)\|_{C^{m,\alpha},r} = \left\|A \subset \left(M, \lambda^2 g\right)\right\|_{C^{m,\alpha},\lambda r}$ *for all* $\lambda > 0$.

(2) *The function* $r \to \|A \subset (M, g)\|_{C^{m,\alpha},r}$ *is continuous and converges to* 0 *as* $r \to 0$.

(3) *Suppose* $(M_i, p_i, g_i) \to (M, p, g)$ *in* $C^{m,\alpha}$. *Then for* $A \subset M$ *we can find precompact domains* $A_i \subset M_i$ *such that*

$$\|A_i\|_{C^{m,\alpha},r} \to \|A\|_{C^{m,\alpha},r} \quad \text{for all} \quad r > 0.$$

When all the manifolds are closed, we can let $A = M$ *and* $A_i = M_i$.

Proof. (1) If we change the metric g to $\lambda^2 g$, then we can change the charts $\varphi_s : B(0, r) \to M$ to $\varphi_s^\lambda(x) = \varphi_s\left(\lambda^{-1} x\right) : B(0, \lambda r) \to M$. Since we scale the metric at the same time, the conditions n1 to n4 will still hold with the same Q.

(2) Suppose, as above, we change the charts $\varphi_s : B(0, r) \to M$ to $\varphi_s^\lambda(x) = \varphi_s\left(\lambda^{-1} x\right) : B(0, \lambda r) \to M$, without changing the metric g. If we assume that $\|A \subset (M, g)\|_{C^{m,\alpha},r} < Q$, then

$$\|A \subset (M, g)\|_{C^{m,\alpha},\lambda r} \leq \max\left\{Q + |\log \lambda|, Q \cdot \lambda^2\right\}.$$

Denoting $N(r) = \|A \subset (M,g)\|_{C^{m,\alpha},r}$, we therefore obtain

$$N(\lambda r) \le \max\left\{N(r) + |\log\lambda|, N(r)\cdot\lambda^2\right\}.$$

By letting $\lambda = \frac{r_i}{r}$, where $r_i \to r$, we see that this implies

$$\limsup N(r_i) \le N(r).$$

Conversely, we have that

$$\begin{aligned} N(r) &= N\left(\frac{r}{r_i}r_i\right) \\ &\le \max\left\{N(r_i) + \left|\log\frac{r}{r_i}\right|, N(r_i)\cdot\left(\frac{r}{r_i}\right)^2\right\}. \end{aligned}$$

So

$$\begin{aligned} N(r) &\le \liminf N(r_i) \\ &= \liminf \max\left\{N(r_i) + \left|\log\frac{r}{r_i}\right|, N(r_i)\cdot\left(\frac{r}{r_i}\right)^2\right\}. \end{aligned}$$

This shows that $N(r)$ is continuous. To see that $N(r) \to 0$ as $r \to 0$, just observe that any coordinate system around a point $p \in M$ can, after a linear change, be assumed to have the property that the metric $g_{ij} = \delta_{ij}$ at p. Using these coordinates on sufficiently small balls will then give the desired charts.

(3) We fix $r > 0$ in the definition of $\|A \subset (M,g)\|_{C^{m,\alpha},r}$. For the given $A \subset M$, pick a domain $\Omega \supset A$ such that for large i we have embeddings $f_i : \Omega \to M_i$ with the property that: $f_i^*g_i \to g$ in $C^{m,\alpha}$ on Ω. Then let $A_i = f_i(A)$.

For $Q > \|A \subset (M,g)\|_{C^{m,\alpha},r}$, choose appropriate charts $\varphi_s : B(0,r) \to M$ covering A, with the properties n1-n4. Then define charts in M_i by $\varphi_{i,s} = f_i \circ \varphi_s : B(0,r) \to M_i$. Condition n1 will hold just because we have Gromov-Hausdorff convergence and condition n4 is trivial. Conditions n2 and n3 will hold for constants $Q_i \to Q$, since $f_i^*g_i \to g$ in $C^{m,\alpha}$. We can therefore conclude that

$$\limsup \|A_i\|_{C^{m,\alpha},r} \le \|A\|_{C^{m,\alpha},r}.$$

On the other hand, for large i and $Q > \|A_i\|_{C^{m,\alpha},r}$, we can take charts $\varphi_{i,s} : B(0,r) \to M_i$ and then pull them back to M by defining $\varphi_s = f_i^{-1} \circ \varphi_{i,s}$. As before, we then have

$$\|A\|_{C^{m,\alpha},r} \le Q_i,$$

where $Q_i \to Q$. This implies

$$\liminf \|A_i\|_{C^{m,\alpha},r} \ge \|A\|_{C^{m,\alpha},r},$$

and hence the desired result. □

10.3.4 Compact Classes of Riemannian Manifolds

We are now ready to prove the result that is our manifold equivalent of the Arzela-Ascoli lemma. This theorem is essentially due to J. Cheeger, although our use of norms makes the statement look different.

Theorem 3.3 (Fundamental Theorem of Convergence Theory) *For given $Q > 0$, $n \geq 2$, $m \geq 0, \alpha \in (0, 1]$, and $r > 0$ consider the class $\mathcal{M}^{m,\alpha}(n, Q, r)$ of complete, pointed Riemannian n-manifolds (M, p, g) with $\|(M, g)\|_{C^{m,\alpha},r} \leq Q$. $\mathcal{M}^{m,\alpha}(n, Q, r)$ is compact in the pointed $C^{m,\beta}$ topology for all $\beta < \alpha$.*

Proof. We proceed in stages. First, we make some general comments about the charts we use. We then show that $\mathcal{M} = \mathcal{M}^{m,\alpha}(n, Q, r)$ is precompact in the pointed Gromov-Hausdorff topology. Next we prove that $\mathcal{M}$ is compact in the Gromov-Hausdorff topology. The last part is then devoted to the compactness statement.

Setup: First fix $K > Q$. Whenever we select an $M \in \mathcal{M}$, we shall assume that it comes equipped with an atlas of charts satisfying n1 to n4 with K in place of Q. Thus we implicitly assume that all charts under consideration belong to these atlases. We will, in consequence, prove only that limit spaces (M, p, g) satisfy $\|(M, g)\|_{C^{m,\alpha},r} \leq K$, but as K was arbitrary, we still get that $(M, p, g) \in \mathcal{M}$.

(1) Every chart $\varphi : B(0, r) \to U \subset M \in \mathcal{M}$ satisfies

- (a) $d(\varphi(x_1), \varphi(x_2)) \leq e^K |x_1 - x_2|$
- (b) $d(\varphi(x_1), \varphi(x_2)) \geq \min\{e^{-K}|x_1 - x_2|, e^{-K}(2r - |x_1| - |x_2|)\}$
  Here, d is distance measured in M, and $|\cdot|$ is the usual Euclidean norm.

The condition $|D\varphi| \leq e^K$, together with convexity of $B(0, r)$, immediately implies the first inequality. For the other, first observe that if any segment from x_1 to x_2 lies in U, then $|D\varphi^{-1}| \leq e^K$ implies that $d(\varphi(x_1), \varphi(x_2)) \geq e^{-K}|x_1 - x_2|$. So we may assume that $\varphi(x_1)$ and $\varphi(x_2)$ are joined by a segment $\sigma : [0, 1] \to M$ that leaves U. Split σ into $\sigma : [0, t_1) \to U$ and $\sigma : (t_2, 1) \to U$ such that $\sigma(t_i) \notin U$. Then we clearly have

$$\begin{aligned} d(\varphi(x_1), \varphi(x_2)) &= L(\sigma) \geq L(\sigma\,|_{[0,t_1)}) + L(\sigma\,|_{(t_2,1]}) \\ &\geq e^{-K}(L(\varphi^{-1} \circ \sigma\,|_{[0,t_1)}) + L(\varphi^{-1} \circ \sigma\,|_{(t_2,1]})) \\ &\geq e^{-K}(2r - |x_1| - |x_2|). \end{aligned}$$

The last inequality follows from the fact that $\varphi^{-1} \circ \sigma(0) = x_1$ and $\varphi^{-1} \circ \sigma(1) = x_2$, and that $\varphi^{-1} \circ \sigma(t)$ approaches the boundary of $B(0, r)$ as $t \nearrow t_1$ or $t \searrow t_2$.

(2) Every chart $\varphi : B(0, r) \to U \subset M \in \mathcal{M}$, and hence any δ-ball $\delta = \frac{1}{10}e^{-K}r$ in M can be covered by at most N $\delta/4$-balls. Here, N depends only on n, K, r. Clearly, there exists an $N(n, K, r)$ such that $B(0, r)$ can be covered by at most

$N\left(e^{-K}\cdot\delta/4\right)$-balls. Since $\varphi : B(0,r)\to U$ is a Lipschitz map with Lipschitz constant $\le e^K$, we get the desired covering property.

(3) Every ball $B(x,\ell\cdot\delta/2)\subset M$ can be covered by $\le N^\ell$ $\delta/4$-balls. For $\ell=1$ we just proved this. Suppose we know that $B(x,\ell\cdot\delta/2)$ is covered by $B(x_1,\delta/4),\ldots,B(x_{N^\ell},\delta/4)$. Then $B(x,\ell\cdot\delta/2+\delta/2)\subset\cup B(x_i,\delta)$. Now each $B(x_i,\delta)$ can be covered by $\le N$ $\delta/4$-balls, and hence $B(x,(\ell+1)\delta/2)$ can be covered by $\le N\cdot N^\ell=N^{\ell+1}$ $\delta/4$-balls.

(4) $\mathcal{M}$ is precompact in the pointed Gromov-Hausdorff topology. This is equivalent to asserting, that for each $R>0$ the family of metric balls $B(p,R)\subset(M,p,g)\in\mathcal{M}$ is precompact in the Gromov-Hausdorff topology. This claim is equivalent to showing that we can find a function $N(\varepsilon)=N(\varepsilon,R,K,r,n)$ such that each $B(p,R)$ can contain at most $N(\varepsilon)$ disjoint ε-balls. To check this, let $B(x_1,\varepsilon),\ldots,B(x_s,\varepsilon)$ be a collection of disjoint balls in $B(p,R)$. Suppose that $\ell\cdot\delta/2<R\le(\ell+1)\delta/2$. Then

$$\begin{aligned}\mathrm{vol}B(p,R)&\le\left(N^{(\ell+1)}\right)\cdot(\text{maximal volume of }\frac{\delta}{4}\text{-ball})\\&\le\left(N^{(\ell+1)}\right)\cdot(\text{maximal volume of chart})\\&\le N^{(\ell+1)}\cdot e^{nK}\cdot\mathrm{vol}B(0,r)\\&\le F(R)=F(R,n,K,r).\end{aligned}$$

Conversely, each $B(x_i,\varepsilon)$ lies in some chart $\varphi : B(0,r)\to U\subset M$ whose preimage in $B(0,r)$ contains an $e^{-K}\cdot\varepsilon$-ball. Thus $\mathrm{vol}B(p_i,\varepsilon)\ge e^{-2nK}\mathrm{vol}B(0,\varepsilon)$. All in all, we get

$$\begin{aligned}F(R)&\ge\mathrm{vol}B(p,R)\\&\ge\sum\mathrm{vol}B(p_i,\varepsilon)\\&\ge s\cdot e^{-2nK}\cdot\mathrm{vol}B(0,\varepsilon).\end{aligned}$$

Thus, $s\le N(\varepsilon)=F(R)\cdot e^{2nK}\cdot(\mathrm{vol}B(0,\varepsilon))^{-1}$.

Now select a sequence (M_i,g_i,p_i) in $\mathcal{M}$. From the previous considerations we can assume that $(M_i,g_i,p_i)\to(X,d,p)$ converge to some metric space in the Gromov-Hausdorff topology. It will be necessary in many places to pass to subsequences of (M_i,g_i,p_i) using various diagonal processes. Whenever this happens, we shall not reindex the family, but merely assume that the sequence was chosen to have the desired properties from the beginning. For each (M_i,p_i,g_i) choose charts $\varphi_{is}: B(0,r)\to U_{is}\subset M_i$ satisfying n1 to n4. We can furthermore assume that the index set $\{s\}=\{1,2,3,4,\cdots\}$ is the same for all M_i, that $p_i\in U_{i1}$, and that the balls $B\left(p_i,\ell\cdot\delta/2\right)$ are covered by the first N^ℓ charts. Note that these N^ℓ charts will then be contained in $\bar{B}\left(p_i,\ell\cdot\delta/2\ell\cdot\delta/2+[e^K+1]\delta\right)$. Finally, for each ℓ the sequence $\bar{B}\left(p_i,\ell\cdot\delta/2\right)$ converges to $\bar{B}\left(p,\ell\cdot\delta/2\right)\subset X$, so we can choose a metric on the disjoint union $Y_\ell=\left(\bar{B}\left(p,\ell\cdot\delta/2\right)\coprod\left(\coprod_{i=1}^{\infty}\bar{B}\left(p_i,\ell\cdot\delta/2\right)\right)\right)$ such that $p_i\to p$ and $\bar{B}\left(p_i,\ell\cdot\delta/2\right)\to\bar{B}\left(p,\ell\cdot\delta/2\right)$ in the Hausdorff distance inside this metric space.

(5) (X, d, p) is a Riemannian manifold of class $C^{m,\alpha}$ with norm $\leq K$. Obviously, we need to find bijections $\varphi_s : B(0, r) \to U_s \subset X$ satisfying n1 to n4. For each s, consider the maps $\varphi_{is} : B(0, r) \to U_{is} \subset Y_{\ell+2[e^K+1]}$. From 1 we have that this is a family of equicontinuous maps into the compact space $Y_{\ell+2[e^K+1]}$. The Arzela-Ascoli lemma shows that this sequence must subconverge (in the C^0 topology) to a map $\varphi_s : B(0, r) \subset Y_{\ell+2[e^K+1]}$ that also has Lipschitz constant e^K. Furthermore, the inequality

$$d(\varphi(x_1), \varphi(x_2)) \geq \min\{e^{-K}|x_1 - x_2|,\ e^{-K}(2r - |x_1| - |x_2|)\}$$

will also hold for this map, as it holds for all the φ_{is} maps. In particular, φ_s is one-to-one. Finally, since $U_{is} \subset \bar{B}\left(p_i, \ell \cdot \delta/2 + [e^K + 1]\right)$ and $\bar{B}\left(p_i, \ell \cdot \delta/2 + [e^K + 1]\right)$ Hausdorff converges to $\bar{B}\left(p, \ell \cdot \delta/2 + [e^K + 1]\right) \subset X$, we see that $\varphi_s(B(0, r)) = U_s \subset X$. A simple diagonal argument yields that we can pass to a subsequence of (M_i, g_i, p_i) that has the property that $\varphi_{is} \to \varphi_s$ for all s. In this way, we have constructed (topological) charts $\varphi_s : B(0, r) \to U \subset X$, and we can easily check that they satisfy n1. Since the φ_s also satisfy 1(a) and 1(b), they would also satisfy n2 if they were differentiable (equivalent to saying that the transition functions are C^1). Now the transition functions $\varphi_{is}^{-1} \circ \varphi_{it}$ approach $\varphi_s^{-1} \circ \varphi_t$, because $\varphi_{is} \to \varphi_s$. Note that these transition functions are not defined on the same domains, but we do know that the domain for $\varphi_s^{-1} \circ \varphi_t$ is the limit of the domains for $\varphi_{is}^{-1} \circ \varphi_{it}$, so the convergence makes sense on all compact subsets of the domain of $\varphi_s^{-1} \circ \varphi_t$. Now, $\|\varphi_{is}^{-1} \circ \varphi_{it}\|_{C^{m+1,\alpha}} \leq (10 + r)e^K$, so a further application (and subsequent passage to subsequences) of Arzela-Ascoli tells us that $\|\varphi_s^{-1} \circ \varphi_t\|_{C^{m+1,\alpha}} \leq (10 + r)e^K$, and that we can assume $\varphi_{is}^{-1} \circ \varphi_{it} \to \varphi_s^{-1} \circ \varphi_t$ in the $C^{m+1,\beta}$ topology. This then establishes n4. We now construct a compatible Riemannian metric on X that satisfies n2 and n3. For each s, consider the metric $g_{is} = g_{is\cdot\cdot}$ written out in its components on $B(0, r)$ with respect to the chart φ_{is}. Since all of the $g_{is\cdot\cdot}$ satisfy n2 and n3, we can again use Arzela-Ascoli to insure that also $g_{is\cdot\cdot} \to g_{s\cdot\cdot}$ on $B(0, r)$ in the $C^{m,\beta}$ topology to functions $g_{s\cdot\cdot}$ that also satisfy n2 and n3. The local "tensors" $g_{s\cdot\cdot}$ satisfy the right change of variables formulae to make them into a global tensor on X. This is because all the $g_{is\cdot\cdot}$ satisfy these properties, and everything we want to converge, to carry these properties through to the limit, also converges. Recall that the rephrasing of n2 gives the necessary C^0 bounds and also shows that $g_{s\cdot\cdot}$ is positive definite. We have now exhibited a Riemannian structure on X such that the $\varphi_s : B(0, r) \to U_s \subset X$ satisfy n1 to n4 with respect to this structure. This, however, does not guarantee that the metric generated by this structure is identical to the metric we got from X being the pointed Gromov-Hausdorff limit of (M_i, p_i, g_i). However, since Gromov-Hausdorff convergence implies that distances converge, and we know at the same time that the Riemannian metric converges locally in coordinates, it follows that the limit Riemannian structure must generate the "correct" metric, at least locally, and therefore also globally.

(6) $(M_i, p_i, g_i) \to (X, p, d) = (X, p, g)$ in the pointed $C^{m,\beta}$ topology. We assume that the setup is as in (5), where charts φ_{is}, transitions $\varphi_{is}^{-1} \circ \varphi_{it}$, and

metrics $g_{is\cdot\cdot}$ converge to the same items in the limit space. First, let us agree that two maps f, g between subsets in M_i and X are $C^{m+1,\beta}$ close if all the coordinate compositions $\varphi_s^{-1} \circ g \circ \varphi_{it}$, $\varphi_s^{-1} \circ f \circ \varphi_{it}$ are $C^{m+1,\beta}$ close. Thus, we have a well-defined $C^{m+1,\beta}$ topology on maps from M_i to X. Our first observation is that $f_{is} = \varphi_{is} \circ \varphi_s^{-1} : U_s \to U_{is}$ and $f_{it} = \varphi_{it} \circ \varphi_t^{-1} : U_t \to U_{it}$ "converge to each other" in the $C^{m+1,\beta}$ topology. Furthermore, $(f_{is})^* g_i|_{U_{is}} \to g|_{U_s}$ in the $C^{m,\beta}$ topology. These are just restatements of what we already know. In order to finish the proof, we therefore only need to construct diffeomorphisms $F_{i\ell} : \Omega_\ell = \bigcup_{s=1}^{\ell} U_s \to \Omega_{i\ell} = \bigcup_{s=1}^{\ell} U_{is}$ that are closer and closer to the f_{is}, $s = 1, \ldots, \ell$ maps (and therefore all f_{is}) as $i \to \infty$. We will construct $F_{i\ell}$ by induction on ℓ and large i depending on ℓ. For this purpose we shall need a partition of unity $\{\lambda_s\}$ on X subordinate to $\{U_s\}$. We can find such a partition, since the covering $\{U_s\}$ is locally finite by choice, and we can furthermore assume that λ_s is $C^{m+1,\beta}$.

For $\ell = 1$ simply define $F_{i1} = f_{i1}$.

Suppose we have $F_{i\ell} : \Omega_\ell \to \Omega_{i\ell}$ for large i that are arbitrarily close to f_{is}, $s = 1, \ldots, \ell$ as $i \to \infty$. If $U_{\ell+1} \cap \Omega_\ell = \emptyset$, then we just define $F_{i\ell+1} = F_{i\ell}$ on $\Omega_{i\ell}$, and $F_{i\ell+1} = f_{i\ell+1}$ on $U_{\ell+1}$. In case $U_{\ell+1} \subset \Omega_\ell$, we simply let $F_{i\ell+1} = F_{i\ell}$. Otherwise, we know that $F_{i\ell}$ and $f_{i\ell+1}$ are as close as we like in the $C^{m+1,\beta}$ topology as $i \to \infty$. So the natural thing to do is to average them on $U_{\ell+1}$. Define $F_{i\ell+1}$ on $U_{\ell+1}$ as

$$
\begin{aligned}
F_{i\ell+1}(x) &= \varphi_{i\ell+1} \circ \left(\sum_{s=\ell+1}^{\infty} \lambda_s(x) \cdot \varphi_{i\ell+1}^{-1} \circ f_{i\ell+1}(x) + \sum_{s=1}^{\ell} \lambda_s(x) \cdot \varphi_{i\ell+1}^{-1} \circ F_{i\ell}(x) \right) \\
&= \varphi_{i\ell+1} \circ (\mu_1(x) \cdot \varphi_{i\ell+1}^{-1} \circ f_{i\ell+1}(x) + \mu_2(x) \cdot \varphi_{i\ell+1}^{-1} \circ F_{i\ell}(x)).
\end{aligned}
$$

This map is clearly well-defined on $U_{\ell+1}$, since $\mu_2(x) = 0$ on $U_{\ell+1} - \Omega_\ell$ and since $\mu_1(x) = 0$ on Ω_ℓ is a smooth $C^{m+1,\beta}$ extension of $F_{i\ell}$. Now consider this map in coordinates

$$
\begin{aligned}
\varphi_{i\ell+1}^{-1} \circ F_{i\ell+1} \circ \varphi_{\ell+1}(y) &= \mu_1 \circ \varphi_{\ell+1}(y) \cdot \varphi_{\ell+1}^{-1} \circ f_{i\ell+1} \circ \varphi_{\ell+1}(y) \\
&\quad + \mu_2 \circ \varphi_{\ell+1}(y) \cdot \varphi_{\ell+1}(y) \cdot \varphi_{i\ell+1}^{-1} \circ F_{i\ell} \circ \varphi_{\ell+1}(y) \\
&= \tilde{\mu}_1(y) \cdot F_1(y) + \tilde{\mu}_2(y) \cdot F_2(y).
\end{aligned}
$$

Then

$$
\begin{aligned}
\|\tilde{\mu}_1 \cdot F_1 + \tilde{\mu}_2 F_2 - F_1\|_{C^{m+1,\beta}} &= \|\tilde{\mu}_1(F_1 - F_1) + \tilde{\mu}_2(F_2 - F_1)\|_{C^{m+1,\beta}} \\
&\leq \|\tilde{\mu}_2\|_{k+1+\beta} \cdot \|F_2 - F_1\|_{C^{m+1,\beta}}.
\end{aligned}
$$

This inequality is valid on all of $B(0, r)$, despite the fact that F_2 is not defined on all of $B(0, r)$, because $\tilde{\mu}_1 \cdot F_1 + \tilde{\mu}_2 \cdot F_2 = F_1$ on the region where F_2 is undefined. By assumption $\|F_2 - F_1\|_{C^{m+1,\beta}} \to 0$ as $i \to \infty$, so $F_{i\ell+1}$ is $C^{m+1,\beta}$-close to f_{is}, $s = 1, \ldots, \ell + 1$ as $i \to \infty$. It remains to be seen that $F_{i\ell+1}$ is a diffeomorphism. But we know that $\tilde{\mu}_1 F_1 + \tilde{\mu}_2 F_2$ is an embedding, since F_1 is, and the space of embeddings is open in the C^1 topology. Also, the map is one-to-one, as the images $f_{i\ell+1}(U_{\ell+1} - \Omega_\ell)$ and $F_{i\ell}(\Omega_\ell)$ don't intersect. □

Corollary 3.4 *The subclasses of $\mathcal{M}^{m,\alpha}(n, Q, r)$, where the elements in addition satisfy* diam $\le D$, *respectively* vol $\le V$, *are compact in the $C^{m,\beta}$ topology. In particular, they contain only finitely many diffeomorphism types.*

Proof. We use notation as in the fundamental theorem. If $\text{diam}(M, g, p) \le D$, then clearly $M \subset B(p, k \cdot \delta/2)$ for $k > D \cdot 2/\delta$. Hence, each element in $\mathcal{M}^{m,\alpha}(n, Q, r)$ can be covered by $\le N^k$ charts. Thus, $C^{m,\beta}$-convergence is actually in the unpointed topology, as desired.

If instead, $\text{vol}M \le V$, then we can use part 4 in the proof to see that we can never have more than $k = V \cdot e^{2nK} \cdot (\text{vol}B(0, \varepsilon))^{-1}$ disjoint ε-balls. In particular, diam $\le 2\varepsilon \cdot k$, and we can use the above argument.

Clearly, compactness in any $C^{m,\beta}$ topology implies that the class cannot contain infinitely many diffeomorphism types. □

Corollary 3.5 *The norm $\|A \subset (M, g)\|_{C^{m,\alpha},r}$ for compact A is always realized by some charts $\varphi_s : B(0, r) \to U_s$ satisfying n1-n4, with $\|(M, g)\|_{C^{m,\alpha},r}$ in place of Q.*

Proof. Choose appropriate charts $\varphi_s^Q : B(0, r) \to U_s^Q \subset M$ for each $Q > \|(M, g)\|_{C^{m,\alpha},r}$, and let $Q \to \|(M, g)\|_{C^{m,\alpha},r}$. If the charts are chosen to conform with the proof of the fundamental theorem, we will obviously get some limit charts with the desired properties. □

Corollary 3.6 *M is a flat manifold if $\|(M, g)\|_{C^{m,\alpha},r} = 0$ for some r, and M is Euclidean space with the canonical metric if $\|(M, g)\|_{C^{m,\alpha},r} = 0$ for all $r > 0$.*

Proof. Using the previous corollary, M can be covered by charts $\varphi_s : B(0, r) \to U_s \subset M$ satisfying $|D\varphi_s| \equiv 1$. This clearly makes M locally Euclidean and hence flat. If M is not Euclidean space, then the same reasoning clearly shows that $\|(M, g)\|_{C^{m,\alpha},r} > 0$ for $r > \text{inj}(M, g)$. □

10.3.5 Other Norms

Finally, we should mention that all properties of this norm concept would not change if we changed n1 to n4 to say

(n1′) U_s has Lebesgue number $f_1(n, Q, r)$.

(n2′) $|D\varphi_s| \le f_2(n, Q)$, and $|\varphi_s^{-1}| \le f_2(n, Q)$.

(n3′) $r^{|j|+\alpha} \cdot \|\partial^j g_{s\cdot\cdot}\|_\alpha \le f_3(n, Q), \quad 0 \le |j| \le m.$

(n4′) $\|\varphi_s^{-1} \circ \varphi_t\|_{C^{m+1,\alpha}} \le f_4(n, Q, r)$.

As long as the f_i are all continuous, $f_1(n, 0, r) = 0$, and $f_2(n, 0) = 1$. The key properties we want to preserve are continuity of $\|(M, g)\|$ with respect to r, the fundamental theorem, and the characterization of flat manifolds and Euclidean space.

Another interesting thing happens if in the definition of $\|(M, g)\|_{C^{m,\alpha},r}$ we let $k = \alpha = 0$. Then n3 no longer makes sense, because $\alpha = 0$, but aside from that, we still have a C^0-norm concept. Note also that n4 is an immediate consequence of n2 in this case. The class $\mathcal{M}^0(n, Q, r)$ is now only precompact in the pointed Gromov-Hausdorff topology, but the characterization of flat manifolds is still valid. The subclasses with bounded diameter, or volume, are also only precompact with respect to the Gromov-Hausdorff topology, and the finiteness of diffeomorphism types apparently fails. It is, however, possible to say more. If we investigate the proof of the fundamental theorem, we see that the problem lies in constructing the maps $F_{ik} : \Omega_k \to \Omega_{ik}$, because we now have convergence of the coordinates only in the C^0 (actually $C^\alpha, \alpha < 1$) topology, and so the averaging process fails as it is described. We can, however, use a deep theorem from topology about local contractibility of homeomorphism groups (see [31]) to conclude that two C^0-close topological embeddings can be "glued" together in some way without altering them too much in the C^0 topology. This makes it possible to exhibit topological embeddings $F_{ik} : \Omega \hookrightarrow M_i$ such that the pullback metrics (not Riemannian metrics) converge. As a consequence, we see that the classes with bounded diameter or volume contain only finitely many homeomorphism types. This is exactly the content of the original version of Cheeger's finiteness theorem, including the proof as we have outlined it. But, as we have pointed out earlier, Cheeger also considered the easier to prove finiteness theorem for diffeomorphism types given better bounds on the coordinates.

Notice that we cannot use the fact that the charts converge in $C^\alpha(\alpha < 1)$, because there is no theory of $C^\alpha(\alpha < 1)$ manifolds. In fact, such a theory is probably meaningless, as the composition of two $C^\alpha(\alpha < 1)$ maps is in general only C^{α^2}. But some more can be done; namely, we can develop a norm that is more natural in this context called the *Lipschitz norm* and denoted by $\|\cdots\|_{L,r}$. It is defined for complete metric spaces X that have the property that any two points are joined by a curve whose length is the distance between the points. In addition, we assume that X has a compatible manifold structure, where the charts are locally bi-Lipschitz (the transition functions are therefore also locally bi-Lipschitz). Such spaces are called *Lipschitz manifolds*. We now say that $\|(X, d)\|_{L,r} \leq Q$ if there are homeomorphisms $\varphi_s : B(0, r) \subset \mathbb{R}^n \to U_s \subset X$ such that

(Ln1) $\delta = \frac{1}{10} r e^Q$ is a Lebesgue number for the covering U_s.

(Ln2) $d(\varphi_s(x), \varphi_s(y)) \leq e^Q |x - y|$ for $x, y \in B(0, r)$ and $\left|\varphi_s^{-1}(x) - \varphi_s^{-1}(y)\right| \leq e^Q d(x, y)$ for all $x, y \in U_s$ sufficiently close to each other.

One can now easily show that the class of spaces with $\|(X, d)\|_{L,r} \leq Q$, for fixed r and Q, is compact (not just precompact) in the pointed Gromov-Hausdorff topology, and that $\|(X, d)\|_{L,r} = 0$ iff (X, d) is a flat manifold. Moreover, if we bound the diameter, then the class will contain only finitely many homeomorphism types. With Sullivan's work (see papers by L. Siebenmann and D. Sullivan in [17]) one can improve this to finitely many Lipschitz homeomorphism types.

10.4 Geometric Applications

We shall now study the relationship between volume, injectivity radius, sectional curvature, and the norm.

First let us see what exponential coordinates can do for us. Let (M, g) be a Riemannian manifold with $|\sec M| \leq K$ and $\text{inj}M \geq i_0$. Now, on $B(0, i_0)$ we have from Chapter 6 that

$$\max\left\{\left|D\exp_p\right|, \left|D\exp_p^{-1}\right|\right\} \leq \exp\left(f(n, K, i_0)\right)$$

for some function $f(n, K, i_0)$ that depends only on the dimension, K, and i_0. Moreover, as $K \to 0$ we have that $f(n, K, i_0) \to 0$. This implies

Theorem 4.1 *For every $Q > 0$ there exists $r > 0$ depending only on i_0 and K such that any complete (M, g) with $|\sec M| \leq K$, $\text{inj}M \geq i_0$ has $\|(M, g)\|_{C^0, r} \leq Q$. Furthermore, if (M_i, p_i, g_i) satisfy $\text{inj}M_i \geq i_0$ and $|\sec M_i| \leq K_i \to 0$, then a subsequence will converge in the pointed Gromov-Hausdorff topology to a flat manifold with $\text{inj} \geq i_0$.*

The proof follows immediately from our previous constructions.

This theorem does not seem very satisfactory, because even though we have assumed a C^2 bound on the Riemannian metric, locally we get only a C^0 bound. To get better bounds under the same circumstances, we must look for different coordinates. Our first choice for alternative coordinates is distance coordinates.

Lemma 4.2 *Given a Riemannian manifold (M, g) with $\text{inj} \geq i_0$, $|\sec| \leq K$, and $p \in M$, then the distance function $d(x) = d(x, p)$ is smooth on $B(p, i_0)$, and the Hessian is bounded in absolute value on the annulus $B(p, i_0) - B(p, i_0/2)$ by a function $F(n, K, i_0)$.*

Proof. We know that the Hessian is always zero when evaluated on the gradient, perpendicular to the gradient, we know from Chapter 6 that

$$\sqrt{K}\cot\left(\sqrt{K}r\right) \cdot I \leq \left(\nabla^2 d\right) \leq \sqrt{K}\coth\left(\sqrt{-K}r\right) \cdot I.$$

Thus, we get the desired estimate as long as $r \in (i_0/2, i_0)$. □

Now fix (M, g), $p \in M$, as in the lemma, and choose an orthonormal basis $e_1, \ldots, e_n$ for T_pM. Then consider the geodesics $\gamma_i(t)$ with $\gamma_i(0) = p$, $\dot{\gamma}_i(0) = e_i$, and together with those, the distance functions $d_i(x) = d\left(x, \gamma_i\left(i_0 \cdot \left(4\sqrt{K}\right)^{-1}\right)\right)$.

These distance functions will then have uniformly bounded Hessians on $B(p,\delta)$, $\delta = i_0 \cdot \left(8\sqrt{K}\right)^{-1}$. Define $\varphi(x) = (d_1(x),\ldots,d_n(x))$ and recall that $g^{ij} = g\left(\nabla d_i, \nabla d_j\right)$.

Theorem 4.3 (The Convergence Theorem of Riemannian Geometry) *Given i_0, $K>0$, there exist $Q, r>0$ such that any (M,g) with*

$$\text{inj} \geq i_0,$$
$$|\sec| \leq K$$

has $\|(M,g)\|_{C^1,r} \leq Q$. In particular, this class is compact in the pointed C^α topology for all $\alpha < 1$.

Proof. The inverse of φ is our potential chart. First, observe that $g_{ij}(p) = \delta_{ij}$, so the uniform Hessian estimate shows that $|D\varphi_p| \leq e^Q$ on $B(p,\varepsilon)$ and $|D\varphi_p^{-1}| \leq e^Q$ on $B(0,\varepsilon)$, where Q, ε depend only on i_0, K. The proof of the inverse function theorem then tells us that there is $\hat{\varepsilon} > 0$ depending only on Q, n such that $\varphi : B(0,\hat{\varepsilon}) \to \mathbb{R}^n$ is one-to-one. We can then easily find r such that $\varphi^{-1} : B(0,r) \to U_p \subset B(p,\varepsilon)$ satisfies n2. The conditions n3 and n4 now immediately follow from the Hessian estimates, except we might have to increase Q somewhat. Finally, n1 holds since we have coordinates centered at every $p \in M$. □

Notice that Q cannot be chosen arbitrarily small, as our Hessian estimates cannot be improved by going to smaller balls. This will be taken care of in the next section by using even better coordinates. This convergence result, as stated, was first proven by M. Gromov. The reader should be aware that what Gromov refers to as a $C^{1,1}$-manifold is in our terminology a manifold with $\|(M,h)\|_{C^{0,1},r} < \infty$, i.e., $C^{0,1}$-bounds on the Riemannian metric.

Using Bonnet's diameter bound and Klingenberg's estimate for the injectivity radius from Chapter 6 we get

Corollary 4.4 (J. Cheeger, 1967) *For given $n \geq 1$ and $k > 0$, the class of Riemannian $2n$-manifolds with $k \leq \sec \leq 1$ is compact in the C^α topology and consequently contains only finitely many diffeomorphism types.*

Our next result shows that one can bound the injectivity radius provided that one has lower volume bounds and bounded curvature. This result is usually referred to as Cheeger's lemma. With a little extra work one can actually prove this lemma for complete manifolds. This requires that one work with pointed spaces and also to some extent incomplete manifolds, as one does not know from the beginning that the complete manifolds in question have lower bounds for the injectivity radius.

Lemma 4.5 (J. Cheeger, 1967) *Given $n \geq 2$ and $v, K \in (0,\infty)$ and a compact n-manifold (M,g) with*

$$|\sec| \leq K,$$
$$\text{vol} B(p,1) \geq v,$$

for all $p \in M$, *then* $\text{inj}M \geq i_0$, *where* i_0 *depends only on* n, K, *and* v.

Proof. The proof goes by contradiction using the previous theorem. So assume we have (M_i, g_i) with $\text{inj}M_i \to 0$ and satisfying the assumptions of the lemma. Find $p_i \in M_i$ such that $\text{inj}_{p_i} = \text{inj}\,(M_i, g_i)$, and consider the pointed sequence (M_i, p_i, h_i), where $h_i = (\text{inj}M_i)^{-2} g_i$ is rescaled so that

$$\begin{aligned} \text{inj}(M_i, h_i) &= 1, \\ |\sec(M_i, h_i)| &\leq (\text{inj}(M_i, g_i))^2 \cdot K = K_i \to 0. \end{aligned}$$

The previous theorem, together with the fundamental theorem, then implies that some subsequence of (M_i, p_i, h_i) will converge in the pointed C^α, $\alpha < 1$, topology to a flat manifold (M, p, g).

The first observation about (M, p, g) is that $\text{inj}(p) \leq 1$. This follows because the conjugate radius for $(M_i, h_i) \geq \pi/\sqrt{K_i} \to \infty$, so Klingenberg's estimate for the injectivity radius implies that there must be a geodesic loop of length 2 at $p_i \in M_i$. Since $(M_i, p_i, h_i) \to (M, p, g)$ in the pointed C^α topology, the geodesic loops must converge to a geodesic loop in M based at p of length 2. Hence, $\text{inj}(M) \leq 1$.

The other contradictory observation is that $(M, g) = (\mathbb{R}^n, \text{can})$. Recall that $\text{vol}B(p_i, 1) \geq v$ in (M_i, g_i), so relative volume comparison shows that there is a $v'(n, K, v)$ such that $\text{vol}B(p_i, r) \geq v' \cdot r^n$, for $r \leq 1$. The rescaled manifold (M_i, h_i) therefore satisfies $\text{vol}B(p_i, r) \geq v' \cdot r^n$, for $r \leq (\text{inj}(M_i, g_i))^{-1}$. Using again that $(M_i, p_i, h_i) \to (M, p, g)$ in the pointed C^α topology, we get $\text{vol}B(p, r) \geq v' \cdot r^n$ for all r. Since (M, g) is flat, this shows that it must be Euclidean space.

This last statement requires some justification. Let M be a complete flat manifold. As the elements of the fundamental group act by isometries on Euclidean space, we know that they must have infinite order (any isometry of finite order is a rotation around a point and therefore has a fixed point). Therefore, if M is not simply connected, then there is an intermediate covering $\hat{M}$:

$$\mathbb{R}^n \to \hat{M} \to M,$$

where $\pi_1\left(\hat{M}\right) = \mathbb{Z}$. This means that $\hat{M}$ looks like a cylinder. Hence, for any $p \in \hat{M}$ we must have

$$\lim_{r \to \infty} \frac{\text{vol}B\,(p, r)}{r^{n-1}} < \infty.$$

The same must then also hold for M itself, contradicting our volume growth assumption. □

This lemma was proved by a more direct method by Cheeger, but we have included this, perhaps more convoluted, proof in order to show how our convergence theory can be used. The lemma also shows that the convergence theorem of Riemannian geometry remains true if the injectivity radius bound is replaced by a lower bound on the volume of 1-balls. The following result is now immediate.

Corollary 4.6 (J. Cheeger, 1967) *Let $n \geq 2$, $\Lambda, D, v \in (0, \infty)$ be given. The class of closed Riemannian n-manifolds with*

$$|\sec| \leq \Lambda,$$
$$\mathrm{diam} \leq D,$$
$$\mathrm{vol} \geq v$$

is precompact in the C^α topology for any $\alpha \in (0, 1)$ and in particular, contains only finitely many diffeomorphism types.

This convergence theorem of Riemannian geometry can be generalized in another interesting direction, as observed by S.-h. Zhu.

Theorem 4.7 *Given $i_0, k > 0$, there exist Q, r depending on i_0, k such that any manifold (M, g) with*

$$\sec \geq -k^2,$$
$$\mathrm{inj} \geq i_0$$

satisfies $\|(M, g)\|_{C^{1},r} \leq Q$.

Proof. It suffices to get some Hessian estimate for distance functions $d(x) = d(x, p)$. We have, as before, that $\nabla^2 d(x)$ has eigenvalues $\leq k \cdot \coth(k \cdot d(x))$. Conversely, if $r(x_0) < i_0$, then $r(x)$ is supported from below by $f(x) = i_0 - d(x, y_0)$, where $y_0 = \gamma(i_0)$ and γ is the unique unit speed geodesic that minimizes the distance from p to x_0. Thus, $\nabla^2 d(x) \geq \nabla^2 f(x)$ at x_0. But $\nabla^2 f(x)$ has eigenvalues $\geq -k \cdot \coth(d(x_0, y_0) \cdot k) = -k \cdot \coth(k(i_0 - r(x_0)))$ at x_0. Hence, we have two-sided bounds for $\nabla^2 d(x)$ on appropriate sets. The proof can then be finished as before. □

This theorem is, interestingly enough, optimal. Consider rotationally symmetric metrics $dr^2 + f_\varepsilon^2(r)d\theta^2$, where f_ε is concave and satisfies $f_\varepsilon(r) = r,\ 0 \leq r \leq 1-\varepsilon$, and $f_\varepsilon(r) = \frac{3}{4}r,\ 1 + \varepsilon \leq r$. These metrics have $\sec \geq 0$ and $\mathrm{inj} = \infty$. As $\varepsilon \to 0$, we get a $C^{1,1}$ manifold with a $C^{0,1}$ Riemannian metric (M, g). In particular, $\|(M, g)\|_{C^{0,1},r} < \infty$ for all r. Limit spaces of sequences with $\mathrm{inj} \geq i_0,\ \sec \geq k$ can therefore not in general be assumed to be smoother than the above example.

With a more careful construction, we can also find g_ε with $g_\varepsilon(r) = \sin(r),\ 0 \leq r \leq \frac{\pi}{2} - \varepsilon$, and $g_\varepsilon(r) = 1,\ r \geq 1 + \varepsilon$. Then the metric $dr^2 + g_\varepsilon^2(r)d\theta^2$ satisfies $|\sec| \leq 4$ and $\mathrm{inj} \geq \frac{1}{4}$. As $\varepsilon \to 0$, we get a limit metric that is $C^{1,1}$. So while we may suspect (this is still unknown) that limit metrics from the convergence theorem are $C^{1,1}$, we prove only that they are $C^{0,1}$. In the next section we shall show that they are in fact $C^{1,\alpha}$ for all $\alpha < 1$.

10.5 Harmonic Norms and Ricci Curvature

To get better estimates on the norms, we must use some more analysis. The idea of using harmonic coordinates for similar purposes goes back to [30]. In [51] it was shown that manifolds with bounded sectional curvature and lower bounds for the injectivity radius admit harmonic coordinates on balls of an a priori size. This result was immediately seized by the geometry community and put to use in improving the theorems from the previous section. At the same time, Nikolaev developed a different, more synthetic approach to these ideas. For the whole story we refer the reader to R. Greene's survey in [41]. Here we shall develop these ideas from a different point of view initiated by Anderson.

10.5.1 The Harmonic Norm

We shall now define another norm, called the *harmonic norm* and denoted

$$\|A \subset (M, g)\|^{harm}_{C^{m,\alpha},r} .$$

The only change in our previous definition is that condition n4 is replaced by the requirement that $\varphi_s^{-1} : U_s \to \mathbb{R}^n$ be harmonic with respect to the Riemannian metric g on M. We can use the elliptic estimates to compare this norm with our old norm. Namely, recall that in harmonic coordinates $\Delta = g^{ij}\partial_i\partial_j$, conditions n2 and n3 insure that these coefficients are bounded in the required way. Therefore, if $u : U \to \mathbb{R}$ is any harmonic function, then we get that on compact subsets $K \subset U \cap U_s$,

$$\|u\|_{C^{m+1,\alpha},K} \leq C\,\|u\|_{C^{\alpha},U} .$$

Using a coordinate function φ_t^{-1} as u then shows that we can get bounds for the transition functions on compact subsets of their domains. Changing the scale will then allow us to conclude that for each $r_1 < r_2$, there is a constant $C = C(n, m, \alpha, r_1, r_2)$ such that

$$\|A \subset (M, g)\|_{C^{m,\alpha},r_1} \leq C\,\|A \subset (M, g)\|^{harm}_{C^{m,\alpha},r_2} .$$

We can then show the harmonic analogue to the fundamental theorem.

Corollary 5.1 *For given $Q > 0$, $n \geq 2$, $m \geq 0, \alpha \in (0, 1]$, and $r > 0$ consider the class of complete, pointed Riemannian n-manifolds (M, p, g) with $\|(M, g)\|^{harm}_{C^{m,\alpha},r} \leq Q$. This class is closed in the pointed $C^{m,\alpha}$ topology and compact in the pointed $C^{m,\beta}$ topology for all $\beta < \alpha$.*

The only issue to worry about is whether it is really true that limit spaces have $\|(M, g)\|^{harm}_{C^{m,\alpha},r} \leq Q$. But one can easily see that harmonic charts converge to harmonic charts.

We shall also need to use the corresponding properties for the harmonic norm.

Proposition 5.2 (M. Anderson, 1990) *If $A \subset (M, g)$ is precompact, then:*

(1) $\|A \subset (M,g)\|^{harm}_{C^{m,\alpha},r} = \left\|A \subset \left(M, \lambda^2 g\right)\right\|^{harm}_{C^{m,\alpha},\lambda r}$ *for all* $\lambda > 0$.

(2) *The function* $r \to \|A \subset (M,g)\|^{harm}_{C^{m,\alpha},r}$ *is continuous. Moreover, when* $m \geq 1$, *it converges to* 0 *as* $r \to 0$.

