Graduate Texts in Mathematics

GTM

Peter Petersen

Riemannian Geometry

Third Edition

Springer

Graduate Texts in Mathematics 171

Graduate Texts in Mathematics

Graduate Texts in Mathematics bridge the gap between passive study and creative understanding, offering graduate-level introductions to advanced topics in mathematics. The volumes are carefully written as teaching aids and highlight characteristic features of the theory. Although these books are frequently used as textbooks in graduate courses, they are also suitable for individual study.

More information about this series at http://www.springer.com/series/136

Peter Petersen

Riemannian Geometry

Third Edition

Springer

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

ISSN 0072-5285 ISSN 2197-5612 (electronic)
Graduate Texts in Mathematics
ISBN 978-3-319-26652-7 ISBN 978-3-319-26654-1 (eBook)
DOI 10.1007/978-3-319-26654-1

Library of Congress Control Number: 2015960754

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

To my wife, Laura

Preface

This book is intended as a comprehensive introduction to Riemannian geometry. The reader is assumed to have basic knowledge of standard manifold theory, including the theory of tensors, forms, and Lie groups. At times it is also necessary to have some familiarity with algebraic topology and de Rham cohomology. Specifically, we recommend that the reader be familiar with texts such as [15, 72] or [97, vol. 1]. On my web page, there are links to lecture notes on these topics as well as classical differential geometry (see [90] and [89]). It is also helpful if the reader has a nodding acquaintance with ordinary differential equations. For this, a text such as [74] is more than sufficient. More basic prerequisites are real analysis, linear algebra, and some abstract algebra. Differential geometry is and always has been an "applied discipline" within mathematics that uses many other parts of mathematics for its own purposes.

Most of the material generally taught in basic Riemannian geometry as well as several more advanced topics is presented in this text. The approach we have taken occasionally deviates from the standard path. Alongside the usual variational approach, we have also developed a more function-oriented methodology that likewise uses standard calculus together with techniques from differential equations. Our motivation for this treatment has been that examples become a natural and integral part of the text rather than a separate item that is sometimes minimized. Another desirable by-product has been that one actually gets the feeling that Hessians and Laplacians are intimately related to curvatures.

The book is divided into four parts:

Part I: Tensor geometry, consisting of chapters 1, 2, 3, and 4

Part II: Geodesic and distance geometry, consisting of chapters 5, 6, and 7

Part III: Geometry à la Bochner and Cartan, consisting of chapters 8, 9, and 10

Part IV: Comparison geometry, consisting of chapters 11 and 12

There are significant structural changes and enhancements in the third edition, so chapters no longer correspond to those of the first two editions. We offer a brief outline of each chapter below.

Chapter 1 introduces Riemannian manifolds, isometries, immersions, and submersions. Homogeneous spaces and covering maps are also briefly mentioned.

There is a discussion on various types of warped products. This allows us to give both analytic and geometric definitions of the basic constant curvature geometries. The Hopf fibration as a Riemannian submersion is also discussed in several places. Finally, there is a section on tensor notation.

Chapter 2 discusses both Lie and covariant derivatives and how they can be used to define several basic concepts such as the classical notions of Hessian, Laplacian, and divergence on Riemannian manifolds. Iterated derivatives and abstract derivations are discussed toward the end and used later in the text.

Chapter 3 develops all of the important curvature concepts and discusses a few simple properties. We also develop several important formulas that relate curvature and the underlying metric. These formulas can be used in many places as a replacement for the second variation formula.

Chapter 4 is devoted to calculating curvatures in several concrete situations such as spheres, product spheres, warped products, and doubly warped products. This is used to exhibit several interesting examples. In particular, we explain how the Riemannian analogue of the Schwarzschild metric can be constructed. There is a new section that explains warped products in general and how they are characterized. This is an important section for later developments as it leads to an interesting characterization of both local and global constant curvature geometries from both the warped product and conformal view point. We have a section on Lie groups. Here two important examples of left invariant metrics are discussed as well as the general formulas for the curvatures of biinvariant metrics. It is also explained how submersions can be used to create new examples with special focus on complex projective space. There are also some general comments on how submersions can be constructed using isometric group actions.