(3) *Suppose* $(M_i, p_i, g_i) \to (M, p, g)$ *in* $C^{m,\alpha}$ *and in addition that* $m \geq 1$. *Then for* $A \subset M$ *we can find precompact domains* $A_i \subset M_i$ *such that*

$$\|A_i\|^{harm}_{C^{m,\alpha},r} \to \|A\|^{harm}_{C^{m,\alpha},r}$$

for all $r > 0$. *When all the manifolds are closed, we can let* $A = M$ *and* $A_i = M_i$.

(4) $\|A \subset (M,g)\|^{harm}_{C^{m,\alpha},r} = \sup_{p \in A} \|\{p\} \subset (M,g)\|^{harm}_{C^{m,\alpha},r}$.

Proof. Properties (1) and (2) are proved as before. For the statement that the norm goes to zero as the scale decreases, just solve the Dirichlet problem as we did when existence of harmonic coordinates was established. Here it was necessary to have coordinates around every point $p \in M$ such that in these coordinates the metric satisfies $g_{ij} = \delta_{ij}$ and $\partial_k g_{ij} = 0$ at p. If $m \geq 1$, then it is easy to show that any coordinate system around p can be changed in such a way that the metric has the desired properties.

(3) The proof of this statement is necessarily somewhat different, as we must use and produce harmonic coordinates. Let the setup be as before. First we show the easy part:

$$\liminf \|A_i\|^{harm}_{C^{m,\alpha},r} \geq \|A\|^{harm}_{C^{m,\alpha},r}.$$

To this end, select $Q > \liminf \|A_i\|^{harm}_{C^{m,\alpha},r}$. For large i we can then select charts $\varphi_{i,s} : (0,r) \to M_i$ with the requisite properties. After passing to a subsequence, we can make these charts converge to charts $\varphi_s = \lim f_i^{-1} \circ \varphi_{i,s} : B(0,r) \to M$. Since the metrics converge in $C^{m,\alpha}$, the Laplacians of the inverse functions must also converge. Hence, the limit charts are harmonic as well. We can then conclude that $\|A\|^{harm}_{C^{m,\alpha},r} \leq Q$.

For the reverse inequality

$$\limsup \|A_i\|^{harm}_{C^{m,\alpha},r} \leq \|A\|^{harm}_{C^{m,\alpha},r},$$

select $Q > \|A\|^{harm}_{C^{m,\alpha},r}$. Then, from the continuity of the norm we can find $\varepsilon > 0$ such that also $\|A\|^{harm}_{C^{m,\alpha},r+\varepsilon} < Q$. For this scale, select charts $\varphi_s : B(0, r+\varepsilon) \to U_s \subset M$ satisfying the usual conditions. Now define $U_{i,s} = f_i(\varphi_s(B(0, r+))) \subset M_i$. This is clearly a closed disc with smooth boundary $\partial U_{i,s} = f_i(\varphi_s(\partial B(0, r+\varepsilon/2)))$. On each $U_{i,s}$ solve the Dirichlet problem

$$\begin{aligned} \psi_{i,s} &: U_{i,s} \to \mathbb{R}^n, \\ \Delta_{g_i} \psi_{i,s} &= 0, \\ \psi_{i,s} &= \varphi_s \circ f_i^{-1} \text{ on } \partial U_{i,s}. \end{aligned}$$

The inverse of $\psi_{i,s}$, if it exists, will then be a coordinate map $B(0,r) \to U_{i,s}$. On the set $B(0, r+\varepsilon/2)$ we can now compare $\psi_{i,s} \circ f_i \circ \varphi_s$ with the identity map I. Note that these maps agree on the boundary of $B(0, r+\varepsilon/2)$. We know that $f_i^* g_i \to g$ in the fixed coordinate system φ_s. Now pull these metrics back to $B\left(0, r+\frac{\varepsilon}{2}\right)$ and refer to them as $g\left(=\varphi_s^* g\right)$ and $g_i\left(=\varphi_s^* f_i^* g_i\right)$. In this way the harmonicity conditions read $\Delta_g I = 0$ and $\Delta_{g_i} \psi_{i,s} \circ f_i \circ \varphi_s = 0$. In these coordinates we have the correct bounds for the operator

$$\Delta_{g_i} = g_i^{kl} \partial_k \partial_l + \frac{1}{\sqrt{\det g_i}} \partial_k \left(\sqrt{\det g_i} \cdot g_i^{kl}\right) \partial_l$$

to use the elliptic estimates for domains with smooth boundary. Note that this is where the condition $m \geq 1$ becomes important, so that we can bound

$$\frac{1}{\sqrt{\det g_i}} \partial_k \left(\sqrt{\det g_i} \cdot g_i^{kl}\right)$$

in C^α. The estimates then imply

$$\begin{aligned} \left\| I - \psi_{i,s} \circ f_i \circ \varphi_s \right\|_{C^{m+1,\alpha}} &\leq C \left\| \Delta_{g_i} \left(I - \psi_{i,s} \circ f_i \circ \varphi_s\right) \right\|_{C^{m-1,\alpha}} \\ &= C \left\| \Delta_{g_i} I \right\|_{C^{m-1,\alpha}}. \end{aligned}$$

However, we have that

$$\begin{aligned} \left\| \Delta_{g_i} I \right\|_{C^{m-1,\alpha}} &= \left\| \frac{1}{\sqrt{\det g_i}} \partial_k \left(\sqrt{\det g_i} \cdot g_i^{kl}\right) \right\|_{C^{m-1,\alpha}} \\ &\to \left\| \frac{1}{\sqrt{\det g}} \partial_k \left(\sqrt{\det g} \cdot g^{kl}\right) \right\|_{C^{m-1,\alpha}} \\ &= \left\| \Delta_g I \right\|_{C^{m-1,\alpha}} = 0. \end{aligned}$$

In particular, we must have

$$\left\| I - \psi_{i,s} \circ f_i \circ \varphi_s \right\|_{C^{m+1,\alpha}} \to 0.$$

It is now evident that $\psi_{i,s}$ must become coordinates for large i. Also, these coordinates will show that $\|A_i\|^{harm}_{C^{m,\alpha},r} < Q$ for large i.

(4) Since there is no transition function condition to be satisfied in the definition of $\|A\|^{harm}_{C^{m,\alpha},r}$, it is obvious that

$$\|A \cup B\|^{harm}_{C^{m,\alpha},r} = \max\left\{\|A\|^{harm}_{C^{m,\alpha},r}, \|B\|^{harm}_{C^{m,\alpha},r}\right\}.$$

This shows that the norm is always realized locally. □

10.5.2 Ricci Curvature and the Harmonic Norm

The most important feature about harmonic coordinates is that when one uses them, it looks as though the metric can be controlled by the Ricci curvature. This is exploited in the next lemma, where we show how one can bound the harmonic $C^{1,\alpha}$ norm in terms of the C^1 norm and Ricci curvature.

Lemma 5.3 (M. Anderson, 1990) *Suppose we have that a Riemannian manifold (M, g) has bounded Ricci curvature $|\mathrm{Ric}| \le \Lambda$. Then for any $r_1 < r_2$, $K \ge \|A \subset (M, g)\|^{harm}_{C^1, r_2}$, and $\alpha \in (0, 1)$ we can find $C(n, \alpha, K, r_1, r_2, \Lambda)$ such that*

$$\|A \subset (M, g)\|^{harm}_{C^{1,\alpha}, r_1} \le C(n, \alpha, K, r_1, r_2, \Lambda).$$

Moreover, if g is an Einstein metric $\mathrm{Ric} = kg$, then for each integer m we can find a constant $C(n, \alpha, K, r_1, r_2, k, m)$ such that

$$\|A \subset (M, g)\|^{harm}_{C^{m+1,\alpha}, r_1} \le C(n, \alpha, K, r_1, r_2, k, m).$$

Proof. We just need to bound the metric components g_{ij} in some fixed harmonic coordinates. In these coordinates we have that $\Delta = g^{ij}\partial_i\partial_j$. Given that $\|A \subset (M, g)\|^{harm}_{C^1, r_2} \le K$, we can conclude that we have the necessary conditions on the coefficients of $\Delta = g^{ij}\partial_i\partial_j$ to use the elliptic estimates

$$\left\|g_{ij}\right\|_{C^{1,\alpha}, B(0,r_1)} \le C(n, \alpha, K, r_1, r_2)\left(\left\|\Delta g_{ij}\right\|_{C^0, B(0,r_2)} + \left\|g_{ij}\right\|_{C^\alpha, B(0,r_2)}\right).$$

Now use that

$$\Delta g_{ij} = -2\mathrm{Ric} + 2Q(g, \partial g)$$

to conclude that

$$\left\|\Delta g_{ij}\right\|_{C^0, B(0,r_2)} \le 2\Lambda\left\|g_{ij}\right\|_{C^0, B(0,r_2)} + \hat{C}\left\|g_{ij}\right\|_{C^1, B(0,r_2)}.$$

Using this we then have

$$\begin{aligned}\left\|g_{ij}\right\|_{C^{1,\alpha}, B(0,r_1)} &\le C(n, \alpha, K, r_1, r_2)\left(\left\|\Delta g_{ij}\right\|_{C^0, B(0,r_2)} + \left\|g_{ij}\right\|_{C^\alpha, B(0,r_2)}\right) \\ &\le C(n, \alpha, K, r_1, r_2)\left(2\Lambda + \hat{C} + 1\right)\left\|g_{ij}\right\|_{C^1, B(0,r_2)}.\end{aligned}$$

For the Einstein case we can use a bootstrap method, as we get $C^{1,\alpha}$ bounds on the Ricci tensor from the Einstein equation $\mathrm{Ric} = kg$. Thus, we have that Δg_{ij} is bounded in C^α rather than just C^0. Hence,

$$\begin{aligned}\left\|g_{ij}\right\|_{C^{2,\alpha}, B(0,r_1)} &\le C(n, \alpha, K, r_1, r_2)\left(\left\|\Delta g_{ij}\right\|_{C^\alpha, B(0,r_2)} + \left\|g_{ij}\right\|_{C^\alpha, B(0,r_2)}\right) \\ &\le C(n, \alpha, K, r_1, r_2, k) \cdot C \cdot \left\|g_{ij}\right\|_{C^{1,\alpha}, B(0,r_2)}.\end{aligned}$$

This gives $C^{2,\alpha}$ bounds on the metric. Then, of course, Δg_{ij} is bounded $C^{1,\alpha}$, and thus the metric will be bounded in $C^{3,\alpha}$. Clearly, one can iterate this until one gets $C^{m+1,\alpha}$ bounds on the metric. □

This result combined with the fundamental theorem gives a very interesting compactness result.

Corollary 5.4 *For given $n \geq 2, Q, r, \Lambda \in (0, \infty)$ consider the class of Riemannian n-manifolds with*

$$\|(M, g)\|^{harm}_{C^1, r} \leq Q,$$
$$|\mathrm{Ric}| \leq \Lambda.$$

This class is precompact in the $C^{1,\alpha}$ topology for any $\alpha \in (0, 1)$. Moreover, if we take the subclass of Einstein manifolds, then this class is compact in the $C^{m,\alpha}$ topology for any $m \geq 0$ and $\alpha \in (0, 1)$.

We can now prove our generalizations of the convergence theorems from the last section.

Theorem 5.5 (M. Anderson, 1990) *Given $n \geq 2$ and $\alpha \in (0, 1)$, $\Lambda \in (0, \infty)$, $i_0 > 0$, one can for each $Q > 0$ find $r(n, \alpha, \Lambda, i_0) > 0$ such any complete Riemannian n-manifold (M, g) with*

$$|\mathrm{Ric}| \leq \Lambda,$$
$$\mathrm{inj} \geq i_0$$

satisfies $\|(M, g)\|^{harm}_{C^{1,\alpha}, r} \leq Q$.

Proof. The proof goes by contradiction. So suppose that there is a $Q > 0$ such that for each $i \geq 1$ there is a Riemannian manifold (M_i, g_i) with

$$|\mathrm{Ric}| \leq \Lambda,$$
$$\mathrm{inj} \geq i_0,$$
$$\|(M_i, g_i)\|^{harm}_{C^{1,\alpha}, i^{-1}} > Q.$$

Using that the norm goes to zero as the scale goes to zero, and that it is continuous as a function of the scale, we can for each i find $r_i \in (0, i^{-1})$ such that $\|(M_i, g_i)\|^{harm}_{C^{1,\alpha}, r_i} = Q$. Now rescale these manifolds: $h_i = r_i^{-2} g_i$. Then we have that (M_i, h_i) satisfies

$$|\mathrm{Ric}| \leq r_i^{-2}\Lambda,$$
$$\mathrm{inj} \geq r_i^{-1} i_0,$$
$$\|(M_i, h_i)\|^{harm}_{C^{1,\alpha}, 1} = Q.$$

We can then select $p_i \in M_i$ such that $\|p_i \in (M_i, h_i)\|^{harm}_{C^{1,\alpha}, 1} \in \left[\frac{Q}{2}, Q\right]$.

The first important step is now to use the bounded Ricci curvature of (M_i, h_i) to conclude that in fact the $C^{1,\gamma}$ norm must be bounded for any $\gamma \in (\alpha, 1)$. Then we can assume by the fundamental theorem that the sequence (M_i, p_i, h_i) converges in the pointed $C^{1,\alpha}$ topology, to a Riemannian manifold (M, p, g) of class $C^{1,\gamma}$. Since the $C^{1,\alpha}$ norm is continuous in the $C^{1,\alpha}$ topology we can conclude that $\|p \in (M, g)\|^{harm}_{C^{1,\alpha}, 1} \in \left[\frac{Q}{2}, Q\right]$.

The second thing we can prove is that $(M, g) = (\mathbb{R}^n, \text{can})$. This clearly violates what we just established about the norm of the limit space. To see that the limit space is Euclidean space, recall that the manifolds in the sequence (M_i, h_i) are covered by harmonic coordinates that converge to harmonic coordinates in the limit space. In these harmonic coordinates the metric components satisfy

$$\frac{1}{2}\Delta h_{kl} + Q(h, \partial h) = -\text{Ric}_{kl}.$$

But we know that

$$|-\text{Ric}_{kl}| \leq r_i^{-2}\Lambda h_{kl}$$

and that the h_{kl} converge in the $C^{1,\alpha}$ topology to the metric coefficients g_{kl} for the limit metric. We can therefore conclude that the limit manifold is covered by harmonic coordinates and that in these coordinates the metric satisfies:

$$\frac{1}{2}\Delta g_{kl} + Q(g, \partial g) = 0.$$

The limit metric is therefore a weak solution to the Einstein equation $\text{Ric} = 0$ and must therefore be a smooth Ricci flat Riemannian manifold. It is now time to use that: $\text{inj}(M_i, h_i) \to \infty$. In the limit space we have that any geodesic is a limit of geodesics from the sequence (M_i, h_i), since the Riemannian metrics converge in the $C^{1,\alpha}$ topology. If a geodesic in the limit is a limit of segments, then it must itself be a segment. We can then conclude that since $\text{inj}(M_i, h_i) \to \infty$, any finite length geodesic must be a segment. This, however, implies that $\text{inj}(M, g) = \infty$. The splitting theorem then shows that the limit space is Euclidean space. □

From this theorem we immediately get

Corollary 5.6 (M. Anderson, 1990) *Let $n \geq 2$ and $\Lambda, D, i \in (0, \infty)$ be given. The class of closed Riemannian n-manifolds satisfying*

$$|\text{Ric}| \leq \Lambda,$$
$$\text{diam} \leq D,$$
$$\text{inj} \geq i$$

is precompact in the $C^{1,\alpha}$ topology for any $\alpha \in (0, 1)$ and in particular contains only finitely many diffeomorphism types.

Notice how the above theorem depended on the characterization of Euclidean space we obtained from the splitting theorem. There are, however, several other similar characterizations of Euclidean space. One of the most interesting ones uses volume pinching.

10.5.3 Volume Pinching

The idea is to use the relative volume comparison theorem rather than the splitting theorem. We know form the exercises to Chapter 9 that Euclidean space is the only space with

$$\begin{aligned} \text{Ric} &\geq 0, \\ \lim_{r\to\infty} \frac{\text{vol}B\,(p,r)}{\omega_n r^n} &= 1, \end{aligned}$$

where $\omega_n r^n$ is the volume of a Euclidean ball of radius r. This result has a very interesting gap phenomenon associated with it, when one assumes the stronger hypothesis that the space is Ricci flat.

Lemma 5.7 (M. Anderson, 1990) *For each $n \geq 2$ there is an $\varepsilon(n) > 0$ such that any complete Ricci flat manifold (M, g) that satisfies*

$$\text{vol}B\,(p,r) \geq (\omega_n - \varepsilon)\, r^n$$

for some $p \in M$ is isometric to Euclidean space.

Proof. First observe that on any complete Riemannian manifold with Ric ≥ 0, relative volume comparison can be used to show that

$$\text{vol}B\,(p,r) \geq (1-\varepsilon)\,\omega_n r^n$$

as long as

$$\lim_{r\to\infty} \frac{\text{vol}B\,(p,r)}{\omega_n r^n} \geq (1-\varepsilon)\,.$$

It is then easy to see that if this holds for one p, then it must hold for all p. Moreover, if we scale the metric to $(M, \lambda^2 g)$, then the same volume comparison still holds, as the lower curvature bound Ric ≥ 0 can't be changed by scaling.

If our assertion were not true, then we could for each integer i find Ricci flat manifolds (M_i, g_i) with

$$\begin{aligned} \lim_{r\to\infty} \frac{\text{vol}B\,(p_i,r)}{\omega_n r^n} &\geq \left(1 - i^{-1}\right), \\ \|(M_i, g_i)\|^{harm}_{C^{1,\alpha},r} &\neq 0 \quad \text{for all} \quad r > 0. \end{aligned}$$

By scaling these metrics suitably, it is then possible to arrange it so that we have a sequence of Ricci flat manifolds (M_i, q_i, h_i) with

$$\begin{aligned} \lim_{r\to\infty} \frac{\text{vol}B\,(q_i,r)}{\omega_n r^n} &\geq \left(1 - i^{-1}\right), \\ \|(M_i, g_i)\|^{harm}_{C^{1,\alpha},1} &\leq 1, \\ \|q_i \in (M_i, g_i)\|^{harm}_{C^{1,\alpha},1} &\in [0.5, 1]\,. \end{aligned}$$

From what we already know, we can then extract a subsequence that converges in the $C^{m,\alpha}$ topology to a Ricci flat manifold (M, q, g). In particular, we must have that metric balls of a given radius converge and that the volume forms converge. Thus, the limit space must satisfy

$$\lim_{r\to\infty} \frac{\mathrm{vol}B(q,r)}{\omega_n r^n} = 1.$$

This means that we have maximal possible volume for all metric balls, and thus the manifold must be Euclidean. This, however, violates the continuity of the norm in the $C^{1,\alpha}$ topology, as the norm for the limit space would then have to be zero. □

Corollary 5.8 *Let $n \geq 2$, $-\infty < \lambda \leq \Lambda < \infty$, and $D, i_0 \in (0,\infty)$ be given. There is a $\delta = \delta\left(n, \lambda \cdot i_0^2\right)$ such that the class of closed Riemannian n-manifolds satisfying*

$$\begin{aligned}(n-1)\Lambda &\geq \mathrm{Ric} \geq (n-1)\lambda,\\ \mathrm{diam} &\leq D,\\ \mathrm{vol}B(p,i_0) &\geq (1-\delta)\,v(n,\lambda,i_0)\end{aligned}$$

is precompact in the $C^{1,\alpha}$ topology for any $\alpha \in (0,1)$ and in particular contains only finitely many diffeomorphism types.

Proof. We use the same techniques as when we had an injectivity radius bound. Instead, we observe that if we have a sequence (M_i, p_i, h_i) where $h_i = k_i^2 g_i$, $k_i \to \infty$, and the (M_i, g_i) lie in the above class, then the volume condition now reads

$$\begin{aligned}\mathrm{vol}B_{h_i}(p_i, i_0\cdot k_i) &= k_i^n \mathrm{vol}B_{g_i}(p_i, i_0)\\ &\geq k_i^n(1-\delta)\,v(n,\lambda,i_0)\\ &= (1-\delta)\,v\left(n, \lambda\cdot k_i^{-2}, i_0\cdot k_i\right).\end{aligned}$$

From relative volume comparison we can then conclude that for $r \leq i_0 \cdot k_i$ and very large i,

$$\mathrm{vol}B_{h_i}(p_i, r) \geq (1-\delta)\,v\left(n, \lambda\cdot k_i^{-2}, r\right) \sim (1-\delta)\,\omega_n r^n.$$

In the limit space we must therefore have

$$\mathrm{vol}B(p,r) \geq (1-\delta)\,\omega_n r^n \text{ for all } r.$$

This limit space is also Ricci flat and is therefore Euclidean space. The rest of the proof goes as before, by getting a contradiction with the continuity of the norms. □

10.5.4 Curvature Pinching

Let us now turn our attention to some applications of these compactness theorems. One natural subject to explore is that of *pinching* results. Recall that we showed earlier that complete constant curvature manifolds have a uniquely defined universal covering. It is natural to ask whether one can in some topological sense still expect this to be true when one has close to constant curvature. Now, any Riemannian manifold (M, g) has curvature close to zero if we multiply the metric by a large scalar. Thus, some additional assumptions must come into play.

We start out with the simpler problem of considering Ricci pinching and then use this in the context of curvature pinching below. The results are very simple consequences of the convergence theorem we have already presented.

Theorem 5.9 *Given $n \geq 2$, $i, D \in (0, \infty)$, and $\lambda \in \mathbb{R}$, there is an $\varepsilon = \varepsilon(n, \lambda, D, i) > 0$ such that any closed Riemannian n-manifold (M, g) with*

$$\begin{aligned} \mathrm{diam} &\leq D, \\ \mathrm{inj} &\geq i, \\ |\mathrm{Ric} - \lambda g| &\leq \varepsilon \end{aligned}$$

is $C^{1,\alpha}$ close to an Einstein metric with Einstein constant λ.

Proof. We already know that this class is precompact in the $C^{1,\alpha}$ topology no matter what ε we choose. If the result were not true, we could therefore find a sequence $(M_i, g_i) \to (M, g)$ that converges in the $C^{1,\alpha}$ topology to a closed Riemannian manifold of class $C^{1,\alpha}$, where in addition, $\left|\mathrm{Ric}_{g_i} - \lambda g_i\right| \to 0$. Using harmonic coordinates as usual we can therefore conclude that the metric on the limit space must be a weak solution to

$$\frac{1}{2}\Delta g + Q(g, \partial g) = -\lambda g.$$

But this means that the limit space is actually Einstein, with Einstein constant λ, thus, contradicting that the spaces (M_i, g_i) were not close to such Einstein metrics. □

Using the compactness theorem for manifolds with almost maximal volumes we see that the injectivity radius condition could have been replaced with an almost maximal volume condition.

Theorem 5.10 *Given $n \geq 2$, $v, D \in (0, \infty)$, and $\lambda \in \mathbb{R}$, there is an $\varepsilon = \varepsilon(n, \lambda, D, i) > 0$ such that any closed Riemannian n-manifold (M, g) with*

$$\begin{aligned} \mathrm{diam} &\leq D, \\ \mathrm{vol} &\geq v, \\ |\sec - \lambda| &\leq \varepsilon \end{aligned}$$

is $C^{1,\alpha}$ close to a metric of constant curvature λ.

Proof. In this case we first observe that Cheeger's lemma gives us a lower bound for the injectivity radius. The previous theorem then shows that such metrics must be close to Einstein metrics. We now have to check that if $(M_i, g_i) \to (M, g)$, where $|\sec_{g_i} - \lambda| \to 0$ and $\mathrm{Ric}_g = (n-1)\lambda g$, then in fact (M, g) has constant curvature λ. To see this, it is perhaps easiest to observe that if $M_i \ni p_i \to p \in M$, then we can use polar coordinates around these points, and write out the metric $(g_{i,\alpha\beta})$ around p_i and $(g_{\alpha\beta})$ around p. Since the metrics converge in $C^{1,\alpha}$, we certainly have that the matrices $(g_{i,\alpha\beta})$ converge to $(g_{\alpha\beta})$. Using the curvature pinching, we conclude from Chapter 6 that

$$\mathrm{sn}^2_{\lambda+\varepsilon_i}(r) \cdot (\delta_{\alpha\beta})_{2\leq\alpha,\beta\leq n} \leq (g_{i,\alpha\beta}(r,\theta))_{2\leq\alpha,\beta\leq n} \leq \mathrm{sn}^2_{\lambda-\varepsilon_i}(r) \cdot (\delta_{\alpha\beta})_{2\leq\alpha,\beta\leq n},$$

where $\varepsilon_i \to 0$. In the limit we therefore have

$$\mathrm{sn}^2_{\lambda}(r) \cdot (\delta_{\alpha\beta})_{2\leq\alpha,\beta\leq n} \leq (g_{\alpha\beta}(r,\theta))_{2\leq\alpha,\beta\leq n} \leq \mathrm{sn}^2_{\lambda}(r) \cdot (\delta_{\alpha\beta})_{2\leq\alpha,\beta\leq n}.$$

This implies that the limit metric has constant curvature k. □

It is interesting that we had to go back and use the more geometric estimates for distance functions in order to prove the curvature pinching, while the Ricci pinching could be handled more easily with analytic techniques using harmonic coordinates. One can actually prove the curvature result with purely analytic techniques, but this requires that we study convergence in a more general setting where one uses L^p norms and estimates. This has been developed rigorously and can be used to improve the above results to situations were one has only L^p curvature pinching rather than the L^∞ pinching we use here (see [70] and [71]).

When the curvature λ is positive, some of the assumptions in the above theorems are in fact not necessary. For instance, Myers' estimate for the diameter makes the diameter hypothesis superfluous. For the Einstein case this seems to be as far as we can go. In the positive curvature case we can do much better. In even dimensions, we already know from Chapter 6, that manifolds with positive curvature have both bounded diameter and lower bounds for the injectivity radius, provided that there is an upper curvature bound. We can therefore show

Corollary 5.11 *Given $2n \geq 2$, and $\lambda > 0$, there is an $\varepsilon = \varepsilon(n, \lambda) > 0$ such that any closed Riemannian $2n$-manifold (M, g) with*

$$|\sec - \lambda| \leq \varepsilon$$

is $C^{1,\alpha}$ close to a metric of constant curvature λ.

This corollary is, in fact, also true in odd dimensions. This was proved by Grove-Karcher-Ruh in [45]. Notice that convergence techniques are not immediately applicable because there are no lower bounds for the injectivity radius. Their pinching constant is also independent of the dimension.

There is a very nice result of Micallef-Moore in [?] that says that any manifold with positive isotropic curvature has the property that the universal cover is homeomorphic to the sphere. However, this doesn't generalize the above theorem, for it is not necessarily true that two manifolds with identical fundamental groups and universal covers are homotopy equivalent.

In negative curvature some special things also happen. Namely, Heintze has proved that any complete manifold with $-1 \leq \sec < 0$ has a lower volume bound when the dimension ≥ 4 (see also [42] for a more general statement). The lower volume bound is therefore an extraneous condition when doing pinching in negative curvature. Unlike the situation in positive curvature, the upper diameter bound is, however, crucial. See, e.g., [44] and [34] for counterexamples.

This leaves us with pinching around 0. As any compact Riemannian manifold can be scaled to have curvature in $[-\varepsilon, \varepsilon]$ for any ε, we do need that the diameter is bounded. The volume condition is also necessary, as the Heisenberg group from the exercises to Chapter 3 has a quotient, where there are metrics with bounded diameter and arbitrarily pinched curvature. This quotient, however, does not admit a flat metric. Gromov was nevertheless able to classify all n-manifolds with

$$\begin{aligned} |\sec| &\leq \varepsilon(n), \\ \operatorname{diam} &\leq 1 \end{aligned}$$

for some very small $\varepsilon(n) > 0$. More specifically, they all have a finite cover that is a quotient of a nilpotent Lie group by a discrete subgroup. For more on this and collapsing in general, the reader can start by reading [35].

10.6 Further Study

Cheeger first proved his finiteness theorem and put down the ideas of C^k convergence for manifolds in [22]. They later appeared in journal form [23], but not all ideas from the thesis were presented in this paper. Also the idea of general pinching theorems as described here are due to Cheeger [24]. For more generalities on convergence and their uses we recommend the surveys by Anderson, Fukaya, Petersen, and Yamaguchi in [41]. Also for more on norms and convergence theorems the survey by Petersen in [46] might prove useful. We must also of necessity mention the enigmatic text [43] again. It was probably this book that really spread the ideas of Gromov-Hausdorff distance and the stronger convergence theorems to a wider audience. Also, the convergence theorem of Riemannian geometry, as stated here, appeared for the first time in this book.

We should also mention that S. Peters in [69] obtained an explicit estimate for the number of diffeomorphism classes in Cheeger's finiteness theorem. This also seems to be the first place where the modern statement of Cheeger's finiteness theorem is proved.

10.7 Exercises

1. Find a sequence of 1-dimensional metric spaces that Hausdorff converge to the unit cube $[0, 1]^3$ endowed with the metric coming from the maximum norm on $\mathbb{R}^3$. Then find surfaces (jungle gyms) converging to the same space.

2. C. Croke has shown that there is a universal constant $c(n)$ such that any n-manifold with $\text{inj} \geq i_0$ satisfies $\text{vol}B(p, r) \geq c(n) \cdot r^n$ for $r \leq i_0/2$. Use this to show that the class of n-dimensional manifolds satisfying $\text{inj} \geq i_0$ and $\text{vol} \leq V$ is precompact in the Gromov-Hausdorff topology.

3. Develop a Bochner formula for $\nabla^2 \left(\frac{1}{2}g(X, Y)\right)$ and $\Delta \frac{1}{2}g(X, Y)$, where X and Y are vector fields with symmetric ∇X and ∇Y. Discuss whether it is possible to devise coordinates where $\nabla^2 \left(g_{ij}\right)$ are bounded in terms of the full curvature tensor. How could this be used to show $C^{1,1}$ regularity of limit spaces from the convergence theorem of Riemannian geometry?

4. Show that in contrast with the elliptic estimates, it is not possible to find C^α bounds for a vector field X in terms of C^0 bounds on X and $\text{div}X$.

5. Define $C^{m,\alpha}$ convergence for incomplete manifolds. On such manifolds define the boundary ∂ as the set of points that lie in the completion but not in the manifold itself. Show that the class of incomplete spaces with $|\text{Ric}| \leq \Lambda$ and $\text{inj}(p) \geq \min\{i_0, i_0 \cdot d(p, \partial)\}$, $i_0 < 1$, is precompact in the $C^{1,\alpha}$ topology.

6. Define a *weighted norm* concept. That is, fix a positive function $\rho(R)$, and assume that in a pointed manifold (M, p, g) the distance spheres $S(p, R)$ have norm $\leq \rho(R)$. Prove the corresponding fundamental theorem.

7. Suppose we have a class that is compact in the $C^{m,\alpha}$ topology. Show that there is a function $f(r)$ depending on the class such that $\|(M, g)\|_{C^{m,\alpha},r} \leq f(r)$ for all elements in this class, and also, $f(r) \to 0$ as $r \to 0$.

8. The *local models* for a class of Riemannian manifolds are the types of spaces one obtains by scaling the elements of the class by a constant $\to \infty$. For example, if we consider the class of manifolds with $|\sec| \leq K$ for some K, then upon rescaling the metrics by a factor of λ^2, we have the condition $|\sec| \leq \lambda^{-2}K$, as $\lambda \to \infty$, we therefore arrive at the condition $|\sec| = 0$. This means that the local models are all the flat manifolds. Notice that we don't worry about any type of convergence here. If, in this example, we additionally assume that the manifolds have $\text{inj} \geq i_0$, then upon rescaling and letting $\lambda \to \infty$ we get the extra condition $\text{inj} = \infty$. Thus, the local model is Euclidean space. It is natural to suppose that any class that has Euclidean space as it only local model must be compact in some topology.

   Show that a class of spaces is compact in the $C^{m,\alpha}$ topology if when we rescale a sequence in this class by constants that $\to \infty$, the sequence subconverges in the $C^{m,\alpha}$ topology to Euclidean space.

9. Consider the singular Riemannian metric $dt^2 + a^2 d\theta^2, a > 1$, on $\mathbb{R}^2$. Show that there is a sequence of rotationally symmetric metrics on $\mathbb{R}^2$ with $\sec \leq 0$ and $\text{inj} = \infty$ that converge to this metric in the Gromov-Hausdorff topology.

10. Show that the class of spaces with $\text{inj} \geq i$ and $|\nabla^k \text{Ric}| \leq \Lambda$ for $k = 0, \ldots, m$ is compact in the $C^{m+1,\alpha}$ topology.

11. (S.-h. Zhu) Consider the class of complete or compact n-dimensional Riemannian manifolds with

$$\begin{aligned} \text{conj rad} &\geq r_0, \\ |\text{Ric}| &\leq \Lambda, \\ \text{vol} B(p, 1) &\geq v. \end{aligned}$$

Using the techniques from Cheeger's lemma, show that this class has a lower bound for the injectivity radius. Conclude that it is compact in the $C^{1,\alpha}$ topology.

12. Using the Eguchi-Hanson metrics from the exercises to Chapter 3, show that one cannot in general expect a compactness result for the class

$$\begin{aligned} |\text{Ric}| &\leq \Lambda, \\ \text{vol} B(p, 1) &\geq v. \end{aligned}$$

Thus, one must assume either that v is large as we did before or that there a lower bound for the conjugate radius.

13. The *weak (harmonic) norm* $\|(M, g)\|^{weak}_{C^{m,\alpha},r}$ is defined in almost the same way as the norms we have already worked with, except that we only insist that the charts $\varphi_s : B(0, r) \to U_s$ are *immersions*. The inverse is therefore only locally defined, but it still makes sense to say that it is harmonic.

    (a) Show that if (M, g) has bounded sectional curvature, then for all $Q > 0$ there is an $r > 0$ such that $\|(M, g)\|^{weak}_{C^{1,\alpha},r} \leq Q$. Thus, the weak norm can be thought of as a generalized curvature quantity.

    (b) Show that the class of manifolds with bounded weak norm is precompact in the Gromov-Hausdorff topology.

    (c) Show that (M, g) is flat iff the weak norm is zero at all scales.

11

Sectional Curvature Comparison II

In the first section we explain how one can find generalized gradients for distance functions in situations where the function might not be smooth. This critical point technique is used in the proofs of all the big theorems in this chapter. The other important technique comes from Toponogov's theorem, which we prove in the next section. The first applications of these new ideas are to sphere theorems. We then prove the soul theorem of Cheeger and Gromoll. Next, we discuss Gromov's finiteness theorem for bounds on Betti numbers and generators for the fundamental group. Finally, we show that these techniques can be adapted to prove the Grove-Petersen homotopy finiteness theorem.

Toponogov's theorem is a very useful refinement of Gauss's early realization that curvature and angle excess of triangles are related. The fact that Toponogov's theorem can be used to get information about the topology of a space seems to originate with Berger's and Toponogov's proofs of the quarter pinched sphere theorem. Toponogov himself proved these theorems in order to establish the splitting theorem for manifolds with nonnegative sectional curvature and the maximal diameter theorem for manifolds with a positive lower bound for the sectional curvature. As we saw in Chapter 9, these results can now be obtained in the Ricci curvature setting. The next use of Toponogov was to the soul theorem of Cheeger-Gromoll-Meyer. However, Toponogov's theorem is not truly needed for any of the results mentioned so far. With little effort one can actually establish these theorems with simpler comparison techniques. Still, it is convenient to have and use a workhorse theorem of universal use. It wasn't until Grove and Shiohama developed critical point theory to prove their diameter sphere theorem that Toponogov's theorem was put to serious use. Shortly after that, Gromov's Betti number estimate put these two ideas to even more nontrivial use, with his Betti number estimate for manifolds with nonnegative sectional curvature. After that, it became clear that in working with manifolds that have lower sectional curvature bounds, the two key techniques are Toponogov's theorem and the critical point theory of Grove-Shiohama. These two very geometric techniques are still used to prove many interesting and nontrivial results.

11.1 Critical Point Theory

In this generalized critical point theory, the object is to define generalized gradients of continuous functions and then use these gradients to conclude that certain regions of a manifold have no topology. The motivating basic lemma is the following:

Lemma 1.1 *Let (M, g) be a Riemannian manifold and $f : M \to \mathbb{R}$ a proper function that is C^1. If f has no critical values in the closed interval $[a, b]$, then the preimages $f^{-1}([-\infty, b])$ and $f^{-1}([-\infty, a])$ are diffeomorphic. Furthermore, there is a deformation retraction of $f^{-1}([-\infty, b])$ onto $f^{-1}([-\infty, a])$, so the inclusion $f^{-1}([-\infty, a]) \hookrightarrow f^{-1}([-\infty, b])$ is a homotopy equivalence.*

Proof. The idea is simply to move the level sets via the gradient of f. Since there are no critical points of f, we have that the gradient ∇f is nonzero everywhere on $f^{-1}([a, b])$. We then construct a bump function $\psi : M \to [0, 1]$ that is 1 on the compact set $f^{-1}([a, b])$ and zero outside some compact neighborhood of $f^{-1}([a, b])$. Finally consider the vector field

$$X = \psi \cdot \frac{\nabla f}{|\nabla f|^2}.$$

This vector field has compact support and must therefore be complete (integral curves are defined for all time). Let φ^t denote the flow for this vector field (see Figure 11.1).

For fixed $q \in M$ consider the function $t \to f\left(\varphi^t(q)\right)$. The derivative of this function is $g(X, \nabla f)$, so as long as the integral curve $t \to \varphi^t(q)$ remains in $f^{-1}([a, b])$, the function $t \to f\left(\varphi^t(q)\right)$ is linear with derivative 1. In particular, the diffeomorphism $\varphi^{b-a} : M \to M$ must carry $f^{-1}([-\infty, a])$ diffeomorphically onto $f^{-1}([-\infty, b])$.

Moreover, by flowing backwards we can define the desired retraction:

$$r_t : f^{-1}([-\infty, b]) \to f^{-1}([-\infty, b]),$$

$$r_t(p) = \begin{cases} p & \text{if} \quad f(p) \le a, \\ \varphi^{t(a-f(p))}(p) & \text{if} \quad a \le f(p) \le b. \end{cases}$$

Then $r_0 = id$, and r_1 maps $f^{-1}([-\infty, b])$ diffeomorphically onto $f^{-1}([-\infty, a])$. □

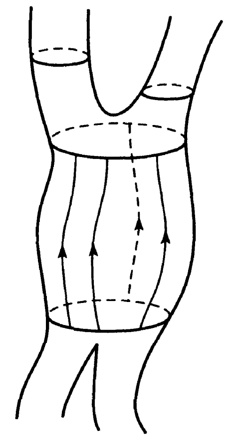


FIGURE 11.1.

Notice that we used in an essential way that the function is proper to conclude that the vector field is complete. In fact, if we delete a single point from the region $f^{-1}([a, b])$, then the function still won't have any critical values, but clearly the conclusion of the lemma is false.

We shall now try to generalize this lemma to functions that are not necessarily C^1. To minimize technicalities we shall work exclusively with distance functions. Suppose (M, g) is complete and $K \subset M$ a compact subset. Then the distance function $f(x) = d(x, K) = \min\{d(x, p) : p \in K\}$ is proper. Wherever this function is smooth, we know that it has unit gradient and must therefore be noncritical at such points. However, it might also have local maxima, and at such points we certainly wouldn't want the function to be noncritical. To define the generalized gradient for such functions, let us list all the possible values it could have. Define $\Gamma(x, K)$, or simply $\Gamma(x)$, as the set of unit vectors in T_xM that are tangent to a segment from K to x. That is, $v \in \Gamma(x, K) \subset T_xM$ if there is a unit speed segment $\sigma : [0, d(x, K)] \to M$ such that $\sigma(0) \in K$, $\sigma(d(x, K)) = x$, and $v = \dot{\sigma}(d(x, K))$. Note that σ is chosen such that no shorter curve from x to K exists. There might, however, be several such segments. In the case where f is smooth at x, we clearly have that $\{\nabla f\} = \Gamma(x, K)$. At other points, $\Gamma(x, K)$ might contain more vectors. We say that f is *regular*, or *noncritical*, at x if the set $\Gamma(x, K)$ is contained in an open hemisphere of the unit sphere in T_xM. The center of such a hemisphere is then a possible direction for the gradient of f at x. Stated differently, we have that f is regular at x iff there is a vector $v \in T_xM$ such that the angles $\angle(v, w) < \pi/2$ for all $w \in \Gamma(x, K)$. If v is a unit vector, then it will be the center of the desired hemisphere. We can quantify being regular by saying that f is α-regular at x if there exists $v \in T_xM$ such that $\angle(v, w) < \alpha$ for all $w \in \Gamma(x, K)$. Thus, f is regular at x iff it is $\pi/2$-regular. The set of vectors v that can be used in the definition of α-regularity is denoted by $G_\alpha f(x)$, where G stands for *generalized* gradient.

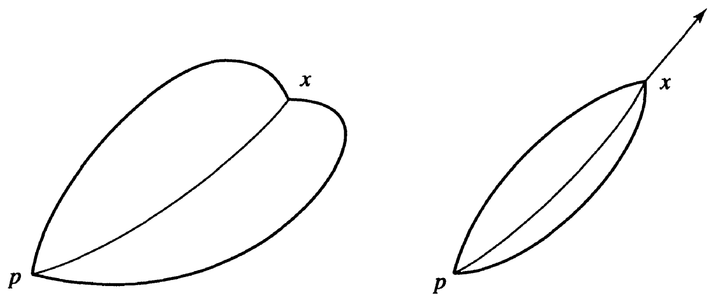


FIGURE 11.2.

Evidently, a point x is critical for $d(\cdot, p)$ if the segments from p to x spread out at x, while it is regular if they more or less point in the same direction (see Figure 11.2).

Proposition 1.2 *Suppose (M, g) and $f = d(\cdot, K)$ are as above. Then:*

(1) *$\Gamma(x, K)$ is closed and therefore compact for all x.*

(2) *The set of α-regular points is open in M.*

(3) *$G_\alpha f(x)$ is convex for all $\alpha \leq \pi/2$.*

(4) *If U is an open set of α-regular points for f, then there is a unit vector field X on U such that $X(x) \in G_\alpha f(x)$ for all $x \in U$. Furthermore, if γ is an integral curve for X and $s < t$, then*

$$f(\gamma(t)) - f(\gamma(s)) > \cos(\alpha)(t - s).$$

Proof. (1) Let $\sigma_i : [0, d(x, K)] \to M$ be a sequence of unit speed segments with $\dot{\sigma}_i(d(x, K))$ converging to some unit vector $v \in T_x M$. Clearly, $\sigma(t) = \exp_x((d(x, K) - t)v)$ is the limit of the segments σ_i and must therefore be a segment itself. Furthermore, since K is closed $\sigma(0) \in K$.

(2) Suppose $x_i \to x$, and x_i is not α-regular. We shall show that x is not α-regular. This means that for each $v \in T_x M$, we can find $w \in \Gamma(x, K)$ such that $\angle(v, w) \geq \alpha$. Now, for some fixed $v \in T_x M$, choose a sequence $v_i \in T_{x_i} M$ converging to v. For each i we can, by assumption, find $w_i \in \Gamma(x_i, K)$ with $\angle(v_i, w_i) \geq \alpha$. The sequence of unit vectors w_i must now subconverge to a vector $w \in T_x M$. Furthermore, the sequence of segments σ_i that generate w_i must also subconverge to a segment that is tangent to w. Thus, $w \in \Gamma(x, K)$.

(3) First observe that if $\alpha \leq \pi/2$, then for each $w \in T_x M$, the open cone $C_\alpha(w) = \{v \in T_x M : \angle(v, w) < \alpha\}$ is convex. Then observe that $G_\alpha f(x)$ is the intersection of the cones $C_\alpha(w)$, $w \in \Gamma(x, K)$, and is therefore itself convex.

(4) For each $p \in U$ we can find $v_p \in G_\alpha f(p)$. For each p, extend v_p to a vector field V_p. It now follows from the proof of (2) that $V_p(x) \in G_\alpha f(x)$ for x near p. We can then assume that V_p is defined on a neighborhood U_p on which

it is a generalized gradient. We can now select a locally finite collection $\{U_i\}$ of U_p's and a corresponding partition of unity φ_i. Then property (3) tells us that the vector field $V = \sum \varphi_i V_i \in G_\varepsilon f$. In particular, it is nonzero and can therefore be normalized to a unit vector field.

The last property is clearly true at points where f is smooth, because in that case the derivative of $t \to f \circ \gamma$ is $g(X, \nabla f) = \cos\angle(X, \nabla f) > \cos\alpha$. Now observe that since f is Lipschitz continuous, this function is at least absolutely continuous. This implies that $f \circ \gamma$ is differentiable a.e. and is the integral of its derivative. Whenever $f \circ \gamma$ is differentiable, we still have that its derivative is $g(X, \nabla f) > \cos\alpha$. Thus, the desired property holds. □

We can now generalize the above lemma.

Lemma 1.3 *Let (M, g) and $f = d(\cdot, K)$ be as above. Suppose that all points in $f^{-1}([a, b])$ are ε-regular. Then $f^{-1}([-\infty, a])$ is homeomorphic to $f^{-1}([-\infty, b])$, and $f^{-1}([-\infty, b])$ deformation retracts onto $f^{-1}([-\infty, a])$.*

Proof. The construction is similar but a little more involved. We can construct a compactly supported vector field X such that the flow φ^t for X satisfies

$$f\left(\varphi^t(p)\right) - f(p) > t \cdot \cos(\varepsilon), \quad t \geq 0 \quad \text{if} \quad p, \varphi^t(p) \in f^{-1}([a, b]).$$

For each $p \in f^{-1}(b)$ we can therefore find a first time $t_p \leq (b-a)/\cos\varepsilon$ for which $\varphi^{t_p}(p) \in f^{-1}(a)$. The function $p \to t_p$ is continuous and thus we get the desired retraction

$$r_t : f^{-1}([-\infty, b]) \to f^{-1}([-\infty, b]),$$

$$r_t(p) = \begin{cases} p & \text{if} \quad f(p) \leq a, \\ \varphi^{-t \cdot t_p}(p) & \text{if} \quad a \leq f(p) \leq b. \end{cases}$$ □

Note that as the level sets for f are not smooth, we can't expect to get diffeomorphic sublevels. So we essentially have the best possible situation. It is now a question of how this can be used. As a very simple result let us mention

Corollary 1.4 *Suppose K is a compact submanifold of a complete Riemannian manifold (M, g) and suppose the distance function $f = d(\cdot, K)$ is regular everywhere on $M - K$. Then M is diffeomorphic to the normal bundle of K in M. In particular, if $K = \{p\}$, then M is diffeomorphic to $\mathbb{R}^n$.*

Proof. We know that $M - K$ admits a vector field X, such that f is strictly increasing along the integral curves for X. Moreover, near K the distance function is smooth, and therefore X can be assumed to be equal to ∇f near K.

If $\nu(K) = \{v \in T_pM : p \in K \text{ and } v \perp T_pK\}$, then we have the normal exponential map

$$\exp : \nu(K) \to M.$$

On a neighborhood of the zero section in $\nu(K)$ we know that this gives a diffeomorphism onto a neighborhood of K. Also, the curves $t \to \exp(tv)$ are, for small t, integral curves for X. In particular, we have for each $v \in \nu(K)$ a unique integral curve $\gamma_v(t) : (0, \infty) \to M$ such that $\lim_{t\to 0} \dot{\gamma}_v(t) = v$. Now define our diffeomorphism $F : \nu(K) \to M$ by

$$F(0_p) = p \quad \text{for the origin in } \nu_p(K),$$
$$F(tv) = \gamma_v(t) \quad \text{where} \quad |v| = 1.$$

This clearly defines a differentiable map. For small t this is just the exponential map. The map is one-to-one since integral curves for X can't intersect. It is onto, since f is proper, and therefore integral curves for X are defined for all time and must leave every compact set (since f is increasing along integral curves). Finally, one can see that its differential is nonsingular, since the flow of a vector field always acts by local diffeomorphisms. □

11.2 Distance Comparison

In this section we shall introduce the main results that will make it possible to conclude that various distance functions are noncritical. This obviously requires some sort of angle comparison. The most important step in this direction is supplied by the Toponogov theorem (or the *hinge version* of Toponogov's theorem; there are triangle and angle versions as well). The proof we present is probably the simplest available; and is based upon an idea by H. Karcher (see [27]).

Some preparations are necessary. Let (M, g) be a Riemannian manifold. We define two very natural geometric objects:

Hinge: A *hinge* consists of two segments σ_1 and σ_2 emanating from a common point p and forming an angle α. We shall always parametrize the geodesics by arc length and assume that $\sigma_1(\ell(\sigma_1)) = p = \sigma_2(0)$. The angle α is then defined as $\alpha = \pi - \angle(\dot{\sigma}_1(\ell(\sigma_1)), \dot{\sigma}_2(0))$. Thus, the first segment ends at p, while the second begins there. The angle is the *interior* angle. See also Figure 11.3.

Triangle: A *triangle* consists of three segments that meet pairwise at three different points. In both definitions one could just use geodesics. It is then possible to have degenerate triangles where some vertices coincide without the joining geodesics being trivial. We shall not need to use such general objects here, so we confine ourselves to just using segments. In Figure 11.4 we have depicted a triangle consisting of segments, and a degenerate triangle where one of the sides is a geodesic loop and two of the vertices coincide.

Given a hinge (or a triangle), we can construct *comparison* hinges (or triangles) in the constant-curvature spaces S^n_k.

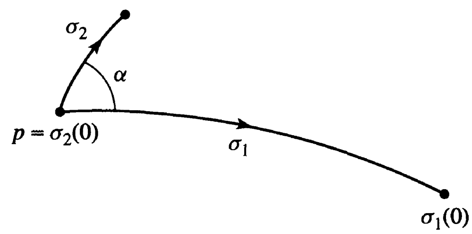


FIGURE 11.3.

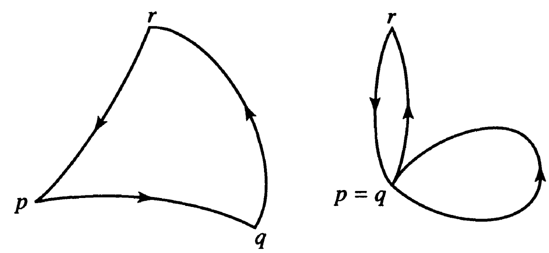


FIGURE 11.4.

Lemma 2.1 *Suppose (M, g) is complete and has* sec $\geq k$. *Then for each hinge (or triangle) in M we can find a comparison hinge (or triangle) in S_k^n where the corresponding segments have the same length and the angle is the same (all corresponding segments have the same length).*

Proof. Suppose we have three points $p, q, r \in M$. First, we know that in case $k > 0$, Myers' theorem implies $\operatorname{diam} M \leq \pi/\sqrt{k} = \operatorname{diam} S_k^n$. Thus, any segments between these three points have length $\leq \pi/\sqrt{k}$.

First the hinge case. Here we have segments from p to q and from q to r forming an angle α at q. In the space form we can first choose $\bar{p}$ and $\bar{q}$ such that $d(\bar{p}, \bar{q}) = d(p, q)$ and then join them by a segment. This is possible because $d(p, q) \leq \pi/\sqrt{k}$. At $\bar{q}$ we can then choose a direction that forms an angle α with the chosen segment. Then we take the unique geodesic going in this direction, and using the arc length parameter we go out distance $d(q, r)$ along this geodesic. This will now be a segment, as $d(q, r) \leq \pi/\sqrt{k}$. We have then found the desired hinge.

The triangle case is similar. First, pick $\bar{p}$ and $\bar{q}$ as above. Then, consider the two distance spheres $\partial B(\bar{p}, d(p, r))$ and $\partial B(\bar{q}, d(q, r))$. Since all possible triangle inequalities between p, q, r hold and $d(q, r), d(p, r) \leq \pi/\sqrt{k}$, these distance spheres are nonempty and they intersect. Then, let $\bar{r}$ be any point in the intersection.

To be honest here, we must use Cheng's diameter theorem in case any of the distances is $\pi/\sqrt{k}$. In this case there is nothing to prove as $(M, g) = S_k^n$. □

We can now state the Toponogov comparison theorem.

Theorem 2.2 (Toponogov, 1959) *Let (M, g) be a complete Riemannian manifold with* sec $\geq k$.

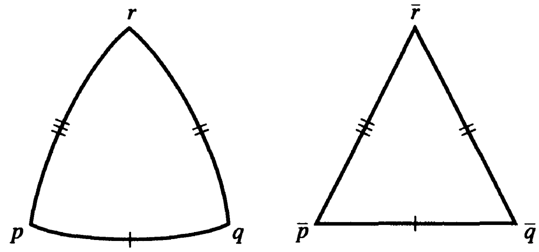


FIGURE 11.5.

Hinge Version: *Given any hinge with vertices $p, q, r \in M$ forming an angle α at q, it follows that for any comparison hinge in S^n_k with vertices $\bar{p}, \bar{q}, \bar{r}$ we have: $d(p,r) \leq d(\bar{p},\bar{r})$.*

Triangle Version: *Given any triangle with vertices $p, q, r \in M$, it follows that for any comparison triangle the interior angles are larger in M than in the comparison triangle in S^n_k. See also Figure 11.5.*

The proof requires a little preparation. First, we claim that the hinge version implies the triangle version. This follows from the *law of cosines* in constant curvature. This law implies that if we have $p, q, r \in S^n_k$ and increase the distance $d(p,r)$ while keeping $d(p,q)$ and $d(q,r)$ fixed, then the angle at q increases as well. For simplicity, we shall only look at the cases where $k = 1, 0, -1$.

Proposition 2.3 (Law of Cosines) *Let a triangle be given in S^n_k with side lengths a, b, c. If α denotes the angle opposite to a, then*

$$\begin{array}{ll} k=0 & a^2 = b^2 + c^2 - 2bc\cos\alpha. \\ k=-1 & \cosh a = \cosh b \cosh c - \sinh b \sinh c \cos\alpha. \\ k=1 & \cos a = \cos b \cos c + \sin b \sin c \cos\alpha. \end{array}$$

Proof. The general setup is the same in all cases. Namely, we suppose that a point $p \in S^n_k$ and a unit speed segment $\sigma : [0, c] \to S^n_k$ are given. We then investigate the restriction of the distance function from p to σ. See also Figure 11.6.

Case $k = 0$: Note that $t \to d(p, \sigma(t))$ is not a very nice function, as it is the square root of a quadratic polynomial. This, however, indicates that the function will become more manageable if we square it. Thus, we consider $\varphi(t) = \frac{1}{2}(d(p,\sigma(t)))^2 = \frac{1}{2}|p - \sigma(t)|^2$. We wish to compute the first and second derivatives of this function. This requires that we know $\nabla\frac{1}{2}d^2$ and also $\nabla^2\frac{1}{2}d^2$:

$$\begin{aligned} \nabla\frac{1}{2}d^2 &= \nabla\frac{1}{2}\left(\left(x^1\right)^2 + \cdots + \left(x^n\right)^2\right) \\ &= x^i\partial_i \\ &= d\nabla d; \end{aligned}$$

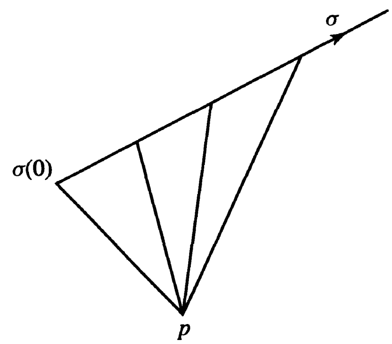


FIGURE 11.6.

$$\begin{aligned}\nabla^2 \frac{1}{2} d^2 &= \nabla \left(x^i \partial_i\right) \\ &= d\left(x^i\right) \partial_i \\ &= I.\end{aligned}$$

Then, since σ is a geodesic, we get

$$\begin{aligned}\varphi'(t) &= g\left(\dot{\sigma}, \nabla \frac{1}{2} d^2\right), \\ \varphi''(t) &= g\left(\left(\nabla^2 \frac{1}{2} d^2\right)(\dot{\sigma}), \dot{\sigma}\right) = 1.\end{aligned}$$

So if we define $b = d(p, \sigma(0))$ and let α be the interior angle between σ and the line joining p with $\sigma(0)$, then we have $\cos(\pi - \alpha) = -\cos\alpha = g(\dot{\sigma}(0), \nabla d)$. After integration of φ'', we then get

$$\begin{aligned}\varphi(t) &= \varphi(0) + \varphi'(0) \cdot t + \frac{1}{2} t^2 \\ &= \frac{1}{2} b^2 - b \cdot \cos\alpha \cdot t + \frac{1}{2} t^2.\end{aligned}$$

Now set $t = c$ and define $a = d(p, \sigma(c))$. Then we get:

$$\frac{1}{2} a^2 = \frac{1}{2} b^2 + b \cdot c \cdot \cos\alpha + \frac{1}{2} c^2,$$

from which the law of cosines follows.

Case $k = -1$: This time we must modify the distance function in a different way. Namely, consider $\varphi(t) = \cosh(d(p, \sigma(t))) - 1$. Then

$$\begin{aligned}\varphi'(t) &= \sinh(d(p, \sigma(t)))\, g(\nabla d, \dot{\sigma}), \\ \varphi''(t) &= \cosh(d(p, \sigma(t))) = \varphi(t) + 1.\end{aligned}$$

As before, we have $b = d(p, \sigma(0))$, and the interior angle satisfies $\cos(\pi - \alpha) = -\cos\alpha = g(\dot{\sigma}(0), \nabla d)$. Thus, we must solve the initial value problem

$$\begin{aligned}\varphi'' - \varphi &= 1,\\ \varphi(0) &= \cosh(b) - 1,\\ \varphi'(0) &= -\sinh(b)\cos\alpha.\end{aligned}$$

The general solution is

$$\begin{aligned}\varphi(t) &= C_1\cosh t + C_2\sinh t - 1\\ &= (\varphi(0) + 1)\cosh t + \varphi'(0)\sinh t - 1.\end{aligned}$$

So if we let $t = c$ and $a = d(p, \sigma(c))$ as before, we arrive at

$$\cosh a - 1 = \cosh b\cosh c - \sinh b\sinh c\cos\alpha - 1,$$

which implies the law of cosines again.

Case $k = 1$: This case is completely analogous to the case $k = -1$. We set $\varphi = 1 - \cos(d(p, \sigma(t)))$ and arrive at the initial value problem

$$\begin{aligned}\varphi'' + \varphi &= 1,\\ \varphi(0) &= 1 - \cos(b),\\ \varphi'(0) &= -\sin b\cos\alpha.\end{aligned}$$

Then,

$$\begin{aligned}\varphi(t) &= C_1\cos t + C_2\sin t + 1\\ &= (\varphi(0) - 1)\cos t + \varphi'(0)\sin t + 1,\end{aligned}$$

and consequently

$$1 - \cos a = -\cos b\cos c - \sin b\sin c\cos\alpha + 1,$$

which implies the law of cosines for the last time. □

This proof of the law of cosines suggests that in working in space forms it is easier to work with a modified distance function, the main advantage being that the Hessian is much simpler. Something similar can be done in variable curvature.

Lemma 2.4 *Let (M, g) be a complete Riemannian manifold $p \in M$ and $f(x) = d(x, p)$. If $\sec M \geq k$, then the Hessian of f satisfies*

$k = 0$: *The function $f_0 = \frac{1}{2}f^2$ satisfies $\nabla^2 f_0 \leq 1$ in the support sense everywhere.*
$k = -1$: *The function $f_{-1} = \cosh f - 1$ satisfies $\nabla^2 f_{-1} \leq \cosh f = f_{-1} + 1$ in the support sense everywhere.*
$k = 1$: *The function $f_1 = 1 - \cos f$ satisfies $\nabla^2 f_1 \leq \cos f = -f_1 + 1$ in the support sense everywhere.*

Proof. All three proofs are, of course, similar so let us concentrate just on the first case. The first comparison estimate from Chapter 6 implies that whenever f is smooth and w is perpendicular to ∇f, then

$$g\left(\nabla^2 f(w), w\right) \leq \frac{1}{f} g(w, w).$$

For such w one can therefore immediately see that

$$g\left(\nabla^2 f_0(w), w\right) \leq g(w, w).$$

If instead, $w = \nabla f$, then it is trivial that this holds, whence we have established the Hessian estimate at points where f is smooth. At all other points we just use the same trick by which we obtained the Laplacian estimates with lower Ricci curvature bounds in Chapter 9. □

We are now ready to prove the hinge version of Toponogov's theorem. The proof is divided into the three cases: $k = 0, -1, 1$. But the setup is the same in all cases. We shall assume that a point $p \in M$ and a *geodesic* $\sigma : [0, c] \to M$ are given. Correspondingly, we assume that a point $\bar{p} \in S^n_k$ and segment $\bar{\sigma} : [0, c] \to S^n_k$ are given. Given the appropriate initial conditions, we claim that

$$d(p, \sigma(t)) \leq d(\bar{p}, \bar{\sigma}(t)).$$

We shall for simplicity assume that $d(x, p)$ is smooth at $\sigma(0)$. Then the initial conditions are

$$\begin{aligned} d(p, \sigma(0)) &\leq d(\bar{p}, \bar{\sigma}(0)), \\ g(\nabla d, \dot{\sigma}(0)) &\geq g_k\left(\nabla \bar{d}, \dot{\bar{\sigma}}(0)\right). \end{aligned}$$

In case d is not smooth at $\sigma(0)$, we can just slide σ down along a segment joining p with $\sigma(0)$ and use a continuity argument. This also shows that we can use the stronger initial condition

$$d(p, \sigma(0)) < d(\bar{p}, \bar{\sigma}(0)).$$

In Figure 11.7 we have shown how σ can be changed by moving it down along a segment joining p and $\sigma(0)$. We have also shown how the angles can be slightly decreased. This will be important in the last part of the proof.

Case $k = 0$: We consider the modified functions

$$\begin{aligned} \varphi(t) &= \frac{1}{2}(d(p, \sigma(t)))^2, \\ \bar{\varphi}(t) &= \frac{1}{2}(d(\bar{p}, \bar{\sigma}(t)))^2. \end{aligned}$$

For small t these functions are smooth and satisfy

$$\begin{aligned} \varphi(0) &< \bar{\varphi}(0), \\ \varphi'(0) &\leq \bar{\varphi}'(0). \end{aligned}$$

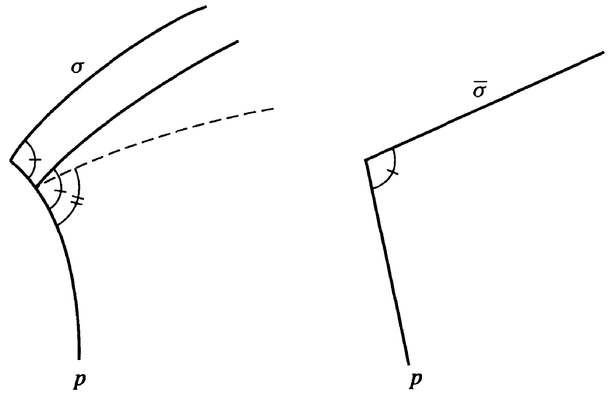


FIGURE 11.7.

Moreover, for the second derivatives we have

$$\varphi'' \leq 1 \quad \text{in the support sense,}$$
$$\bar{\varphi}'' = 1,$$

whence the difference $\psi(t) = \bar{\varphi}(t) - \varphi(t)$ satisfies

$$\psi(0) > 0,$$
$$\psi'(0) \geq 0,$$
$$\psi''(t) \geq 0 \quad \text{in the support sense.}$$

This shows that ψ is a convex function that is positive and increasing for small t, and hence increasing, and in particular positive, for all t. This proves the hinge version.

Case $k = -1$: Consider

$$\varphi(t) = \cosh d(p, \sigma(t)) - 1,$$
$$\bar{\varphi}(t) = \cosh d(\bar{p}, \bar{\sigma}(t)) - 1.$$

Then

$$\varphi(0) < \bar{\varphi}(0),$$
$$\varphi'(0) \leq \bar{\varphi}'(0),$$
$$\varphi'' \leq \varphi + 1 \quad \text{in the support sense,}$$
$$\bar{\varphi}'' = \bar{\varphi} + 1.$$

Then the difference $\psi = \bar{\varphi} - \varphi$ satisfies

$$\psi(0) > 0,$$
$$\psi'(0) \geq 0,$$
$$\psi''(t) \geq \psi(t) \quad \text{in the support sense.}$$

The first condition again implies that ψ is positive for small t. The last condition shows that as long as ψ is positive, it is also convex. The second condition then shows that ψ is increasing to begin with. It must now follow that ψ keeps increasing. Otherwise, there would be a positive maximum, and that violates convexity at points where ψ is positive.

Case $k = 1$: This case is considerably harder. We begin as before by defining

$$\begin{aligned}\varphi(t) &= 1 - \cos(d(p, \sigma(t))),\\ \bar{\varphi}(t) &= 1 - \cos(d(\bar{p}, \bar{\sigma}(t)))\end{aligned}$$

and then observing that the difference $\psi = \bar{\varphi} - \varphi$ satisfies

$$\begin{aligned}&\psi(0) > 0,\\ &\psi'(0) \geq 0,\\ &\psi''(t) \geq -\psi(t) \quad \text{in the support sense.}\end{aligned}$$

That, however, doesn't look very promising. Even though the function starts out being positive, the last condition gives just a *negative* lower bound for the second derivative. At this point we might then recall that perhaps Sturm-Liouville theory could save us. But for that to work well it is necessary as well to have $\psi'(0) > 0$. Thus, another little continuity argument is necessary as we need to perturb σ again to decrease the interior angle. If the interior angle is positive, this can clearly be done, and in the case where this angle is zero the hinge version is trivially true anyway. As long as ψ is smooth, Sturm–Liouville theory then ensures us that

$$\psi(t) \geq \zeta(t)$$

for as long as $\zeta(t)$ remains positive, where $\zeta(t)$ is defined as

$$\begin{aligned}\zeta'' &= -(1+\eta)\zeta,\\ \zeta(0) &= \psi(0) = \varepsilon > 0,\\ \zeta'(0) &= \psi'(0) = \delta > 0\end{aligned}$$

(the η will be used later in the proof). This means that

$$\zeta(t) = \sqrt{\varepsilon^2 + \frac{\delta^2}{1+\eta}} \cdot \sin\left(\sqrt{1+\eta} \cdot t + \arctan\left(\frac{\varepsilon \cdot \sqrt{1+\eta}}{\delta}\right)\right).$$

Now, $\zeta(t)$ actually remains positive for

$$t < \frac{\pi - \arctan\left(\frac{\varepsilon \cdot \sqrt{1+\eta}}{\delta}\right)}{\sqrt{1+\eta}},$$

so we should be able to extend the inequality $\psi(t) \geq \zeta(t)$ that far. To see this, consider the quotient

$$h = \frac{\psi}{\zeta}.$$

So far, we know that this function satisfies

$$h(0) = 1,$$
$$h(t) \geq 1 \quad \text{for small } t.$$

Should it therefore go below 1 before $\pi - \arcsin \varepsilon$, then h would have a positive maximum at some $t_0 \in (0, \pi - \arcsin \varepsilon)$. At this point we can use support functions ψ_ϵ for ψ from below, and conclude that also $\frac{\psi_\epsilon}{\zeta}$ has a maximum at t_0. Thus, we have

$$\begin{aligned}
0 &\geq \frac{d^2}{dt^2}\left(\frac{\psi_\epsilon}{\zeta}\right)(t_0) \\
&= \frac{\psi_\epsilon''(t_0)}{\zeta(t_0)} - 2\frac{\zeta'(t_0)}{\zeta(t_0)} \cdot \frac{d}{dt}\left(\frac{\psi_\epsilon}{\zeta}\right)_{t=t_0} - \frac{\psi_\epsilon(t_0)}{\zeta^2(t_0)}\zeta''(t_0) \\
&\geq \frac{-\psi_\epsilon(t_0) - \epsilon}{\zeta(t_0)} + \frac{\psi_\epsilon(t_0)}{\zeta(t_0)}(1+\eta) \\
&= \frac{\eta \cdot \psi_\epsilon(t_0) - \epsilon}{\zeta(t_0)}.
\end{aligned}$$

But this becomes positive as $\epsilon \to 0$, since we assumed $\psi_\epsilon(t_0) > 0$, and so we have a contradiction. Next, we can let $\eta \to 0$ and finally, let $\varepsilon \to 0$ to get the desired estimate for all $t \leq \pi$ using continuity.

Note that we never really use in the proof that we work with segments. The only thing that must hold is that the geodesics in the space form are segments. For $k \leq 0$ this is of course always true, but when $k = 1$ this means that the geodesic must have length $\leq \pi$. This was precisely the important condition in the last part of the proof.

11.3 Sphere Theorems

Our first applications of the Toponogov theorem are to the case of positively curved manifolds. Using scaling, we shall assume throughout this section that we work with a closed Riemannian n-manifold (M, g) with $\sec \geq 1$. For such spaces we have proved

(1) $\operatorname{diam}(M, g) \leq \pi$, with equality holding only if $M = S^n(1)$.

(2) If n is odd, then M is orientable.

(3) If n is even and M is orientable, then M is simply connected and $\operatorname{inj}(M) \geq \pi/\sqrt{\max \sec}$.

(4) If n is even and $\max \sec$ is close to 1, then (M, g) is close to a constant curvature metric. In particular, M must be a sphere when it is simply connected.

(5) It has also been mentioned that Klingenberg has shown that if M is simply connected and $\max \sec < 4$, then $\operatorname{inj}(M) \geq \pi/\sqrt{\max \sec}$.

This last result is quite subtle and is beyond what we can prove here. Gromov (see [32]) has a proof of this that in spirit goes as follows: One considers $p \in M$. If the upper curvature bound is $4-\delta$, then we know that if we pull the metric back to the tangent bundle, then there are no conjugate points on the disc $B\left(0, \pi/\sqrt{4-\delta}\right)$. Now recall our Hessian estimates for the distance function when $\sec \geq 1$. We use the modified distance f_1 to the origin in T_pM. This function is smooth on $B\left(0, \pi/\sqrt{4-\delta}\right)$ and satisfies

$$\nabla^2 f_1 \leq 1 - f_1 = \cos f.$$

On the region $B\left(0, \pi/\sqrt{4-\delta}\right) - B\left(0, \frac{\pi}{2} - \varepsilon\right)$ this function will therefore have strictly negative Hessian. In particular, the level sets for f or f_1 that lie in that region are strictly concave. Now map these level sets down into M via the exponential map. As this map is nonsingular they will be mapped to strictly concave, possibly immersed, hypersurfaces in M. In the case M is simply connected, one can prove an analogue to the Hadamard theorem for immersed convex hypersurfaces, namely, that they must be embedded spheres (this also uses that M has nonnegative curvature). However, if these hypersurfaces are embedded, then the exponential map must be an embedding on $B\left(0, \pi/\sqrt{4-\delta}\right)$, and in particular, we obtain the desired injectivity radius estimate.

We can now prove the celebrated Rauch-Berger-Toponogov-Klingenberg sphere theorem, also known as the quarter pinched sphere theorem.

Theorem 3.1 (1951–1961) *If M is a simply connected closed Riemannian manifold with $1 \leq \sec \leq 4 - \delta$, then M is homeomorphic to a sphere.*

Proof. We have shown that the injectivity radius is $\geq \pi/\sqrt{4-\delta}$. Thus, we have large discs around every point in M. Now select two points $p, q \in M$ such that $d(p,q) = \operatorname{diam} M$. Note that $\operatorname{diam} M \geq \operatorname{inj} M > \pi/2$. We now claim that every point $x \in M$ lies in one of the two balls $B\left(p, \pi/\sqrt{4-\delta}\right)$, or $B\left(q, \pi/\sqrt{4-\delta}\right)$, and thus M is covered by two discs. From this one can check using Meyer-Vietoris that all homology groups up to n vanish. This will imply that any degree 1 map $M \to S^n$ is an isomorphism in homology. As M is simply connected Whitehead's theorem then tells us that this map is a homotopy equivalence. Actually, an explicit homeomorphism can be constructed. This will be done below in a more general setting.

Now take $x \in M$. Let $d = \operatorname{diam} M = d(p,q)$, $a = d(p,x)$, and $b = d(x,q)$. If, for instance, $b > \pi/2$, then we claim that $a < \pi/2$. First, observe that since q is at maximal distance from p, it must follow that q cannot be a regular point for the distance function to p. Therefore, if we select any segment σ_1 from x to q, then we can find a segment σ_2 from p to q that forms an angle $\alpha \leq \pi/2$ with σ_1 at q. Then we can consider the hinge σ_1, σ_2 with angle α. The hinge version of

Toponogov's theorem now implies

$$\begin{aligned}\cos a &\geq \cos b \cos d + \sin b \sin d \cos \alpha \\ &\geq \cos b \cos d.\end{aligned}$$

Now, both $b, d > \pi/2$, so the left hand side is positive. This implies that $a < \pi/2$, as desired. □

Note that the theorem says nothing about nonsimply connected manifolds other than that their universal covering must a sphere. We also know that Micallef and Moore proved a better theorem for manifolds that merely have positive isotropic curvature. The above proof suggests, perhaps, that the conclusion of the theorem should hold as long as the manifold has large diameter. This is the content of the next theorem. This theorem was first proved by Berger in 1962 with a different proof; however, he only concluded that the manifold was a homotopy sphere. The present version is known as the Grove-Shiohama diameter sphere theorem. It was for the purpose of proving this theorem that Grove and Shiohama invented critical point theory.

Theorem 3.2 (Grove-Shiohama, 1977) *If (M, g) is a closed Riemannian manifold with* $\sec \geq 1$ *and* $\operatorname{diam} > \pi/2$, *then M is homeomorphic to a sphere.*

Proof. Fix $p, q \in M$ with $d(p, q) = \operatorname{diam} M = d > \pi/2$. We claim that the distance function from p has only q as a critical point. To see this, let $x \in M - \{p, q\}$ and let α be the angle between any two geodesics from x to p and q. If we suppose that $\alpha \leq \pi/2$ and set $a = d(p, x)$ and $b = d(x, q)$, then the hinge version of Toponogov's theorem implies

$$\begin{aligned}0 > \cos d &\geq \cos a \cos b + \sin a \sin b \cos \alpha \\ &\geq \cos a \cos b.\end{aligned}$$

But then $\cos a$ and $\cos b$ have opposite signs. If, for example, $\cos a \in (0, 1)$, then we have $\cos d > \cos b$, which implies $b > d = \operatorname{diam} M$. Thus we have arrived at a contradiction, and hence we must have $\alpha > \pi/2$. See also Figure 11.8.

We can now construct a vector field X such that X is the gradient field for $x \to d(x, p)$ near p and the negative of the gradient field for $x \to d(x, q)$ near q. Furthermore, the distance to p increases along integral curves for X. For each $x \in M - \{p, q\}$ there is a unique integral curve for X through x. Suppose that x varies over a small distance sphere $\partial B(p, \varepsilon)$ that is diffeomorphic to S^{n-1}. Then the length of this integral curve varies continuously with x. We can then multiply X by a function such that all integral curves have length π. Then $M - \{p, q\}$ is parametrized as $(t, x) \in (0, \pi) \times \partial B(p, \varepsilon)$. Using polar coordinates $(r, \theta) \in (0, \pi) \times S^{n-1}$ on S^n, we can then construct a homeomorphism $M \to S^n$. Note

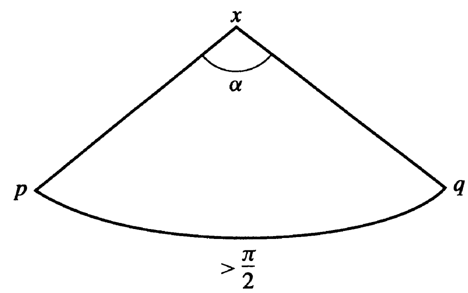


FIGURE 11.8.

that while this map could be differentiable, we have no way of knowing whether it is a diffeomorphism, as we can't compute its differential. □

Aside from the fact that the conclusions in the above theorems could possibly be strengthened to diffeomorphism, we have optimal results. For the first case, complex projective space has curvatures in $[1, 4]$, and for the second theorem, we can use real projective space in all dimensions as a space with $\sec = 1$ and $\mathrm{diam} = \pi/2$. If one relaxes the conditions slightly, it is, however, still possible to say something.

Theorem 3.3 *Suppose (M, g) is simply connected of dimension n with* $1 \leq \sec \leq 4 + \varepsilon$.

(1) (Berger, 1983) *If n is even, then there is $\varepsilon(n) > 0$ such that M must be homeomorphic to a sphere or diffeomorphic to one of the spaces* $\mathbb{C}P^{n/2}, \mathbb{H}P^{n/4}, CaP^2$.

(2) (Abresch-Meyer, 1994) *If n is odd, then there is a universal $\varepsilon > 0$ such that M is homeomorphic to a sphere.*

The spaces $\mathbb{C}P^{n/2}$, $\mathbb{H}P^{n/4}$, or CaP^2 are known as the compact rank 1 symmetric spaces (CROSS). We already know the complex projective space. The quaternionic projective space is $\mathbb{H}P^n = S^{4n+3}/S^3$, and the Cayley plane comes form a Riemannian submersion: $S^{23} \to CaP^2$ (see also Chapter 8 for more on these spaces). The proof of (1) uses convergence theory. First, it is shown that if $\varepsilon = 0$, then M is either homeomorphic to a sphere or isometric to one of the CROSSs. Then using the injectivity radius estimate in even dimensions, we can apply the convergence machinery.

Theorem 3.4 (Gromoll-Grove, 1987) *Suppose (M, g) is closed and satisfies* $\sec \geq 1$, $\mathrm{diam} \geq \frac{\pi}{2}$ *Then one of the following cases holds:*

(1) *M is homeomorphic to a sphere.*

(2) *M is isometric to a finite quotient $S^n(1)/\Gamma$, where the action of Γ is reducible (has an invariant subspace).*

(3) *M is isometric to one of* $\mathbb{C}P^{n/2}$, $\mathbb{H}P^{n/4}$, $\mathbb{C}P^{n/2}/\mathbb{Z}_2$ *for* $n = 2\text{mod}4$.

(4) *M has the cohomology ring of* CaP^2.

It is conjectured, but still unproved, that (4) can be improved to say that M is isometric to the Cayley plane.

11.4 The Soul Theorem

Let us commence by stating the theorem we are aiming to prove and then slowly work our way through the proof.

Theorem 4.1 (Cheeger-Gromoll-Meyer, 1969, 1972) *If* (M, g) *is a complete Riemannian manifold with* sec ≥ 0, *then* M *contains a* soul $S \subset M$, *which is a closed totally convex submanifold, such that M is diffeomorphic to the normal bundle over S. Moreover, when* sec > 0, *the soul is a point and M is diffeomorphic to* $\mathbb{R}^n$.

The history is briefly that Gromoll-Meyer first showed that sec > 0 implies that M is diffeomorphic to $\mathbb{R}^n$. Soon after Cheeger-Gromoll established the full theorem. The Gromoll-Meyer theorem is in itself rather remarkable.

We shall use critical point theory to establish this theorem. The problem lies in finding the soul. When this is done, it will be easy to see that the distance function to the soul has only regular points, and then we can use the results from the first section.

Before embarking on the proof, it might be instructive to show the following much easier result, whose proof will be used in the next section.

Lemma 4.2 (Critical Point Estimate) *If* (M, g) *is a complete open manifold of nonnegative sectional curvature, then for every* $p \in M$ *the distance function* $d(\cdot, p)$ *has no critical points outside some ball* $B(p, R)$. *In particular, M must have the topology of a compact manifold with boundary.*

Proof. We shall use a contradiction argument. So suppose we have a sequence p_k of critical points for $d(\cdot, p)$, where $d(p_k, p) \to \infty$. We can without loss of generality assume that

$$d(p_{k+1}, p) \geq 2d(p_k, p).$$

Now select segments σ_k from p to p_k. The above inequality implies that the angle at p between any two segments is $\geq 1/6$. To see this, suppose σ_k and σ_{k+l} form an angle $< 1/6$ at p. The hinge version of Toponogov's theorem then implies

$$\begin{aligned}(d(p_k, p_{k+l}))^2 &< (d(p, p_{k+l}))^2 + (d(p_k, p))^2 - 2d(p, p_{k+l})\, d(p_k, p)\cos\frac{1}{6} \\ &\leq \left(d(p, p_{k+l}) - \frac{3}{4}d(p_k, p)\right)^2.\end{aligned}$$

Now use that p_k is critical for p to conclude that there are segments from p to p_k and p_{k+1} to p_k that from an angle $\leq \pi/2$ at p_k. Then use the hinge version again to conclude

$$\begin{aligned}(d(p, p_{k+l}))^2 &\leq (d(p_k, p))^2 + (d(p_k, p_{k+l}))^2 \\ &\leq (d(p_k, p))^2 + \left(d(p, p_{k+l}) - \frac{3}{4}d(p_k, p)\right)^2 \\ &= \frac{25}{16}(d(p_k, p))^2 + (d(p, p_{k+l}))^2 - \frac{3}{2}d(p, p_{k+l})\,d(p_k, p),\end{aligned}$$

which implies

$$d(p, p_{k+l}) \leq \frac{25}{24}d(p_k, p).$$

But this contradicts our assumption that

$$d(p, p_{k+l}) \geq d(p_{k+1}, p) \geq 2d(p_k, p).$$

Now that all the unit vectors $\dot{\sigma}_k(0)$ form angles of at least $1/6$ with each other, we can conclude there can't be infinitely many such vectors. Hence, there cannot be critical points infinitely far away from p.

Observe that the vectors $\dot{\sigma}_k(0)$ lie on the unit sphere in T_pM and are distance $1/6$ away form each other. Thus, the balls $B(\dot{\sigma}_k(0), 1/12)$ are disjoint in the unit sphere and hence there are at most

$$\frac{v(n-1, 1, \pi)}{v\left(n-1, 1, \frac{1}{12}\right)} \leq 100^n$$

such points. □

We now have to explain what it means for a submanifold, or more generally a subset, to be totally convex. A subset $C \subset M$ of a Riemannian manifold is said to be *totally convex* if any geodesic in M joining two points in A actually lies in A. There are actually several different kinds of convexity, but as they are not important for any other developments, we shall confine ourselves to total convexity. The first observation is that this definition agrees with the usual definition for convexity in Euclidean space. Other than that, it is not clear that any totally convex sets exist at all. For example, if $A = \{p\}$, then A is totally convex only if there are no geodesic loops based at p. This means that points will almost never be totally convex. In fact, if M is closed, then M is the only totally convex subset as there must be a geodesic loop based at every point (to establish this requires something like the curve shortening process as explained in [50]). On complete manifolds it is possible to find totally convex sets.

Example 4.3 Let (M, g) be the flat cylinder $\mathbb{R} \times S^1$. All of the circles $\{p\} \times S^1$ are geodesics and actually totally convex. This also means that no point in M can be totally convex. In fact, all of those circles are souls. See also Figure 11.9.

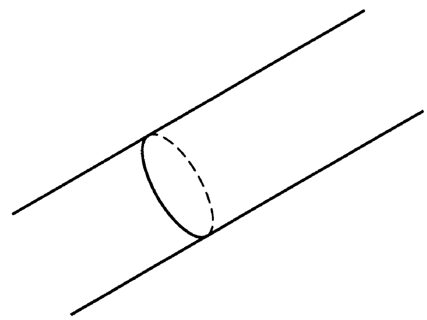


FIGURE 11.9.

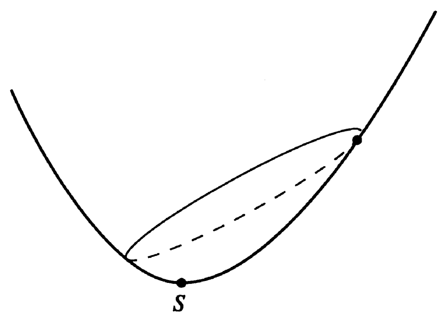


FIGURE 11.10.

Example 4.4 Let (M, g) be a smooth rotationally symmetric metric on $\mathbb{R}^2$ of the form $dr^2 + \varphi^2(r)\, d\theta^2$, where $\varphi'' < 0$. Thus, (M, g) looks like a parabola of revolution. The radial symmetry implies that all geodesics emanating from the origin $r = 0$ are rays going to infinity. Thus the origin is a soul and totally convex. Most other points, however, will have geodesic loops based there. See also Figure 11.10.

The way to find totally convex sets is via

Lemma 4.5 *If $f : (M, g) \to \mathbb{R}$ is concave, in the sense that the Hessian is weakly nonpositive everywhere, then every superlevel set $A = \{x \in M : f(x) \geq a\}$ is totally convex.*

Proof. Given any geodesic γ in M, we have that the function $f \circ \gamma$ has nonpositive weak second derivative. Thus, $f \circ \gamma$ is concave as a function on $\mathbb{R}$. In particular, the minimum of this function on any compact interval is obtained at one of the endpoints. This finishes the proof. □

Now the important discovery we must make is that of the existence of proper concave functions on any complete manifold of nonnegative sectional curvature.

Lemma 4.6 *Suppose (M, g) is as in the theorem and that $p \in M$. If we take all rays γ_α emanating from p and construct*

$$f = \inf_\alpha b_{\gamma_\alpha},$$

where b_γ denotes the Busemann function, then f is both proper and concave.

Proof. First we show that in nonnegative sectional curvature all Busemann functions are concave. Using that, we can then show that the given function is concave and proper.

Recall that in nonnegative Ricci curvature Busemann functions are superharmonic. The proof of concavity is almost identical. But instead of the Laplacian estimate for distance functions, we must use a similar Hessian estimate. If $h = d(\cdot, p)$, then we know

$$g\left(\nabla^2 h(w), w\right) \le \begin{cases} 0 & w = a\nabla h \\ \frac{1}{h} & g(w, \nabla h) = 0 \end{cases}.$$

Thus, we always have that

$$\nabla^2 h \le \frac{1}{h}.$$

We can now proceed as in the Ricci curvature case to show that Busemann functions have nonpositive Hessians in the weak sense and are therefore concave.

The infimum of a collection of concave functions is clearly concave. So we must now show that the superlevel sets for f are compact. Suppose, on the contrary, that some superlevel set $A = \{x \in M : f(x) \ge a\}$ is noncompact. As all of the Busemann functions b_{γ_α} are zero at p, also $f(p) = 0$. So $p \in A$. Now, using noncompactness, select a sequence $p_n \in A$ that goes to infinity. Then join p_n to p by a segment, and as in the construction of rays, choose a subsequence of these segments converging to a ray emanating from p. As A is convex, all of these segments lie in A. Then, since A is also closed, the ray must lie in A as well and therefore be one of the rays γ_α. But $f(\gamma_\alpha(t)) \le b_{\gamma_\alpha}(\gamma_\alpha(t)) = -t \to -\infty$, so we have a contradiction. □

We now need to establish a few properties of totally convex sets.

Lemma 4.7 *If $A \subset (M, g)$ is totally convex, then A has an interior, denoted by* $\operatorname{int} A$, *and a boundary ∂A. The interior is a totally convex submanifold of M, and the boundary has the property that for each $x \in \partial A$ there is a vector $v \in T_x M$ such that if $\gamma(t) : [0, a] \to A$ is a geodesic with $\gamma(0) = x$ and $\gamma(a) \in \operatorname{int} A$, then $\angle(v, \dot\gamma(0)) < \pi/2$.*

Some comments are in order before the proof. The words *interior* and *boundary*, while describing fairly accurately what the sets look like, are not meant in

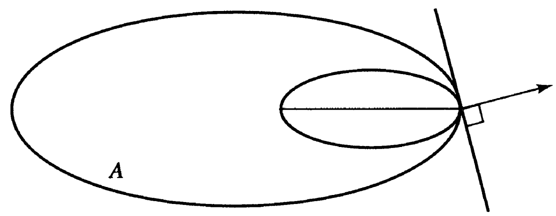


FIGURE 11.11.

the topological sense. Most convex sets will, of course, not have any topological interior at all. The property about the boundary is what is often called the *supporting hyperplane property*. Namely, the interior of the convex set is supposed to lie strictly on one side of a hyperplane at any of the boundary points. The vector v is the normal to this hyperplane and can be taken to be tangent to some geodesic that goes into the interior. It is important to note that the supporting hyperplane property shows that the distance function to a subset of $\text{int}A$ cannot have any critical points on ∂A. See also Figure 11.11.

Proof. We can without loss of generality assume that A is closed. The convexity radius estimate from Chapter 6 will also be used in many places. Namely, we shall use that there is a positive function $\varepsilon(p) : M \to (0, \infty)$ such that the distance function $d_p(x) = d(x, p)$ is smooth and strictly convex on $B(p, \varepsilon(p))$.

First, let us identify points in the interior and on the boundary. Find the maximal integer k such that A contains a k-dimensional submanifold of M. If $k = 0$, then A must be a point. For if A contains two points, then A also contains a segment joining these points and therefore a 1-dimensional manifold. Now define $N \subset A$ as being the union of all k-dimensional submanifolds in M that are contained in A. We claim that N is a k-dimensional totally geodesic submanifold whose closure is A. We shall thus identify $\text{int}A$ with N and ∂A with $A - N$.

To see that it is a submanifold, pick $p \in N$ and let $N_p \subset A$ be a k-dimensional submanifold of M containing p. Since N_p is a submanifold, we can assume that for some small $\delta \in (0, \varepsilon(p))$ we have $B(p, \delta) \cap N_p = N_p$. We now claim that also $B(p, \delta) \cap A = N_p$. If this were not true, then we could find $q \in A \cap B(p, \delta) - N_p$. Now assume that δ is so small that also $\delta < \text{inj}_p$. Then we can join each point in N_p to q by a unique segment. The union of these segments will, away from q, form a cone that is a $k+1$-dimensional submanifold which is contained in A (see Figure 11.12), thus contradicting maximality of k. In particular, N must be a submanifold as we have $B(p, \delta) \cap N = N_p$.

What we have just proved can easily be modified to show that for points $p \in N$ and $q \in A$ with the property that $d(p, q) < \text{inj}_q$ there is a k-dimensional submanifold $N_p \subset N$ such that $q \in \bar{N}_p$, namely, just take a $(k-1)$-dimensional submanifold through p in N perpendicular to the segment from p to q and consider the cone over this submanifold with vertex q. From this statement we get the property that if $\gamma : [0, a] \to A$ is a geodesic, then $\gamma(0, a) \subset N$ provided that, say, $\gamma(0) \in N$. In particular, N is dense in A.

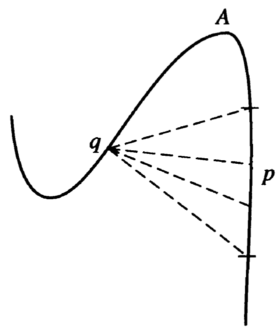


FIGURE 11.12.

Having identified the interior and boundary, we now have to establish the supporting hyperplane property. Define $A_\varepsilon = \{x \in A : d(x, \partial A) \geq \varepsilon\}$. For $p \in \partial A$ select $q \in A_\varepsilon$ such that $d(q, p) = \varepsilon$. Then $x \to d(q, x)$ has a local minimum at p when restricted to ∂A. Therefore, if $\gamma : [0, a] \to A$ is a geodesic with $\gamma(0) = p$, then we must have that $d(q, \gamma(t)) \leq \varepsilon$ for small t. But then, $g(\nabla d, \dot{\gamma}(0)) \leq 0$, or in other words, $\angle(\nabla d, \dot{\gamma}(0)) \geq \pi/2$. This almost gives us the supporting plane property, but a little more work is needed.