Chapter 5 further develops the foundational topics for Riemannian manifolds. These include the first variation formula, geodesics, Riemannian manifolds as metric spaces, exponential maps, geodesic completeness versus metric completeness, and maximal domains on which the exponential map is an embedding. The chapter includes a detailed discussion of the properties of isometries. This naturally leads to the classification of simply connected space forms. At a more basic level, we obtain metric characterizations of Riemannian isometries and submersions. These are used to show that the isometry group is a Lie group and to give a proof of the slice theorem for isometric group actions.

Chapter 6 contains three more foundational topics: parallel translation, Jacobi fields, and the second variation formula. Some of the classical results we prove here are the Hadamard-Cartan theorem, Cartan's center of mass construction in nonpositive curvature and why it shows that the fundamental group of such spaces is torsion-free, Preissman's theorem, Bonnet's diameter estimate, and Synge's lemma. At the end of the chapter, we cover the ingredients needed for the classical quarter pinched sphere theorem including Klingenberg's injectivity radius estimates and Berger's proof of this theorem. Sphere theorems are revisited in chapter 12.

Chapter 7 focuses on manifolds with lower Ricci curvature bounds. We discuss volume comparison and its uses. These include proofs of how Poincaré and Sobolev constants can be bounded and theorems about restrictions on fundamental groups for manifolds with lower Ricci curvature bounds. The strong maximum principle for continuous functions is developed. This result is first used in a warm-up exercise to prove Cheng's maximal diameter theorem. We then proceed to cover the Cheeger-Gromoll splitting theorem and its consequences for manifolds with nonnegative Ricci curvature.

Chapter 8 covers various aspects of symmetries on manifolds with emphasis on Killing fields. Here there is a further discussion on why the isometry group is a Lie group. The Bochner formulas for Killing fields are covered as well as a discussion on how the presence of Killing fields in positive sectional curvature can lead to topological restrictions. The latter is a fairly new area in Riemannian geometry.

Chapter 9 explains both the classical and more recent results that arise from the Bochner technique. We start with harmonic 1-forms as Bochner did and move on to general forms and other tensors such as the curvature tensor. We use an approach that considerably simplifies many of the tensor calculations in this subject (see, e.g., the first and second editions of this book). The idea is to consistently use how derivations act on tensors instead of using Clifford representations. The Bochner technique gives many optimal 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. More generally, we also show how the Bochner technique can be used to control the topology with more general curvature bounds. This requires a little more analysis, but is a fascinating approach that has not been presented in book form yet.

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 philosophy.

Chapter 10 develops part of the theory of symmetric spaces and holonomy. The standard representations of symmetric spaces as homogeneous spaces or via Lie algebras are explained. There are several concrete calculations both specific and more general examples to get a feel for how curvatures behave. Having done this, we define holonomy for general manifolds and discuss the de Rham decomposition theorem and several corollaries of it. In particular, we show that holonomy irreducible symmetric spaces are Einstein and that their curvatures have the same sign as the Einstein constant. This theorem and the examples are used to indicate how one can classify symmetric spaces. Finally, we present a brief overview of how holonomy and symmetric spaces are related to the classification of holonomy groups. This is used, together with most of what has been learned up to this point, to give the Gallot and Meyer classification of compact manifolds with nonnegative curvature operator.

Chapter 11 focuses on the convergence theory of metric spaces and manifolds. First, we introduce the most general 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 here is to figure out what weak convergence implies in the presence of stronger side conditions. There is a section with a quick overview of 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. The main focus of the chapter is to prove the Cheeger-Gromov convergence theorem, which is called the Convergence Theorem of Riemannian Geometry, as well as Anderson's generalizations of this theorem to manifolds with bounded Ricci curvature.