For $p \in \partial A$ let C_p be the cone of vectors $v \in T_pM$ such that the geodesic $\exp_p(tv) \in N$ for some $t > 0$, and hence all small $t > 0$. Using that N is a submanifold, it is easy to show that C_p is an open subset of $\mathrm{span} C_p$ and that $\dim \mathrm{span} C_p = k$. For $\varepsilon > 0$ small, suppose we can select $q \in A_\varepsilon$ such that $d(q, p) = \varepsilon$. The set of such points is clearly 2ε-dense in ∂A. So the set of points $p \in \partial A$ for which we can find an $\varepsilon > 0$ and $q \in A_\varepsilon$ such that $d(q, p) = \varepsilon$ is dense in ∂A. As the supporting plane property is an open property, it suffices to prove it for such p (this follows from critical point theory). We can also suppose ε is so small that $d_q = d(\cdot, q)$ is smooth and convex on a neighborhood containing p. We claim that $\angle\left(-\nabla d_q, v\right) < \pi/2$ for all $v \in C_p$. To see this, observe that we have a convex set $A \cap B(q, \varepsilon) \subset N$, and $p \in \partial A \cap \bar{B}(q, \varepsilon)$. Then C_p contains the set of vectors $\left\{v \in T_pM : \exp_p(tv) \in A \cap B(q, \varepsilon) \text{ for small } t > 0\right\}$. This set is clearly also open in $\mathrm{span} C_p$ and furthermore can be identified with the set of vectors $\left\{v \in \mathrm{span} C_p : \angle\left(v, -\nabla d_q\right) < \pi/2\right\}$. If now $C_p \neq \left\{v \in \mathrm{span} C_p : \angle\left(v, -\nabla d_q\right) < \pi/2\right\}$, then openness of C_p in $\mathrm{span} C_p$ implies that there must be a $v \in C_p$ such that also $-v \in C_p$ (because the weaker inequality $\leq$ already holds). But this implies that $p \in N$, as it becomes a point on a geodesic whose endpoints lie in N. (See Figure 11.13.) □

The last lemma we need is

Lemma 4.8 *Let (M, g) have $\sec \geq 0$. If $A \subset M$ is totally convex, then distance function $d : A \to \mathbb{R}$ defined by $d(x) = d(x, \partial A)$ is concave on A, and strictly concave if $\sec > 0$.*

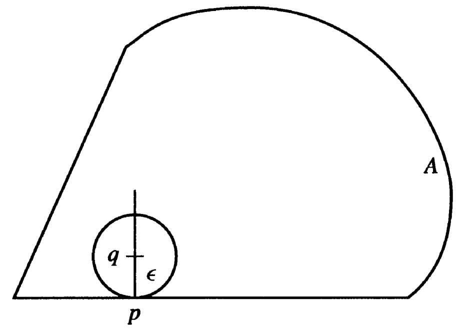


FIGURE 11.13.

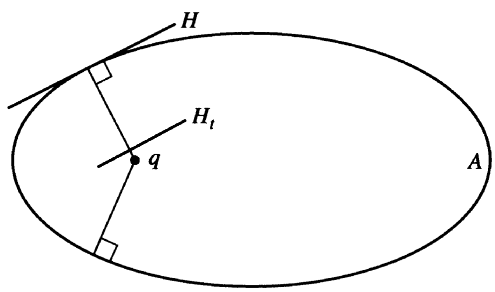


FIGURE 11.14.

Proof. We shall show that the Hessian is nonpositive in the support sense. So fix $q \in \text{int}A$, and find $p \in \partial A$ so that $d(p,q) = d(q, \partial A)$. Then select $v \in T_pM$ as in the above lemma. We can now find a small hypersurface H, as in Chapter 6, such that v is normal to H, the shape operator for H at p is zero and $H \cap \text{int}A = \emptyset$. (See Figure 11.14.)

Then $d(\cdot, H)$ is a support function from above for $d(\cdot, \partial A)$ near q. Thus, we only need to show that it has nonpositive Hessian in the support sense. If $\sigma : [0, a] \to A$ is a segment from p to q, then it is also an integral curve for $\nabla d(\cdot, H)$ as long as $d(\cdot, H)$ is smooth. As $\nabla^2 d(\cdot, H) = 0$ initially along σ, the radial curvature equation

$$\begin{aligned} L_{\dot\sigma} \nabla^2 d(\cdot, H) &= -R_{\dot\sigma} - \left(\nabla^2 d(\cdot, H)\right)^2 \\ &\leq -\left(\nabla^2 d(\cdot, H)\right)^2 \end{aligned}$$

tells us that $\nabla^2 d(\cdot, H) \leq 0$ along σ (and < 0 if sec > 0). Thus, we are finished if $d(\cdot, H)$ is smooth at q. Otherwise, we can find $t < a$ and a hypersurface H_t with normal $\dot\sigma(t)$ at $\sigma(t)$ and zero shape operator at $\sigma(t)$ such that $d(\cdot, H_t)$ is smooth at q and therefore also has nonpositive (negative) Hessian at q. In this case we claim that $t + d(\cdot, H_t)$ is a support function for $d(\cdot, \partial A)$. Clearly, the functions are equal at q. As the hypersurfaces H and H_t can be supposed to be of the form where they contain the geodesics perpendicular to σ, we have that the length comparison from Chapter 6 implies

$$t \geq d(x, H) \geq d(x, \partial A) \quad \text{for all} \quad x \in H_t.$$

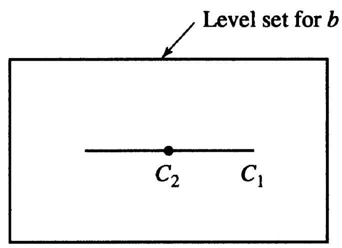


FIGURE 11.15.

This shows the support property. □

We are now ready to prove the soul theorem. Start with the proper concave function f constructed from the Busemann functions. The maximum level set $C_1 = \{x \in M : f(x) = \max f\}$ is nonempty and convex since f is proper and concave. Moreover, it follows from the previous lemma that C_1 is a point if $\sec > 0$. This is because the superlevel sets $A = \{x \in M : f(x) \geq a\}$ are convex with $\partial A = f^{-1}(a)$, so $f = d(\cdot, \partial A)$ on A. Now, a strictly concave function (Hessian in support sense is negative) must have a unique maximum or no maximum, thus showing that C_1 is a point. If C_1 is a submanifold, then we are also done, for $d(\cdot, C_1)$ has no critical points, as any point lies on the boundary of a convex superlevel set. Otherwise, C_1 is a convex set with nonempty boundary. But then $d(\cdot, \partial C_1)$ is concave. The maximum set C_2 is again nonempty, since C_1 is compact and convex. If it is a submanifold, then we again claim that we are done. For the distance function $d(\cdot, C_2)$ has no critical points, as any point lies on the boundary for a superlevel set for either f or $d(\cdot, \partial C_1)$. We can now iterate to get a sequence of convex sets $C_1 \supset C_2 \supset \cdots \supset C_k$. We claim that in at most $n = \dim M$ steps we arrive at a point or submanifold, which we write as S and call the soul (see Figure 11.15). This is because $\dim C_i > \dim C_{i+1}$. To see this suppose $\dim C_i = \dim C_{i+1}$; then $\text{int}C_{i+1}$ will be an open subset of $\text{int}C_i$. So if $p \in \text{int}C_{i+1}$, then we can find δ such that $B(p, \delta) \cap \text{int}C_{i+1} = B(p, \delta) \cap \text{int}C_i$. Now choose a segment σ from p to ∂C_i. Clearly $d(\cdot, \partial C_i)$ is strictly increasing along this curve. This curve, however, runs through $B(p, \delta) \cap \text{int}C_i$, thus showing that $d(\cdot, \partial C_i)$ must be constant on the part of the curve close to p.

Much more can be said about complete manifolds with nonnegative sectional curvature. A rather complete account can be found in Greene's survey in [46]. We briefly mention two important results:

Theorem 4.9 *Let S be a soul of a complete Riemannian manifold with* $\sec \geq 0$, *arriving from the above construction.*

(1) (Sharafudtinov, 1978) *There is a distance-nonincreasing map $sh : M \to S$ such that $sh_{|S} = id$. In particular, all souls must be isometric to each other.*

(2) (Perel'man, 1993) *The map $sh : M \to S$ is a submetry. From this it follows that S must be a point if all sectional curvatures based at just one point are positive.*

Having reduced all complete nonnegatively curved manifolds to bundles over closed nonnegatively curved manifolds, it is natural to ask the converse question: Given a closed manifold S with nonnegative curvature, which bundles over S admit complete metrics with $\sec \geq 0$? Clearly, the trivial bundles do. When $S = T^2$ Özaydın-Walschap in [67] have shown that this is the only 2-dimensional vector bundle that admits such a metric. Still, there doesn't seem to be a satisfactory general answer. If, for instance, we let $S = S^2$, then any 2-dimensional bundle is of the form $\left(S^3 \times \mathbb{C}\right)/S^1$, where S^1 is the Hopf action on S^3 and acts by rotations on $\mathbb{C}$ in the following way: $\omega \times z = \omega^k z$ for some integer k. This integer is the Euler number of the bundle. As we have a complete metric of nonnegative curvature on $S^3 \times \mathbb{C}$, the O'Neill formula from the exercises to Chapter 2 shows that these bundles admit metrics with $\sec \geq 0$.

There are some interesting examples of manifolds with positive and zero Ricci curvature that show how badly the soul theorem fails for such manifolds. In 1978, Gibbons-Hawking in [39] constructed Ricci flat metrics on quotients of $\mathbb{C}^2$ blown up at any finite number of points. Thus, one gets a Ricci flat manifold with arbitrarily large second Betti number. About ten years later Sha-Yang showed that the infinite connected sum $\left(S^2 \times S^2\right) \sharp \left(S^2 \times S^2\right) \sharp \cdots \sharp \left(S^2 \times S^2\right) \sharp \cdots$ admits a metric with positive Ricci curvature, thus putting to rest any hopes for general theorems in this direction. Sha-Yang have a very nice survey in [41] describing these and other examples. The construction uses doubly warped product metrics on $I \times S^2 \times S^1$ as described in Chapter 3, and then some topology.

11.5 Finiteness of Betti Numbers

The theorem we wish to prove is

Theorem 5.1 (Gromov, 1978, 1981) *There is a constant $C(n)$ such that any complete manifold (M, g) with* $\sec \geq 0$ *satisfies*

(1) *$\pi_1(M)$ can be generated by $\leq C(n)$ generators.*

(2) *For any field F of coefficients the Betti numbers are bounded:*

$$\sum_{i=0}^{n} b_i(M, F) = \sum_{i=0}^{n} \dim H_i(M, F) \leq C(n).$$

Part (2) of this result is considered one of the deepest and most beautiful results in Riemannian geometry. Before embarking on the proof, let us put it in context. First, we should note that the Gibbons-Hawking and Sha-Yang examples show that a similar result cannot hold for manifolds with nonnegative Ricci curvature. Sha-Yang also exhibited metrics with positive Ricci curvature on the connected sums

$$\underbrace{\left(S^2 \times S^2\right) \sharp \left(S^2 \times S^2\right) \sharp \cdots \sharp \left(S^2 \times S^2\right)}_{k \text{ times}}.$$

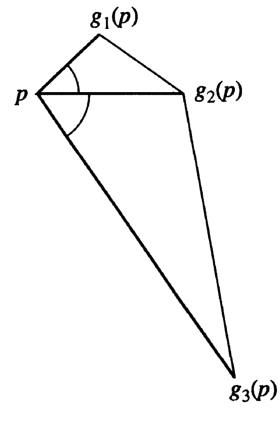


FIGURE 11.16.

For large k, the Betti number bound shows that these connected sums cannot have a metric with nonnegative sectional curvature. Thus, we have simply connected manifolds that admit positive Ricci curvature but not nonnegative sectional curvature. The reader should also consult our discussion of manifolds with nonnegative curvature operator at the end of Chapters 7 and 8. Let us list the open problems that were posed there and settled for manifolds with nonnegative curvature operator:

(1) (H. Hopf) Does $S^2 \times S^2$ admit a metric with positive sectional curvature?

(2) (H. Hopf) If M is even-dimensional, does $\sec \geq 0$ (> 0) imply $\chi(M) \geq 0$ (> 0)?

(3) (Gromov) If $\sec \geq 0$, is $\sum_{i=0}^{n} b_i(M, F) \leq 2^n$?

First we establish part (1) of the above theorem:

Proof of (1). We shall construct what is called a *short set of generators* for $\pi_1(M)$. We consider $\pi_1(M)$ as acting by deck transformations on the universal covering $\tilde{M}$ and fix $p \in \tilde{M}$. We then inductively select a generating set $\{g_1, g_2, \ldots\}$ such that

(a) $d(p, g_1(p)) \leq d(p, g(p))$ for all $g \in \pi_1(M)$.

(b) $d(p, g_k(p)) \leq d(p, g(p))$ for all $g \in \pi_1(M) - \langle g_1, \ldots, g_{k-1} \rangle$.

Now join p and $g_k(p)$ by segments σ_k (see Figure 11.16). We claim that the angle between any two such segments is $\geq \pi/3$.

Otherwise, the hinge version of Toponogov's theorem would imply

$$\begin{aligned}(d(g_{k+l}(p), g_k(p)))^2 &< (d(p, g_k(p)))^2 \\ &\quad + (d(p, g_{k+l}(p)))^2 - d(p, g_k(p))\, d(p, g_{k+l}(p)) \\ &\leq (d(p, g_{k+l}(p)))^2 .\end{aligned}$$

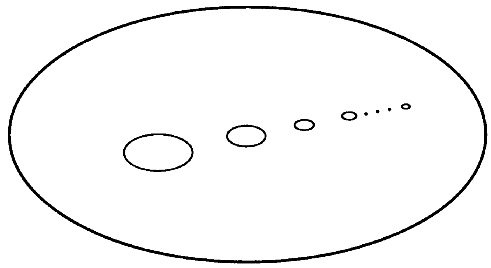


FIGURE 11.17.

But then

$$d\left(g_{k+l}^{-1} \circ g_k(p), p\right) < d(p, g_{k+l}(p)),$$

which contradicts our choice of g_{k+l}.

It now follows that there can be at most

$$\frac{v(n-1, 1, \pi)}{v\left(n-1, 1, \frac{\pi}{6}\right)}$$

elements in the set $\{g_1, g_2, \ldots\}$. We have therefore produced a generating set with a bounded number of elements.

Observe how closely this proof resembles that of the critical point estimate lemma from the previous section. □

The proof of the Betti number estimate is established through several lemmas. First, we need to make three definitions for metric balls. Throughout, we fix a Riemannian n-manifold M with sec ≥ 0 and a field F of coefficients for our homology theory $H_*(\cdot, F) = H_0(\cdot, F) \oplus \cdots \oplus H_n(\cdot, F)$.

Content: The *content* of a metric ball $B(p, r) \subset M$ is

$$\mathrm{cont}B(p, r) = \mathrm{rank}\left(H_*\left(B\left(p, \frac{1}{5}r\right), F\right) \to H_*(B(p, r), F)\right).$$

The reason for working with content, rather that just the rank of $H_*(B(p, r), F)$ itself, is that metric balls might not have finitely generated homology. However, if $O_1 \subset M$ is any bounded subset of a manifold and $\bar{O}_1 \subset O_2 \subset M$, then the image of $H_*(O_1, F)$ in $H_*(O_2, F)$ is finitely generated. In Figure 11.17 we have taken a plane domain and extracted infinitely many discs of smaller and smaller size. This yields a compact set with infinite topology. Nevertheless, this set has finitely generated topology when mapped into any neighborhood of itself, as that has the effect of canceling all of the smallest holes.

Corank: The *corank* of a set $A \subset M$ is defined as the largest integer k such that we can find k metric balls $B(p_1, r_1), \ldots, B(p_k, r_k)$ with the properties

(a) There is a critical point x_i for p_i with $d(p_i, x_i) = 10r_i$.

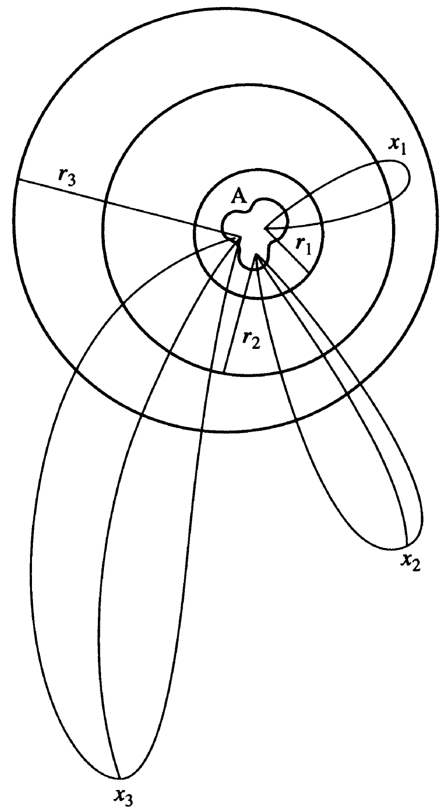


FIGURE 11.18.

(b) $r_i \geq 3r_{i-1}$ for $i = 2, \ldots, k$.
(c) $A \subset \bigcap_{i=1}^{k} B(p_i, r_i)$.

In Figure 11.18 we have a picture of how the set A and the larger circles might be situated relative to each other.

Compressibility: We say that a ball $B(p, r)$ is *compressible* if it contains a ball $B(q, r') \subset B(p, r)$ such that

(a) $r' \leq \frac{r}{2}$.
(b) $\text{cont}B(q, r') \geq \text{cont}B(p, r)$.

If a ball is not compressible we call it *incompressible*. Note that any ball with content > 1, can be successively compressed to an incompressible ball. Figure 11.19 gives a scematic picture of a ball that can be compressed into a smaller ball.

We shall now tie these three concepts together through some lemmas that will ultimately lead us to the proof of the Betti number estimate. Observe that for large r, the ball $B(p, r)$ contains all the topology of M, so $\text{cont}B(p, r) = \sum_i b_i(M, F)$. Also, the corank of such a ball must be zero, as there can't be any critical points

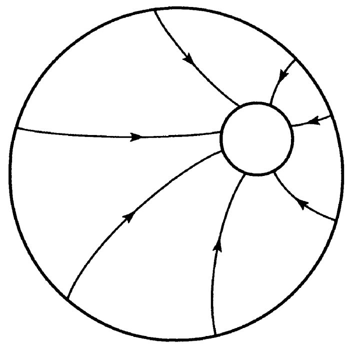


FIGURE 11.19.

outside this ball. The idea is now to compress this ball until it becomes incompressible and then estimate its content in terms of balls that have corank 1. We shall in this way successively be able to estimate the content of balls of fixed corank in terms of the content of balls with one higher corank. The proof is then finished first, by showing that the corank of a ball is uniformly bounded by 100^n, and second, by observing that balls of maximal corank must be contractible and therefore have content 1 (otherwise they would contain critical points for the center, and the center would have larger corank).

Lemma 5.2 *The corank of any set $A \subset M$ is bounded by* 100^n.

Proof. Suppose that A has corank larger than 100^n. Select balls $B(p_1, r_1), \ldots, B(p_k, r_k)$ with corresponding critical points $x_1, \ldots, x_k$, where $k > 100^n$. Now choose $z \in A$ and join z to x_i by segments σ_i. As in the critical point estimate lemma, we can then find two of these segments σ_i and σ_j that form an angle $< 1/6$ at z.

For simplicity, suppose $i < j$ and define

$$\begin{aligned} a_i &= \ell(\sigma_i) = d(z, x_i), \\ a_j &= \ell(\sigma_j) = d(z, x_j), \\ l &= d(x_i, x_j), \end{aligned}$$

and observe that

$$\begin{aligned} b_i &= d(z, p_i) \leq r_i, \\ b_j &= d(z, p_j) \leq r_j. \end{aligned}$$

Figure 11.20 gives pictures explaining the notation in the proof.

The triangle inequality implies

$$\begin{aligned} a_i &\leq 10 r_i + b_i \leq 11 r_i, \\ a_j &\geq 10 r_j - r_j \geq 9 r_j. \end{aligned}$$

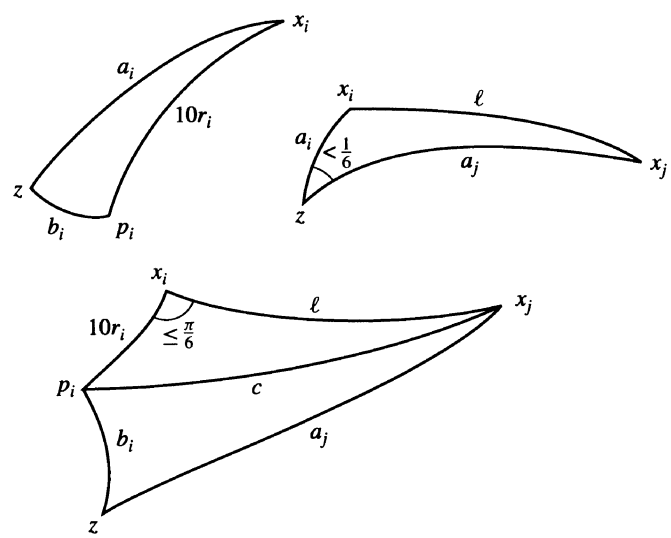


FIGURE 11.20.

Also, $r_j \geq 3r_i$, so we see that $a_j > a_i$. As in the critical point estimate lemma, we can conclude that

$$l \leq a_j - \frac{3}{4}a_i.$$

Now use the triangle inequality to conclude

$$\begin{aligned} c = d\left(p_i, x_j\right) &\geq a_j - b_i \\ &\geq 10r_j - b_j - b_i \\ &\geq 8r_j \\ &\geq 24r_i \\ &\geq 20r_i = 2d\left(p_i, x_i\right). \end{aligned}$$

Yet another application of the triangle inequality will then imply

$$l \geq d\left(x_i, p_i\right).$$

Since x_i is critical for p_i, we can now use the hinge version of Toponogov's theorem to conclude

$$\begin{aligned} c^2 &\leq \left(d\left(p_i, x_i\right)\right)^2 + l^2 \\ &\leq \left(l + \frac{1}{2}d\left(p_i, x_i\right)\right)^2. \end{aligned}$$

Thus,

$$\begin{aligned} c &\leq l + \frac{1}{2}d\left(p_i, x_i\right) \\ &\leq l + 5r_i. \end{aligned}$$

The triangle inequality then implies

$$a_j \leq c + b_i \leq c + r_i \leq l + 6r_i.$$

However, we also have

$$a_i \geq 10r_i - b_i \geq 9r_i,$$

which together with

$$l \leq a_j - \frac{3}{4}a_i$$

implies

$$l \leq a_j - \frac{27}{4}r_i.$$

Thus, we have a contradiction:

$$l + \frac{27}{4}r_i \leq a_j \leq l + 6r_i. \qquad \square$$

Having bounded the corank, let us see how the topology changes when we pass from balls of lower corank to balls of higher corank. Let $\mathcal{C}(k)$ denote the set of balls in M of corank $\geq k$, and $\mathcal{B}(k)$ the largest content of any ball in $\mathcal{C}(k)$.

Lemma 5.3 *There is a constant $C(n)$ depending only on dimension such that*

$$\mathcal{B}(k) \leq C(n)\,\mathcal{B}(k+1).$$

Proof. The number $\mathcal{B}(k)$ is, of course, realized by some incompressible ball $B(p, R)$. Now consider a ball $B(x, r)$ where $x \in B(p, R/4)$ and $r \leq R/10$. We claim that this ball lies in $\mathcal{C}(k+1)$. To see this, consider the ball $B(x, R/2) \subset B(p, R) \subset B(x, 5R)$. Since $B(p, R)$ is assumed to be incompressible, there must be a critical point for x in the annulus $B(x, 5R) - B(x, R/2)$. For otherwise we could deform $B(p, R)$ to $B(x, R/2)$ inside $B(x, 5R)$. This would imply that $\operatorname{cont}B(p, R) \geq \operatorname{cont}B(x, R/2)$ and thus contradict incompressibility of $B(p, R)$. We can now show that $B(x, r) \in \mathcal{C}(k+1)$. Using that $B(p, R) \in \mathcal{C}(k)$, select $B(p_1, r_1), \ldots, B(p_l, r_l)$, $l \geq k$, as in the definition of corank. Then pick a critical point y for x in $B(x, 5R) - B(x, R/2)$ and consider the ball $B(x, d(x, y)/10)$. Then the balls $B(p_1, r_1), \ldots, B(p_l, r_l), B(x, d(x, y)/10)$ can be used to show that $B(x, r)$ has corank $\geq l + 1 > k$.

Now cover $B(p, R/5)$ by balls $B(p_i, R/50)$, $i = 1, \ldots, m$. If we suppose that the balls $B(p_i, R/100)$ are pairwise disjoint, then we must have:

$$m \leq \frac{v(n, 0, 2R)}{v\left(n, 0, \frac{1}{100}R\right)} = 200^n.$$

Now consider the sets $B\left(p_i, \frac{1}{2}R\right) \subset B(p, R)$. First, we claim that

$$\operatorname{cont}B(p,R) \le \operatorname{rank}\left(H_*\left(\bigcup_{i=1}^{m} B\left(p_i, \frac{1}{50}R\right), F\right)\right.$$
$$\left.\to H_*\left(\bigcup_{i=1}^{m} B\left(p_i, \frac{1}{2}R\right), F\right)\right).$$

This follows from the simple observation that if $A \subset B \subset C \subset D$, then

$$\operatorname{rank}\left(H_*(A, F) \to H_*(D, F)\right) \le \operatorname{rank}\left(H_*(B, F) \to H_*(C, F)\right).$$

To estimate the right-hand side of the above inequality, it is natural to suppose that we can use a Meyer-Vietoris argument, together with induction on m, to show

$$\operatorname{rank}\left(H_*\left(\bigcup_{i=1}^{m} B\left(p_i, \frac{1}{50}R\right), F\right) \to H_*\left(\bigcup_{i=1}^{m} B\left(p_i, \frac{1}{2}R\right), F\right)\right)$$
$$\le \sum_{\substack{i_1<\cdots<i_s \\ 1\le s\le m}} \operatorname{rank}\left(H_*\left(\bigcap_{t=1}^{s} B\left(p_{i_t}, \frac{1}{50}R\right), F\right) \to H_*\left(\bigcap_{t=1}^{s} B\left(p_{i_t}, \frac{1}{2}R\right), F\right)\right).$$

We then observe that if $\bigcap_{t=1}^{s} B\left(p_{i_t}, \frac{1}{50}R\right) \neq \emptyset$, then the triangle inequality implies

$$\bigcap_{t=1}^{s} B\left(p_{i_t}, \frac{1}{50}R\right) \subset B\left(p_{i_1}, \frac{1}{50}R\right) \subset B\left(p_{i_1}, \frac{1}{10}R\right) \subset \bigcap_{t=1}^{s} B\left(p_{i_t}, \frac{1}{2}R\right).$$

As each of the balls $B(p_i, R/10) \in \mathcal{C}(k+1)$, and there can be at most 2^m nonempty intersections, we then arrive at the estimate

$$\operatorname{cont}B(p,R) = \mathcal{B}(k) \le 2^{200^n} \cdot \mathcal{B}(k+1).$$

This is the desired inequality. □

We now claim that

$$\operatorname{cont}M \le 2^{20000^n},$$

which will, of course, prove the theorem. The above lemma clearly yields that

$$\begin{aligned}\operatorname{cont}M &= \mathcal{B}(0) \\ &\le \mathcal{B}(k) \cdot \left(2^{200^n}\right)^k \\ &= \mathcal{B}(k) \cdot 2^{k \cdot 200^n} \\ &\le \mathcal{B}(k) \cdot 2^{20000^n},\end{aligned}$$

where $k \le 100^n$ is the largest possible corank in M. It then remains to check that $\mathcal{B}(k) = 1$. However, it follows from the above that if $\mathcal{C}(k)$ contains an incompressible ball, then $\mathcal{C}(k+1) \neq \emptyset$. Thus, all balls in $\mathcal{C}(k)$ are compressible, but then they must have minimal content $= 1$.

The above estimate on the rank of the inclusion

$$H_*\left(\bigcup_{i=1}^{m} B\left(p_i, \frac{1}{50}R\right), F\right) \to H_*\left(\bigcup_{i=1}^{m} B\left(p_i, \frac{1}{2}R\right), F\right),$$

in terms of the ranks of all the intersections, is in fact not quite right. One really needs to consider the doubly indexed family $B\left(p_i, 1\left(5\cdot 10^j\right)^{-1}\cdot R\right)$, $j = 1, \dots, n+2$, where we assume that for each fixed j the family covers $B\left(p, \frac{1}{5}R\right)$. The correct estimate is then that the rank of the inclusion

$$H_*\left(\bigcup_{i=1}^{m} B\left(p_i, \frac{1}{5\cdot 10^{n+1}}\cdot R\right), F\right) \to H_*\left(\bigcup_{i=1}^{m} B\left(p_i, \frac{1}{2}R\right), F\right)$$

is bounded by the rank of all of the possible intersections

$$H_*\left(\bigcap_{t=1}^{s} B\left(p_{i_t}, \frac{1}{5\cdot 10^j}\cdot R\right), F\right) \to H_*\left(\bigcap_{t=1}^{s} B\left(p_{i_t}, \frac{2}{10^j}\cdot R\right), F\right).$$

Whenever such an intersection $\bigcap_{t=1}^{s} B\left(p_{i_t}, \cdot R\right) \neq \emptyset$, we still have the inclusions

$$\begin{aligned}\bigcap_{t=1}^{s} B\left(p_{i_t}, \frac{1}{5\cdot 10^j}\cdot R\right) &\subset B\left(p_{i_1}, \frac{1}{5\cdot 10^j}\cdot R\right)\\ &\subset B\left(p_{i_1}, \frac{1}{10^j}\cdot R\right)\\ &\subset \bigcap_{t=1}^{s} B\left(p_{i_t}, \frac{2}{10^j}\cdot R\right).\end{aligned}$$

So we can still estimate those ranks by the content of balls in $C(k+1)$. We have, however, more intersections and also more balls, as this time the smaller balls $B\left(p_i, 10^{-n-1}\cdot R\right)$ have to cover. One can easily compute the correct Betti number estimate with these modifications. The reader should consult the survey by Cheeger in [25] for the complete story.

The Betti number theorem can easily be proved in the more general context of manifolds with lower sectional curvature bounds, but one must then also assume an upper diameter bound. Otherwise, the ball covering arguments, and also the estimates using Toponogov's theorem, won't work. Thus, there is a constant $C(n, D, k)$ such that any closed Riemannian n-manifold (M, g) with $\sec \geq k$ and $\operatorname{diam} \leq D$ has

(1) $\pi_1(M)$ can be generated by $\leq C(n, k, D)$ elements.
(2) $\sum_{i=0}^{n} b_i(M, F) \leq C(n, D, k)$.

11.6 Homotopy Finiteness

We shall prove in this section a result that interpolates between Cheeger's finiteness theorem and Gromov's Betti number estimate. We know that in Gromov's theorem the class under investigation contains infinitely many homotopy types, while if we have a lower volume bound and an upper curvature bound as well, Cheeger's result says that we have finiteness of diffeomorphism types.

Theorem 6.1 (Grove-Petersen, 1988) *Given an integer $n > 1$ and numbers $v, D, k \in (0, \infty)$, the class of Riemannian n-manifolds with*

$$\begin{aligned} \text{diam} &\leq D, \\ \text{vol} &\geq v, \\ \sec &\geq -k^2 \end{aligned}$$

contains only finitely many homotopy types.

As with the other proofs in this chapter, we need to proceed in stages. First, we present the main technical result.

Lemma 6.2 *For a manifold as in the above theorem, we can find $\alpha = \alpha(n, D, v, k) \in \left(0, \frac{\pi}{2}\right)$ and $\delta = \delta(n, D, v, k) > 0$ such that if $p, q \in M$ satisfy $d(p, q) \leq \delta$, then either p is α-regular for q or q is α-regular for p.*

Proof. The proof is by contradiction. So assume we have a pair of points $p, q \in M$ that are not α-regular with respect to each other, and set $l = d(p, q) \leq \delta$. Let $\Gamma(p, q)$ denote the set of unit speed segments from p to q, and define

$$\begin{aligned} \dot{\Gamma}_{pq} &= \left\{ v \in T_pM : v = \dot{\sigma}(0), \sigma \in \Gamma(p, q) \right\}, \\ \dot{\Gamma}_{qp} &= \left\{ -v \in T_qM : v = \dot{\sigma}(r), \sigma \in \Gamma(p, q) \right\}. \end{aligned}$$

Then the two sets $\dot{\Gamma}_{pq}$ and $\dot{\Gamma}_{qp}$ of unit vectors are by assumption $(\pi - \alpha)$-dense in the unit sphere. Now recall from the exercises to Chapter 9 that if $A \subset S^{n-1}$, then the function

$$t \to \frac{\text{vol}B(A, t)}{v(n-1, 1, t)}$$

is decreasing. In particular, we must have for any $(\pi - \alpha)$-dense set $A \subset S^{n-1}$ that

$$\begin{aligned} \text{vol}\left(S^{n-1} - B(A, \alpha)\right) &= \text{vol}S^{n-1} - \text{vol}B(A, \alpha) \\ &\leq \text{vol}S^{n-1} - \text{vol}S^{n-1} \cdot \frac{v(n-1, 1, \alpha)}{v(n-1, 1, \pi - \alpha)} \\ &= \text{vol}S^{n-1} \cdot \frac{v(n-1, 1, \pi - \alpha) - v(n-1, 1, \alpha)}{v(n-1, 1, \pi - \alpha)}. \end{aligned}$$

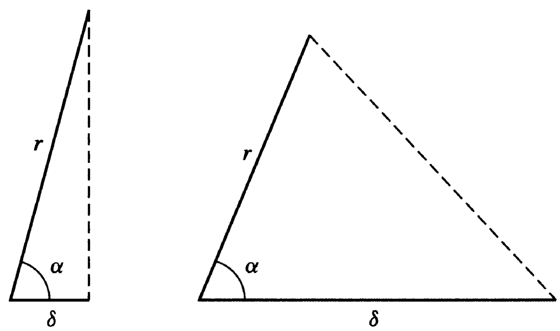


FIGURE 11.21.

Now choose $\alpha < \pi/2$ such that

$$\mathrm{vol}S^{n-1} \cdot \frac{v(n-1,1,\pi-\alpha) - v(n-1,1,\alpha)}{v(n-1,1,\pi-\alpha)} \cdot \int_0^D (\mathrm{sn}_k(t))^{n-1}\, dt = \frac{v}{6}.$$

Thus, the two cones (see exercises to Chapter 9) satisfy

$$\mathrm{vol}B^{S^{n-1}-B(\dot{\Gamma}_{pq},\alpha)}(p,D) \le \frac{v}{6},$$
$$\mathrm{vol}B^{S^{n-1}-B(\dot{\Gamma}_{qp},\alpha)}(q,D) \le \frac{v}{6}.$$

We now use Toponogov's theorem to choose δ such that any point in M that does not lie in one of these two cones must be close to either p or q (Figure 11.21 shows how a small δ will force the other leg in the triangle to be smaller than r). To this end, pick $r > 0$ such that

$$v\left(n, -k^2, r\right) = \frac{v}{6}.$$

We now claim that if δ is sufficiently small, then

$$M = B(p,r) \cup B(q,r) \cup B^{S^{n-1}-B(\dot{\Gamma}_{pq},\alpha)}(p,D) \cup B^{S^{n-1}-B(\dot{\Gamma}_{qp},\alpha)}(q,D).$$

This will, of course, lead to a contradiction, as we would then have

$$\begin{aligned} v &\le \mathrm{vol}M \\ &\le \mathrm{vol}\left(B(p,r) \cup B(q,r) \cup B^{S^{n-1}-B(\dot{\Gamma}_{pq},\alpha)}(p,D) \cup B^{S^{n-1}-B(\dot{\Gamma}_{qp},\alpha)}(q,D)\right) \\ &\le 4 \cdot \frac{v}{6} < v. \end{aligned}$$

To see that these sets cover M, observe that if $x \notin \mathrm{vol}B^{S^{n-1}-B(\dot{\Gamma}_{pq},\alpha)}(p,D)$, then there is a segment form x to p and a segment form p to q that form an angle $\le \alpha$. (See Figure 11.22.)

Thus, we have from Toponogov's theorem that

$$\cosh d(x,q) \le \cosh l \cosh d(x,p) - \sinh l \sinh d(x,p) \cos(\alpha).$$

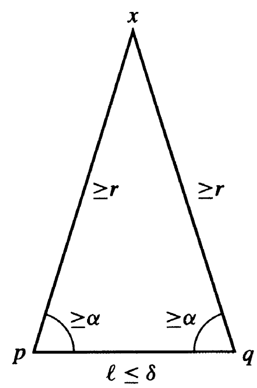


FIGURE 11.22.

If also $x \notin \mathrm{vol}B^{S^{n-1}-B(\dot{\Gamma}_{qp},\alpha)}(p, D)$, we have in addition,

$$\cosh d(x, p) \leq \cosh l \cosh d(x, q) - \sinh l \sinh d(x, q) \cos(\alpha).$$

If in addition $d(x, p) > r$ and $d(x, q) > r$, we get

$$\begin{aligned}\cosh d(x, q) &\leq \cosh l \cosh d(x, p) - \sinh l \sinh d(x, p) \cos(\alpha)\\ &\leq \cosh d(x, p) + (\cosh l - 1)\cosh D - \sinh l \sinh r \cos(\alpha)\end{aligned}$$

and

$$\cosh d(x, p) \leq \cosh d(x, q) + (\cosh l - 1)\cosh D - \sinh l \sinh r \cos(\alpha).$$

However, as $l \to 0$, we see that the quantity

$$\begin{aligned}f(l) &= (\cosh l - 1)\cosh D - \sinh l \sinh r \cos(\alpha)\\ &= (-\sinh r \cos\alpha)\, l + O\left(l^2\right)\end{aligned}$$

becomes negative. Thus, we can find $\delta(D, r, \alpha) > 0$ such that for $l \leq \delta$ we have

$$(\cosh l - 1)\cosh D - \sinh l \sinh r \cos(\alpha) < 0.$$

We have then arrived at another contradiction, as this would imply

$$\cosh d(x, q) < \cosh d(x, p)$$

and

$$\cosh d(x, p) < \cosh d(x, q)$$

at the same time. Thus, the sets cover as we claimed. As this covering is also impossible, we are lead to the conclusion that under the assumption that $d(p, q) \leq \delta$, we must have that either p is α-regular for q or q is α-regular for p. □

As it stands, this lemma seems rather strange and unmotivated. A little analysis will, however, enable us to draw some very useful conclusions from it.

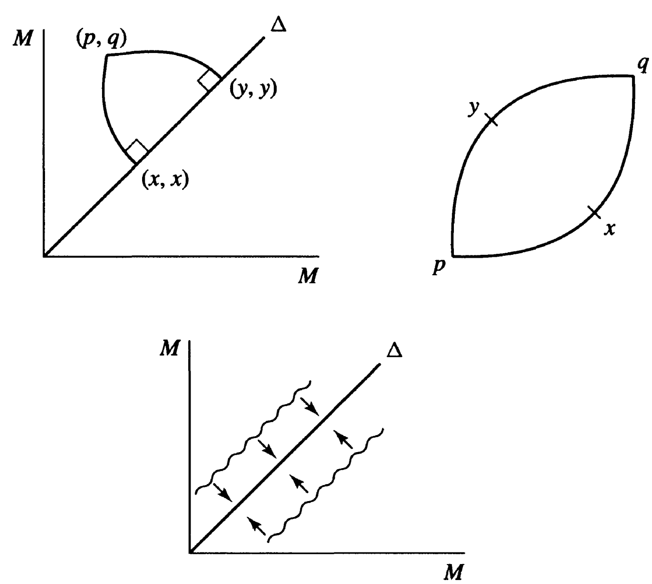


FIGURE 11.23.

Consider the product $M \times M$ with the product metric. Geodesics in this space are of the form (γ_1, γ_2), where both γ_1, γ_2 are geodesics in M. In $M \times M$ we have the diagonal $\Delta = \{(x, x) : x \in M\}$, which is a compact submanifold. Note that $T_{(p,p)}\Delta = \left\{(v, v) : v \in T_pM\right\}$, and consequently, the normal bundle is $\nu(\Delta) = \left\{(v, -v) : v \in T_pM\right\}$. Therefore, if $(\sigma_1, \sigma_2) : [a, b] \to M \times M$ is a segment from (p, q) to Δ, then we must have that $\dot{\sigma}_1(b) = -\dot{\sigma}_2(b)$. Thus these two segments can be joined at the common point $\sigma_1(b) = \sigma_2(b)$ to form a geodesic from p to q in M. This geodesic is, in fact, a segment, for otherwise, we could find a shorter curve from p to q. Dividing this curve in half would then produce a shorter curve from (p, q) to Δ. Thus, we have a bijective correspondence between segments from p to q and segments from (p, q) to Δ. Moreover, $\sqrt{2} \cdot d((p, q), \Delta) = d(p, q)$. The above lemma now implies

Corollary 6.3 *Any point within distance $\delta/\sqrt{2}$ of Δ is α-regular for Δ.*

Figure 11.23 shows how the contraction onto the diagonal works and also how segments to the diagonal are related to segments in M.

Thus, we can find a curve of length $\leq 1/\cos\alpha \cdot d((p, q), \Delta)$ from any point in this neighborhood to Δ. Moreover, this curve depends continuously on (p, q). We can translate this back into M. Namely, if $d(p, q) < \delta$, then p and q are joined by a curve $t \to H(p, q, t), 0 \leq t \leq 1$, whose length is $\leq 1/\cos\alpha \cdot d(p, q)$. Furthermore, the map $(p, q, t) \to H(p, q, t)$ is continuous. For simplicity, we let $C = 1/\cos\alpha$ in the constructions below.

We now have the first ingredient of our proof.

Corollary 6.4 *If $f_0, f_1 : X \to M$ are two continuous maps such that $d(f_0(x), f_1(x)) < \delta$ for all $x \in X$, then f_0 and f_1 are homotopy equivalent.*

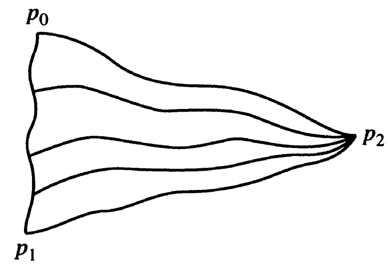


FIGURE 11.24.

For the next construction, recall that a k-simplex Δ^k can be thought of as the set of affine linear combinations of all the basis vectors in $\mathbb{R}^{k+1}$, i.e.,

$$\Delta^k = \left\{ \left(x^0, \dots, x^k \right) : x^0 + \dots + x^k = 1 \text{ and } x^0, \dots, x^k \in [0,1] \right\}.$$

The basis vectors $e_i = \left(\delta_i^1, \dots, \delta_i^n \right)$ are called the vertices of the simplex.

Lemma 6.5 *Suppose we have $k+1$ points $p_0, \dots, p_k \in B(p, r) \subset M$. If*

$$2r \frac{C^k - 1}{C - 1} < \delta,$$

then we can find a continuous map

$$f : \Delta^k \to B\left(p, r + 2r \cdot C \cdot \frac{C^k - 1}{C - 1} \right),$$

where $f(e_i) = p_i$.

Proof. Figure 11.24 gives the essential idea of the proof. The proof goes by induction on k. For $k = 0$ there is nothing to show.

Suppose now that the statement holds for k and that we have $k+2$ points $p_0, \dots, p_{k+1} \in B(p, r)$. First, we find a map

$$f : \Delta^k \to B\left(p, 2r \cdot C \cdot \frac{C^k - 1}{C - 1} + r \right)$$

with $f(e_i) = p_i$ for $i = p_0, \dots, p_k$. We then define

$$\bar{f} : \Delta^{k+1} \to B\left(p, r + 2r \cdot C \cdot \frac{C^{k+1} - 1}{C - 1} \right),$$

$$\bar{f}\left(x^0, \dots, x^k, x^{k+1} \right) = H\left(f\left(\frac{x^0}{\sum_{i=1}^k x^i}, \dots, \frac{x^k}{\sum_{i=1}^k x^i} \right), p_{k+1}, x^{k+1} \right).$$

This clearly gives a well-defined continuous map as long as

$$d\left(f\left(\frac{x^0}{\sum_{i=1}^k x^i}, \dots, \frac{x^k}{\sum_{i=1}^k x^i} \right), p_{k+1} \right)$$

$$
\begin{aligned}
&\leq d\left(f\left(\frac{x^0}{\sum_{i=1}^k x^i},\ldots,\frac{x^k}{\sum_{i=1}^k x^i}\right),p\right)+d\left(p,p_{k+1}\right)\\
&\leq\left(2r\cdot C\cdot\frac{C^k-1}{C-1}+r\right)+r\\
&=2r\cdot\frac{C^{k+1}-1}{C-1}\\
&<\delta
\end{aligned}
$$

with the property that

$$
\begin{aligned}
d\left(p,\bar{f}(\cdot)\right)&\leq d\left(p,p_{k+1}\right)+d\left(p_{k+1},\bar{f}(\cdot)\right)\\
&\leq r+2r\cdot C\cdot\frac{C^{k+1}-1}{C-1}.
\end{aligned}
$$

This concludes the induction step. □

Note that if we select a face spanned by, say, $(e_1,\ldots,e_k)$ of the simplex Δ^k, then we could, of course, construct a map in the above way by mapping e_i to p_i. The resulting map will, however, be the same as if we constructed the map on the entire simplex and restricted it to the selected face.

We can now prove finiteness of homotopy types. First, observe that the class we work with is precompact in the Gromov-Hausdorff distance, as we have an upper diameter bound and a lower bound for the Ricci curvature. It therefore suffices to prove

Lemma 6.6 *There is an* $\varepsilon=\varepsilon(n,k,v,D)>0$ *such that if two Riemannian* n*-manifolds* (M,g_1) *and* (N,g_2) *satisfy*

$$
\begin{aligned}
\operatorname{diam}&\leq D,\\
\operatorname{vol}&\geq v,\\
\sec&\geq -k^2,
\end{aligned}
$$

and

$$
d_{G-H}(M,N)<\varepsilon,
$$

then they are homotopy equivalent.

Proof. Suppose M and N are given as in the lemma, together with a metric d on $M\amalg N$, inside which the two spaces are ε Hausdorff close. The size of ε will be found through the construction.

First, triangulate both manifolds in such a way that any simplex of the triangulation lies in a ball of radius ε. Using the triangulation on M, we can now construct a continuous map $f:M\to N$ as follows. First we use the Hausdorff approximation to map all the vertices $\{p_\alpha\}\subset M$ of the triangulation to points $\{q_\alpha\}\subset N$ such that $d(p_\alpha, q_\alpha) < \varepsilon$. If now $(p_{\alpha_0}, \dots, p_{\alpha_n})$ forms a simplex in the triangulation of M, then we constructed the triangulation such that $(p_{\alpha_0}, \dots, p_{\alpha_n}) \subset B(x, \varepsilon)$ for some $x \in M$. Thus $(q_{\alpha_0}, \dots, q_{\alpha_n}) \subset B(q_{\alpha_0}, 4\varepsilon)$. Therefore, if

$$8\varepsilon \frac{C^n - 1}{C - 1} < \delta,$$

then we can use the above lemma to define f on the simplex spanned by $(p_{\alpha_0}, \dots, p_{\alpha_n})$. In this way we get a map $f : M \to N$ by constructing it on each simplex as just described. To see that it is continuous, we must check that the construction agrees on common faces of simplices. But this follows, as the construction is natural with respect to restriction to faces of simplices. We now need to estimate how good a Hausdorff approximation f is. To this end, select $x \in M$ and suppose that it lies in the face spanned by the vertices $(p_{\alpha_0}, \dots, p_{\alpha_n})$. Then we have

$$\begin{aligned} d(x, f(x)) &\le d(x, p_{\alpha_0}) + d(p_{\alpha_0}, f(x)) \\ &\le 2\varepsilon + \varepsilon + d(q_{\alpha_0}, f(x)) \\ &\le 3\varepsilon + 4\varepsilon + 8\varepsilon \cdot C \cdot \frac{C^n - 1}{C - 1} \\ &= 7\varepsilon + 8\varepsilon \cdot C \cdot \frac{C^n - 1}{C - 1}. \end{aligned}$$

We can now construct $g : N \to M$ in the same manner. This map will, of course, also satisfy

$$d(y, g(y)) \le 7\varepsilon + 8\varepsilon \cdot C \cdot \frac{C^n - 1}{C - 1}.$$

It is now possible to estimate how close the compositions $f \circ g$ and $g \circ f$ are to the identity maps on N and M, respectively, as follows:

$$\begin{aligned} d(y, f \circ g(y)) &\le d(y, g(y)) + d(g(y), f \circ g(y)) \\ &\le 14\varepsilon + 16\varepsilon \cdot C \cdot \frac{C^n - 1}{C - 1}; \\ d(x, g \circ f(x)) &\le 14\varepsilon + 16\varepsilon \cdot C \cdot \frac{C^n - 1}{C - 1}. \end{aligned}$$

As long as

$$14\varepsilon + 16\varepsilon \cdot C \cdot \frac{C^n - 1}{C - 1} < \delta,$$

we can then conclude that these compositions are homotopy equivalent to the respective identity maps. In particular, the two spaces are homotopy equivalent. □

Note that as long as

$$16\varepsilon \cdot \frac{C^{n+1} - 1}{C - 1} < \delta,$$

the two spaces are homotopy equivalent. Thus, ε depends in an explicit way on $C = 1/\cos\alpha$ and δ. It is possible, in turn, to estimate α and δ from n, k, v, and D. We can therefore get an explicit estimate for how close spaces must be to ensure that they are homotopy equivalent. Given this explicit ε, it is then possible, using our work from the section on Gromov-Hausdorff distance, to find an explicit estimate for the number of homotopy types.

To conclude, let us compare the three finiteness theorems by Cheeger, Gromov, and Grove-Petersen. We have inclusions of classes of closed Riemannian n-manifolds

$$\left\{\begin{array}{lll} \text{diam} & \leq & D \\ \sec & \geq & -k^2 \end{array}\right\} \supset \left\{\begin{array}{lll} \text{diam} & \leq & D \\ \text{vol} & \geq & v \\ \sec & \geq & -k^2 \end{array}\right\} \supset \left\{\begin{array}{lll} \text{diam} & \leq & D \\ \text{vol} & \geq & v \\ |\sec| & \leq & k^2 \end{array}\right\}$$

with strengthenings of conclusions from bounded Betti numbers to finitely many homotopy types to compactness in the $C^{1,\alpha}$ topology. In the special case of nonnegative curvature Gromov's estimate actually doesn't depend on the diameter, thus yielding obstructions to the existence of such metrics on manifolds with complicated topology. For the other two results the diameter bound is still necessary. Consider for instance the family of lens spaces $\{S^3/\mathbb{Z}_p\}$ with curvature $= 1$. Now rescale these metrics so that they all have the same volume. Then we get a class which contains infinitely many homotopy types and also satisfies

$$\text{vol} = v,$$
$$1 \geq \sec > 0.$$

The family of lens spaces $\{S^3/\mathbb{Z}_p\}$ with curvature $= 1$ also shows that the lower volume bound is necessary in two of the theorems.

Some further improvements are possible in the conclusion of the homotopy finiteness result. Namely, one can strengthen the conclusion to state that the class contains finitely many homeomorphism types. This was proved for $n \neq 3$ in [47] and in a more general case in [68]. One can also prove many of the above results for manifolds with certain types of integral curvature bounds, see for instance [70] and [71]. The volume [46] also contains complete discussions of generalizations to the case where one has merely Ricci curvature bounds.

11.7 Further Study

There are many texts that partially cover or expand the material in this chapter. We wish to attract attention to the surveys by Grove in [41], by Abresch-Meyer, Colding, Greene, and Zhu in [46], by Cheeger in [25], and by Karcher in [27]. The most glaring omission from this chapter is probably that of the Abresch-Gromoll theorem and other uses of the excess function. The above-mentioned articles by Zhu and Cheeger cover this material quite well.

11.8 Exercises

1. Let (M, g) be a closed positively curved manifold. Show that if M contains a totally geodesic closed hypersurface (i.e., the shape operator is zero), then M is homeomorphic to a sphere. (Hint: first show that the hypersurface is orientable, and then show that the signed distance function to this hypersurface has only two critical points—a maximum and a minimum.)

2. Show that the converse of Toponogov's theorem is also true. In other words, if for some k the conclusion to Toponogov's theorem holds when hinges (or triangles) are compared to the same objects in S_k^2, then $\sec \geq k$.

3. (Heintze-Karcher) Let $\gamma \subset (M, g)$ be a geodesic in a Riemannian n-manifold with $\sec \geq -k^2$. Let $T(\gamma, R)$ be the normal tube around γ of radius R, i.e., the set of points in M that can be joined to γ by a segment of length $\leq R$ that is perpendicular to γ. The last condition is superfluous when γ is a closed geodesic, but if it is a loop or a segment, then not all points in M within distance R of γ will belong to this tube. On this tube introduce coordinates (r, s, θ), where r denotes the distance to γ, s is the arc-length parameter on γ, and $\theta = \left(\theta^1, \ldots, \theta^{n-2}\right)$ are spherical coordinates normal to γ. These give adapted coordinates for the distance r to γ. Show that as $r \to 0$ the metric looks like
$$g(r) = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \cdot r^2 + O\left(r^3\right).$$
Using the lower sectional curvature bound, find an upper bound for the volume density on this tube. Conclude that
$$\operatorname{vol} T(\gamma, R) \leq f(n, k, R, \ell(\gamma)),$$
for some continuous function f depending on dimension, lower curvature bound, radius, and length of γ. Moreover, as $\ell(\gamma) \to 0$, $f \to 0$. Use this estimate to prove Cheeger's lemma from Chapter 10 and the main lemma on mutually critical points from the homotopy finiteness theorem. This shows that Toponogov's theorem is not needed for the latter result.

4. Show that any vector bundle over a 2-sphere admits a complete metric of nonnegative sectional curvature. Hint: You need to know something about the classification of vector bundles over spheres. In this case k-dimensional vector bundles are classified by homotopy classes of maps from S^1, the equator of the 2-sphere, into $SO(k)$. This is the same as $\pi_1(SO(k))$, so there is only one 1-dimensional bundle, the 2-dimensional bundles are parametrized by $\mathbb{Z}$, and two higher-dimensional bundles.

Appendix A
de Rham Cohomology

We shall explain in this appendix the main ideas surrounding de Rham cohomology. This is done as a service to the reader who has learned about tensors and algebraic topology but had only sporadic contact with Stokes' theorem. First we give a digest of forms and important operators on forms. Then we explain how one integrates forms and prove Stokes' theorem for manifolds without boundary. Finally, we define de Rham cohomology and show how the Poincaré lemma and the Meyer-Vietoris lemma together imply that de Rham cohomology is simply standard cohomology. The cohomology theory that comes closest to de Rham cohomology is Čech cohomology. As this cohomology theory often is not covered in standard courses on algebraic topology, we define it here and point out that it is easily seen to satisfy the same properties as de Rham cohomology.

A.1 Elementary Properties

On a manifold M we let $\Omega^p(M)$ denote the collection of p-forms. On forms we have the *wedge product* operation

$$\begin{aligned} \Omega^p(M) \times \Omega^q(M) &\to \Omega^{p+q}(M), \\ (\omega, \psi) &\to \omega \wedge \psi. \end{aligned}$$

This operation is bilinear and antisymmetric in the sense that:

$$\omega \wedge \psi = (-1)^{pq}\, \psi \wedge \omega.$$

This product is defined as follows. The wedge product of a function and a form is simply standard multiplication. Given two 1-forms $\omega, \psi \in \Omega^1$, we define

$$(\omega \wedge \psi)(v, w) = \omega(v)\psi(w) - \psi(v)\omega(w)$$

and then extend this to all forms using associativity and linearity.

There are three other important operations defined on forms: the *exterior derivative* $d : \Omega^p(M) \to \Omega^{p+1}(M)$, the *Lie derivative* $L_X : \Omega^p(M) \to \Omega^p(M)$, and the *interior product* $i_X : \Omega^p(M) \to \Omega^{p-1}(M)$.

The exterior derivative of a function is simply its usual differential, while if we are given a form $\omega = f_0 df_1 \wedge \cdots \wedge df_p$, then we declare that $d\omega = df_0 \wedge df_1 \wedge \cdots \wedge df_p$.

The Lie derivative is defined by

$$(L_X\omega)(Y_1, \ldots, Y_p) = L_X(\omega(Y_1, \ldots, Y_p)) - \sum_{i=1}^{p} \omega(Y_1, \ldots, [X, Y_i], \ldots, Y_p)$$

and the interior product

$$(i_X\omega)(Y_1, \ldots, Y_{p-1}) = \omega(X, Y_1, \ldots, Y_{p-1}).$$

These operators satisfy the following properties:

$$\begin{aligned}
d \circ d &= 0, \\
d(\omega \wedge \psi) &= (d\omega) \wedge \psi + (-1)^p \omega \wedge (d\psi), \\
i_X \circ i_X &= 0, \\
i_X(\omega \wedge \psi) &= (i_X\omega) \wedge \psi + (-1)^p \omega \wedge (i_X\psi), \\
L_X(\omega \wedge \psi) &= (L_X\omega) \wedge \psi + \omega \wedge (L_X\psi), \\
(d\omega)(Y_0, Y_1, \ldots, Y_p) & \\
= \sum_{i=0}^{p} (-1)^i L_{Y_i}&\left(\omega\left(Y_0, \ldots, \hat{Y}_i, \ldots, Y_p\right)\right) \\
+ \sum_{i<j} (-1)^{i+j} \omega&\left(\left[Y_i, Y_j\right], Y_0, \ldots, \hat{Y}_i, \ldots, \hat{Y}_j, \ldots, Y_p\right), \\
L_X &= d \circ i_X + i_X \circ d, \\
L_X \circ d &= d \circ L_X, \\
i_X \circ L_X &= L_X \circ i_X.
\end{aligned}$$

A.2 Integration of Forms

We shall assume that M is an oriented n-manifold. Thus, M comes with a covering of charts $\varphi_\alpha = \left(x_\alpha^1, \ldots, x_\alpha^n\right) : U_\alpha \longleftrightarrow B(0,1) \subset \mathbb{R}^n$ such that the transition functions $\varphi_\alpha \circ \varphi_\beta^{-1}$ preserve the usual orientation on Euclidean space, i.e.,

$\det\left(D\left(\varphi_\alpha \circ \varphi_\beta^{-1}\right)\right) > 0$. In addition, we shall also assume that a partition of unity with respect to this covering is given. In other words, we have smooth functions $\phi_\alpha : M \to [0, 1]$ such that $\phi_\alpha = 0$ on $M - U_\alpha$ and $\sum_\alpha \phi_\alpha = 1$. For the last condition to make sense, it is obviously necessary that the covering be also locally finite.

Given an n-form ω on M we wish to define:

$$\int_M \omega.$$

When M is not compact, it might be necessary to assume that the form has compact support, i.e., it vanishes outside some compact subset of M.

In each chart we can write

$$\omega = f_\alpha dx_\alpha^1 \wedge \cdots \wedge dx_\alpha^n.$$

Using the partition of unity, we then obtain

$$\begin{aligned}\omega &= \sum_\alpha \phi_\alpha \omega \\ &= \sum_\alpha \phi_\alpha f_\alpha dx_\alpha^1 \wedge \cdots \wedge dx_\alpha^n,\end{aligned}$$

where each of the forms $\phi_\alpha f_\alpha dx_\alpha^1 \wedge \cdots \wedge dx_\alpha^n$ has compact support in U_α. Since U_α is identified with $B(0, 1)$, we simply declare that

$$\int_{U_\alpha} \phi_\alpha f_\alpha dx_\alpha^1 \wedge \cdots \wedge dx_\alpha^n = \int_{B(0,1)} \phi_\alpha f_\alpha dx^1 \cdots dx^n.$$

Here the right-hand side is simply the integral of the function $\phi_\alpha f_\alpha$ viewed as a function on $B(0, 1)$. Then we define

$$\int_M \omega = \sum_\alpha \int_{U_\alpha} \phi_\alpha f_\alpha dx_\alpha^1 \wedge \cdots \wedge dx_\alpha^n$$

whenever this sum converges. Using the standard change of variables formula for integration on Euclidean space, we see that indeed this definition is independent of the choice of coordinates.

With these definitions behind us, we can now state and prove Stokes' theorem for manifolds without boundary.

Theorem 2.1 *For any $\omega \in \Omega^{n-1}(M)$ with compact support we have*

$$\int_M d\omega = 0.$$

Proof. If we use the trick

$$d\omega = \sum_\alpha d(\phi_\alpha \omega),$$

then we see that it suffices to prove the theorem in the case $M = B(0,1)$ and ω has compact support on $B(0,1)$. Then write

$$\omega = \sum_{i=1}^{n} f_i dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n,$$

where the functions f_i are zero near the boundary of $B(0,1)$. The differential of ω is now easily computed:

$$\begin{aligned} d\omega &= \sum_{i=1}^{n} (df_i) \wedge dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n \\ &= \sum_{i=1}^{n} \left(\frac{\partial f_i}{\partial x^i}\right) dx^i \wedge dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n \\ &= \sum_{i=1}^{n} (-1)^{i-1} \left(\frac{\partial f_i}{\partial x^i}\right) dx^1 \wedge \cdots \wedge dx^i \wedge \cdots \wedge dx^n. \end{aligned}$$

Thus,

$$\begin{aligned} \int_{B(0,1)} \omega &= \int_{B(0,1)} \sum_{i=1}^{n} (-1)^{i-1} \left(\frac{\partial f_i}{\partial x^i}\right) dx^1 \cdots dx^n \\ &= \sum_{i=1}^{n} (-1)^{i-1} \int_{B(0,1)} \left(\frac{\partial f_i}{\partial x^i}\right) dx^1 \cdots dx^n \\ &= \sum_{i=1}^{n} (-1)^{i-1} \int \left(\int \left(\frac{\partial f_i}{\partial x^i}\right) dx^i \right) dx^1 \cdots \widehat{dx^i} \cdots dx^n. \end{aligned}$$

The fundamental theorem of calculus tells us that

$$\int \left(\frac{\partial f_i}{\partial x^i}\right) dx^i = 0,$$

as f_i is zero near the boundary of the range of x^i. In particular, the entire integral must be zero. □

Stokes' theorem leads to some important formulae on Riemannian manifolds.

Corollary 2.2 (The Divergence Theorem) *If X is a vector field on (M, g) with compact support, then*

$$\int_M \operatorname{div} X \cdot d\text{vol} = 0.$$

Proof. Just observe

$$\begin{aligned} \operatorname{div} X \cdot d\text{vol} &= L_X d\text{vol} \\ &= i_X d\,(d\text{vol}) + d\,(i_X d\text{vol}) \\ &= d\,(i_X d\text{vol}) \end{aligned}$$

and use Stokes' theorem. □

Corollary 2.3 (Green's Formulae) *If f_1, f_2 are two compactly supported functions on (M, g), then*

$$\int_M (\Delta f_1) \cdot f_2 \cdot d\mathrm{vol} = -\int_M g(\nabla f_1, \nabla f_2) = \int_M f_1 \cdot (\Delta f_2) \cdot d\mathrm{vol}.$$

Proof. Just use that

$$\mathrm{div}(f_1 \cdot \nabla f_2) = g(\nabla f_1, \nabla f_2) + f_1 \cdot \Delta f_2,$$

and apply the divergence theorem to get the desired result. □

Corollary 2.4 (Integration by Parts) *If S, T are two $(1, p)$ tensors with compact support on (M, g), then*

$$\int_M g\left(S^\flat, \nabla \mathrm{div} T\right) \cdot d\mathrm{vol} = -\int_M g(\mathrm{div} S, \mathrm{div} T) \cdot d\mathrm{vol},$$

where $S^\flat$ denotes the $(0, p+1)$-tensor defined by

$$S^\flat(X, Y, Z, \ldots) = g(X, S(Y, Z, \ldots)).$$

Proof. For simplicity, first assume that S and T are vector fields X and Y. Then the formula can be interpreted as

$$\int_M g(X, \nabla \mathrm{div} Y) \cdot d\mathrm{vol} = -\int_M \mathrm{div} X \cdot \mathrm{div} Y \cdot d\mathrm{vol}.$$

We can then use that

$$\mathrm{div}(f \cdot X) = g(\nabla f, X) + f \cdot \mathrm{div} X.$$

Therefore, if we define $f = \mathrm{div} Y$ and use the divergence theorem, we get the desired formula.

In general, choose an orthonormal frame E_i, and observe that we can define a vector field by

$$X = \sum_{i_1, \ldots, i_p} S\left(E_{i_1}, \ldots, E_{i_p}\right) \mathrm{div} T\left(E_{i_1}, \ldots, E_{i_p}\right).$$

In other words, if we think of $g\left(V, S\left(X_1, \ldots, X_p\right)\right)$ as a $(0, p)$-tensor, then X is implicitly defined by

$$g(X, V) = g(g(V, S), \mathrm{div} T).$$

Then we have
$$\mathrm{div}X = g\left(\mathrm{div}S, \mathrm{div}T\right) - g\left(S^\flat, \nabla\mathrm{div}T\right),$$
and the formula is established as before. □

It is worthwhile pointing out that it is NOT in general true that
$$\int_M g\left(S^\flat, \mathrm{div}\nabla T\right) = -\int_M g\left(\nabla S, \nabla T\right),$$
even when the tensors are vector fields. On Euclidean space, for example, simply define $S = T = x^1\partial_1$. Then
$$\begin{aligned} \nabla\left(x^1\partial_1\right) &= dx^1\partial_1, \\ \left|dx^1\partial_1\right| &= 1, \\ \mathrm{div}\left(dx^1\partial_1\right) &= 0. \end{aligned}$$
Of course, the tensors in this example do not have compact support, but that can easily be fixed by multiplying with a compactly supported function.

A.3 Čech Cohomology

Before defining de Rham cohomology, we shall briefly mention how Čech cohomology is defined. This is the cohomology theory that seems most natural from a geometric point of view. Also, it is the cohomology that is most naturally associated with de Rham cohomology.

For a manifold M, suppose that we have a covering of contractible open sets U_α such that all possible nonempty intersections $U_{\alpha_1} \cap \cdots \cap U_{\alpha_k}$ are also contractible. Such a covering is called a *good cover*. Now let I^k be the set of ordered indices that create nontrivial intersections
$$I^k = \left\{(\alpha_0, \ldots, \alpha_k) : U_{\alpha_0} \cap \cdots \cap U_{\alpha_k} \neq 0\right\}.$$
Čech cycles with values in a ring R are defined as a space of alternating maps
$$\check{Z}^k = \left\{f : I^k \to R : f \circ \tau = -f \text{ where } \tau \text{ is a transposition of two indices}\right\}.$$
The differential, or coboundary operator, is now defined by
$$\begin{aligned} \check{Z}^k &\to \check{Z}^{k+1}, \\ df\left(\alpha_0, \ldots, \alpha_{k+1}\right) &= \sum_{i=0}^{k} (-1)^i f\left(\alpha_0, \ldots, \hat{\alpha}_i, \ldots, \alpha_{k+1}\right). \end{aligned}$$
Čech cohomology is then defined as
$$H^k(M, R) = \frac{\ker\left(d : \check{Z}^k \to \check{Z}^{k+1}\right)}{\mathrm{im}\left(d : \check{Z}^{k-1} \to \check{Z}^{k+1}\right)}.$$

The standard arguments with refinements of covers can be used to show that this cohomology theory is independent of the choice of good cover. Below, we shall define de Rham cohomology for forms and prove several properties for that cohomology theory. At each stage one can easily see that Čech cohomology satisfies those same properties. Note that Čech cohomology seems almost purely combinatorial. This feature makes it very natural to work with in many situations.

A.4 de Rham Cohomology

Throughout we let M be an n-manifold. Using that $d \circ d = 0$, we trivially get that the exact forms $B^p(M) = d\left(\Omega^{p-1}(M)\right)$ are a subset of the closed forms $Z^p(M) = \{\omega \in \Omega^p(M) : d\omega = 0\}$. The de Rham cohomology is then defined as

$$H^p(M) = \frac{Z^p(M)}{B^p(M)}.$$

Given a closed form ψ, we let $[\psi]$ denote the corresponding cohomology class.

The first simple property comes form the fact that any function with zero differential must be locally constant. On a connected manifold we therefore have

$$H^0(M) = \mathbb{R}.$$

Given a smooth map $f : M \to N$, we get an induced map in cohomology:

$$\begin{aligned} H^p(N) &\to H^p(M), \\ f^*([\psi]) &= [f^*\psi]. \end{aligned}$$

This definition is independent of the choice of ψ, since the pullback f^* commutes with d.

The two key results that are needed for a deeper understanding of de Rham cohomology are the Meyer-Vietoris sequence and the Poincaré lemma.

Lemma 4.1 (The Meyer-Vietoris Sequence) *If $M = A \cup B$ for open sets $A, B \subset M$, then there is a long exact sequence*

$$\cdots \to H^p(M) \to H^p(A) \oplus H^p(B) \to H^p(A \cap B) \to H^{p+1}(M) \to \cdots.$$

Proof. The proof is given in outline, as it is exactly the same as the corresponding proof in algebraic topology.

First, we need to define the maps. We clearly have inclusions

$$\begin{aligned} H^p(M) &\to H^p(A), \\ H^p(M) &\to H^p(B), \\ H^p(A) &\to H^p(A \cap B), \\ H^p(B) &\to H^p(A \cap B). \end{aligned}$$

By adding the first two, we get

$$\begin{aligned} H^p(M) &\to H^p(A) \oplus H^p(B), \\ [\psi] &\to \left([\psi_{|A}], [\psi_{|B}]\right). \end{aligned}$$

Subtraction of the last two, yields

$$\begin{aligned} H^p(A) \oplus H^p(B) &\to H^p(A \cap B), \\ ([\omega], [\psi]) &\to \left[\omega_{|A\cap B}\right] - \left[\psi_{|A\cap B}\right]. \end{aligned}$$

With these definitions it is not hard to see that the sequence is exact at $H^p(A) \oplus H^p(B)$.

The coboundary operator $H^p(A \cap B) \to H^{p+1}(M)$ is as usual defined by considering the exact diagram

$$\begin{array}{ccccccccc} 0 & \to & \Omega^{p+1}(M) & \to & \Omega^{p+1}(A) \oplus \Omega^{p+1}(B) & \to & \Omega^{p+1}(A \cap B) & \to & 0 \\ & & \uparrow d & & \uparrow d & & \uparrow d & & \\ 0 & \to & \Omega^p(M) & \to & \Omega^p(A) \oplus \Omega^p(B) & \to & \Omega^p(A \cap B) & \to & 0. \end{array}$$

If we take a closed form $\omega \in Z^p(A \cap B)$, then we have $\psi \in \Omega^p(A) \oplus \Omega^p(B)$, which is mapped to ω. Then $d\psi$ is zero when mapped to $\Omega^{p+1}(A \cap B)$, as we assumed that $d\omega = 0$. But then exactness tells us that $d\psi$ must come from an element in $\Omega^{p+1}(M)$. It is now easy to see that in cohomology, this element is well defined and gives us a linear map

$$H^p(A \cap B) \to H^{p+1}(M)$$

that makes the Meyer-Vietoris sequence exact. □

Lemma 4.2 (The Poincaré Lemma) *The cohomology of the open unit disk $B(0,1) \subset \mathbb{R}^n$ is*

$$\begin{aligned} H^0(B(0,1)) &= \mathbb{R}, \\ H^p(B(0,1)) &= \{0\} \text{ for } p > 0. \end{aligned}$$

Proof. Evidently, the proof hinges on showing that any closed p-form ω is exact when $p > 0$. Using that the form is closed, we see that for any vector field

$$L_X\omega = di_X\omega.$$

We shall use the radial field $X = \sum x^i \partial_i$ to construct a map

$$H : \Omega^p \to \Omega^p$$

that satisfies

$$\begin{aligned} H \circ L_X &= id, \\ d \circ H &= H \circ d. \end{aligned}$$

This is clearly enough, as we would then have

$$\omega = d\left(H\left(i_X \omega\right)\right).$$

Since L_X is differentiation in the direction of the radial field, the map H should be integration in the same direction. Motivated by this, define

$$H\left(f dx^{i_1} \wedge \cdots \wedge dx^{i_p}\right) = \left(\int_0^1 t^{p-1} f(tx)\, dt\right) dx^{i_1} \wedge \cdots \wedge dx^{i_p}$$

and extend it to all forms using linearity. We now need to check the two desired properties. This is done by direct calculations:

$$\begin{aligned}
H \circ L_X\left(f dx^{i_1} \wedge \cdots \wedge dx^{i_p}\right) &= H\left(x^i \partial_i f dx^{i_1} \wedge \cdots \wedge dx^{i_p}\right.\\
&\quad \left. + f L_X\left(dx^{i_1} \wedge \cdots \wedge dx^{i_p}\right)\right)\\
&= H\left(x^i \partial_i f dx^{i_1} \wedge \cdots \wedge dx^{i_p}\right.\\
&\quad \left. + p f dx^{i_1} \wedge \cdots \wedge dx^{i_p}\right)\\
&= \left(\left(\int_0^1 t^{p-1}\left(t x^i\right) \partial_i f(tx)\, dt\right)\right.\\
&\quad \left. + p\left(\int_0^1 p t^{p-1} f(tx)\, dt\right)\right) \cdot dx^{i_1} \wedge \cdots \wedge dx^{i_p}\\
&= \left(\int_0^1 \frac{d}{dt}\left(t^p \cdot f(tx)\right) dt\right) dx^{i_1} \wedge \cdots \wedge dx^{i_p}\\
&= f(x)\, dx^{i_1} \wedge \cdots \wedge dx^{i_p};\\
H \circ d\left(f dx^{i_1} \wedge \cdots \wedge dx^{i_p}\right) &= H\left(\partial_i f \cdot dx^i \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_p}\right)\\
&= \left(\left(\int_0^1 t^p \partial_i f(tx)\, dt\right) dx^i\right) \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_p}\\
&= d\left(\int_0^1 t^{p-1} f(tx)\, dt\right) \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_p}\\
&= d\left(\left(\int_0^1 t^{p-1} f(tx)\, dt\right) \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_p}\right)\\
&= d \circ H\left(f dx^{i_1} \wedge \cdots \wedge dx^{i_p}\right).
\end{aligned}$$

This finishes the proof. □

We can now prove de Rham's theorem.

Theorem 4.3 (de Rham, 1931) *If M is a closed manifold, then the de Rham cohomology groups $H^p(M)$ are the same as the Čech, or singular, cohomology groups $H^p(M, \mathbb{R})$ with real coefficients. In particular, all the cohomology groups are finitely generated.*

Proof. We first observe that both theories have natural Meyer-Vietoris sequences. Therefore, if M has a finite covering by open sets U_α with the property that

$$H^p\left(U_{\alpha_1}\cap\cdots\cap U_{\alpha_k}\right)=H^p\left(U_{\alpha_1}\cap\cdots\cap U_{\alpha_k},\mathbb{R}\right)$$

for all p and intersections $U_{\alpha_1}\cap\cdots\cap U_{\alpha_k}$, then using induction on the number of elements in the covering, we see that the two cohomologies of M are the same.

To find such a covering, take a Riemannian metric on M. Then find a covering of convex balls $B(p_\alpha, r)$. The intersections of convex balls are clearly diffeomorphic to the unit ball. Thus, the Poincaré lemma ensures that the two cohomology theories are the same on all intersections.

It also follows from this proof that the cohomology groups are finitely generated. □

Note that the proof hinges on the fact that we have good coverings, which were also at the heart of the definition of Čech cohomology.

Suppose now we have two manifolds M and N with good coverings $\{U_\alpha\}$ and $\{V_\beta\}$. A map $f : M \to N$ is said to preserve these coverings if for each α we can find $\beta(\alpha)$ such that

$$f(U_\alpha)\subset V_{\beta(\alpha)}.$$

Given a good cover of N and a map $f : M \to N$, we can clearly always find a good covering of M such that f preserves these covers. The induced map: $f^* : H^p(N) \to H^p(M)$ is now completely determined by the combinatorics of the map $\alpha \to \beta(\alpha)$. This makes it possible to define f^* for all continuous maps. Moreover, since the set of maps that satisfy

$$f(U_\alpha)\subset V_{\beta(\alpha)}$$

is open, we see that any map close to f induces the same map in cohomology. Consequently, homotopic maps must induce the same map in cohomology. This gives a very important result.

Theorem 4.4 *If two manifolds, possibly of different dimension, M and N are homotopy equivalent, then they have the same cohomology.*

A.5 Poincaré Duality

The last piece of information we need to understand is how the wedge product acts on cohomology. It is easy to see that we have a map

$$\begin{aligned}H^p(M)\times H^q(M)&\to H^{p+q}(M),\\ ([\psi],[\omega])&\to[\psi\wedge\omega].\end{aligned}$$

We are interested in understanding what happens in case $p+q=n$. This requires a surprising amount of preparatory work. First we have

Theorem 5.1 *If M is an oriented closed n-manifold, then we have a well-defined isomorphism*

$$H^n(M) \to \mathbb{R},$$
$$[\omega] \to \int_M \omega.$$

Proof. That the map is well-defined follows from Stokes' theorem. It is also onto, since any form with the property that it is positive when evaluated on a positively oriented frame is integrated to a positive number. Thus, we must show that any form with $\int_M \omega = 0$ is exact. This is not easy to show, and in fact, it is more natural to show this in a more general context: If M is an oriented n-manifold that can be covered by finitely many charts, then any compactly supported n-form ω with $\int_M \omega = 0$ is exact.

The proof of this result is by induction on the number of charts it takes to cover M. But before we can start the inductive procedure, we must establish the result for the n-sphere.

Case 1: $M = S^n$. Cover M by two open discs whose intersection is homotopy equivalent to S^{n-1}. Then use induction on n together with the Meyer-Vietoris sequence to show that for each $n > 0$,

$$H^p(S^n) = \begin{cases} 0, & p \neq 0, n, \\ \mathbb{R}, & p = 0, n. \end{cases}$$

The induction apparently starts at $n = 0$ and S^0 consists of two points and therefore has $H^0(S^0) = \mathbb{R} \oplus \mathbb{R}$. Having shown that $H^n(S^n) = \mathbb{R}$, it is then clear that the map $\int : H^n(S^n) \to \mathbb{R}$ is an isomorphism.

Case 2: $M = B(0, 1)$. We can think of M as being an open hemisphere of S^n. Any compactly supported form ω on M therefore yields a form on S^n. Given that $\int_M \omega = 0$, we therefore also get that $\int_{S^n} \omega = 0$. Thus, ω must be exact on S^n. Let $\psi \in \Omega^{n-1}(S^n)$ be chosen such that $d\psi = \omega$. Use again that ω is compactly supported to find an open disc N such that ω vanishes on N and $N \cup M = S^n$. Then ψ is clearly closed on N and must by the Poincaré lemma be exact. Thus, we can find $\theta \in \Omega^{n-2}(N)$ with $d\theta = \psi$ on N. Now observe that $\psi - d\theta$ is actually defined on all of S^n, as it vanishes on N. But then we have found a form $\psi - d\theta$ with support in M whose differential is ω.

Case 3: $M = A \cup B$ where the result holds on A, B, and $A \cap B$. Select a partition of unity $\phi_A + \phi_B$ subordinate to the cover $\{A, B\}$. Given an n-form ω with $\int_M \omega = 0$, we get two forms $\phi_A \cdot \omega$ and $\phi_B \cdot \omega$ with support in A and B, respectively. Using our assumptions, we see that

$$\begin{aligned} 0 &= \int_M \omega \\ &= \int_A \phi_A \cdot \omega + \int_B \phi_B \cdot \omega. \end{aligned}$$

On $A \cap B$ we can by assumption (orientability is used here) select an n-form ψ with compact support inside $A \cap B$ such that

$$\int_{A\cap B} \tilde{\omega} = \int_A \phi_A \cdot \omega.$$

Using $\tilde{\omega}$ we can create two forms,

$$\begin{aligned} &\phi_A \cdot \omega - \tilde{\omega}, \\ &\phi_B \cdot \omega + \tilde{\omega}, \end{aligned}$$

with support in A and B, respectively. From our assumptions it follows that they both have integral zero. Thus, we can by assumption find ψ_A and ψ_B with support in A and B, respectively, such that

$$\begin{aligned} d\psi_A &= \phi_A \cdot \omega - \tilde{\omega}, \\ d\psi_B &= \phi_B \cdot \omega + \tilde{\omega}. \end{aligned}$$

Then we get a globally defined form $\psi = \psi_A + \psi_B$ with

$$\begin{aligned} d\psi &= \phi_A \cdot \omega - \tilde{\omega} + \phi_B \cdot \omega + \tilde{\omega} \\ &= (\phi_A + \phi_B) \cdot \omega \\ &= \omega. \end{aligned}$$

The theorem now follows by using induction on the number of charts it takes to cover M. □

The above proof indicates that it is really more convenient to work with compactly supported forms. This leads us to *compactly supported cohomology*, which is defined as follows: Let $\Omega_c^p(M)$ denote the compactly supported p-forms. With this we have the compactly supported exact and closed forms $B_c^p(M) \subset Z_c^p(M)$ (note that $d : \Omega_c^p(M) \to \Omega_c^{p+1}(M)$). Then define

$$H_c^p(M) = \frac{Z_c^p(M)}{B_c^p(M)}.$$

Needless to say, for closed manifolds the two cohomology theories are identical. For open manifolds, on the other hand, we have that the closed 0-forms must be zero, as they also have to have compact support. Thus $H_c^0(M) = \{0\}$ if M is not closed.

Note that only proper maps $f : M \to N$ have the property that they map $f^* : \Omega_c^p(N) \to \Omega_c^p(M)$. In particular, if $A \subset M$ is open, we do not have a map $H_c^p(M) \to H_c^p(A)$. Instead we observe that there is a natural inclusion $\Omega_c^p(A) \to \Omega_c^p(M)$, which induces

$$H_c^p(A) \to H_c^p(M).$$

The above proof, stated in our new terminology, states that

$$H_c^n(M) \to \mathbb{R},$$
$$[\omega] \to \int_M \omega$$

is an isomorphism for oriented n-manifolds. Moreover, using that $B(0,1) \subset S^n$, we can easily prove the following version of the Poincaré lemma:

$$H_c^p(B(0,1)) = \begin{cases} 0, & p \neq n, \\ \mathbb{R}, & p = n. \end{cases}$$

In order to carry out induction proofs effectively with this cohomology theory, we also need a Meyer-Vietoris sequence:

$$\cdots \leftarrow H_c^p(M) \leftarrow H_c^p(A) \oplus H_c^p(B) \leftarrow H_c^p(A \cap B) \leftarrow H_c^{p+1}(M) \leftarrow \cdots.$$

This is established in the same way using the diagram

$$\begin{array}{ccccccccc} 0 & \leftarrow & \Omega_c^{p+1}(M) & \leftarrow & \Omega_c^{p+1}(A) \oplus \Omega_c^{p+1}(B) & \leftarrow & \Omega_c^{p+1}(A \cap B) & \leftarrow & 0 \\ & & \uparrow d & & \uparrow d & & \uparrow d & & \\ 0 & \leftarrow & \Omega_c^p(M) & \leftarrow & \Omega_c^p(A) \oplus \Omega_c^p(B) & \leftarrow & \Omega_c^p(A \cap B) & \leftarrow & 0. \end{array}$$

Theorem 5.2 *Let M be an oriented n-manifold that can be covered by finitely many charts. The pairing*

$$H^p(M) \times H_c^{n-p}(M) \to \mathbb{R},$$
$$([\omega],[\psi]) \to \int_M \omega \wedge \psi$$

is well-defined and nondegenerate. In particular, the two cohomology groups $H^p(M)$ and $H_c^{n-p}(M)$ are dual to each other and therefore have the same dimension as finite-dimensional vector spaces.

Proof. We proceed by induction on the number of charts it takes to cover M. For the case $M = B(0,1)$, this theorem follows from the two versions of the Poincaré lemma. In general suppose $M = A \cup B$, where the theorem is true for A, B, and $A \cap B$. Note that the pairing gives a natural map

$$H^p(N) \to \left(H_c^{n-p}(N)\right)^* = \mathrm{Hom}\left(H_c^{n-p}(N), \mathbb{R}\right)$$

for any manifold N. We apparently assume that this map is an isomorphism for $N = A, B, A \cap B$. Using that taking duals reverses arrows, we obtain a diagram where the left- and rightmost columns have been eliminated

$$\begin{array}{ccccccc} \to & H^p(A \cap B) & \to & H^{p+1}(M) & \to & H^{p+1}(A) \oplus H^p(B) & \to \\ & \downarrow & & \downarrow & & \downarrow & \\ \to & \left(H_c^p(A \cap B)\right)^* & \to & \left(H_c^{p+1}(M)\right)^* & \to & \left(H^{p+1}(A)\right)^* \oplus \left(H^p(B)\right)^* & \to . \end{array}$$

Each square in this diagram is either commutative or anticommutative (i.e., commutes with a minus sign.) As all vertical arrows, except for the middle one, are assumed to be isomorphisms, we see by a simple diagram chase (the five lemma) that the middle arrow is also an isomorphism. □

Corollary 5.3 *On a closed oriented n-manifold M we have that $H^p(M)$ and $H^{n-p}(M)$ are isomorphic.*

A.6 Degree Theory

Given the simple nature of the top cohomology class of a manifold, we see that maps between manifolds of the same dimension can act only by multiplication on the top cohomology class. We shall see that this multiplicative factor is in fact an integer, called the *degree* of the map.

To be precise, suppose we have two oriented n-manifolds M and N and also a proper map $f : M \to N$. Then we get a diagram

$$\begin{array}{ccc} H_c^n(N) & \stackrel{f^*}{\to} & H_c^n(M) \\ \downarrow \int & & \downarrow \int \\ \mathbb{R} & \stackrel{d}{\to} & \mathbb{R}. \end{array}$$

Since the vertical arrows are isomorphisms, the induced map f^* yields a unique map $d : \mathbb{R} \to \mathbb{R}$. This map must be multiplication by some number, which we call the degree of f, denoted by $\deg f$. Clearly, the degree is defined by the property

$$\int_M f^*\omega = \deg f \cdot \int_N \omega.$$

Lemma 6.1 *If $f : M \to N$ is a diffeomorphism between oriented n-manifolds, then* $\deg f = \pm 1$, *depending on whether f preserves or reverses orientation.*

Proof. Note that our definition of integration of forms is independent of coordinate changes. It relies only on a choice of orientation, and if this choice is changed then the integral changes by a sign. This clearly establishes the lemma. □

Theorem 6.2 *If $f : M \to N$ is a proper map between oriented n-manifolds, then* $\deg f$ *is an integer.*

Proof. The proof will also give a recipe for computing the degree. First, we must appeal to Sard's theorem. This theorem ensures that we can find $y \in N$ such that for each $x \in f^{-1}(y)$ the differential $Df : T_xM \to T_yN$ is an isomorphism. The inverse function theorem then tells us that f must be a diffeomorphism in a neighborhood of each such x. In particular, the preimage $f^{-1}(y)$ must be a discrete set. As we also assumed the map to be proper, we can conclude that the preimage is finite: $\{x_1, \ldots, x_k\} = f^{-1}(y)$. We can then find a neighborhood U of y in N, and neighborhoods U_i of x_i in M, such that $f : U_i \to U$ is a diffeomorphism for each i. Now select $\omega \in \Omega_c^n(U)$ with $\int \omega = 1$. Then we can write

$$f^*\omega = \sum_{i=1}^{k} f^*\omega_{|U_i},$$

where each $f^*\omega_{|U_i}$ has support in U_i. The above lemma now tells us that

$$\int_{U_i} f^*\omega_{|U_i} = \pm 1.$$

Hence,

$$\begin{aligned} \deg f &= \deg f \cdot \int_N \omega \\ &= \deg f \cdot \int_U \omega \\ &= \int_M f^*\omega \\ &= \sum_{i=1}^{k} \int_{U_i} f^*\omega_{|U_i} \end{aligned}$$

is an integer. □

Note that $\int_{U_i} f^*\omega_{|U_i}$ is ± 1, depending simply on whether f preserves or reverses the orientations at x_i. Thus, the degree simply counts the number of preimages for regular values with sign. In particular, a finite covering map has degree equal to the number of sheets in the covering.

On an oriented Riemannian manifold (M, g) we always have a canonical volume form denoted by $d\text{vol}_g$. Using this form, we see that the degree of a map between closed Riemannian manifolds $f : (M, g) \to (N, h)$ can be computed as

$$\deg f = \frac{\int_M f^*(d\text{vol}_h)}{\text{vol}(N)}.$$

In case f is locally a Riemannian isometry, we must have that

$$f^*(d\text{vol}_h) = \pm d\text{vol}_g.$$

Hence,

$$\deg f = \pm \frac{\text{vol} M}{\text{vol} N}.$$

This gives the well-known formula for the relationship between the volumes of Riemannian manifolds that are related by a finite covering map.

A.7 Further Study

There are several texts that expand on the material covered here. The book by Warner [82] is more than sufficient for most purposes. There is also a very nice book by Bott and Tu [15] that in addition covers characteristic classes. This book only has the small defect that it doesn't mention how one can compute characteristic classes using curvature forms. This can, however, be found in [76, vol. V].

Appendix B
Principal Bundles

We shall here give a more sophisticated version of parts of Riemannian geometry. The goal is to understand, in a unified way, how all tensor bundles are constructed and then see how the covariant derivative acts on tensors. The story begins with Cartan formalism. This is simply a different way of keeping track of covariant derivatives and the curvature tensor using the language of differential forms. Cartan formalism is at first defined only in terms of frames. In order to make the theory invariant, we need to work on the frame bundle. After the theory has been transformed to the frame bundle, we then observe that all tensors are $O(n)$ invariant maps on this frame bundle. This enables us to define covariant differentiation, in one fell swoop, on all tensors at the same time.

This new point of view easily generalizes to a formalism of principal bundles with a given Lie group as structure group. One can define connections on such bundles, and then covariant differentiation on sections of vector bundles, that are associated to the given principal bundle. At first, this all just seems like generalization for the sake of generalization. There are, however, some very important bundles that can really only be understood in this general context. This will be further investigated in the next appendix.

B.1 Cartan Formalism

The original form of Cartan formalism has indeed caused many headaches over the years. We shall here try to explain it as best we can. The central idea is to develop expressions for the connection and curvature using orthonormal frames.

Thus, the entire approach is frame dependent rather than invariant. The advantage is that many calculations become simpler and also that the usual properties for the curvature tensor simply become consequences of this new setup.

An infinitesimal movement on a manifold is simply a tangent vector. Thus, the notation dp or $p+dp$ denotes an element of T_pM, i.e., an infinitesimal movement of $p \in M$. Given a frame E_i and dual coframe θ^i, we can then write

$$v = \theta^i (v) E_i,$$
$$dp = \theta \cdot E,$$

where

$$E = (E_1, \ldots, E_n),$$
$$\theta = \begin{pmatrix} \theta^1 \\ \vdots \\ \theta^n \end{pmatrix}.$$

Thus, $\theta \cdot E$ is nothing but the identity map on T_pM. Note that frames are consistently written as row vectors, while coframes are column vectors. The idea is that θ carries metric information in case the frame is assumed to be orthonormal. Conversely, any choice of frame, at least locally, induces a Riemannian metric by declaring it to be orthonormal.