Chapter 12 proves some of the more general finiteness theorems that do not fall into the philosophy developed in Chapter 11. To begin, we discuss generalized critical point theory and Toponogov's theorem. These two techniques are used throughout the chapter to establish 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, and Klingenberg. Critical point theory was introduced by Grove and Shiohama to prove the diameter sphere theorem. Following 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.

At the end of most chapters, there is a short 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 strongly urged to go from those sources back to the original papers as ideas are often lost in the modernization of most subjects. 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 [12]. Other valuable texts that expand or complement much of the material covered here are [77, 97] and [99]. There is also a historical survey by Berger (see [11]) that complements this text very well.

Each chapter ends with a collection of exercises that are designed to reinforce the material covered, to establish some simple results that will be needed later, and also to offer alternative proofs of several results. The first six chapters have about 30 exercises each and there are 300+ in all. The reader should at least read and think about all of the exercises, if not actually solve all of them. There are several exercises that might be considered very challenging. These have been broken up into more reasonable steps and with occasional hints. Some instructors might want to cover some of the exercises in class.

A first course should definitely cover Chapters 3, 5, and 6 together with whatever one feels is necessary from Chapters 1, 2, and 4. I would definitely not recommend teaching every single topic covered in Chapters 1, 2, and 4. A more advanced course could consist of going through Chapter 7 and parts III or IV as defined earlier. These two parts do not depend in a serious way on each other. One can probably not cover the entire book in two semesters, but it should be possible to cover parts I, II, and III or alternatively I, II, and IV depending on one's inclination.

There are many people I would like to thank. First and foremost are those students who suffered through my continuing pedagogical experiments over the last 25 years. While using this text I always try different strategies every time I teach. Special thanks go to Victor Alvarez, Igor Belegradek, Marcel Berger, Timothy Carson, Gil Cavalcanti, Edward Fan, Hao Fang, John Garnett, or Hershkovits, Ilkka Holopainen, Michael Jablonski, Lee Kennard, Mayer Amitai Landau, Peter Landweber, Pablo Lessa, Ciprian Manolescu, Geoffrey Mess, Jiayin Pan, Priyanka Rajan, Jacob Rooney, Yanir Rubinstein, Semion Shteingold, Jake Solomon, Chad Sprouse, Marc Troyanov, Gerard Walschap, Nik Weaver, Burkhard Wilking, Michael Williams, and Hung-Hsi Wu for their constructive criticism of parts of the book and mentioning various typos and other deficiencies in the first and second editions. I would especially like to thank Joseph Borzellino for his very careful reading of this text. Finally, I would like to thank Robert Greene, Karsten Grove, Gregory Kallo, and Fred Wilhelm for all the discussions on geometry we have had over the years.

Los Angeles, CA, USA Peter Petersen

Contents

Chapter 1
Riemannian Metrics

In this chapter we introduce the spaces and maps that pervade the subject. Without discussing any theory we present several examples of basic Riemannian manifolds and Riemannian maps. All of these examples will be at the heart of 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 as it wasn't really until Whitney's work in 1936 that mathematicians obtained a clear understanding of what abstract manifolds were other than just being submanifolds of Euclidean space. Riemann himself defined Riemannian metrics only on domains in Euclidean space. Riemannian manifolds where then metric objects that locally looked like a Riemannian metric on a domain in Euclidean space. It is, however, important to realize that this local approach to a global theory of Riemannian manifolds is as honest as the modern top-down approach.

Prior to Riemann, other famous mathematicians such as Euler, Monge, and Gauss only worked with 2-dimensional curved geometry. Riemann's invention of multi-dimensional geometry is quite curious. The story goes that Gauss was on Riemann's defense committee for his Habilitation (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." Evidently, he was hoping that the committee would select from the first two topics, which were on material he had already developed. 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. Thus, 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, apparently a very rare occurrence for him.