The *connection form* is now defined as follows:

Proposition 1.1 (E. Cartan) *If we define* $d\theta = \left(d\theta^1, \ldots, d\theta^n\right)$, *then there is a unique matrix of 1-forms* $\omega = \left(\omega^i_j\right)$ *such that*

$$d\theta = -\omega \wedge \theta,$$
$$d\theta^i = -\omega^i_j \wedge \theta^j,$$

and

$$\omega^t = -\omega,$$
$$\omega^i_j = -\omega^j_i.$$

Proof. Observe that since ω is a matrix and θ is a column, we put θ on the right. By assuming that we start out with an orthonormal frame, we can generate the connections forms by declaring

$$\nabla_X E_j = \omega^i_j (X) E_i.$$

This certainly defines some 1-forms $\left(\omega^i_j\right)$. The skew-symmetry property comes from the connection being metric:

$$\begin{aligned} 0 &= \nabla_X g\left(E_i, E_j\right) \\ &= g\left(\nabla_X E_i, E_j\right) + g\left(E_i, \nabla_X E_j\right) \\ &= \omega^j_i (X) + \omega^i_j (X). \end{aligned}$$

The proof of the formula $d\theta = -\omega \wedge \theta$ uses that the connection is torsion free. Apparently, we must show

$$d\theta^i (E_k, E_l) = -\left(\omega^i_j \wedge \theta^j\right)(E_k, E_l).$$

The left-hand side is by definition

$$\begin{aligned} d\theta^i (E_k, E_l) &= D_{E_k}\theta^i (E_l) - D_{E_l}\theta^i (E_k) - \theta^i ([E_k, E_l]) \\ &= -\theta^i ([E_k, E_l]). \end{aligned}$$

The right-hand side, on the other hand, is

$$\begin{aligned} -\left(\omega^i_j \wedge \theta^j\right)(E_k, E_l) &= \omega^i_j (E_l)\theta^j (E_k) - \omega^i_j (E_k)\theta^j (E_l) \\ &= \omega^i_k (E_l) - \omega^i_l (E_k) \\ &= \theta^i \left(\nabla_{E_l} E_k\right) - \theta^i \left(\nabla_{E_k} E_l\right) \\ &= -\theta^i ([E_k, E_l]). \end{aligned}$$

Thus, they must be equal to each other. □

The equations

$$\begin{aligned} d\theta &= -\omega \wedge \theta, \\ \omega &= -\omega^t \end{aligned}$$

are called the *first structural equations*. Apparently, they define a unique connection that is torsion free and metric. The frame version of the first structural equations can also be written in a more compact matrix version just as for forms:

$$\begin{aligned} \nabla E &= E \cdot \omega, \\ \nabla E_i &= E_j \omega^j_i. \end{aligned}$$

We now come to the curvature tensor.

Proposition 1.2 (E. Cartan) *The equations*

$$\begin{aligned} \Omega &= d\omega + \omega \wedge \omega, \\ \Omega^i_j &= d\omega^i_j + \omega^i_k \wedge \omega^k_j \end{aligned}$$

define a skew-symmetric matrix of 2-forms that also gives us the curvature tensor via

$$R(X, Y) E_j = \Omega^i_j (X, Y) \cdot E_i.$$

Proof. To see why this is true, we simply compute both sides using how they are defined:

$$R(X, Y) E_j = \nabla_X \nabla_Y E_j - \nabla_Y \nabla_X E_j - \nabla_{[X,Y]} E_j;$$

$$\begin{aligned}
\left(d\omega_j^i + \omega_k^i \wedge \omega_j^k\right)(X, Y) \cdot E_i &= \left(d\omega_j^i (X, Y) + \omega_k^i \wedge \omega_j^k (X, Y)\right) \cdot E_i \\
&= \left(\nabla_X \omega_j^i (Y)\right) \cdot E_i \left(\nabla_Y \omega_j^i (X)\right) \cdot E_i \\
&\quad - \left(\omega_j^i ([X, Y])\right) \cdot E_i + \omega_k^i (X) \cdot E_i \cdot \omega_j^k (Y) \\
&\quad - \omega_k^i (Y) \cdot E_i \cdot \omega_j^k (X) \\
&= \nabla_X \left(\nabla_Y E_j\right) - \omega_j^i (Y) \nabla_X E_i - \nabla_Y \left(\nabla_X E_j\right) \\
&\quad + \omega_j^i (X) \nabla_Y E_i - \nabla_{[X,Y]} E_j + \omega_j^k (Y) \nabla_X E_k \\
&\quad - \omega_j^k (X) \nabla_Y E_k \\
&= R(X, Y) E_i + \omega_j^i (X) \nabla_Y E_i - \omega_j^k (X) \nabla_Y E_k \\
&\quad + \omega_j^k (Y) \nabla_X E_k - \omega_j^i (Y) \nabla_X E_i \\
&= R(X, Y) E_i. \qquad \square
\end{aligned}$$

These new equations

$$\Omega = d\omega + \omega \wedge \omega$$

are called the *second structural equations*, and the skew-symmetric matrix Ω of 2-forms is called the *curvature form*.

At this point it, would probably be useful to see how this works in action.

Example 1.3 Let $M = \mathbb{R}^n$. The usual Cartesian coordinate frame $(\partial_1, \ldots, \partial_n)$ can then be used as an orthonormal frame. The structural equations will now look like

$$\begin{aligned}
dp &= dx^i \partial_i, \\
\theta &= \left(dx^1, \ldots, dx^n\right), \\
d\theta &= 0, \\
\omega &= 0, \\
\Omega &= 0.
\end{aligned}$$

For an arbitrary orthonormal frame E_i with dual coframe θ^i, we have

$$\begin{aligned}
dp &= \theta \cdot E, \\
\theta &= \left(\theta^1, \ldots, \theta^n\right), \\
d\theta &= -\omega \wedge \theta, \\
0 &= d\omega + \omega \wedge \omega.
\end{aligned}$$

Here, ω must be computed from the first structural equation, but since we know from above that the curvature is zero, we get for free the special form of the second structural equation.

These last structure equations can be used to derive the various fundamental equations for a submanifold or hypersurface in Euclidean space. Namely, for a k-dimensional submanifold $M \subset \mathbb{R}^n$, suppose that the frame E_i is chosen such that on M the first k vector fields are tangent to M, while the others are perpendicular to M. One can then both compute connection and curvature on M and also see how they relate to Euclidean space through the structural equations there. For more details see also [76, vol. V].

Example 1.4 Let $M = S^3$ and take as usual the left invariant frame $\{X_1, X_2, X_3\}$ with $[X_i, X_{i+1}] = 2X_{i+2}$, where indices are mod3. For the dual frame $\theta = (\theta^1, \theta^2, \theta^3)$ we have

$$\begin{aligned} d\theta^1 &= -2\theta^2 \wedge \theta^3, \\ d\theta^2 &= -2\theta^3 \wedge \theta^1, \\ d\theta^3 &= -2\theta^1 \wedge \theta^2. \end{aligned}$$

To find the connection forms, we must in addition to simply computing $d\theta$ also make sure that the skew-symmetry condition holds. Therefore, if

$$\omega = \begin{pmatrix} 0 & \omega_2^1 & \omega_3^1 \\ -\omega_2^1 & 0 & \omega_3^2 \\ -\omega_3^1 & -\omega_3^2 & 0 \end{pmatrix},$$

then we have to solve

$$\begin{aligned} 2\theta^2 \wedge \theta^3 &= \omega_2^1 \wedge \theta^2 + \omega_3^1 \wedge \theta^3, \\ 2\theta^3 \wedge \theta^1 &= -\omega_2^1 \wedge \theta^1 + \omega_3^2 \wedge \theta^3, \\ 2\theta^1 \wedge \theta^2 &= -\omega_3^1 \wedge \theta^1 - \omega_3^2 \wedge \theta^2. \end{aligned}$$

Thus, we have

$$\omega = \begin{pmatrix} 0 & -\theta^3 & \theta^2 \\ \theta^3 & 0 & -\theta^1 \\ -\theta^2 & \theta^1 & 0 \end{pmatrix}.$$

To find the curvature, we first compute

$$\begin{aligned} d\omega &= \begin{pmatrix} 0 & -d\theta^3 & d\theta^2 \\ d\theta^3 & 0 & -d\theta^1 \\ -d\theta^2 & d\theta^1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 2\theta^1 \wedge \theta^2 & -2\theta^3 \wedge \theta^1 \\ -2\theta^1 \wedge \theta^2 & 0 & 2\theta^2 \wedge \theta^3 \\ 2\theta^3 \wedge \theta^1 & -2\theta^2 \wedge \theta^3 & 0 \end{pmatrix}, \end{aligned}$$

$$\begin{aligned} \omega \wedge \omega &= \begin{pmatrix} 0 & -\theta^3 & \theta^2 \\ \theta^3 & 0 & -\theta^1 \\ -\theta^2 & \theta^1 & 0 \end{pmatrix} \wedge \begin{pmatrix} 0 & -\theta^3 & \theta^2 \\ \theta^3 & 0 & -\theta^1 \\ -\theta^2 & \theta^1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & \theta^2 \wedge \theta^1 & \theta^3 \wedge \theta^1 \\ -\theta^2 \wedge \theta^1 & 0 & \theta^3 \wedge \theta^2 \\ -\theta^3 \wedge \theta^1 & -\theta^3 \wedge \theta^2 & 0 \end{pmatrix}. \end{aligned}$$

Hence,

$$\Omega = \begin{pmatrix} 0 & \theta^1 \wedge \theta^2 & -\theta^3 \wedge \theta^1 \\ -\theta^1 \wedge \theta^2 & 0 & \theta^2 \wedge \theta^3 \\ \theta^3 \wedge \theta^1 & -\theta^2 \wedge \theta^3 & 0 \end{pmatrix}.$$

Since we know that

$$g\left(\mathfrak{R}\left(E_i \wedge E_j\right), \left(E_k \wedge E_l\right)\right) = \Omega_l^k\left(E_i, E_j\right),$$

it follows immediately that the curvature operator is the identity map.

One can also compute the curvatures for the Berger spheres in a similar fashion.

From these two simple examples, it is certainly clear that in many contexts the formalism we have developed for computing the connection and curvature tensor is quite convenient.

The last equations we have for the curvature tensor, namely, the Bianchi identities, can also be expressed via this formalism.

Proposition 1.5 *The curvature forms satisfy*

$$\begin{aligned} 0 &= \theta \wedge \Omega, \\ d\Omega &= \Omega \wedge \omega - \omega \wedge \Omega, \end{aligned}$$

or more precisely,

$$\begin{aligned} \theta^j \wedge \Omega_j^i &= 0, \\ d\Omega_j^i &= \Omega_k^i \wedge \omega_j^k - \omega_k^i \wedge \Omega_j^k. \end{aligned}$$

Moreover, these two equations are exactly the first and second Bianchi identities.

Proof. Note that the usual anticommutation laws don't hold for these wedge matrix products. To see the first identity, simply apply d to the first structural equation:

$$\begin{aligned} 0 &= dd\theta^j \\ &= d\left(\theta^j \wedge \omega_j^i\right) \\ &= d\theta^j \wedge \omega_j^i - \theta^j \wedge d\omega_j^i \\ &= \theta^k \wedge \omega_k^j \wedge \omega_j^i - \theta^j \wedge \left(\Omega_j^i - \omega_k^i \wedge \omega_j^k\right) \\ &= -\theta^j \wedge \Omega_j^i + \theta^j \wedge \omega_k^i \wedge \omega_j^k + \theta^k \wedge \omega_k^j \wedge \omega_j^i \\ &= -\theta^j \wedge \Omega_j^i + \theta^j \wedge \omega_k^i \wedge \omega_j^k - \theta^k \wedge \omega_j^i \wedge \omega_k^j \\ &= -\theta^j \wedge \Omega_j^i. \end{aligned}$$

For the second identity, we use the second structural equation:

$$d\Omega_j^i = d\left(d\omega_j^i + \omega_k^i \wedge \omega_j^k\right)$$

$$
\begin{aligned}
&= d\omega_k^i \wedge \omega_j^k - \omega_k^i \wedge d\omega_j^k \\
&= \left(\Omega_k^i - \omega_l^i \wedge \omega_k^l\right) \wedge \omega_j^k - \omega_k^i \wedge \left(\Omega_j^k - \omega_l^k \wedge \omega_j^l\right) \\
&= \Omega_k^i \wedge \omega_j^k - \omega_k^i \wedge \Omega_j^k - \omega_l^i \wedge \omega_k^l \wedge \omega_j^k + \omega_k^i \wedge \omega_l^k \wedge \omega_j^l \\
&= \Omega_k^i \wedge \omega_j^k - \omega_k^i \wedge \Omega_j^k.
\end{aligned}
$$

We leave it as an exercise to check that these two equations are indeed equivalent to the first and second Bianchi identities. □

We shall also at times have use for the transformation formulae for the connection and curvature forms between different frames. A little notational comment is in order before we proceed. Given a map $F : M \to G$ from a manifold to a Lie group, we have two different differentials: $DF : TM \to TG$ and $dF : TM \to \mathfrak{g}$. The latter is related to the former by the trivialization $TG = G \times \mathfrak{g}$. Unless we are in the case where $G = \mathbb{R}^n$, this trivialization is, however, not uniquely defined, as one can use either right- or left-invariant vector fields to trivialize the tangent bundle on G. One usually decides from the context which of the two trivializations should be used.

Assume that we have two frames E and $\tilde{E}$ with corresponding coframes θ and $\tilde{\theta}$. Assuming that both frames are orthonormal, we can find a function g, defined on the intersection of the domains of the frames with values in $O(n)$, such that

$$
\begin{aligned}
\tilde{E} &= E \cdot g, \\
\left(\tilde{E}_1, \ldots, \tilde{E}_n\right) &= (E_1, \ldots, E_n) \cdot \begin{pmatrix} g_1^1 & \cdots & g_n^1 \\ \vdots & \ddots & \vdots \\ g_1^n & \cdots & g_n^n \end{pmatrix}.
\end{aligned}
$$

Conversely, the coframes are related through

$$\tilde{\theta} = g^t \cdot \theta.$$

Proposition 1.6 *The connection and curvature forms satisfy the transformation rules*

$$
\begin{aligned}
\tilde{\omega} &= g^t \cdot \omega \cdot g - dg^t \cdot g, \\
\tilde{\Omega} &= g^t \cdot \Omega \cdot g.
\end{aligned}
$$

Proof. First, we compute

$$
\begin{aligned}
d\tilde{\theta} &= dg^t \cdot \theta + g^t \cdot d\theta \\
&= dg^t \cdot \theta - g^t \cdot \omega \wedge \theta \\
&= \left(dg^t - g^t \cdot \omega\right) \wedge \theta \\
&= \left(dg^t - g^t \cdot \omega\right) \wedge g \cdot \tilde{\theta} \\
&= \left(dg^t \cdot g - g^t \cdot \omega \cdot g\right) \wedge \tilde{\theta},
\end{aligned}
$$

which shows that the formula for the connection forms is correct.

For the curvature forms, we need to use that $g \cdot g^t = I$ and $0 = d\left(g \cdot g^t\right) = g \cdot dg^t + dg \cdot g^t$. Also, recall how one takes d of products, using the skew Leibnitz rule for forms. With this in mind we can compute

$$\begin{aligned}
\tilde{\Omega} &= d\tilde{\omega} + \tilde{\omega} \wedge \tilde{\omega} \\
&= d\left(g^t \cdot \omega \cdot g - dg^t \cdot g\right) + \left(g^t \cdot \omega \cdot g - dg^t \cdot g\right) \wedge \left(g^t \cdot \omega \cdot g - dg^t \cdot g\right) \\
&= dg^t \wedge \omega \cdot g + g^t \cdot d\omega \cdot g - g^t \cdot \omega \wedge dg - ddg^t \cdot g + dg^t \wedge dg \\
&\quad + g^t \cdot \omega \cdot g \wedge g^t \cdot \omega \cdot g - g^t \cdot \omega \cdot g \wedge dg^t \cdot g \\
&\quad - dg^t \cdot g \wedge g^t \cdot \omega \cdot g + dg^t \cdot g \wedge dg^t \cdot g \\
&= g^t \cdot (d\omega + \omega \wedge \omega) \cdot g \\
&\quad + dg^t \wedge \omega \cdot g - g^t \cdot \omega \wedge dg + dg^t \wedge dg - g^t \cdot \omega \wedge g \cdot dg^t \cdot g \\
&\quad - dg^t \wedge g \cdot g^t \cdot \omega \cdot g + dg^t \cdot g \wedge dg^t \cdot g \\
&= g^t \cdot \Omega \cdot g + dg^t \wedge \omega \cdot g - g^t \cdot \omega \wedge dg - dg^t \wedge dg \\
&\quad + g^t \cdot \omega \wedge dg - dg^t \wedge \omega \cdot g - dg^t \wedge dg \\
&= g^t \cdot \Omega \cdot g.
\end{aligned}$$

□

B.2 The Frame Bundle

The idea of the frame bundle is to find a more invariant approach to the above described Cartan formalism. This is done in the usual fashion adopted by mathematicians. Namely, one makes the problem into the solution. Let us first give some general instances of this:

(1) The problem is to find roots of real polynomials. One realizes quickly that $X^2 + 1 = 0$ cannot be solved over the real numbers. The solution is to consider the ring $\mathbb{R}[X]$ and divide out by ideal $\left(X^2 + 1\right)$. This gives us the complex numbers $\mathbb{C}$. Thus, the unsolvable equation $X^2 + 1 = 0$ can be solved over the field $\mathbb{R}[X] / \left(X^2 + 1\right)$.

(2) The problem is to define the rational numbers from the integers. Again, one cannot solve integer equations of the form: $ax = b$. Two such equations $ax = b$ and $cx = d$ are equivalent if $ad = bc$. Thus, let $\mathbb{Q}$ denote the set of such equations modulo this equivalence. Over this new field the above equations are again the solutions to themselves.

(3) The problem is to compute limits of sequences of rational numbers. Define Cauchy sequences and then construct the real numbers as the set of all Cauchy sequences of rational numbers modulo sequences whose difference converges to zero. Then one gets a complete space.

So guess how one makes a frame-dependent approach to geometry frame independent. Yes, you take the space of all frames and then start translating Cartan formalism to this setting.

B.3 Construction of the Frame Bundle

As already discussed in Chapter 7, we define the frame bundle FM as the set of ordered orthonormal bases of the tangent space:

$$FM = \coprod_{p \in M} F_p M,$$
$$F_p M = \left\{ e : e = (e_1, \ldots, e_n) \text{ is an orthonormal basis for } T_p M \right\}.$$

With the frame bundle we have the projection

$$\pi : FM \to M$$

that takes the fibers $F_p M$ to p.

The orthogonal group $O(n)$ acts from the right on each of the fibers and therefore also on all of FM in the following simple way:

$$\begin{aligned} e \cdot g &= (e_1, \ldots, e_n) \cdot \begin{pmatrix} g_1^1 & \cdots & g_n^1 \\ \vdots & \ddots & \vdots \\ g_1^n & \cdots & g_n^n \end{pmatrix} \\ &= \left(e_i g_1^i, \ldots, e_i g_n^i \right). \end{aligned}$$

Via this action we see that each fiber can by identified with $O(n)$, as the correspondence $g \to e \cdot g$ is bijective for each fixed choice of basis. We can use this to give FM the structure of a locally trivial bundle. Namely, if E is an orthonormal frame over the open set $U \subset M$, then we have a bijection

$$\begin{aligned} U \times O(n) &\to FU, \\ (p, g) &\to E(p) \cdot g \end{aligned}$$

that clearly commutes with the actions of $O(n)$ on both $U \times O(n)$ and FU.

This local trivialization can be used in reverse to define the frame bundle in the following way. On M pick a covering U_α such that we have orthonormal frames E_α on each U_α. On the intersections $U_\alpha \cap U_\beta$ we can now find maps

$$\begin{aligned} \tau_{\alpha\beta} &: U_\alpha \cap U_\beta \to O(n), \\ E_\beta &= E_\alpha \cdot \tau_{\alpha\beta}. \end{aligned}$$

These maps satisfy

$$\tau_{\alpha\beta} \cdot \tau_{\beta\gamma} = \tau_{\alpha\gamma}.$$

Thus, we have an equivalence relation by identifying

$$\begin{aligned} (p, g_\alpha) &\in U_\alpha \times O(n), \\ \left(p, g_\beta\right) &\in U_\beta \times O(n) \end{aligned}$$

if

$$g_\alpha = \tau_{\alpha\beta} \cdot g_\beta.$$

This identification clearly respects the natural right actions of $O(n)$ on the trivial bundles. Thus, we get a locally trivial bundle with a natural right action of $O(n)$. It is clear that this reconstructs the frame bundle.

B.4 Construction of Tensor Bundles

Having now constructed the frame bundle from the Riemannian metric and the tangent bundle, we wish to reverse this process and construct the tangent bundle, and also all of the other tensor bundles, from the frame bundle.

First, let us attack the tangent bundle. The orthogonal group acts by isometries from the left on $\mathbb{R}^n$. Thus, we have an action

$$\begin{aligned}\left(FM \times \mathbb{R}^n\right) \times O(n) &\to \left(FM \times \mathbb{R}^n\right),\\ (e, v) \cdot g &= \left(e \cdot g, g^{-1} \cdot v\right).\end{aligned}$$

The orbit space of this action is denoted by $FM \times_{O(n)} \mathbb{R}^n$. This indicates that we start with a product and then divide out by an action of $O(n)$ on each of the spaces. Observe that the projection $\pi : FM \to M$ clearly induces a projection $\pi : FM \times_{O(n)} \mathbb{R}^n \to M$, and that the preimage of a point under this projection can be identified with $\mathbb{R}^n$. Thus, $FM \times_{O(n)} \mathbb{R}^n$ looks like a vector bundle with a Euclidean metric on each of its fibers. We contend that this is the tangent bundle with the given Riemannian metric. This is seen by trivializing both bundles. First, select an open set $U \subset M$ on which we have an orthonormal frame E. Then

$$\begin{aligned}U \times \mathbb{R}^n &\to FU \times_{O(n)} \mathbb{R}^n,\\ (p, v) &\to \left\{\left(E(p) \cdot g, g^{-1} \cdot v\right) : g \in O(n)\right\}\end{aligned}$$

defines a bijection, which generates the commutative diagram

$$\begin{array}{ccc} U \times \mathbb{R}^n & \to & FU \times_{O(n)} \mathbb{R}^n \\ & \searrow \quad \swarrow & \\ & U & \end{array}$$

The top horizontal arrow in this diagram is also a fiberwise isometry. This actually gives us the desired vector bundle structure on $FM \times_{O(n)} \mathbb{R}^n$, together with the inner product on the fibers. The key is now to observe that we have a similar diagram for the tangent bundle:

$$\begin{array}{ccc} U \times \mathbb{R}^n & \to & TU \\ & \searrow \quad \swarrow & \\ & U & \end{array}$$

where

$$\begin{aligned} U \times \mathbb{R}^n &\to TU, \\ (p, v) &\to E(p) \cdot v \\ &= E_1 v^1 + \cdots + E_n v^n. \end{aligned}$$

This is again a natural trivialization of the tangent bundle that also preserves the inner products on the fibers. Thus, we have a locally defined fiber-preserving and inner-product-preserving isomorphism:

$$TU \leftarrow U \times \mathbb{R}^n \to FU \times_{O(n)} \mathbb{R}^n.$$

It is now a simple matter to see that this defines an inner-product-preserving isomorphism:

$$\begin{array}{ccc} FM \times_{O(n)} \mathbb{R}^n & \leftrightarrow & TM \\ & \searrow \quad \swarrow & \\ & M & \end{array}$$

We now turn our attention to tensor bundles. Abstractly, they all come by having a Euclidean space V and a homomorphism $\rho : O(n) \to O(V)$, where $O(V)$ denotes the linear maps that preserve the Euclidean structure on V. In this way we have that $O(n)$ acts by isometries from the left on V. We can then, as before, construct a vector bundle

$$FM \times_{O(n)} V \to M$$

with inner products on the fibers. The local trivialization is defined by

$$\begin{aligned} U \times V &\to FU \times_{O(n)} V, \\ (p, v) &\to \left\{ \left(E(p) \cdot g, g^{-1} \cdot v \right) : g \in O(n) \right\}. \end{aligned}$$

It is perhaps not entirely obvious why this recreates all tensor bundles. To see how this works, let us construct $\Lambda^2 TM$. We start with the vector space of Euclidean bivectors $\Lambda^2\mathbb{R}^n$. Given an orthonormal basis $(e_1, \ldots, e_n)$ for $\mathbb{R}^n$, we get an orthonormal basis $e_i \wedge e_j$, $i < j$, for $\Lambda^2\mathbb{R}^n$. This makes $\Lambda^2\mathbb{R}^n$ into a vector space with an inner product, as we already know. But how does $O(n)$ act on $\Lambda^2\mathbb{R}^n$? In fact, any linear map $L : \mathbb{R}^n \to \mathbb{R}^n$ induces a linear map $\Lambda^2 L : \Lambda^2\mathbb{R}^n \to \Lambda^2\mathbb{R}^n$ by the rule

$$\Lambda^2 L (v \wedge w) = Lv \wedge Lw.$$

In particular, we see that if L is an isometry, then the orthonormal basis $(e_1, \ldots, e_n)$ is mapped to an orthonormal basis $(Le_1, \ldots, Le_n)$, and consequently, the orthonormal basis $e_i \wedge e_j$, $i < j$, is mapped to an orthonormal basis $Le_i \wedge Le_j$, $i < j$. Hence, $\Lambda^2 L$ is really an isometry. Thus, we have the desired homomorphism:

$$\Lambda^2 : O(n) \to O\left(\Lambda^2\mathbb{R}^n\right).$$

One can now in the usual manner see that $\Lambda^2 TM = FM \times_{O(n)} \Lambda^2\mathbb{R}^n$.

B.5 Tensors

Having found a new description of tensor bundles, including the tangent bundle itself, it is natural to see how we can interpret sections of these bundles. Sections of tensor bundles are, of course, called tensors.

We fix a tensor bundle $FM \times_{O(n)} V$ coming from a homomorphism $\rho : O(n) \to O(V)$. A section of a tensor bundle $\pi : FM \times_{O(n)} V \to M$ is simply a map $T : M \to FM \times_{O(n)} V$ such that $\pi \circ T = id$. Let us use the notation $[\cdot, \cdot] : FM \times V \to FM \times_{O(n)} V$ for the quotient map and $\pi : FM \to M$ for the projection. Then we claim that for any $e \in FM$, there must be a unique $v \in V$ such that $[e, v] = T \circ \pi(e)$. If $[e, v] = [e, w]$, then we know that there must exist an orthogonal transformation $g \in O(n)$ such that $e \cdot g = e$ and $\rho\left(g^{-1}\right) v = w$, but this implies that $g = I$ and hence $v = w$. Thus, we have constructed a map $T : FM \to V$. This map also commutes with the actions of $O(n)$, as $w = \rho\left(g^{-1}\right) v$ must be the unique vector corresponding to $e \cdot g$. Conversely we see that any map $T : FM \to V$ which commutes with the $O(n)$ actions yields a tensor $T : M \to FM \times_{O(n)} V$ by sending $p \in M$ to $[e, T(e)] \in FM \times_{O(n)} V$, where $\pi(e) = p$. Note that this is well-defined exactly because $T : FM \to V$ is invariant under the $O(n)$ actions. Thus, we have shown

Proposition 5.1 *Sections of $\pi : FM \times_{O(n)} V \to M$ are in one-to-one correspondence with $O(n)$-equivariant maps $FM \to V$.*

The differential of a tensor $T : FM \to V$ is a map $dT : TFM \to V$. Note that $\mathfrak{so}(n)$ acts on both of these spaces by the differential of the $O(n)$ actions, so dT will be an $\mathfrak{so}(n)$-equivariant map.

This new way of viewing tensors enables us to give a new description of the components of a tensor in a given orthonormal frame E. Since such a frame is simply a section $E : U \subset M \to FM$, we see that any tensor $T : M \to FM \times_{O(n)} V$, when viewed as a map $T : FM \to V$, can be composed with E to yield a map

$$T \circ E : U \to V.$$

Let us see how all of this works for vector fields. Fix a vector field $X : M \to TM$ and a frame $E : U \to FM$. First, we have the trivialization

$$\begin{aligned} U \times \mathbb{R}^n &\to TU, \\ (p, v) &\to E(p) \cdot v. \end{aligned}$$

The vector field X yields a section $X : U \to U \times \mathbb{R}^n$ by sending p to $(p, x(p))$, where $E(p) \cdot x(p) = X(p)$. This gives us X as a section of the alternative definition $FM \times_{O(n)} \mathbb{R}^n$ for TM. To get the map $FM \to \mathbb{R}^n$ we simply map $E(p) \cdot g$ to $g \cdot x(p)$. In this way, we see that $X \circ E(p) = x(p)$, thus giving us the components of the vector field in the locally given frame E. The differential of X under the identifications $TFU = TU \times \mathfrak{so}(n)$ is given by

$$dX(E \cdot v, s) = s \cdot dx(E \cdot v),$$

where $dx : TU \to \mathbb{R}^n$.

B.6 The Connection on the Frame Bundle

We have seen that the frame bundle with its $O(n)$ action reconstructs the tangent bundle and the Riemannian metric. Thus, all information about the Riemannian manifold (M, g) should somehow be encoded into the frame bundle. We shall now see how the connection and curvature live on this bundle. The Cartan formalism we developed above will be a help, as it was intimately connected with orthonormal frames and at the same time contained the information we are seeking.

First, let try to decide what a connection on FM should be. From Cartan formalism we have the equation

$$\nabla E = E \cdot \omega$$

for a given frame field. The meaning of this is that an infinitesimal rotation of a frame is given by a skew-symmetric matrix ω. This is quite natural, for any two frames are related by a function of orthogonal transformations. An infinitesimal change therefore corresponds to a function into the Lie algebra over the orthogonal transformations. But this Lie algebra $\mathfrak{o}(n) = \mathfrak{so}(n)$ is simply the set of $n \times n$ skew-symmetric matrices. Guided by this, we imagine that the connection on FM is an $\mathfrak{so}(n)$ valued 1-form

$$\omega : T(FM) \to \mathfrak{so}(n),$$

with the condition that it becomes the identity map on the fibers of FM. The meaning of this is simply, any element $X \in \mathfrak{so}(n)$ generates a vector field $\sigma(X)$ on FM that is tangent to the fibers. It is defined by assuming $X = \dot{c}(0)$ for some curve $c : I \to O(n)$ with $c(0) = I$, and then declaring its value at $e \in FM$ to be $\sigma(X)(e) = d/dt_{|t=0}(e \cdot c(t))$. Thus the connection form must satisfy

$$\omega(\sigma(X)) = X.$$

This means that ω accurately measures how frames with the same base point are moved. The question is then how it acts on tangent vectors that are not tangent to the fibers. This is where we need our connection forms from above. Namely, we also want to make sure that the connection ω on FM gives us back the connection forms. There is a natural way for this to happen. For an orthonormal frame field E let the corresponding connection forms be denoted by ω_E. Note that $E : U \to FM$ is a section. So if we pull back a form on FM by this section, we get a form on M. The desired condition we require should then be

$$\omega_E = E^* \omega.$$

Note that pulling back via a choice of frame has the effect of ignoring the fiber directions, thus, the need for separately explaining the action of ω on those vectors.

On the other hand, this pullback action is an isomorphism on a complement of the fiber, so the two conditions together uniquely define ω. The only problem is that we haven't shown that it gives a well-defined mapping. For that, we have to use how ω_E transforms under change of frame. We established that if $\tilde{E} = E \cdot g$ on U for some function $g : U \to O(n)$, then

$$\begin{aligned}\omega_{\tilde{E}} &= g^t \cdot \omega_E \cdot g - dg^t \cdot g \\ &= g^t \cdot \omega_E \cdot g + g^t \cdot dg.\end{aligned}$$

For the pullback of a form, we have on the other hand

$$\begin{aligned}\tilde{E}^*\omega(X) &= \omega\left(D\tilde{E}(X)\right) \\ &= \omega(D(E \cdot g)(X)).\end{aligned}$$

We view $E \cdot g$ as a composition of the two functions

$$\begin{aligned}E &: U \to FM, \\ R_g &: FM \to FM.\end{aligned}$$

As g is also a function, we need to be careful when computing the derivative of

$$(p, g) \to R_g \circ E.$$

The chain rule yields for $v \in T_pM$,

$$\begin{aligned}D(E \cdot g)(v) &= D\left(R_{g(p)}\right)(DE(v)) + \sigma\left(g^t \cdot dg(v)\right)(E(p) \cdot g(p)) \\ &= g^t \cdot DE(v) \cdot g + \sigma\left(g^t \cdot dg(v)\right)(E \cdot g),\end{aligned}$$

implying that

$$\tilde{E}^*\omega(X) = g^t \cdot E^*\omega(X) \cdot g + g^t \cdot dg(DE(X)).$$

Hence, we get a unique and well-defined connection form ω on FM. Note that the last term in this formula actually comes from the fact that $\omega \circ \sigma = id$, so one really only needs to check the transformation formula in case $g : U \to SO(n)$ is constant. This leads us to a very nice and abstract characterization of the connection on the frame bundle.

Proposition 6.1 *There is one and only one 1-form $\omega : TFM \to \mathfrak{so}(n)$ such that*

$$\begin{aligned}\omega \circ \sigma &= id, \\ \omega \circ DR_g &= g^t \cdot \omega \cdot g \\ &= Ad\left(g^t\right) \circ \omega,\end{aligned}$$

where $Ad(g) : \mathfrak{g} \to \mathfrak{g}$ is the differential of the inner automorphism $x \to g \cdot x \cdot g^{-1}$.

The connection on the frame bundle is also often called the *Ehresmann connection*. The 1-form, as we have seen, is never zero on nonzero tangent vectors to fibers. Since it maps $T(FM)$ to $\mathfrak{so}(n)$, it must therefore have n-dimensional kernel at each $e \in FM$. The distribution generated by this construction is called the *horizontal distribution*. It is denoted $\mathcal{H}$ and the fibers $\mathcal{H}_e = \ker(\omega : T_eFM \to \mathfrak{so}(n))$. Evidently this distribution completely determines the connection form ω. As $\mathcal{H}_e \cap T_e(F_pM) = \{0\}$, where $p = \pi(e)$, we must have that $D\pi : \mathcal{H}_e \to T_pM$ is an isomorphism. The inverse of this isomorphism $T_pM \to \mathcal{H}_e$ is called the *horizontal lift* of $v \in T_pM$ to $v^h \in \mathcal{H}_e \subset T_eFM$.

B.7 Covariant Differentiation of Tensors

Let us now see how the connection on the frame bundle can give us back covariant differentiation of tensors. Let a tensor bundle $FM \times_{O(n)} V \to M$ coming from a representation $\rho : O(n) \to O(V)$ be given. The differential of ρ is represented by a linear map $d\rho : \mathfrak{so}(n) \to \mathfrak{so}(V)$. If we think of a tensor as a map $T : FM \to V$, then we have the differential $dT : TFM \to V$, which commutes with the actions of $\mathfrak{so}(n)$ on TFM and of $\mathfrak{so}(n)$ on V via $d\rho$. The covariant derivative ∇T is supposed to be a new tensor with one extra vector variable. This means that it should be a section of $FM \times_{O(n)} End(\mathbb{R}^n, V)$, or in other words, a map $FM \to End(\mathbb{R}^n, V)$. The definition is

$$\begin{aligned} \nabla T &: FM \to End(\mathbb{R}^n, V), \\ e &\to (x \to dT((e \cdot x)^h)). \end{aligned}$$

In other words, we think of ∇T as the restriction of dT to the horizontal distribution. Since $d\pi : TFM \to TM$ is an isomorphism when restricted to the horizontal distribution, we can reinterpret this as follows. Think of $T : FM \to V$ and $\nabla T : TM \to FM \times_{O(n)} V$. Then

$$\nabla_v T = [e, dT(v^h)], \; v^h \in \mathcal{H}_e \subset T_eFM.$$

Therefore, if we take a frame $E : U \to FM$, then we arrive at

$$\begin{aligned} \nabla_v T &= [E, d(T \circ E)(v) - d\rho(\omega_E(v))(T \circ E)] \\ &= [E, d(T \circ E)(v) - \omega_E(v) \cdot (T \circ E)] \\ &= [E, dT(DE(v)) - \omega(DE(v)) \cdot (T \circ E)]. \end{aligned}$$

It is perhaps not immediately clear why this is the same as the invariant definition. Using that $D(\pi \circ E)(v) = v$, we can write $DE(v) = v^h + \sigma(s) \in \mathcal{H} \oplus \mathfrak{so}(n) = TFM$. With this we obtain

$$\begin{aligned} dT(DE(v)) - \omega(DE(v)) \cdot (T \circ E) &= dT(v^h) - \omega(v^h) \cdot (T \circ E) \\ &\quad + dT(\sigma(s)) - \omega(\sigma(s)) \cdot (T \circ E) \\ &= dT(v^h) + dT(\sigma(s)) - s \cdot (T \circ E) \\ &= dT(v^h). \end{aligned}$$

Here, the term $dT(\sigma(s)) - s \cdot (T \circ E) = 0$, since we assumed that T is invariant with respect to the $O(n)$ actions.

In the special case of $X = E \cdot x$, $x : U \to \mathbb{R}^n$, being a vector field, we already know from Cartan formalism that

$$\begin{aligned}\nabla X &= \nabla(E \cdot x)\\ &= (\nabla E) \cdot x + E \cdot dx\\ &= -E \cdot \omega_E \cdot x + E \cdot dx\\ &= E \cdot (-\omega_E \cdot x + dx).\end{aligned}$$

Since this agrees with the local formula we derived from the invariant definition, we see that the above definition of covariant differentiation gives us back the old definition.

We now have to see how the curvature enters the picture. The *curvature form* is defined as:

$$\begin{aligned}\Omega &: \Lambda^2 TM \to \mathfrak{so}(n),\\ \Omega(X, Y) &= \Omega(X \wedge Y) = -\omega\left(\left[X^h, Y^h\right]\right).\end{aligned}$$

Thus, the curvature vanishes iff $\left[X^h, Y^h\right]$ is horizontal. This implies that the curvature is zero iff the horizontal distribution is integrable. Thus, the curvature somehow measures how close the horizontal distribution is to being integrable.

It still remains to be seen what this has to do with our curvature forms from above. For that we must show

$$\Omega(X, Y) = (d\omega + \omega \wedge \omega)\left(X^h, Y^h\right).$$

But ω vanishes on horizontal vectors, and thus the only term left on the right-hand side is

$$\begin{aligned}d\omega\left(X^h, Y^h\right) &= -\omega\left(\left[X^h, Y^h\right]\right)\\ &= \Omega(X, Y).\end{aligned}$$

We now wish to connect this version of the curvature tensor with our alternative definition $R(X, Y)T = \nabla^2_{X,Y}T - \nabla^2_{Y,X}T$ for the curvature of a tensor $T : FM \to V$. It is natural to guess that the relationship should be

$$\begin{aligned}R(X, Y)T &= d\rho(\Omega(X \wedge Y))(T)\\ &= \Omega(X \wedge Y) \cdot T.\end{aligned}$$

It follows from our definition of Ω and ∇ that this is true for vector fields. Thus, it must also be true for all tensors.

B.8 Principal Bundles in General

In the previous section we saw how the principal bundle, together with its connection, contains all information about the underlying Riemannian metric, the associated tensor bundles and covariant derivatives. At the same time, it also gave us an invariant approach to the Cartan formalism. While this approach might at first seem unnecessarily abstract and useless, it really gives us a very general method for dealing with many different subjects. In this section we shall describe how this theory can be further axiomatized and generalized.

A *principal G-bundle* over a manifold M consists of a submersion $f : P \to M$ such that each fiber is isomorphic to G. We shall further assume that G is a compact Lie group that acts from the right on P, with the property that orbits are exactly the fibers of the bundle. We shall also assume that there are local trivializations

$$\begin{array}{ccc} f^{-1}(U) & \to & U \times G \\ & \searrow \quad \swarrow & \\ & U & \end{array}$$

that commute with the right actions of G on both $f^{-1}(U)$ and $U \times G$. Clearly, we can construct for each $X \in \mathfrak{g}$ a vector field $\sigma(X)$ tangent to the fibers of the fibration.

A connection on a principal G-bundle consists of a $\mathfrak{g}$-valued 1-form $\omega : TP \to \mathfrak{g}$ with the properties that

$$\begin{aligned} \omega \circ \sigma &= id, \\ \omega \circ DR_g &= \mathrm{Ad}\left(g^{-1}\right) \circ \omega. \end{aligned}$$

Again, we have a horizontal distribution $\mathcal{H}$ with fibers $\mathcal{H}_e = \ker(\omega : T_eP \to \mathfrak{g})$. This leads, as before, to horizontal lifts from TM. The curvature form is also defined as above:

$$\begin{aligned} &\Lambda^2 TM \to \mathfrak{g}, \\ &\Omega(X \wedge Y) = -\omega\left(\left[X^h, Y^h\right]\right). \end{aligned}$$

Given a vector space E and a representation of G on E, i.e., a homomorphism $\rho : G \to End(E)$, we can construct a vector bundle $P \times_G E \to M$ with fibers isomorphic to E. Given a connection ω on P, we have a natural connection on sections of this fiber bundle defined by the formula

$$\nabla_v s = ds\left(v^h\right).$$

Moreover, we have the important curvature identity

$$\begin{aligned} R(X, Y)s &= d\rho(\Omega(X \wedge Y))(s) \\ &= \Omega(X \wedge Y) \cdot s \end{aligned}$$

for all sections $s : FM \to E$ of this bundle.

Suppose now that the vector space E comes with an inner product structure $\langle \cdot, \cdot \rangle$. We then get an inner product on each of the fibers in the vector bundle $P \times_G E \to M$. However, we don't necessarily have that Leibniz's rule holds:

$$D_v \langle s_1, s_2 \rangle = \langle \nabla_v s_1, s_2 \rangle + \langle s_1, \nabla_v s_2 \rangle,$$

where $s_1, s_2 : M \to P \times_G E$. For this to be true, we must assume that the inner product is invariant under the action of G on V. This is not hard to check. Note also that for all the tensor bundles it was assumed that the inner products were invariant, as we always insisted on having a representation of the form $O(n) \to O(V)$.

B.9 Further Study

For more on principal bundles the reader can consult [76, vol. II] and the comprehensive account in [53, vol. I]. The reader should be aware that we have developed everything in the framework of Riemannian manifolds. One can on any manifold construct the principal Gl_n bundle of frames (not just orthonormal frames) and with this develop tensor bundles.

Cartan formalism is very useful in submanifold theory, where one can use adapted frames for a submanifold together with the structural equations to develop all of the fundamental equations of submanifold geometry. A good place to learn about all this is in [76, vol. IV].

Appendix C
Spinors

Spinors have for decades been useful in physics, and many proofs in geometry can be considerably simplified using spin geometry. Without writing a new book on this, we cannot explain these ideas in any detail. The point here is to give as quick and elementary an account of spin geometry as possible. Hopefully, this will make it much easier to get into this subject, without getting bogged down in the rather massive preliminary work that authors usually go through before mentioning spinors. We shall give a very quick, but detailed, introduction to the idea of spin manifolds and why they have special spinor bundles. With that behind us, it will become clear that these bundles naturally come with connections induced from a Riemannian connection on the underlying manifold. After this preliminary material, we define the Dirac operator on spinors in the same way as we defined it for forms. The corresponding Weitzenböck formula is then established and used to prove various rigidity results for manifolds with nonnegative scalar curvature.

One of the magic uses of the Dirac operator on both forms and spinors is that its square is a natural Laplace operator. In both cases, Clifford multiplication gives us the algebra that is needed in order to take a sum of squares and make it the square of a sum:

$$-a_1^2 - \cdots - a_n^2 = \left(e^1 \cdot a_1 + \cdots + e^n \cdot a_n\right)^2,$$
$$-\partial_1^2 - \cdots - \partial_n^2 = \left(e^1 \cdot \partial_1 + \cdots + e^n \cdot \partial_n\right)^2,$$

where the e^i generate an algebra over $\mathbb{R}$ subject to the relations

$$e^i \cdot e^j + e^j \cdot e^i = 0 \quad \text{if} \quad i \neq j,$$
$$e^i \cdot e^i = -1.$$

Even though Clifford studied these algebras in the nineteenth century and Cartan shortly afterwards discovered the spin representations, it wasn't until Dirac and Pauli saw a use for these matters in the description of the electron that spinors became interesting to mathematicians and physicists alike. The physicists were, and still are, mostly interested in spinors when the underlying space is either 3- or 4-dimensional, and in the latter situation they of course use Minkowski space rather than Euclidean space as the infinitesimal model. This has the effect of making things a little more concrete. In dimension 3, for instance, one can use the following choices for e^1, e^2, e^3:

$$\text{Quarternion model:} \quad e^1 = i, \qquad e^2 = j, \qquad e^3 = k,$$

$$\mathfrak{su}(2) \text{ model:} \quad e^1 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad e^2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad e^3 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}.$$

It is also remarkable that even for many mathematical theorems that mathematicians have proven by "traditional" methods, physicists have found very simple proofs using spinors. This is particularly interesting, as it was in fact Atiyah and Singer who introduced spinor bundles as they appear in our account here. A great example of this is the positive mass conjecture, which was first established by Schoen and Yau using very delicate analytical techniques about minimal surfaces. Witten then found a very simple proof using spinors. We shall give an indication of how this works below. In order to understand this result, it is also necessary to get a clear picture of what the square of a spinor is. This is a very interesting concept that links forms and spinors in a very concrete way. Unfortunately, there doesn't seem to be any broad agreement on what exactly this square should be. We present several possible definitions and use some of these to prove a couple of important theorems.

C.1 Spin Structures

First we shall explain how orientability and spin structures are related to cohomology and the frame bundle. We assume that we have a Riemannian manifold (M, g), and suppose in addition that we have a good covering $\{U_\alpha\}$. On each U_α select an orthonormal frame $E_\alpha = \left(E_{\alpha,1}, \dots, E_{\alpha,n}\right)$. On nonempty intersections we then have the transition matrix functions $\tau_{\alpha\beta} : U_\alpha \cap U_\beta \to O(n)$ defined by $E_\beta = E_\alpha \cdot \tau_{\alpha\beta}$. The important property of these transition matrices is that

$$\tau_{\alpha\beta} \cdot \tau_{\beta\gamma} = \tau_{\alpha\gamma}.$$

The maps $\tau_{\alpha\beta} : U_\alpha \cap U_\beta \to O(n)$ can be reduced to maps

$$\det \tau_{\alpha\beta} : U_\alpha \cap U_\beta \to O(n) / SO(n) = \{1, -1\} = \mathbb{Z}_2,$$

by taking the the determinant of $\tau_{\alpha\beta}$. In the language of Čech cohomology, we have then constructed a map $f : I^2 \to \mathbb{Z}_2$. This map is alternating, since $\left(\tau_{\alpha\beta}\right)^{-1} =$

$(\tau_{\alpha\beta})^t = \tau_{\beta\alpha}$. We can compute the differential as

$$\begin{aligned} df(\alpha_0, \alpha_1, \alpha_2) &= f(\alpha_1, \alpha_2) \cdot (f(\alpha_0, \alpha_2))^{-1} \cdot f(\alpha_0, \alpha_1) \\ &= \det \tau_{\alpha_1\alpha_2} \left(\det \tau_{\alpha_0\alpha_2}\right)^{-1} \det \tau_{\alpha_0\alpha_1} \\ &= \det \left(\tau_{\alpha_1\alpha_2} \cdot \left(\tau_{\alpha_0\alpha_2}\right)^{-1} \cdot \tau_{\alpha_0\alpha_1}\right) \\ &= \det \left(\tau_{\alpha_1\alpha_2} \cdot \tau_{\alpha_2\alpha_0} \cdot \tau_{\alpha_0\alpha_1}\right) \\ &= \det I \\ &= 1. \end{aligned}$$

Thus, we get a cohomology class denoted by $w_1(M) \in H^1(M, \mathbb{Z}_2)$. It is very easy to see that for the fixed good covering given, this class is independent of the metric we used, since we reduce mod $SO(n)$. In fact, the mod $SO(n)$ reduction of τ is -1 or 1, depending on whether the coordinate transitions reverse or preserve the standard orientation on Euclidean space. The cohomology class $w_1(M)$ is called the *first Stiefel-Whitney class*. It obviously measures whether or not the manifold is orientable.

Now assume that the manifold is orientable, i.e., $w_1(M) = 0$. Then we can suppose that the coordinates are chosen such that $\tau_{\alpha\beta} \in SO(n)$. If $n > 2$, we know that $\pi_1(SO(n)) = \mathbb{Z}_2$, so there is a Lie group called $Spin(n)$ and a nontrivial 2-fold covering map $\pi : Spin(n) \to SO(n)$. We shall give a more concrete description of this group below. For now, we just need the abstract setup. Each of the maps $\tau_{\alpha\beta} : U_\alpha \cap U_\beta \to SO(n)$ has two lifts to $Spin(n)$, as the intersection is assumed to be simply connected. Let $\tilde{\tau}_{\alpha\beta} : U_\alpha \cap U_\beta \to Spin(n)$ denote one of these lifts. We can certainly assume that these choices at least satisfy $(\tilde{\tau}_{\alpha\beta})^{-1} = \tilde{\tau}_{\beta\alpha}$. But it might happen that the relation $\tilde{\tau}_{\alpha\beta} \cdot \tilde{\tau}_{\beta\gamma} = \tilde{\tau}_{\alpha\gamma}$ doesn't hold on the intersection $U_\alpha \cap U_\beta \cap U_\gamma$. However, as $\tilde{\tau}_{\alpha\beta}\tilde{\tau}_{\beta\gamma}$ and $\tilde{\tau}_{\alpha\gamma}$ differ by an element in $\mathbb{Z}_2 = \ker(Spin(n) \to SO(n))$, we can define a map by

$$\begin{aligned} f &: I^3 \to \mathbb{Z}_2 = \{1, -1\}, \\ f(\alpha, \beta, \gamma) &= \tilde{\tau}_{\alpha\beta}\tilde{\tau}_{\beta\gamma}\left(\tilde{\tau}_{\alpha\gamma}\right)^{-1} \\ &= \tilde{\tau}_{\alpha\beta}\tilde{\tau}_{\beta\gamma}\tilde{\tau}_{\gamma\alpha} \\ &\in \ker(Spin(n) \to SO(n)). \end{aligned}$$

It is easily verified that this map is alternating, thus it defines a cocycle. To make it onto a cohomology class we must check that it is also coclosed. Using that $f(\alpha, \beta, \gamma)$ is always central (this will become clear below when we give an explicit description of $Spin(n)$) and does not change under cyclic permutations of (α, β, γ), we can compute:

$$\begin{aligned} df(\alpha_0, \alpha_1, \alpha_2, \alpha_3) &= f(\alpha_1, \alpha_2, \alpha_3) \cdot (f(\alpha_0, \alpha_2, \alpha_3))^{-1} \\ &\quad \cdot f(\alpha_0, \alpha_1, \alpha_3) \cdot (f(\alpha_0, \alpha_1, \alpha_2))^{-1} \\ &= \left(\tilde{\tau}_{\alpha_1\alpha_2}\tilde{\tau}_{\alpha_2\alpha_3}\tilde{\tau}_{\alpha_3\alpha_1}\right)\left(\tilde{\tau}_{\alpha_0\alpha_2}\tilde{\tau}_{\alpha_2\alpha_3}\tilde{\tau}_{\alpha_3\alpha_0}\right)^{-1} \end{aligned}$$

$$
\begin{aligned}
&\cdot\left(\tilde{\tau}_{\alpha_0\alpha_1}\tilde{\tau}_{\alpha_1\alpha_3}\tilde{\tau}_{\alpha_3\alpha_0}\right)\left(\tilde{\tau}_{\alpha_0\alpha_1}\tilde{\tau}_{\alpha_1\alpha_2}\tilde{\tau}_{\alpha_2\alpha_0}\right)^{-1}\\
&=\left(\tilde{\tau}_{\alpha_1\alpha_2}\tilde{\tau}_{\alpha_2\alpha_3}\tilde{\tau}_{\alpha_3\alpha_1}\right)\left(\tilde{\tau}_{\alpha_0\alpha_3}\tilde{\tau}_{\alpha_3\alpha_2}\tilde{\tau}_{\alpha_2\alpha_0}\right)\\
&\quad\cdot\left(\tilde{\tau}_{\alpha_0\alpha_1}\tilde{\tau}_{\alpha_1\alpha_3}\tilde{\tau}_{\alpha_3\alpha_0}\right)\left(\tilde{\tau}_{\alpha_0\alpha_2}\tilde{\tau}_{\alpha_2\alpha_1}\tilde{\tau}_{\alpha_1\alpha_0}\right)\\
&=\left(\tilde{\tau}_{\alpha_1\alpha_2}\tilde{\tau}_{\alpha_2\alpha_3}\tilde{\tau}_{\alpha_3\alpha_1}\right)\left(\tilde{\tau}_{\alpha_0\alpha_3}\tilde{\tau}_{\alpha_3\alpha_2}\tilde{\tau}_{\alpha_2\alpha_0}\right)\\
&\quad\cdot\left(\tilde{\tau}_{\alpha_0\alpha_2}\tilde{\tau}_{\alpha_2\alpha_1}\tilde{\tau}_{\alpha_1\alpha_0}\right)\left(\tilde{\tau}_{\alpha_0\alpha_1}\tilde{\tau}_{\alpha_1\alpha_3}\tilde{\tau}_{\alpha_3\alpha_0}\right)\\
&=\left(\tilde{\tau}_{\alpha_1\alpha_2}\tilde{\tau}_{\alpha_2\alpha_3}\tilde{\tau}_{\alpha_3\alpha_1}\right)\tilde{\tau}_{\alpha_0\alpha_3}\left(\tilde{\tau}_{\alpha_3\alpha_2}\tilde{\tau}_{\alpha_2\alpha_1}\tilde{\tau}_{\alpha_1\alpha_3}\right)\tilde{\tau}_{\alpha_3\alpha_0}\\
&=\left(\tilde{\tau}_{\alpha_1\alpha_2}\tilde{\tau}_{\alpha_2\alpha_3}\tilde{\tau}_{\alpha_3\alpha_1}\right)\left(\tilde{\tau}_{\alpha_3\alpha_2}\tilde{\tau}_{\alpha_2\alpha_1}\tilde{\tau}_{\alpha_1\alpha_3}\right)\\
&=\left(\tilde{\tau}_{\alpha_3\alpha_1}\tilde{\tau}_{\alpha_1\alpha_2}\tilde{\tau}_{\alpha_2\alpha_3}\right)\left(\tilde{\tau}_{\alpha_3\alpha_2}\tilde{\tau}_{\alpha_2\alpha_1}\tilde{\tau}_{\alpha_1\alpha_3}\right)\\
&=1.
\end{aligned}
$$

Thus, we have found a cohomology class $w_2(M) \in H^2(M, \mathbb{Z}_2)$ called the *second Stiefel-Whitney class*. This class is zero iff one can choose the lifts to $Spin(n)$ such that the conditions $\tilde{\tau}_{\alpha\beta} \cdot \tilde{\tau}_{\beta\gamma} = \tilde{\tau}_{\alpha\gamma}$ hold whenever it makes sense. It is again easy to show that $w_2(M)$ is invariant of the metric and therefore a manifold invariant. One can also show that w_2 is the image of the first Chern class under the map $H^2(M, \mathbb{Z}) \to H^2(M, \mathbb{Z}_2)$. A manifold such that $w_1(M) = 0$ and $w_2(M) = 0$ is called a *spin manifold*.

On a spin manifold it is now possible to lift the frame bundle to a principal $Spin(n)$ bundle. The construction imitates our local construction of the frame bundle. Thus, we assume that a good covering is given, where $\tau_{\alpha\beta} \in SO(n)$, and that the lifts $\tilde{\tau}_{\alpha\beta} \in Spin(n)$ satisfy the compatibility condition $\tilde{\tau}_{\alpha\beta} \cdot \tilde{\tau}_{\beta\gamma} = \tilde{\tau}_{\alpha\gamma}$. Over U_α we declare the bundle to be trivial: $U_\alpha \times Spin(n)$. On the intersections $U_\alpha \cap U_\beta$ we then identify points $(p_\alpha, \sigma) \in U_\alpha \times Spin(n)$ and $\left(p_\alpha, \sigma \cdot \tilde{\tau}_{\alpha\beta}\right) \in U_\beta \times Spin(n)$. This gives a well-defined equivalence relation, because we have assumed that

$$
\begin{aligned}
\left(\tilde{\tau}_{\alpha\beta}\right)^{-1} &= \tilde{\tau}_{\beta\alpha},\\
\tilde{\tau}_{\alpha\beta} \cdot \tilde{\tau}_{\beta\gamma} &= \tilde{\tau}_{\alpha\gamma}.
\end{aligned}
$$

This bundle is denoted by SM or $SpinM$ and is called the *spin bundle* over M. Over a given coordinate chart, we have the commutative diagram:

$$
\begin{array}{ccc}
U_\alpha \times Spin(n) & \to & U_\alpha \times SO(n)\\
\downarrow & & \downarrow\\
U_\alpha & \to & U_\alpha,
\end{array}
$$

where the vertical arrows are projections to the first factor and the upper horizontal arrow is simply the covering map $\pi : Spin(n) \to SO(n)$. If in this F^+M denotes the $SO(n)$ principal bundle of positively oriented frames, then we get a fiber-preserving covering map: $\pi : SM \to F^+M$, such that for all $\sigma \in SM$ and $g \in Spin(n)$

$$
\pi(\sigma \cdot g) = \pi(\sigma) \cdot \pi(g),
$$

and when restricted to the fibers, it is simply the covering map $\pi : Spin(n) \to SO(n)$.

Since $O(n)$, $SO(n)$, and $Spin(n)$ all have $\mathfrak{so}(n)$ as a Lie algebra, we see that F^+M has a natural connection induced from FM. By pulling back this 1-form to SM, we also get a natural connection form on SM.

The hope now is that this new principal bundle contains more information than the frame bundle. This is by no means clear, but it turns out that $Spin(n)$ does have some new representations that are not just induced from an $SO(n)$ action. This is the subject of the next section. First, we should mention some examples of spin manifolds.

Example 1.1 We have kept 2-manifolds in the background in the above discussion, as $SO(2)$ has infinite fundamental group. In that case, one can define $Spin(2)$ to be the unique 2-fold covering of $SO(2)$. Thus, everything goes through without change. Using the classification of 2-manifolds one can, by inspection, check that all the orientable ones are in fact also spin. For the 2-sphere, for instance, the frame bundle is isomorphic to the unit sphere bundle, which in turn is $\mathbb{R}P^3 \to S^2$. Thus, it is not hard to guess that $SS^2 = S^3 \to S^2$ is the Hopf fibration.

Example 1.2 Any parallelizable manifold must be spin, as the frames on the individual sets of the covering can be chosen to agree on intersections.

Example 1.3 One can show that any orientable 3-manifold has $w_2 = 0$ and is therefore spin (see [61]). In fact, all orientable 3-manifolds are parallelizable. In particular, $\mathbb{R}P^3$ is a spin manifold.

Example 1.4 For a manifold M we can construct the class $w(M) = 1+w_1(M)+w_2(M) \in H^0(M,\mathbb{Z}_2)\oplus H^1(M,\mathbb{Z}_2)\oplus H^2(M,\mathbb{Z}_2)$. (Normally, one would continue this series with the higher Stiefel-Whitney classes. Also, we are being a little loose, as w_2 was only defined when $w_1 = 0$.) Using the above characterizations, we note that

$$\begin{aligned} w(M \times N) &= w(M)\, w(N) \\ &= 1 + w_1(M) + w_1(N) \\ &\quad + w_2(M) + w_1(M)\, w_1(N) + w_2(N). \end{aligned}$$

In particular, the product of two spin manifolds is again a spin manifold. This is also easily checked directly from the definition.

Example 1.5 One can also show, using a Meyer-Vietoris argument, that the connected sum $M\sharp N$ is spin, provided that both M and N are spin.

Example 1.6 In dimension 4 we have that S^4 and $S^2 \times S^2$ are spin. On the other hand, in analogy with dimension 2, $\mathbb{C}P^2$ is not spin.

C.2 Spinor Bundles

In analogy with our construction of tensor bundles from the frame bundle, we shall now construct some spin bundles. In order to construct these bundles, we need to recall Clifford multiplication. On $\mathbb{R}^n$ we construct the space $(\Lambda^p\mathbb{R}^n)^*$ of alternating p-linear maps from $\mathbb{R}^n$ to $\mathbb{R}$ and the space $(\Lambda^*\mathbb{R}^n)^* = \left(\Lambda^0\mathbb{R}^n\right)^* \oplus \cdots \oplus (\Lambda^n\mathbb{R}^n)^*$of all alternating multilinear maps. Note that $\left(\Lambda^0\mathbb{R}^n\right)^* = \mathbb{R}$ and $\left(\Lambda^1\mathbb{R}^n\right)^* = (\mathbb{R}^n)^*$. Given an orthonormal basis e_i for $\mathbb{R}^n$ and dual basis e^i for $(\mathbb{R}^n)^*$, we declare $e^{i_1} \wedge \cdots \wedge e^{i_p}$, $i_1 < \cdots < i_p$, $p = 0, \ldots, n$, an orthonormal basis for $(\Lambda^*\mathbb{R}^n)^*$. On this space we have, as already mentioned, defined Clifford multiplication. Aside from being associative and distributive with respect to addition, it satisfies the important relation

$$-2\,\langle\theta_1, \theta_2\rangle = \theta_1 \cdot \theta_2 + \theta_2 \cdot \theta_1$$

for all $\theta_1, \theta_2 \in (\mathbb{R}^n)^*$, where $\langle\theta_1, \theta_2\rangle$ denotes the inner product. It will be convenient to use the notation Cl_n for the space $(\Lambda^*\mathbb{R}^n)^*$ when thought of as an algebra with respect to Clifford multiplication. Again we see that $e^{i_1} \cdot \cdots \cdot e^{i_p}$, $i_1 < \cdots < i_p$, $p = 0, \ldots, n$, forms an orthonormal basis for Cl_n. We shall call Cl_n the n-dimensional *Clifford algebra.*

We have the multiplicative group $\mathrm{Cl}_n^\times$ consisting of all invertible elements of Cl_n. Note that all nonzero $\theta \in (\mathbb{R}^n)^*$ have $-\theta \cdot |\theta|^{-2}$ as a multiplicative inverse. Thus, $(\mathbb{R}^n)^* - \{0\} \subset \mathrm{Cl}^\times(E)$, and in particular, it follows that $\mathrm{Cl}_n^\times$ is an open dense subset of Cl_n and therefore also a Lie group.

The important observation, for our purposes here, is that $Spin\,(n)$ is a subgroup of $\mathrm{Cl}_n^\times$.

Theorem 2.1 *The spin group $Spin\,(n)$ is the group G generated by elements of the form $\theta_1 \cdot \theta_2$, where $\theta_1, \theta_2 \in (\mathbb{R}^n)^*$ have unit length.*

Proof. The proof of this is actually quite nice and geometric. First, we realize that each $\theta \in (\mathbb{R}^n)^*$ of unit length generates a reflection on $\mathbb{R}^n$ by the formula

$$\phi\,(\theta)\,(v) = v - 2\theta\,(v)\,\theta^\flat.$$

In other words, $\phi\,(\theta)$ is the reflection in the hyperplane orthogonal to the dual $\theta^\flat \in \mathbb{R}^n$ of θ. From this we see that $\phi\,(\theta_1) = \phi\,(\theta_2)$ iff $\theta_1 = \pm\theta_2$. Also, ϕ generates a homomorphism from the group generated by all elements θ of unit length. In particular, $\phi\,(\theta_1) \circ \phi\,(\theta_2)$ is an orientation-preserving isometry and therefore an element of $SO\,(n)$. Thus, we have a homomorphism: $\phi : G \to SO\,(n)$.

It is a classic theorem of Cartan that any element of the orthogonal group is a composition of reflections. Since all reflections are orientation reversing, we see that each element of $SO\,(n)$ is the composition of an even number of reflections. This means that ϕ is onto. The maps is two-to-one because only elements that are the negative of each other are identified. To make G into $Spin\,(n)$, it only remains to be seen that it is connected. Since $SO\,(n)$ is connected and the kernel of ϕ is $\{1, -1\}$, it suffices to find a path connecting these two points. This is given by choosing two perpendicular unit elements $\theta_1, \theta_2 \in (\mathbb{R}^n)^*$ and then considering the path

$$\left[0, \frac{\pi}{2}\right] \to G,$$
$$\begin{aligned}\gamma(t) &= (\theta_1 \cos t + \theta_2 \sin t) \cdot (-\theta_1 \cos t + \theta_2 \sin t) \\ &= \cos^2 t - \sin^2 t + (\theta_1 \cdot \theta_2)(2 \cos t \sin t) \\ &= \cos(2t) + \theta_1 \cdot \theta_2 \sin(2t).\end{aligned}$$

A twofold covering map $G \to SO(n)$ having been constructed and G shown to be connected, it must follow that $G = Spin(n)$. □

From this we see that $Spin(n)$ acts on Cl_n from the left by Clifford multiplication. This action is not just an action that comes from an action of $SO(n)$, as ± 1 clearly act differently on Cl_n. The representation theory of $Spin(n)$ is therefore richer than that of $SO(n)$. This obviously has the consequence that we can construct some natural spin bundles over spin manifolds that are not just tensor bundles.

Note that $SO(n)$ also acts on Cl_n from the left in the following way:

$$\begin{aligned}g \cdot (\theta_1 \cdot \cdots \cdot \theta_p) &= \theta_1 \circ g^{-1} \cdot \cdots \cdot \theta_p \circ g^{-1} \\ &= \theta_1 \circ g^t \cdot \cdots \cdot \theta_p \circ g^t.\end{aligned}$$

This action is clearly very different from the spin action.

On any Riemannian manifold (M, g) we can now construct the *Clifford bundle*

$$\mathrm{Cl}_O(M) = FM \times_{O(n)} \mathrm{Cl}_n.$$

In case M is oriented, this bundle is the same as

$$\mathrm{Cl}_{SO}(M) = F^+M \times_{SO(n)} \mathrm{Cl}_n.$$

This bundle is easily seen to be the same as the bundle $\Omega^*(M)$ of forms with the Clifford multiplication induced from the Riemannian metric. Still, it is convenient to have two different notations for the these spaces depending on how we view them.

On a Riemannian spin manifold we can now construct a *spinor bundle*

$$\mathrm{Cl}_{Spin}(M) = SM \times_{Spin(n)} \mathrm{Cl}_n.$$

Sections of this bundle are called *spinors*. Due to the different representation we use, these sections cannot be thought of as tensors. In this definition one must, of course, watch out that the correct action of $Spin(n)$ on Cl_n is chosen, as we have the spin action and then the trivial one induced from the $SO(n)$ action. In the latter case we merely reproduce the Clifford bundle, so we shall always think of $Spin(n)$ as acting by Clifford multiplication on Cl_n. This spinor bundle has a very important module structure. Starting with the fact that Cl_n acts on itself from the left by Clifford multiplication, we see that the Clifford bundle must act on the spinor bundle from the left as well. Thus, we can form the Clifford product $\omega \cdot \sigma$ for sections ω of $\mathrm{Cl}_{SO}(M)$ and σ of $\mathrm{Cl}_{Spin}(M)$.

Given a left Cl_n module V (i.e., a vector space V and a ring homomorphism $\mathrm{Cl}_n \to End(V)$), we can more generally construct a *spinor bundle*

$$S(V) = SM \times_{Spin} V$$

using that $Spin(n)$ acts on V via the action of Cl_n on V. On this bundle we clearly also have an action of the Clifford bundle from the left. Thus, there is probably no shortage of spinor bundles on a spin manifold. Again, sections of this bundle are called spinors.

It would be nice to a have a canonical spinor bundle just as the tangent bundle is the canonical tensor bundle. Using the Wedderburn theorem, one can show that any Cl_n module is a direct sum of irreducible Cl_n modules. One can then show that when $n \not\equiv 3 \,(\mathrm{mod} 4)$ there is only one such irreducible module, while if $n \equiv 3 \,(\mathrm{mod} 4)$ there are two distinct irreducible modules. In the former case we therefore get a canonical spinor bundle. In the latter situation it turns out that the two inequivalent Cl_n modules have the same actions by $Spin(n)$, so even in this case one gets a unique spinor bundle. While all of this is quite important, if one goes more deeply into the theory, one can still get around these subtleties. Namely, we know that there always is a spinor bundle on a spin manifold, and the theory for all of the spinor bundles is virtually the same. So at a more superficial level there is no reason to separate out the analysis for different spinor bundles. Still, it might be instructive to see what happens in low dimensions.

One can easily check that the first two Clifford algebras look like

$$\mathrm{Cl}_1 = \mathbb{C},$$
$$\mathrm{Cl}_2 = \mathbb{H}.$$

The irreducible representations are then just the action of the Clifford algebras on themselves. The corresponding spin actions are, however, not irreducible. This is because $Spin(1)$ is simply defined to be the trivial group, and $Spin(2) = S^1$. This is a phenomenon that happens in many dimensions and yields some very important information.

The first time something really interesting happens is in dimension 3. The Clifford algebra Cl_3 has the orthonormal basis $\{1, e^1, e^2, e^3, e^1 \cdot e^2, e^2 \cdot e^3, e^3 \cdot e^1, e^1 \cdot e^2 \cdot e^3\}$. We claim that Cl_3 has two inequivalent actions on the quaternions $\mathbb{H}$. If we think of $\mathbb{H} = \mathrm{span}\{1, i, j, k\}$, then the first action can be described as

$$1 \cdot v = v,$$
$$e^1 \cdot v = iv,$$
$$e^2 \cdot v = jv,$$
$$e^3 \cdot v = kv.$$

This gives a well-defined action, as Clifford multiplication and quaternion multiplication with $\{i, j, k\}$ have similar relations. The other action is defined the same way, except multiplication is from the right. Clearly, both of these actions are irreducible, and it is not hard to see that these two actions are inequivalent.

Now for the spin representations. Recall that $Sp(1) = SU(2) = Spin(3) = S^3$ and $SO(3) = \mathbb{R}P^3$. It is therefore natural to assume that the action of $Sp(1) = S^3$ by multiplication of unit quaternions on $\mathbb{H}$ is the desired representation. This is indeed true. Just note that any element of $Spin(3)$ looks like

$$\begin{aligned}\left(a + be^1 + ce^2 + de^3\right)\left(a' + b'e^1 + c'e^2 + d'e^3\right) \\ = \left(\alpha + \beta e^1 \cdot e^2 + \gamma e^2 \cdot e^3 + \delta e^3 \cdot e^1\right)\end{aligned}$$

when acting on $\mathbb{H}$. Using the action just described, such an element will act like

$$(\alpha + \beta ij + \gamma jk + \delta ki) = (\alpha + \beta k + \gamma i + \delta j).$$

One then only needs to check that this is an element of unit length. This comes from the fact that both $\left(a + be^1 + ce^2 + de^3\right)$ and $\left(a' + b'e^1 + c'e^2 + d'e^3\right)$ have unit length. Finally, one can use the commutation rules $ij = -ji$, etc., to see that the right action by unit quaternions is equivalent to the left action.

All in all, we have then found a natural spinor bundle:

$$SM \times_{Sp(1)} \mathbb{H}$$

for all orientable 3-manifolds. It comes equipped with a metric and a quaternion structure. In general, one almost always gets a natural complex or quaternion structure on spinor bundles. Note that without knowledge of the Clifford algebra, we could have guessed that the bundle should look like

$$SM \times_{Sp(1)} \mathbb{H} \quad \text{or} \quad SM \times_{SU(2)} \mathbb{C}^2,$$

but then we would have had to reconstruct the action of the Clifford bundle. It is usually better to begin with that action and then restrict to the spin action.

In dimension 4 we see that $Spin(4) = Spin(3) \times Spin(3) = Sp(1) \times Sp(1)$. This group clearly acts irreducibly on $\mathbb{H} \times \mathbb{H}$. So it is certainly a good guess that the irreducible action of Cl_4 in this case is on $\mathbb{H} \times \mathbb{H}$. Again this is indeed true. The details are left to the reader.

It is still not clear what these bundles are good for. Hopefully some justification can be found in the next sections.

Before proceeding, we wish to mention briefly something about the local theory of spinors. Recall that for a tensor $T : FM \to V$ the coefficients in a local frame $E : U \to FM$ are found simply by taking the composition $T \circ E : U \to V$. A similar thing can be done for spinors, as long as the domain U is simply connected. Suppose we a have a positively oriented frame $E : U \to F^+M$. Then there are two lifts:

$$\begin{array}{ccccc} & & SU & = & Spin(n) \times U \\ & \overset{E^{\pm}}{\nearrow} & \downarrow & & \downarrow \\ U & \overset{E}{\to} & F^+U & = & SO(n) \times U. \end{array}$$

The local coordinates for a spinor $\sigma : SM \to V$ are therefore given by the compositions $\sigma \circ E^{\pm}$. Note that if the spinor has constant coefficients in one of the lifts, then they will also be constant in the other. Thus, we have a well-defined notion of a spinor having constant coefficients with respect to a given frame. This local theory of spinors is not nearly as well understood as the similar local theory for tensors, at least by mathematicians. Physicists seem generally to have a much better feel for spinors. A perfect example of this is Witten's proof of the positive mass conjecture, see e.g., [9] and [55]. A key ingredient in this proof is the use of spinors that have constant coefficients in a given frame. Another key ingredient is the Bochner technique for spinors. The next two sections will be devoted to these matters.

C.3 The Weitzenböck Formula for Spinors

Our approach here is exactly the same as we took in Chapter 7, when we developed Clifford multiplication to take care of the Bochner technique for forms. Here the theory is similar, but also different, as we have used a different type of representation to define spinors.

We shall throughout work with a spinor bundle

$$S(V) = SM \times_{Spin(n)} V$$

on a closed oriented Riemannian spin n-manifold (M, g). As already described, sections of such a bundle can be multiplied by sections on the Clifford bundle $\mathrm{Cl}_{SO}(M)$.

On both of these bundles we have natural connections, which are both denoted by ∇, and on the spinor bundle we have the curvature transformation

$$R^{Spin}(X, Y)\sigma = \nabla^2_{X,Y}\sigma - \nabla^2_{Y,X}\sigma.$$

The important property is that we have Leibniz's rule:

$$\nabla_X(\omega \cdot \sigma) = (\nabla_X \omega) \cdot \sigma + \omega \cdot \nabla_X \sigma,$$
$$R(X, Y)(\omega \cdot \sigma) = R(X, Y)(\omega) \cdot \sigma + \omega \cdot R(X, Y)(\sigma).$$

The *Dirac operator* is a first-order operator that acts on sections of $S(V)$, i.e., it acts on spinors. It is defined exactly as we defined the Dirac operator on forms. Namely, for any (not necessarily orthonormal) frame E_i and dual frame θ^i we declare

$$D\sigma = \theta^i \cdot \nabla_{E_i}\sigma$$

for any section σ of the spinor bundle $S(V)$. Note that this formula is independent of the chosen frame and therefore defines an invariant operator. The square of D is the natural Laplacian on spinors. As with forms, we have

Proposition 3.1 *For any frame E_i and dual coframe θ^i we have for any spinor σ the formulae*

$$\begin{aligned} D^2\sigma &= \theta^i \cdot \theta^j \cdot \nabla^2_{E_i,E_j}\sigma \\ &= \nabla^*\nabla\sigma + \frac{1}{2}\sum_{i,j=1}^{n} \theta^i \cdot \theta^j \cdot R^{Spin}\left(E_i, E_j\right)\sigma. \end{aligned}$$

Proof. As the right-hand side is invariant of the chosen frame, we can as usual assume that it is orthonormal and compute

$$\begin{aligned} D^2\sigma &= \theta^i \cdot \left(\nabla_{E_i}\left(\theta^j \cdot \nabla_{E_j}\sigma\right)\right) \\ &= \theta^i \cdot \left(\nabla_{E_i}\left(\theta^j\right) \cdot \nabla_{E_j}\sigma + \theta^j \cdot \nabla_{E_i}\nabla_{E_j}\sigma\right) \\ &= \theta^i \cdot \theta^j \cdot \nabla_{E_i}\nabla_{E_j}\sigma + \theta^i \cdot \nabla_{E_i}\left(\theta^j\right) \cdot \nabla_{E_j}\sigma \\ &= \theta^i \cdot \theta^j \cdot \nabla_{E_i}\nabla_{E_j}\sigma + \theta^i \cdot \theta^k \cdot g\left(E_k, \nabla_{E_i}E_j\right)\nabla_{E_j}\sigma \\ &= \theta^i \cdot \theta^j \cdot \nabla_{E_i}\nabla_{E_j}\sigma - \theta^i \cdot \theta^k \cdot g\left(\nabla_{E_i}E_k, E_j\right)\nabla_{E_j}\sigma \\ &= \theta^i \cdot \theta^j \cdot \nabla_{E_i}\nabla_{E_j}\sigma - \theta^i \cdot \theta^k \cdot \nabla_{g(\nabla_{E_i}E_k, E_j)E_j}\sigma \\ &= \theta^i \cdot \theta^j \cdot \nabla_{E_i}\nabla_{E_j}\sigma - \theta^i \cdot \theta^k \cdot \nabla_{\nabla_{E_i}E_k}\sigma \\ &= \theta^i \cdot \theta^j \cdot \nabla_{E_i}\nabla_{E_j}\sigma - \theta^i \cdot \theta^j \cdot \nabla_{\nabla_{E_i}E_j}\sigma \\ &= \theta^i \cdot \theta^j \cdot \nabla^2_{E_i,E_j}\sigma \\ &= \sum_{i=1}^{n}\theta^i \cdot \theta^i \cdot \nabla^2_{E_i,E_i}\sigma + \sum_{i\neq j}\theta^i \cdot \theta^j \cdot \nabla^2_{E_i,E_j}\sigma \\ &= -\sum_{i=1}^{n}\nabla^2_{E_i,E_i}\sigma + \sum_{i<j}\theta^i \cdot \theta^j \cdot \left(\nabla^2_{E_i,E_j}\sigma - \nabla^2_{E_j,E_i}\sigma\right) \\ &= \nabla^*\nabla\sigma + \sum_{i<j}\theta^i \cdot \theta^j \cdot R^{Spin}\left(E_i, E_j\right)\sigma \\ &= \nabla^*\nabla\sigma + \frac{1}{2}\sum_{i,j=1}^{n}\theta^i \cdot \theta^j \cdot R^{Spin}\left(E_i, E_j\right)\sigma. \end{aligned}$$

□

Finally, we must deconstruct the curvature term. This was first done by Lichnerowicz. He arrived at the surprising result that the curvature term in the above Weitzenböck formula acts simply by multiplication of the scalar curvature on spinors.

Theorem 3.2 (Lichnerowicz, 1963) *For any frame E_i and dual coframe θ^i we have*

$$\frac{1}{4}\mathrm{scal} = \frac{1}{2}\sum_{i,j=1}^{n}\theta^i \cdot \theta^j \cdot R^{Spin}\left(E_i, E_j\right).$$

Proof. First, we must show that the curvature $R^{Spin}(X,Y)$ on spinors can be rewritten in terms of the curvature tensor R on M:

$$R^{Spin}(X,Y)\sigma = \frac{1}{4}\sum_{i,j}^{n} g\left(R(X,Y)E_i, E_j\right)\theta^i\cdot\theta^j\cdot\sigma.$$

This formula is clearly invariant of the frame, so we shall assume that we work with an orthonormal frame. We already know that the curvature form Ω^{Spin} : $\Lambda^2 TM \to \mathfrak{spin}(n)$ satisfies

$$R^{Spin}(X,Y)\sigma = \Omega^{Spin}(X\wedge Y)\cdot\sigma.$$

Note that as the full Clifford algebra Cl_n has itself as tangent space and $Spin(n) \subset \mathrm{Cl}_n$, we also have that $\mathfrak{spin}(n) \subset \mathrm{Cl}_n$. As the spinor bundle is constructed using a Cl_n module, we therefore have that $\Omega^{Spin}(X\wedge Y)\cdot\sigma$ represents Clifford multiplication of the element $\Omega^{Spin}(X\wedge Y) \in \mathfrak{spin}(n) \subset \mathrm{Cl}_n$ on σ. At the same time, we also have that the connection form $\omega^{Spin} : TSM \to \mathfrak{spin}(n)$ simply comes from pulling back the connection form ω on F^+M, via the covering map $SM \to F^+M$, and then lifting $\mathfrak{so}(n)$ to $\mathfrak{spin}(n)$ via the isomorphism $d\phi : \mathfrak{spin}(n) \to \mathfrak{so}(n)$ coming from the covering map $\phi : Spin(n) \to SO(n)$. Thus, the curvature form is simply obtained by the formula

$$\Omega^{Spin} = (d\phi)^{-1}\,\Omega.$$

To compute $d\phi$ we need a concrete description of $\mathfrak{spin}(n)$ as a subset of Cl_n. To this end, we use that we have a curve in $Spin(n)$ with the properties

$$\begin{aligned}\gamma(t) &= \cos(2t) + \theta_1\cdot\theta_2\sin(2t),\\ \gamma(0) &= 1,\\ \dot\gamma(0) &= 2\theta_1\cdot\theta_2.\end{aligned}$$

Thus, we see that the forms $\theta^i\cdot\theta^j \in \mathfrak{spin}(n)$. It is now a simple matter to check that

$$d\phi\left(\theta^i\cdot\theta^j\right)(E_k) = \begin{cases} 2E_j, & k=i,\\ -2E_i, & k=j,\\ 0, & \text{otherwise,}\end{cases}$$

or in other words, $d\phi\left(\theta^i\cdot\theta^j\right) = 2A^{ij}$, where A^{ij} is the skew-symmetric matrix with only two nonzero entries, at positions ij and ji, and those entries are ± 1. As any skew-symmetric matrix $A = \left(a_l^k\right)$ is a linear combination of the matrices A^{ij}, we see that

$$(d\phi)^{-1}(A) = \frac{1}{2}\sum_{k<l} a_l^k\theta^k\cdot\theta^l.$$

We know that the curvature form Ω acts like the skew-symmetric matrix $\left(\Omega_i^j\right)$ defined by

$$\Omega_i^j(X\wedge Y) = g\left(R(X,Y)E_i, E_j\right).$$

Thus, we have the curvature identity

$$\begin{aligned}
\Omega^{Spin}(X \wedge Y) &= \frac{1}{2}\sum_{i<j} \Omega_i^j (X \wedge Y)\theta^i \cdot \theta^j \\
&= \frac{1}{2}\sum_{i<j} g\left(R(X,Y)E_i, E_j\right)\theta^i \cdot \theta^j \\
&= \frac{1}{4}\sum_{i,j=1}^{n} g\left(R(X,Y)E_i, E_j\right)\theta^i \cdot \theta^j.
\end{aligned}$$

We can now reduce the curvature expression, using again that the frame is orthonormal, as follows:

$$\begin{aligned}
&\frac{1}{2}\sum_{i,j=1}^{n} \theta^i \cdot \theta^j \cdot R^{Spin}\left(E_i, E_j\right) \\
&= \frac{1}{8}\sum_{i,j,k,l} \theta^i \cdot \theta^j \cdot g\left(R\left(E_i, E_j\right)E_k, E_l\right)\theta^k \cdot \theta^l \\
&= \frac{1}{8}\sum_{i,j,k,l} g\left(R\left(E_i, E_j\right)E_k, E_l\right)\theta^i \cdot \theta^j \cdot \theta^k \cdot \theta^l \\
&= \frac{1}{8}\sum_{l}\left(\sum_{\substack{i,j,k \\ \text{distinct}}} g\left(R\left(E_i, E_j\right)E_k, E_l\right)\theta^i \cdot \theta^j \cdot \theta^k\right)\cdot \theta^l \\
&\quad + \frac{1}{8}\sum_{l}\sum_{i,j} g\left(R\left(E_i, E_j\right)E_j, E_l\right)\theta^i \cdot \theta^j \cdot \theta^j \cdot \theta^l \\
&\quad + \frac{1}{8}\sum_{l}\sum_{i,j} g\left(R\left(E_i, E_j\right)E_i, E_l\right)\theta^i \cdot \theta^j \cdot \theta^i \cdot \theta^l \\
&= \frac{1}{8}\sum_{l}\left(\sum_{\substack{i,j,k \\ \text{distinct}}} g\left(R\left(E_i, E_j\right)E_k, E_l\right)\theta^i \cdot \theta^j \cdot \theta^k\right)\cdot \theta^l \\
&\quad - \frac{1}{8}\sum_{l}\sum_{i,j} g\left(R\left(E_i, E_j\right)E_j, E_l\right)\theta^i \cdot \theta^l \\
&\quad - \frac{1}{8}\sum_{l}\sum_{i,j} g\left(R\left(E_j, E_i\right)E_i, E_l\right)\theta^j \cdot \theta^l \\
&= \frac{1}{8}\sum_{l}\left(\sum_{\substack{i,j,k \\ \text{distinct}}} g\left(R\left(E_i, E_j\right)E_k, E_l\right)\theta^i \cdot \theta^j \cdot \theta^k\right)\cdot \theta^l
\end{aligned}$$

$$
\begin{aligned}
&-\frac{1}{4}\sum_{i,l}\mathrm{Ric}\left(E_i,E_l\right)\theta^i\cdot\theta^l\\
&=\frac{1}{8}\sum_l\left(\sum_{\substack{i,j,k\\ \text{distinct}}} g\left(R\left(E_i,E_j\right)E_k,E_l\right)\theta^i\cdot\theta^j\cdot\theta^k\right)\cdot\theta^l\\
&+\frac{1}{4}\mathrm{scal}.
\end{aligned}
$$

Finally, we can make the extra term disappear by using Bianchi's first identity and the fact that a cyclic permutation of the indices in $\theta^i\cdot\theta^j\cdot\theta^k$ does not change the expression since the indices are all distinct, as follows:

$$
\begin{aligned}
&\sum_{\substack{i,j,k\\ \text{distinct}}} g\left(R\left(E_i,E_j\right)E_k,E_l\right)\theta^i\cdot\theta^j\cdot\theta^k\\
&\quad=\frac{1}{3}\sum_{\substack{i,j,k\\ \text{distinct}}}\left(\begin{array}{l} g\left(R\left(E_i,E_j\right)E_k,E_l\right)\\ +g\left(R\left(E_k,E_i\right)E_j,E_l\right)\\ +g\left(R\left(E_j,E_k\right)E_i,E_l\right)\end{array}\right)\theta^i\cdot\theta^j\cdot\theta^k\\
&\quad=0.
\end{aligned}
$$

□

Thus, we have the Lichnerowicz formula for the Dirac operator on spinors:

$$D^2=\nabla^*\nabla+\frac{1}{4}\mathrm{scal}.$$

From this formula we can deduce

Theorem 3.3 (Lichnerowicz, 1963) *On a compact Riemannian spin manifold with* scal ≥ 0 *any harmonic spinor is parallel, and if in addition* scal >0 *somewhere, then there are no nontrivial harmonic spinors.*

It is a consequence of this and the Atiyah-Singer index theorem that the $\hat{A}$ genus of such manifolds vanishes in dimensions $n\equiv 0\,(\mathrm{mod}4)$. The full story can be found in [54] and [79, vol. II].

We get a very interesting and strong rigidity result from the above theorem:

Corollary 3.4 *Any compact spin manifold with* scal ≥ 0 *and nonvanishing* $\hat{A}$ *genus must be Ricci flat.*

Proof. The hypotheses guarantee that there are nontrivial harmonic spinors. These harmonic spinors must be parallel as the scalar curvature is nonnegative. The result now follows from a simple calculation similar to the one done for the curvature term in the Bochner formula for spinors. The claim is, in fact, that whenever there is a nontrivial parallel spinor the metric must be Ricci flat.

In analogy with the calculation showing that

$$\frac{1}{4}\text{scal} = \frac{1}{2}\sum_{i,j=1}^{n}\theta^i \cdot \theta^j \cdot R^{Spin}\left(E_i, E_j\right),$$

one can easily see that

$$\sum_{i=1}^{n}\theta^i \cdot R^{Spin}\left(X, E_i\right) = -2\sum_{i=1}^{n} g\left(\text{Ric}\left(X\right), E_i\right)\theta^i.$$

Thus, the action of $\sum_{i=1}^{n}\theta^i \cdot R^{Spin}(X, E_i)$ on a spinor is like the action of the dual of Ric(X) on the spinor. If the spinor σ is parallel, then we must have that $R^{Spin}(X, Y)\sigma = 0$. Thus, the dual of Ric(X) always acts trivially on σ. This implies that Ric$(X) = 0$ in case the spinor is nontrivial. □

It is a very curious and almost unbelievable fact that the scalar curvature is the natural curvature for the Bochner formula on spinors. In all other cases we had to use Ricci curvature or the curvature operator to get control of this term. The reason for this lies in how spinors are defined. Namely, they come from a Clifford module together with the natural inclusion of $Spin\,(n)$ in Cl_n. This gives a very rich structure to spinors, but at the same time also makes it very hard to give expressions that are invariant, as the structure group is bigger. Thus, one can essentially only come up with scalar curvature as the only natural invariant curvature on spinors.

Another strange thing about spinors is that it doesn't seem to be necessary to distinguish between different types of spinors. This is in sharp contrast to tensors, where we must worry about the exact type of tensor we work with each time. In more advanced treatments of spinors one actually introduces a little more structure, but not nearly as much as we have available for tensors or even just forms. The reason again comes from having a Clifford module structure, together with the fact that the bundle is constructed using this structure. The Clifford module structure in itself is not sufficient information, as the space of all forms is the prototypical example of a Clifford module. But in this case, as we know, the structure group does not act by Clifford multiplication on the forms themselves, but rather by pullbacks. Thus, the key fact about spinors is that the structure group acts by Clifford multiplication, rather than by pushforwards or pullbacks, as in the case of tensor bundles.

C.4 The Square of a Spinor

Aside from the topological implications on the $\hat{A}$ genus, it would be nice to have a more geometric feeling for spinors. Even though they are not tensors, one can construct tensors out of spinors by squaring them. As we don't have a way of multiplying spinors, this at first seems meaningless. There are two different ways of squaring spinors. The first uses tensor products, the other inner products.

Suppose we have a spinor bundle $SM \times_{Spin(n)} V$ coming from a Cl_n module V. We shall in addition assume that $Spin(n)$ acts by isometries on V. This is not a severe restriction, as $Spin(n)$ acts by isometries on Cl_n. In fact, we shall assume the seemingly stronger condition

$$\langle \theta \cdot \sigma_1, \sigma_2 \rangle = - \langle \sigma_1, \theta \cdot \sigma_2 \rangle \quad \text{for all} \quad \theta \in \Omega^1(M).$$

Again, we know this to be true on Cl_n. Since $\theta \cdot \theta = -|\theta|^2$, this is equivalent to the condition

$$\langle \theta \cdot \sigma_1, \theta \cdot \sigma_2 \rangle = |\theta|^2 \langle \sigma_1, \sigma_2 \rangle.$$

Hence, unit 1-forms act by isometries. In particular, $Spin(n)$ must also act by isometries.