From Riemann's work it appears that he worked with changing metrics mostly by multiplying them by a function (conformal change). By conformally changing

P. Petersen, *Riemannian Geometry*, Graduate Texts in Mathematics 171,
DOI 10.1007/978-3-319-26654-1_1

the standard Euclidean metric 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 exhibits all constant curvature geometries in a unified way. Of course, Gauss already knew about polar coordinate representations on surfaces, and rotationally symmetric metrics were studied even earlier by Clairaut. But those examples are much simpler than the higher-dimensional analogues. Throughout this book we 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 examples of Riemannian metrics is by using left-invariant metrics on Lie groups. This leads us, among other things, to the Hopf fibration and Berger spheres. Both of these are of fundamental importance and are also at the core of a large number of examples in Riemannian geometry. These will also be defined here and studied further throughout the book.

1.1 Riemannian Manifolds and Maps

A *Riemannian manifold* (M, g) consists of a C^∞-manifold M (Hausdorff and second countable) and a Euclidean inner product g_p or $g|_p$ on each of the tangent spaces T_pM of M. In addition we assume that $p \mapsto g_p$ varies smoothly. This means that for any two smooth vector fields X, Y the inner product $g_p\left(X|_p, Y|_p\right)$ is a smooth function of p. The subscript p will usually be suppressed when it is not needed. Thus we might write $g(X, Y)$ with the understanding that this is to be evaluated at each p where X and Y are defined. When we wish to associate the metric with M we also denote it as g_M. The tensor g is referred to as the *Riemannian metric* or simply the *metric*. Generally speaking the manifold is assumed to be connected. Exceptions do occur, especially when studying level sets or submanifolds defined by constraints.

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 the n-dimensional Euclidean space $\mathbb{R}^n$ with its canonical inner product. Hence, all Riemannian manifolds have the same infinitesimal structure not only as manifolds but also as Riemannian manifolds.

Example 1.1.1. The simplest and most fundamental Riemannian manifold is Euclidean space $(\mathbb{R}^n, g_{\mathbb{R}^n})$. The canonical Riemannian structure $g_{\mathbb{R}^n}$ is defined by the tangent bundle identification $\mathbb{R}^n \times \mathbb{R}^n \simeq T\mathbb{R}^n$ given by the map:

$$(p, v) \mapsto \frac{d\,(p + tv)}{dt}(0)\,.$$

With this in mind the standard inner product on $\mathbb{R}^n$ is defined by

$$g_{\mathbb{R}^n}\left((p,v),(p,w)\right) = v \cdot w.$$

A *Riemannian isometry* between Riemannian manifolds (M, g_M) and (N, g_N) is a diffeomorphism $F : M \to N$ such that $F^* g_N = g_M$, i.e.,

$$g_N\left(DF(v), DF(w)\right) = g_M(v, w)$$

for all tangent vectors $v, w \in T_pM$ and all $p \in M$. In this case F^{-1} is also a Riemannian isometry.

Example 1.1.2. Any finite-dimensional vector space V with an inner product, becomes a Riemannian manifold by declaring, as with Euclidean space, that

$$g\left((p,v),(p,w)\right) = v \cdot w.$$

If we have two such Riemannian manifolds (V, g_V) and (W, g_W) of the same dimension, then they are isometric. A example of a Riemannian isometry $F : V \to W$ is simply any linear isometry between the two spaces. Thus $(\mathbb{R}^n, g_{\mathbb{R}^n})$ 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) $F : M \to N$, where (N, g_N) is a Riemannian manifold. This leads to a pull-back Riemannian metric $g_M = F^* g_N$ on M, where

$$g_M(v, w) = g_N\left(DF(v), DF(w)\right).$$

It is an inner product as $DF(v) = 0$ only when $v = 0$.

A *Riemannian immersion* (or *Riemannian embedding*) is an immersion (or embedding) $F : M \to N$ such that $g_M = F^* g_N$. Riemannian immersions are also called *isometric immersions* , but as we shall see below they are almost never distance preserving.