The key observation is that any action $Spin(n) \to O(V)$ comes from an $SO(n)$ action exactly when $-1 \in Spin(n)$ acts trivially on V. In case $V = \mathrm{Cl}_n$, the action of -1 is simply multiplication by -1, and thus this is not a tensor bundle. In general, however, we have made no particular assumption about how $-1 \in Spin(n) \subset \mathrm{Cl}_n$ acts on V. The fact that V is a Cl_n module only implies that -1 acts like an involution. Suppose that $-1 \in Spin(n) \subset \mathrm{Cl}_n$ really acts by multiplication by -1 on V. This will always be the case when $V = \mathrm{Cl}_n$ and also for any ideal $V \subset \mathrm{Cl}_n$. This, as we mentioned above, comprises all irreducible Cl_n modules, so we haven't restricted our scope too much. On the tensor product $V \otimes V$ taken over $\mathbb{R}$ we have a tensor-product action of $Spin(n)$ defined by

$$s \cdot (v \otimes w) = s \cdot v \otimes s \cdot w.$$

In particular, $-1 \in Spin(n)$ must act trivially as

$$-1 \cdot (v \otimes w) = (-v) \otimes (-w) = v \otimes w.$$

Hence, this action really comes from an $SO(n)$ action. Consequently, the vector bundle

$$SM \times_{Spin(n)} (V \otimes V)$$

is also a tensor bundle

$$F^+ \times_{SO(n)} (V \otimes V).$$

Thus, we see that any spinor σ that is a section of the spinor bundle $SM \times_{Spin(n)} V$, when "squared" in $SM \times_{Spin(n)} (V \otimes V)$, yields a tensor $\sigma \otimes \sigma$, as it must also be a section of $F^+ \times_{SO(n)} (V \otimes V)$.

In order to get any specific information about the tensor-product bundle, we must start with a special spinor bundle. In many cases it turns out that the square of a spinor is actually a form or even a vector field. It is therefore natural to conjecture that having nontrivial parallel spinors reduces the holonomy. This will be proved below using another way of squaring spinors.

Suppose we have an inner product $\langle \cdot, \cdot \rangle$ on V that is preserved by the $Spin(n)$ action on V. Given a spinor $\sigma : M \to SM \times_{Spin(n)} V$, we can consider the linear map

$$\begin{aligned} \mathrm{Cl}_{SO(n)}(M) &\to \mathbb{R}, \\ \varphi &\to \langle \varphi \cdot \sigma, \sigma \rangle . \end{aligned}$$

Using this, we can construct a mixed form $\psi_\sigma \in \mathrm{Cl}_{SO(n)}(M) = \Omega(M)$ by the relation

$$g(\psi_\sigma, \varphi) = \langle \varphi \cdot \sigma, \sigma \rangle .$$

In the situation where the spinor is parallel, we note that

$$\begin{aligned} \nabla_v (g(\psi_\sigma, \varphi)) &= \langle \nabla_v (\varphi \cdot \sigma), \sigma \rangle + \langle \varphi \cdot \sigma, \nabla_v \sigma \rangle \\ &= \langle (\nabla_v \varphi) \cdot \sigma, \sigma \rangle . \end{aligned}$$

As we also have

$$\begin{aligned} \nabla_v (g(\psi_\sigma, \varphi)) &= g(\nabla_v \psi_\sigma, \varphi) + g(\psi_\sigma, \nabla_v \varphi) \\ &= g(\nabla_v \psi_\sigma, \varphi) + \langle (\nabla_v \varphi) \cdot \sigma, \sigma \rangle , \end{aligned}$$

it follows that

$$g(\nabla_v \psi_\sigma, \varphi) = 0 \quad \text{for all} \quad \varphi \in \mathrm{Cl}_{SO(n)}(M),$$

whence ψ_σ is parallel. This is a little less exciting than one would hope, as the mixed form ψ_σ could be just a constant function. In fact, if we assume the relationship

$$\langle \theta \cdot \sigma, \sigma \rangle = -\langle \sigma, \theta \cdot \sigma \rangle \quad \text{for all} \quad \theta \in \Omega^1(M),$$

then one can easily check that (this is also proved below)

$$\langle \varphi \cdot \sigma, \sigma \rangle = 0 \quad \text{for} \quad \varphi \in \Omega^p(M) \quad \text{when} \quad p \equiv 1, 2 \mathrm{mod} 4.$$

In particular, ψ_σ is never interesting when the dimension is 3.

If instead, we assume that the module V comes with an antisymmetric 2-form ζ, then we can extend this 2-form to the spinor bundle and define ψ_σ by the relation

$$g(\psi_\sigma, \varphi) = \zeta(\varphi \cdot \sigma, \sigma).$$

The antisymmetry of ζ then shows that the zeroth-order term in ψ_σ vanishes, while on the other hand, the first-order term is not forced to vanish. In order to make sure that this form induces a parallel structure on the spinor bundle, we have to assume that it is invariant under the $Spin(n)$ action. We shall, as above, insist on the stronger condition

$$\zeta(\theta \cdot \sigma_1, \sigma_2) = -\zeta(\sigma_1, \theta \cdot \sigma_2) \quad \text{for all} \quad \theta \in (\mathbb{R}^n)^*,$$

or in other words, that multiplication with 1-forms is skew-symmetric. This condition is again equivalent to

$$\zeta(\theta \cdot \sigma_1, \theta \cdot \sigma_2) = |\theta|^2 \zeta(\sigma_1, \sigma_2) \quad \text{for all} \quad \theta \in (\mathbb{R}^n)^*.$$

Hence, the $Spin(n)$ action preserves the 2-form. Thus, we can conclude that the mixed form ψ_σ is parallel if the spinor σ is parallel. Note that with this condition we have instead that (this is also proved below)

$$\zeta(\varphi \cdot \sigma, \sigma) = 0 \quad \text{for} \quad \varphi \in \Omega^p(M) \quad \text{when} \quad p \equiv 3, 4 \text{mod} 4.$$

The two constructions therefore give mixed forms that live in different degrees. We can develop both of them simultaneously by using an Hermitian structure on V. In that case, the real part gives a Euclidean inner product and the imaginary part a 2-form. It is convenient to have a name for this type of structure. We call (V, ζ) an *Hermitian* Cl_n *module* if V is a Cl_n module and $\langle \cdot, \cdot \rangle$ an Hermitian inner product on V such that

$$\langle \theta \cdot \sigma_1, \sigma_2 \rangle = -\langle \sigma_1, \theta \cdot \sigma_2 \rangle \quad \text{for all} \quad \theta \in (\mathbb{R}^n)^*.$$

Our first concern is whether such structures exist.

Proposition 4.1 *Any* Cl_n *module* V *admits an Hermitian structure* $\langle \cdot, \cdot \rangle$ *such that* $(V, \langle \cdot, \cdot \rangle)$ *becomes an Hermitian* Cl_n *module.*

Proof. First we must exhibit a complex structure on V in order for our search to make sense. For that, simply pick a 1-form $\bar{\theta} \in \mathrm{Cl}_n$ of unit length. Then the mapping

$$\sigma \to \bar{\theta} \cdot \sigma$$

gives a complex structure, as $\bar{\theta} \cdot \bar{\theta} = -1$. We now have to construct a Hermitian metric $\langle \cdot, \cdot \rangle$ with respect to this complex structure such that

$$\langle \theta \cdot \sigma, \theta \cdot \sigma \rangle = |\theta|^2 \langle \sigma, \sigma \rangle \quad \text{for all} \quad \theta \in (\mathbb{R}^n)^*.$$

In fact, it suffices to check that

$$\langle \theta \cdot \sigma, \theta \cdot \sigma \rangle = \langle \sigma, \sigma \rangle \quad \text{for all} \quad \theta \in (\mathbb{R}^n)^* \quad \text{with} \quad |\theta| = 1.$$

Now the unit 1-forms on $\mathbb{R}^n$ generate a subgroup $Pin(n) \subset \mathrm{Cl}_n^\times$ that via the reflection map defined above yields a double cover $\phi : Pin(n) \to O(n)$ extending the cover $\phi : Spin(n) \to SO(n)$. The group $Pin(n)$ is therefore a (disconnected) compact Lie group. The condition for the Hermitian metric is that it be invariant under the action of $Pin(n)$ on V. Since the group is a compact Lie group, we can simply take any Hermitian structure on V and then average it with respect to some left-invariant volume form on $Pin(n)$. □

As already mentioned, such an Hermitian structure always yields two mixed forms $\psi_\sigma^{\mathrm{Re}}$ and $\psi_\sigma^{\mathrm{Im}}$ for a given spinor σ. The formula is

$$g(\psi_\sigma^{\mathrm{Re}}, \varphi) + ig(\psi_\sigma^{\mathrm{Im}}, \varphi) = \langle \varphi \cdot \sigma, \sigma \rangle.$$

Depending on the dimension of the underlying manifold, one or the other form can be useful. In many physical applications one takes the degree 1 term from the imaginary part. This is a 1-form that can be thought of as a vector field, hence the phrase, the square of a spinor is a vector field. In dimensions 3 and 4 this is essentially the only nontrivial part of these mixed forms. This is why one usually defines only the vector field part of the square of the spinor.

To clarify matters a little, let us see what happens in dimension 3. As we wish to have a Hermitian structure on the spinors, we shall use that $Spin(3) = SU(2)$ and then consider the spinors as lying in $\mathbb{C}^2$. The Clifford algebra Cl_3 is generated by three 1-forms e^1, e^2, e^3, which act as follows:

$$\begin{aligned} e^1 \cdot (z, w) &= (iz, -iw), \\ e^2 \cdot (z, w) &= (-w, z), \\ e^3 \cdot (z, w) &= (-iw, -iz). \end{aligned}$$

This action comes from defining

$$e^1 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \; e^2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \; e^3 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}.$$

and letting them act on $\mathbb{C}^2$ in the standard way. We then use matrix multiplication to define the action by the Clifford algebra. Using the standard Hermitian metric on $\mathbb{C}^2$, we then get

$$\begin{aligned} \langle (z, w), (z, w) \rangle &= |z|^2 + |w|^2, \\ \left\langle e^1 \cdot (z, w), (z, w) \right\rangle &= i \left(|z|^2 - |w|^2 \right), \\ \left\langle e^2 \cdot (z, w), (z, w) \right\rangle &= -2i \mathrm{Im}(z\bar{w}), \\ \left\langle e^3 \cdot (z, w), (z, w) \right\rangle &= -2i \mathrm{Re}(z\bar{w}), \\ \left\langle e^1 \cdot e^2 \cdot (z, w), (z, w) \right\rangle &= -2i \mathrm{Re}(z\bar{w}), \\ \left\langle e^2 \cdot e^3 \cdot (z, w), (z, w) \right\rangle &= i \left(|z|^2 - |w|^2 \right), \\ \left\langle e^3 \cdot e^1 \cdot (z, w), (z, w) \right\rangle &= -2i \mathrm{Im}(z\bar{w}), \\ \left\langle e^1 \cdot e^2 \cdot e^3 \cdot (z, w), (z, w) \right\rangle &= -\left(|z|^2 + |w|^2 \right). \end{aligned}$$

The squares of the spinor (z, w) are therefore given by

$$\begin{aligned} \psi^{\mathrm{Re}}_{(z,w)} &= |z|^2 + |w|^2 - \left(|z|^2 + |w|^2 \right) e^1 \wedge e^2 \wedge e^3, \\ \psi^{\mathrm{Im}}_{(z,w)} &= \left(|z|^2 - |w|^2 \right) e^1 - 2\mathrm{Im}(z\bar{w}) e^2 - 2\mathrm{Re}(z\bar{w}) e^3 \\ &\quad - 2\mathrm{Re}(z\bar{w}) e^1 \wedge e^2 + \left(|z|^2 - |w|^2 \right) e^2 \wedge e^3 - 2\mathrm{Im}(z\bar{w}) e^3 \wedge e^1. \end{aligned}$$

Thus, the real part of the square carries information only about the norm of the spinor. The imaginary part is certainly more interesting. However, all information is carried by the degree 1 term:

$$\left(|z|^2 - |w|^2 \right) e^1 - 2\mathrm{Im}(z\bar{w}) e^2 - 2\mathrm{Re}(z\bar{w}) e^3.$$

This term is usually what is referred to as the square of the spinor. In fact, the norm of the spinor can be computed as

$$\begin{aligned}\left(|z|^2+|w|^2\right)^2 &= \left(|z|^2-|w|^2\right)^2+4|z|^2|w|^2\\ &= \left(|z|^2-|w|^2\right)^2+4|z\bar{w}|^2.\end{aligned}$$

From this it follows that the degree 1 term contains all information about the square. To say that the square of a spinor in dimension 3 is a 1-form or vector is therefore an accurate picture of what is happening. In fact, Cartan in [20] uses this particular squaring operation in dimension 3 as the starting point for his discussion of spinors.

We note some further interesting properties: The two spinors (z, w) and $e^{i\alpha}(z, w)$ have the same squares. Not every mixed form is the sum of the two squares. And finally, any 1-form can be expressed as a square, in the sense that it can be written as

$$\left(|z|^2-|w|^2\right)e^1-2\mathrm{Im}\,(z\bar{w})\,e^2-2\mathrm{Re}\,(z\bar{w})\,e^3$$

for some (z, w). This last property is quite important and can be shown to hold in the abstract setting as well.

The fact that the holonomy is reduced in the presence of parallel spinors is settled by the next result.

Theorem 4.2 *Let (M, g) be a Riemannian n-manifold that is also spin, and $(V, \langle\cdot,\cdot\rangle)$ an irreducible Hermitian Cl_n module. If the vector bundle $SM\times_{Spin(n)}V$ has a nontrivial parallel section or spinor, then (M, g) has reduced holonomy.*

Proof. We already know from Chapter 8 that a Riemannian manifold that admits a nontrivial parallel p-form, for $p = 1, \ldots, n-1$, has reduced holonomy. Thus, our goal is to construct such a form. Given a nontrivial parallel spinor field $\sigma : M \to SM\times_{Spin(n)}V$, we construct the mixed forms $\psi_\sigma^{\mathrm{Re}}$ and $\psi_\sigma^{\mathrm{Im}}$ defined by

$$g\left(\psi_\sigma^{\mathrm{Re}},\varphi\right)+ig\left(\psi_\sigma^{\mathrm{Im}},\varphi\right)=\langle\varphi\cdot\sigma,\sigma\rangle.$$

We then expand these forms as follows:

$$\begin{aligned}\psi_\sigma^{\mathrm{Re}} &= \psi_\sigma^0+\psi_\sigma^3+\cdots\in\Omega^0\oplus\Omega^3\oplus\cdots,\\ \psi_\sigma^{\mathrm{Im}} &= \psi_\sigma^1+\psi_\sigma^2+\cdots\in\Omega^1\oplus\Omega^2\oplus\cdots.\end{aligned}$$

We should check the above claim that these forms live in the degrees indicated. This comes about as follows. Pick p orthogonal 1-forms $\theta^1, \ldots, \theta^p$ and compute

$$\begin{aligned}\langle\varphi\cdot\sigma,\sigma\rangle &= \left\langle\theta^1\cdot\cdots\cdot\theta^p\cdot\sigma,\sigma\right\rangle\\ &= (-1)^p\left\langle\sigma,\left(\theta^p\cdot\cdots\cdot\theta^1\right)\cdot\sigma\right\rangle\\ &= (-1)^p(-1)^{\frac{p(p-1)}{2}}\left\langle\sigma,\left(\theta^1\cdot\cdots\cdot\theta^p\right)\cdot\sigma\right\rangle\\ &= (-1)^{\frac{p(p+1)}{2}}\left\langle\sigma,\left(\theta^1\cdot\cdots\cdot\theta^p\right)\cdot\sigma\right\rangle\\ &= (-1)^{\frac{p(p+1)}{2}}\,\overline{\langle\varphi\cdot\sigma,\sigma\rangle}.\end{aligned}$$

Therefore, if $p \equiv 3, 4 \pmod 4$, we see that $\langle \varphi \cdot \sigma, \sigma \rangle$ must be real. On the other hand, when $p \equiv 1, 2 \pmod 4$, the term $\langle \varphi \cdot \sigma, \sigma \rangle$ is imaginary.

In order to show that the holonomy is reduced, we must show that at least one of the forms $\psi_\sigma^1, \psi_\sigma^2, \ldots, \psi_\sigma^{n-1}$ is nontrivial. If they are all trivial, then it must follow that

$$\langle \varphi \cdot \sigma, \sigma \rangle = 0 \quad \text{for all} \quad \varphi \in \Omega^1 \oplus \Omega^2 \oplus \cdots \oplus \Omega^{n-1}.$$

Recall from the three-dimensional case that this would definitely imply that the spinor itself is zero. A similar argument can be devised in higher dimensions. □

Harvey and Lawson have a more elegant proof that gives a different kind of information about the holonomy group (see, e.g., [54]). The idea is simply that the holonomy group keeps a parallel spinor fixed and must therefore be a subgroup of the elements of $Spin\,(n)$ that also keep the spinor fixed. As $Spin\,(n)$ is never contained in $SO\,(n)$, it must follow that the holonomy is reduced. The advantage of our proof here is that it exhibits a parallel form. This is something that works nicely only for certain spinor bundles and then only in some dimensions according to the theory set forth in [54]. In fact, it seems that the squaring operation using Hermitian structures, while working in more general contexts, doesn't give any less information.

Combining the results of this section we now have:

Corollary 4.3 *Let (M, g) be a closed Riemannian $4k$-manifold that is spin and has $scal_g \geq 0$. If the holonomy is general, i.e., $\mathfrak{hol}_p = \mathfrak{so}\,(4k)$, then $\hat{A} = 0$.*

Corollary 4.4 *Let (M, g) be a closed Riemannian $4k$-manifold that is irreducible, spin, and has $scal_g \geq 0$. If the $\hat{A}$ genus is nonzero, then the holonomy must be $SU\,(2k)$, $Sp\,(k)$, or $Spin\,(7)$.*

Proof. Given that the manifold is irreducible, we can just look at the list of possibilities. Recall that the metric is Ricci flat by the results from the previous section. General holonomy is ruled out by the above corollary. Otherwise, we note that all but the above listed holonomies force the metric to have positive scalar curvature. This is impossible if $\hat{A}$ is nontrivial. □

The $\hat{A}$ genus is defined independently of the spin structure, but it is not always an integer invariant. In dimension 4 one has the nice formula

$$\hat{A} = \frac{1}{16}\sigma,$$

where $\sigma\,(M)$ is the signature of M, i.e., the index of the intersection form on the middle cohomology class. One can compute

$$\sigma\left(S^4\right) = 0,$$
$$\sigma\left(S^2 \times S^2\right) = 0,$$

$$\sigma\left(\mathbb{C}P^2\right) = 1,$$
$$\sigma\,(\text{K3 surface}) = -16.$$

A K3 surface is a complex hypersurface in $\mathbb{C}P^3$ given by the quartic equation $\left(z^1\right)^4 + \cdots + \left(z^4\right)^4 = 0$ in homogeneous complex coordinates. Of these four manifolds all but $\mathbb{C}P^2$ are spin. The first three admit metrics with positive scalar curvature, while the K3 surface cannot admit a metric with positive scalar curvature from the above results. This manifold, however, does admit a scalar flat metric that must then have $SU\,(2)$ as a holonomy group. Thus, all scalar flat metrics on K3 surfaces are Kähler-Einstein and Ricci flat. Such metrics are known to exist only due to Yau's solution to the Calabi conjecture (see, e.g., [11]).

We should also point out that Hitchin has shown that Milnor's α genus (defined for manifolds of dimension $1, 2$ mod8) vanishes for spin manifolds with positive scalar curvature. The proof again consists in using index theory, to show that this invariant must vanish if there are no harmonic spinors.

Another great use of squaring spinors is in the positive mass conjecture. We can here give an indication of how this works in a much simpler setting (which generalizes an exercise from Chapter 7).

Theorem 4.5 *Let (M, g) be a complete spin manifold that is isometric to Euclidean space outside some compact set. If* $\text{scal}_g \geq 0$, *then (M, g) is isometric to Euclidean space.*

Proof. We assume that there are compact sets $K \subset M$ and $C \subset \mathbb{R}^n$ such that $M - K$ is isometric to $\mathbb{R}^n - C$. On $M - K$ we therefore have an orthonormal frame E of parallel vector fields. Suppose now $SM \times_{Spin(n)} V$ is a spin bundle over M. Given any vector $\sigma_0 \in V$, we can find a unique section $\sigma : F\,(M - K) \to V$ of this bundle such that $\sigma \circ E = \sigma_0$. We must now use some analytic facts. The square D^2 of the Dirac operator is elliptic so it is possible to solve Dirichlet boundary value problems. Now observe that since E is parallel, the section σ must also be parallel on $M - K$. In particular, it is harmonic. By solving the boundary value problem

$$D^2\sigma = 0,$$
$$\sigma = \sigma_0 \quad \text{on} \quad \partial K,$$

we therefore get a global harmonic section σ that is constant in the frame E. The Weitzenböck formula now tells us that

$$0 = D^2\sigma = \nabla^*\nabla\sigma + \frac{1}{4}\text{scal}\sigma.$$

Using the Hermitian inner product on the spinor bundle, integrating over a domain Ω with smooth boundary, and using Stokes' theorem yields

$$0 = \int_\Omega \left\langle \nabla^*\nabla\sigma + \frac{1}{4}\text{scal}\sigma, \sigma \right\rangle d\text{vol}$$

$$= \int_{\Omega} |\nabla\sigma|^2 \, d\text{vol} - \int_{\partial\Omega} \langle \nabla_{E_i}\sigma, \sigma \rangle \cdot i_{E_i} d\text{vol} + \frac{1}{4}\int_{\Omega} \text{scal}\, |\sigma|^2 \, d\text{vol}$$
$$\geq -\int_{\partial\Omega} \langle \nabla_{E_i}\sigma, \sigma \rangle \cdot i_{E_i} d\text{vol}.$$

Therefore, if $\partial\Omega \subset M - K$, then the boundary term must vanish, as $\nabla\sigma = 0$. In particular, the spinor σ is globally parallel. By choosing σ_0 judiciously, we can now extend the frame E to a global parallel frame, thus showing that M must be $\mathbb{R}^n$. To see how this is done, let θ^i be the coframe dual to E. We can for each i pick σ_0 such that

$$g\left(\theta^i, \theta^j\right) = \text{Im}\langle \theta^j \cdot \sigma_0, \sigma_0 \rangle \quad \text{for all} \quad j.$$

That this is possible was established above for the 3-dimensional case and is easy to handle in general. It follows that the degree 1 term in ψ_σ^{Im} is a parallel 1-form on all of M that extends θ^i. Thus, we get a globally parallel set of 1-forms that extend the given parallel orthonormal 1-forms on $M - K$. The fact that they are parallel then shows that they must be orthonormal everywhere. □

Note that we didn't assume M to be diffeomorphic to Euclidean space, and thus M could look like the tautological bundle over one of the projective spaces. In Chapter 3 we constructed scalar flat metrics on all of the manifolds $\tau\left(\mathbb{R}P^{n-1}\right)$ with the property that the metric is asymptotically Euclidean. The above result tells us that there can't be any such metrics that are flat outside a compact set, at least when n is even. In the positive mass conjecture, one studies manifolds of this type that are asymptotically Euclidean of a certain order and have nonnegative scalar curvature. It is then shown that a certain quantity, called the mass, is always positive, and can be zero only when the space is Euclidean. The mass is a quantity that can be computed from the geometry at infinity. In fact, given a spinor that is asymptotically constant at infinity (in an appropriate frame of course) and a sequence of domains Ω_i that exhaust M, we have that the mass is proportional to

$$\lim_{i\to\infty} \int_{\partial\Omega_i} \langle \nabla_{E_i}\sigma, \sigma \rangle \cdot i_{E_i} d\text{vol}.$$

If we use a harmonic spinor in this formula, then as Witten observed, this term is the same as

$$\int_M |\nabla\sigma|^2 + \frac{1}{4}\int_M \text{scal}\, |\sigma|^2 .$$

Hence, the mass is positive when scal ≥ 0 and can be zero only when the space is Euclidean.

There are several interesting rigidity phenomena related to the positive mass conjecture that we have ignored here. First, it should be observed that if instead one assumes nonpositive Ricci curvature, then there are warped product examples on $\tau\left(\mathbb{R}P^{n-1}\right)$ that are isometric to Euclidean space outside a compact set. On the other hand, if the metric has nonpositive sectional curvature, then it must be flat if it is flat outside a compact set. One can then refine these results to situations where one has certain types of curvature decay and various other topological conditions at infinity. The interested reader can consult Greene's article in [46] for more on this story.

The idea behind the positive mass conjecture is that gravity cannot be isolated, i.e., the gravitational effects of a massive body can be measured everywhere in the universe. While we haven't developed the machinery for accurately defining and describing this here (see [62] for the most readable account of the physics behind this), we have presented some Riemannian analogues that are very similar in nature. The idea that gravity always attracts is translated into scal ≥ 0, and the isolation phenomenon is that the space becomes flat outside a compact set, thus making it impossible to tell what happens in the compact region. The further generalizations we mentioned give even stronger results along the line that scalar curvature (gravity) cannot decay too fast, depending on the dimension.

There is one more result that is easy to state, but whose proof unfortunately is not nearly so accessible. The theorem was established by Schoen-Yau in low dimensions using complicated analytical machinery and by Gromov-Lawson using spin geometry.

Theorem 4.6 *Any metric on the n-torus T^n with nonnegative scalar curvature is flat.*

From the Bochner technique we know that this is true for any metric with nonnegative Ricci curvature. Unfortunately, a similar argument using harmonic spinors doesn't seem to work. Even though there are plenty of harmonic forms, there doesn't seem to be a way of extracting harmonic spinors. Note that we need only one nontrivial spinor in order to conclude that the metric is Ricci flat. Instead, one must resort to completely different techniques. The reader can find a treatment in [54], but even then it is necessary to consult other papers to get the complete proof. This theorem actually holds for a large class of manifolds, including all closed manifolds that admit metrics with nonpositive sectional curvature.

C.5 Further Study

We have already mentioned the books [54] and [84] as sources for more in-depth discussions of spinors. The first is a very exhaustive guide, while the latter gives a very nice, quick overview of the irreducible Clifford representations and the Bochner technique for spinors. Both of these books use a slightly different notation than ours. Namely, they think of the Clifford algebra as being the alternating algebra of multivectors rather than the algebra of forms. It is our feeling that the approach used here is preferable in many ways. First and foremost, all our formulae are invariant under general frames rather than just orthonormal frames, thus opening up the way for representing the theory in coordinates. For the reader who is interested in learning about spinors and index theory, the text [79, vol. II] might be the best place to start.

The above-mentioned references [9] and [55] for the positive mass conjecture are also good for further study. The reader might also wish to look at the very readable account of spin representations in [83]. The reader who wishes to get a feeling for how spinors are used in physics can start with the comprehensive text [62]. For some of the exciting new developments in 4-manifold theory that use spin geometry, we refer to [64]. This book gives a basic account of spin geometry and how it is used to construct the Seiberg-Witten invariants. Some of the important developments in this subject are intimately related to the things we have discussed above.

References

[1] C.B. Allendoerfer and A. Weil, *The Gauss-Bonnet theorem for Riemannian polyhedra*, TAMS 53 (1943) 101–129.

[2] M. Anderson, *Short Geodesics and Gravitational Instantons*, J. Diff. Geo. 31 (1990) 265–275.

[3] M. Anderson, *Metrics of positive Ricci curvature with large diameter*, Manu. Math. 68 (1990) 405–415.

[4] V.I. Arnol'd, *Mathematical methods of classical mechanics*, New York: Springer-Verlag, 1989.

[5] W. Ballmann, *Non-positively curved manifolds of higher rank*, Ann. Math. 122 (1985) 597–609.

[6] W. Ballmann, *Spaces of non-positive curvature*, Basel: Birkhäuser, 1995.

[7] W. Ballmann and P. Eberlein, *Fundamental groups of manifolds of non-positive curvature*, J. Diff. Geo. 25 (1987) 1–22.

[8] W. Ballmann, V. Schroeder and M. Gromov, *Manifolds of non-positive curvature*, Boston: Birkhäuser, 1985.

[9] R. Bartnik, *The mass of an asymptotically flat manifold*, Comm. Pure Appl. Math. 39 (1986), 661–693.

[10] M. Berger, *Riemannian geometry during the second half of the twentieth century*, to appear.

[11] A.L. Besse, *Einstein Manifolds*, Berlin–Heidelberg: Springer-Verlag, 1978.

[12] A.L. Besse, *Manifolds all of whose geodesics are closed*, Berlin-Heidelberg: Springer-Verlag, 1987.

[13] R.L. Bishop and R.J. Crittenden, *Geometry of Manifolds*, New York: Academic Press, 1964.

[14] R.L. Bishop and S.I. Goldberg, *Tensor analysis on manifolds*, Dover, 1980.

[15] R. Bott and L.W. Tu, *Differential forms in algebraic topology*, New York: Springer-Verlag, 1982.

[16] K. Burns and R. Spatzier, *On topological Tits buildings and their classification*, IHES Publ. Math. 65 (1987) 5–34.

[17] J.C. Cantrell, ed., *Geometric Topology*, New York-London: Academic Press, 1979.

[18] M.P. do Carmo, *Differential forms and applications*, Berlin-Heidelberg: Springer Verlag, 1994.

[19] M.P. do Carmo, *Riemannian Geometry*, Boston: Birkhäuser, 1993.

[20] E. Cartan, *The theory of spinors*, New York: Dover, 1966.

[21] I. Chavel, *Riemannian Geometry, A Modern Introduction*, New York: Cambridge University Press, 1995.

[22] J. Cheeger, *Comparison and finiteness theorems for Riemannian manifolds*, Ph. D. thesis, Princeton University.

[23] J. Cheeger, *Finiteness theorems for Riemannian manifolds*, Am. J. Math. 92 (1970), 61–75.

[24] J. Cheeger, *Pinching theorems for a certain class of Riemannian manifolds*, Am. J. Math. 92 (1970), 807–834.

[25] J. Cheeger et al., *Geometric Topology: Recent developments*, LNM 1504, Berlin-Heidelberg: Springer-Verlag, 1991.

[26] J. Cheeger and D. Gromoll, *The splitting theorem for manifolds of non-negative Ricci curvature,* J. Diff. Geo. 6 (1971) 119–128.

[27] S.S. Chern, ed., *Global Geometry and Analysis*, 2nd edition, MAA Studies 27, Washington: Mathematical Association of America, 1989.

[28] R. Courant and D. Hilbert, *Methods of Mathematical Physics*, vol. II, New York: Wiley Interscience, 1962.

[29] B. Chow and D. Yang, *Rigidity of non-negatively curved compact quaternionic-Kähler manifolds*, J. Diff. Geo 29 (1989) 361–372.

[30] D. DeTurk and J. Kazdan, *Some regularity theorems in Riemannian geometry*, Ann. scient. Éc. Norm. Sup. 14 (1981) 249–260.

[31] R. Edwards and R. Kirby, *Deformations of spaces of embeddings*, Ann. of Math. 93 (1971), 63–88.

[32] J.-H. Eschenburg, *Local convexity and non-negative curvature- Gromov's proof of the sphere theorem*, Invt. Math. 84 (1986) 507–522.

[33] J.-H. Eschenburg and E. Heintze, *An elementary proof of the Cheeger-Gromoll splitting theorem*, Ann. Glob. Ana. and Geo. 2 (1984) 141–151.

[34] T. Farrell and L. Jones, *Negatively curved manifolds with exotic smooth structures*, J. AMS 2 (1989) 899–908.

[35] K. Fukaya, *Hausdorff convergence of Riemannian manifolds and its applications*, Advanced Studies in Pure Math. 18-I (1990) Recent topics in Differential and Analytic Geometry pp143–238.

[36] S. Gallot, *Isoperimetric inequalities based on integral norms of Ricci curvature*, Astérisque, 157–158, (1988) pp191–216.

[37] S. Gallot, D. Hulin and J. Lafontaine, *Riemannian Geometry*, Berlin-Heidelberg: Springer-Verlag, 1987.

[38] S. Gallot and D. Meyer, *Opérateur de courbure et laplacien des formes différentielles d'une variété riemannianne*, J. Math. Pures. Appl. 54 (1975), 259–284.

[39] G. Gibbons and S. Hawking, *Gravitational multi-instantons*, Phys. Lett. B 78 (1978) 430–432.

[40] D. Gilbarg and N.S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, 2nd edition, Berlin-Heidelberg: Springer-Verlag, 1983.

[41] R. Greene and S.T. Yau, eds., Proc. Symp. Pure Math. 54 vol 3 (1994).

[42] M. Gromov, *Manifolds of negative curvature*, J. Diff. Geo. 12 (1978) 223–230.

[43] M. Gromov, J. Lafontaine and P. Pansu, *structures métriques pour les variétés riemannienness*, Paris: Cedic/Fernand Nathan, 1981.

[44] M. Gromov and W. Thurston, *Pinching constants for hyperbolic manifolds*, Invt. Math. 89 (1987) 1–12.

[45] K. Grove, H. Karcher, and E. Ruh, *Group actions and curvature*, Invt. Math. 23 (1974), 31–48.

[46] K. Grove and P. Petersen, eds., *Comparison Geometry*, MSRI publications vol. 30, New York: Cambridge University Press, 1997.

[47] K. Grove, P. Petersen, and J.-Y. Wu, *Geometric finiteness theorems via controlled topology*, Invt. Math. 99 (1990) 205–213, *Erratum* Invt. Math. 104 (1991) 221–222.

[48] R.S. Hamilton, *The formation of singularities in the Ricci flow*, Surveys in Diff. Geo. vol. 2, International Press (1995) 7–136.

[49] S. Helgason, *Differential Geometry, Lie Groups and Symmetric spaces*, New York-London: Academic Press, 1962.

[50] J. Jost, *Riemannian Geometry and Geometric Analysis*, Berlin-Heidelberg: Springer-Verlag, 1995.

[51] J. Jost and H. Karcher, *Geometrische Methoden zur Gewinnung von a-priori-Schranken für harmonische Abbildungen*, Manu. Math. 19 (1982) 27–77.

[52] S. Kobayashi, *Transformation Groups in Differential Geometry*, Berlin-Heidelberg: Springer-Verlag, 1972.

[53] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry*, vols. I, II, New York: Wiley-Interscience, 1963.

[54] H.B. Lawson Jr. and M.-L. Michelsohn, *Spin Geometry*, Princeton: Princeton University Press, 1989.

[55] J.M. Lee and T.H. Parker, *The Yamabe problem*, Bull. AMS 17 (1987), 37–91.

[56] A. Lichnerowicz, *Géométrie des groupes de transformations*, Paris: Dunod, 1958.

[57] A. Lichnerowicz, *Propagateurs et Commutateurs en relativité générale*, Publ. Math. IHES 10 (1961) 293–343.

[58] M.J. Micallef and J.D. Moore, *Minimal 2-spheres and the topology of manifolds with positive curvature on totally isotropic 2-planes*, Ann. of Math. 127 (1988), 199–227.

[59] R.K. Miller and A.N. Michel, *Ordinary differential equations*, New York-London: Academic Press, 1982.

[60] J.W. Milnor, *Morse Theory*, Princeton: Princeton University Press, 1963.

[61] J.W. Milnor and J.D. Stasheff, *Characteristic Classes*, Princeton: Princeton University Press, 1974.

[62] C.W. Misner, K.S. Thorne and J.A. Wheeler, *Gravitation*, New York: Freeman, 1973.

[63] N. Mok, *The uniformization theorem for compact Kähler manifolds of non-negative holomorphic bi-sectional curvature*, J. Diff. Geo. 27 (1988), 179–214.

[64] J.W. Morgan, *The Seiberg-Witten equations and applications to the topology of smooth four-manifolds*, Princeton: Princeton Univ. Press, 1996.

[65] B. O'Neill, *Semi-Riemannian Geometry*, New York-London: Academic Press, 1983.

[66] Y. Otsu, *On manifolds of positive Ricci curvature with large diameters*, Math. Z. 206 (1991) 255–264.

[67] M. Özaydın and G. Walschap, *Vector bundles with no soul*, PAMS 120 (1994) 565–567.

[68] G. Perel'man, *Alexandrov's spaces with curvatures bounded from below II*, preprint.

[69] S. Peters, *Cheeger's finiteness theorem for diffeomorphism classes of manifolds*, J. Reine Angew. Math. 349 (1984) 77–82.

[70] P. Petersen, S. Shteingold and G. Wei, *Comparison geometry with integral curvature bounds*, to appear in GAFA.

[71] P. Petersen and G. Wei, *Relative volume comparison with integral Ricci curvature bounds*, to appear in GAFA.

[72] G. de Rham, *Differentiable Manifolds*, Berlin-Heidelberg: Springer-Verlag, 1984.

[73] X. Rong, *The almost cyclicity of the fundamental groups of positively curved manifolds*, Invt. Math. 126 (1996) 47–64. And *Positive curvature, local and global symmetry, and fundamental groups*, preprint, Rutgers, New Brunswick.

[74] X. Rong, *A Bochner Theorem and Applications*, preprint. To appear in Duke Math. J.

[75] Y. Shen and S.-h. Zhu, Ricci curvature, minimal surfaces, and sphere theorems. To appear in J. Geo. Anal.

[76] M. Spivak, *A Comprehensive Introduction to Differential Geometry*, vols. I–V, Wilmington: Publish or Perish, 1979.

[77] J. J. Stoker, *Differential Geometry*, New York: Wiley-Interscience, 1989.

[78] S. Tachibana, *A theorem on Riemannian manifolds of positive curvature operator*, Proc. Japan Acad. 50 (1974), 301–302.

[79] M.E. Taylor, *Partial differential equations*, vols. I–III, New York: Springer Verlag, 1996.

[80] T.Y. Thomas, *Riemann spaces of class one and their characterization*, Acta Math. 67 (1936) 169–211.

[81] F. Tricerri and L. Vanhecke, C. R. Acad. Sc. Paris t.302 Series I no. 6 (1986) 233–235.

[82] F.W. Warner, *Foundations of Differentiable Manifolds and Lie Groups*, New York: Springer-Verlag, 1983.

[83] H. Weyl, *The classical groups*, Princeton: Princeton University Press, 1966.

[84] H.H. Wu, *The Bochner Technique in Differential Geometry*, Mathematical Reports vol. 3, part 2, London: Harwood Academic Publishers, 1988.

Index

Graduate Texts in Mathematics

continued from page ii

61 WHITEHEAD. Elements of Homotopy Theory.
62 KARGAPOLOV/MERLZJAKOV. Fundamentals of the Theory of Groups.
63 BOLLOBAS. Graph Theory.
64 EDWARDS. Fourier Series. Vol. I 2nd ed.
65 WELLS. Differential Analysis on Complex Manifolds. 2nd ed.
66 WATERHOUSE. Introduction to Affine Group Schemes.
67 SERRE. Local Fields.
68 WEIDMANN. Linear Operators in Hilbert Spaces.
69 LANG. Cyclotomic Fields II.
70 MASSEY. Singular Homology Theory.
71 FARKAS/KRA. Riemann Surfaces. 2nd ed.
72 STILLWELL. Classical Topology and Combinatorial Group Theory. 2nd ed.
73 HUNGERFORD. Algebra.
74 DAVENPORT. Multiplicative Number Theory. 2nd ed.
75 HOCHSCHILD. Basic Theory of Algebraic Groups and Lie Algebras.
76 IITAKA. Algebraic Geometry.
77 HECKE. Lectures on the Theory of Algebraic Numbers.
78 BURRIS/SANKAPPANAVAR. A Course in Universal Algebra.
79 WALTERS. An Introduction to Ergodic Theory.
80 ROBINSON. A Course in the Theory of Groups. 2nd ed.
81 FORSTER. Lectures on Riemann Surfaces.
82 BOTT/TU. Differential Forms in Algebraic Topology.
83 WASHINGTON. Introduction to Cyclotomic Fields. 2nd ed.
84 IRELAND/ROSEN. A Classical Introduction to Modern Number Theory. 2nd ed.
85 EDWARDS. Fourier Series. Vol. II. 2nd ed.
86 VAN LINT. Introduction to Coding Theory. 2nd ed.
87 BROWN. Cohomology of Groups.
88 PIERCE. Associative Algebras.
89 LANG. Introduction to Algebraic and Abelian Functions. 2nd ed.
90 BRØNDSTED. An Introduction to Convex Polytopes.
91 BEARDON. On the Geometry of Discrete Groups.
92 DIESTEL. Sequences and Series in Banach Spaces.
93 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part I. 2nd ed.
94 WARNER. Foundations of Differentiable Manifolds and Lie Groups.
95 SHIRYAEV. Probability. 2nd ed.
96 CONWAY. A Course in Functional Analysis. 2nd ed.
97 KOBLITZ. Introduction to Elliptic Curves and Modular Forms. 2nd ed.
98 BRÖCKER/TOM DIECK. Representations of Compact Lie Groups.
99 GROVE/BENSON. Finite Reflection Groups. 2nd ed.
100 BERG/CHRISTENSEN/RESSEL. Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions.
101 EDWARDS. Galois Theory.
102 VARADARAJAN. Lie Groups, Lie Algebras and Their Representations.
103 LANG. Complex Analysis. 3rd ed.
104 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part II.
105 LANG. $SL_2(\mathbf{R})$.
106 SILVERMAN. The Arithmetic of Elliptic Curves.
107 OLVER. Applications of Lie Groups to Differential Equations. 2nd ed.
108 RANGE. Holomorphic Functions and Integral Representations in Several Complex Variables.
109 LEHTO. Univalent Functions and Teichmüller Spaces.
110 LANG. Algebraic Number Theory.
111 HUSEMÖLLER. Elliptic Curves.
112 LANG. Elliptic Functions.
113 KARATZAS/SHREVE. Brownian Motion and Stochastic Calculus. 2nd ed.
114 KOBLITZ. A Course in Number Theory and Cryptography. 2nd ed.
115 BERGER/GOSTIAUX. Differential Geometry: Manifolds, Curves, and Surfaces.
116 KELLEY/SRINIVASAN. Measure and Integral. Vol. I.
117 SERRE. Algebraic Groups and Class Fields.
118 PEDERSEN. Analysis Now.

119 ROTMAN. An Introduction to Algebraic Topology.
120 ZIEMER. Weakly Differentiable Functions: Sobolev Spaces and Functions of Bounded Variation.
121 LANG. Cyclotomic Fields I and II. Combined 2nd ed.
122 REMMERT. Theory of Complex Functions. *Readings in Mathematics*
123 EBBINGHAUS/HERMES et al. Numbers. *Readings in Mathematics*
124 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part III.
125 BERENSTEIN/GAY. Complex Variables: An Introduction.
126 BOREL. Linear Algebraic Groups. 2nd ed.
127 MASSEY. A Basic Course in Algebraic Topology.
128 RAUCH. Partial Differential Equations.
129 FULTON/HARRIS. Representation Theory: A First Course. *Readings in Mathematics*
130 DODSON/POSTON. Tensor Geometry.
131 LAM. A First Course in Noncommutative Rings.
132 BEARDON. Iteration of Rational Functions.
133 HARRIS. Algebraic Geometry: A First Course.
134 ROMAN. Coding and Information Theory.
135 ROMAN. Advanced Linear Algebra.
136 ADKINS/WEINTRAUB. Algebra: An Approach via Module Theory.
137 AXLER/BOURDON/RAMEY. Harmonic Function Theory.
138 COHEN. A Course in Computational Algebraic Number Theory.
139 BREDON. Topology and Geometry.
140 AUBIN. Optima and Equilibria. An Introduction to Nonlinear Analysis.
141 BECKER/WEISPFENNING/KREDEL. Gröbner Bases. A Computational Approach to Commutative Algebra.
142 LANG. Real and Functional Analysis. 3rd ed.
143 DOOB. Measure Theory.
144 DENNIS/FARB. Noncommutative Algebra.
145 VICK. Homology Theory. An Introduction to Algebraic Topology. 2nd ed.
146 BRIDGES. Computability: A Mathematical Sketchbook.
147 ROSENBERG. Algebraic *K*-Theory and Its Applications.
148 ROTMAN. An Introduction to the Theory of Groups. 4th ed.
149 RATCLIFFE. Foundations of Hyperbolic Manifolds.
150 EISENBUD. Commutative Algebra with a View Toward Algebraic Geometry.
151 SILVERMAN. Advanced Topics in the Arithmetic of Elliptic Curves.
152 ZIEGLER. Lectures on Polytopes.
153 FULTON. Algebraic Topology: A First Course.
154 BROWN/PEARCY. An Introduction to Analysis.
155 KASSEL. Quantum Groups.
156 KECHRIS. Classical Descriptive Set Theory.
157 MALLIAVIN. Integration and Probability.
158 ROMAN. Field Theory.
159 CONWAY. Functions of One Complex Variable II.
160 LANG. Differential and Riemannian Manifolds.
161 BORWEIN/ERDÉLYI. Polynomials and Polynomial Inequalities.
162 ALPERIN/BELL. Groups and Representations.
163 DIXON/MORTIMER. Permutation Groups.
164 NATHANSON. Additive Number Theory: The Classical Bases.
165 NATHANSON. Additive Number Theory: Inverse Problems and the Geometry of Sumsets.
166 SHARPE. Differential Geometry: Cartan's Generalization of Klein's Erlangen Program.
167 MORANDI. Field and Galois Theory.
168 EWALD. Combinatorial Convexity and Algebraic Geometry.
169 BHATIA. Matrix Analysis.
170 BREDON. Sheaf Theory. 2nd ed.
171 PETERSEN. Riemannian Geometry.
172 REMMERT. Classical Topics in Complex Function Theory.
173 DIESTEL. Graph Theory.
174 BRIDGES. Foundations of Real and Abstract Analysis.
175 LICKORISH. An Introduction to Knot Theory.
176 LEE. Riemannian Manifolds.
177 NEWMAN. Analytic Number Theory.
178 CLARKE/LEDYAEV/STERN/WOLENSKI. Nonsmooth Analysis and Control Theory.