Example 1.1.3. Another very important example is the Euclidean sphere of radius R defined by

$$S^n(R) = \left\{x \in \mathbb{R}^{n+1} \mid |x| = R\right\}.$$

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 a round sphere in $\mathbb{R}^3$.

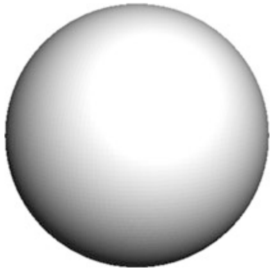


Fig. 1.1 Sphere

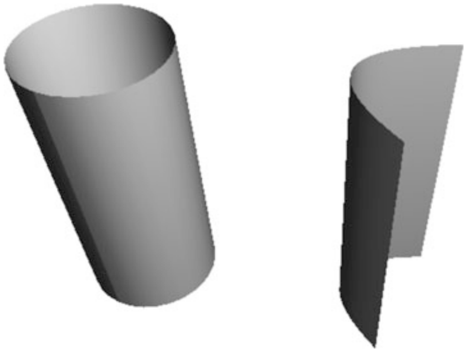


Fig. 1.2 Isometric Immersions

If $k < n$ there are several linear isometric immersions $\left(\mathbb{R}^k, g_{\mathbb{R}^k}\right) \to (\mathbb{R}^n, g_{\mathbb{R}^n})$. Those are, however, not the only isometric immersions. In fact, any unit speed curve $c : \mathbb{R} \to \mathbb{R}^2$, i.e., $|\dot{c}(t)| = 1$ for all $t \in \mathbb{R}$, is an example of an isometric immersion. For example, one could consider

$$t \mapsto (\cos t, \sin t)$$

as an isometric immersion and

$$t \mapsto \left(\log\left(t + \sqrt{1+t^2}\right), \sqrt{1+t^2}\right)$$

as an isometric embedding. A map of the form:

$$F : \mathbb{R}^k \to \mathbb{R}^{k+1}$$
$$F(x^1, \dots, x^k) = (c(x^1), x^2, \dots, x^k),$$

(where c fills up the first two coordinate entries) will then also yield an isometric immersion (or embedding) that is not linear. This initially seems contrary to intuition 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 also a dual concept of a *Riemannian submersion* $F:(M,g_M)\to(N,g_N)$. This is a submersion $F:M\to N$ such that for each $p\in M$, $DF:\ker(DF)^{\perp}\to T_{F(p)}N$ is a linear isometry. In other words, if $v,w\in T_pM$ are perpendicular to the kernel of $DF:T_pM\to T_{F(p)}N$, then

$$g_M(v,w)=g_N\left(DF\left(v\right),DF\left(w\right)\right).$$

This is equivalent to the adjoint $\left(DF_p\right)^*:T_{F(p)}N\to T_pM$ preserving inner products of vectors.

Example 1.1.4. Orthogonal projections $(\mathbb{R}^n,g_{\mathbb{R}^n})\to\left(\mathbb{R}^k,g_{\mathbb{R}^k}\right)$, where $k<n$, are examples of Riemannian submersions.

Example 1.1.5. A much less trivial example is the *Hopf fibration* $S^3(1)\to S^2(1/2)$. As observed by F. Wilhelm this map can be written explicitly as

$$H(z,w)=\left(\frac{1}{2}\left(|w|^2-|z|^2\right),z\bar{w}\right)$$

if we think of $S^3(1)\subset\mathbb{C}^2$ and $S^2(1/2)\subset\mathbb{R}\oplus\mathbb{C}$. Note that the fiber containing (z,w) consists of the points $\left(e^{\mathrm{i}\theta}z,e^{\mathrm{i}\theta}w\right)$, where $\mathrm{i}=\sqrt{-1}$. Consequently, $\mathrm{i}\,(z,w)$ is tangent to the fiber and $\lambda\,(-\bar{w},\bar{z})\,,\ \lambda\in\mathbb{C}$, are the tangent vectors orthogonal to the fiber. We can check what happens to the latter tangent vectors by computing DH. Since H extends to a map $H:\mathbb{C}^2\to\mathbb{R}\oplus\mathbb{C}$ its differential can be calculated as one would do it in multivariable calculus. Alternately note that the tangent vectors $\lambda\,(-\bar{w},\bar{z})$ at $(z,w)\in S^3(1)$ lie in the plane $(z,w)+\lambda\,(-\bar{w},\bar{z})$ parameterized by λ. H restricted to this plane is given by

$$H\left((z-\lambda\bar{w},w+\lambda\bar{z})\right)=\left(\frac{1}{2}\left(|w+\lambda\bar{z}|^2-|z-\lambda\bar{w}|^2\right),(z-\lambda\bar{w})\,\overline{(w+\lambda\bar{z})}\right).$$

To calculate DH we simply expand H in terms of λ and $\bar{\lambda}$ and isolate the first-order terms

$$DH|_{(z,w)}\left(\lambda\,(-\bar{w},\bar{z})\right)=\left(2\mathrm{Re}\left(\bar{\lambda}zw\right),-\lambda\bar{w}^2+\bar{\lambda}z^2\right).$$

Since these have the same length $|\lambda|$ as $\lambda\,(-\bar{w},\bar{z})$ we have shown that the map is a Riemannian submersion. Below we will examine this example more closely. There is a quaternion generalization of this map in exercise 1.6.22.

Finally, we mention a very important generalization of Riemannian manifolds. A *semi-* or *pseudo-Riemannian* manifold consists of a manifold and a smoothly varying symmetric bilinear form g on each tangent space. We assume in addition that g is nondegenerate, i.e., for each nonzero $v\in T_pM$ there exists $w\in T_pM$ such that $g\,(v,w)\neq 0$. This is clearly a generalization of a Riemannian metric where nondegeneracy follows from $g\,(v,v)>0$ when $v\neq 0$. Each tangent space admits a

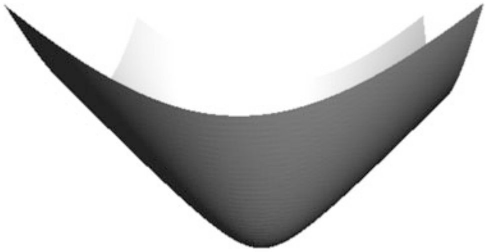


Fig. 1.3 Hyperbolic Space

splitting $T_pM = P \oplus N$ such that g is positive definite on P and negative definite on N. These subspaces are not unique but it is easy to show that their dimensions are well-defined. Continuity of g shows that nearby tangent spaces must have a similar splitting where the subspaces have the same dimension. The *index* of a connected pseudo-Riemannian manifold is defined as the dimension of the subspace N on which g is negative definite.

Example 1.1.6. Let $n = n_1 + n_2$ and $\mathbb{R}^{n_1,n_2} = \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$. We can then write vectors in $\mathbb{R}^{n_1,n_2}$ as $v = v_1 + v_2$, where $v_1 \in \mathbb{R}^{n_1}$ and $v_2 \in \mathbb{R}^{n_2}$. A natural pseudo-Riemannian metric of index n_2 is defined by

$$g\left((p,v),(p,w)\right) = v_1 \cdot w_1 - v_2 \cdot w_2.$$

When $n_1 = 1$ or $n_2 = 1$ this coincides with one or the other version of Minkowski space. This space describes the geometry of Einstein's space-time in special relativity.

Example 1.1.7. We define the family of *hyperbolic spaces* $H^n(R) \subset \mathbb{R}^{n,1}$ using the rotationally symmetric hyperboloids

$$\left(x^1\right)^2 + \cdots + (x^n)^2 - \left(x^{n+1}\right)^2 = -R^2.$$

Each of these level sets consists of two components that are each properly embedded copies of $\mathbb{R}^n$ in $\mathbb{R}^{n+1}$. The branch with $x^{n+1} > 0$ is $H^n(R)$ (see figure 1.3). The metric is the induced Minkowski metric from $\mathbb{R}^{n,1}$. The fact that this defines a Riemannian metric on $H^n(R)$ is perhaps not immediately obvious. Note first that tangent vectors $v = \left(v^1, \cdots, v^n, v^{n+1}\right) \in T_pH^n(R), p \in H^n(R)$, satisfy the equation

$$v^1p^1 + \cdots + v^np^n - v^{n+1}p^{n+1} = 0$$

as they are tangent to the level sets for $\left(x^1\right)^2 + \cdots + (x^n)^2 - \left(x^{n+1}\right)^2$. This shows that

$$\begin{aligned} |v|^2 &= \left(v^1\right)^2 + \cdots + (v^n)^2 - \left(v^{n+1}\right)^2 \\ &= \left(v^1\right)^2 + \cdots + (v^n)^2 - \left(\frac{v^1p^1 + \cdots + v^np^n}{p^{n+1}}\right)^2. \end{aligned}$$

Using Cauchy-Schwarz on the expression in the numerator together with

$$\frac{\left(p^1\right)^2 + \cdots + (p^n)^2}{\left(p^{n+1}\right)^2} = 1 - \left(\frac{R}{p^{n+1}}\right)^2$$

shows that

$$|v|^2 \geq \left(\frac{R}{p^{n+1}}\right)^2 \left(\left(v^1\right)^2 + \cdots + (v^n)^2\right).$$

When $R = 1$ we generally just write H^n and refer to this as *hyperbolic n-space*.

Much of the tensor analysis that we shall develop on Riemannian manifolds can be carried over to pseudo-Riemannian manifolds without further ado. It is only when we start using norm and distances that we have to be more careful.

1.2 The Volume Form

In Euclidean space the inner product not only allows us to calculate norms and angles but also areas, volumes, and more. The key to understanding these definitions better lies in using determinants.

To compute the volume of the parallelepiped spanned by n vectors $v_1, \ldots, v_n \in \mathbb{R}^n$ we can proceed in different ways. There is the usual inductive way where we multiply the height by the volume (or area) of the base parallelepiped. This is in fact a Laplace expansion of a determinant along a column. If the canonical basis is denoted $e_1, \ldots, e_n$, then we define the *signed volume* by

$$\begin{aligned} \operatorname{vol}(v_1, \ldots, v_n) &= \det\left[g\left(v_i, e_j\right)\right] \\ &= \det\left([v_1, \ldots, v_n]\,[e_1, \ldots, e_n]^t\right) \\ &= \det[v_1, \ldots, v_n]\,. \end{aligned}$$

This formula is clearly also valid if we had selected any other positively oriented orthonormal basis $f_1, \ldots f_n$ as

$$\begin{aligned} \det\left[g\left(v_i, f_j\right)\right] &= \det\left([v_1, \ldots, v_n]\,[f_1, \ldots, f_n]^t\right) \\ &= \det\left([v_1, \ldots, v_n]\,[f_1, \ldots, f_n]^t\right) \det\left([f_1, \ldots, f_n]\,[e_1, \ldots, e_n]^t\right) \\ &= \det\left([v_1, \ldots, v_n]\,[e_1, \ldots, e_n]^t\right). \end{aligned}$$

In an oriented Riemannian n-manifold (M, g) we can then define the volume form as an n-form on M by

$$\mathrm{vol}_g(v_1, \ldots, v_n) = \mathrm{vol}(v_1, \ldots, v_n) = \det\left[g\left(v_i, e_j\right)\right],$$

where $e_1, \ldots, e_n$ is any positively oriented orthonormal basis. One often also uses the notation $d\,\mathrm{vol}$ instead of vol, however, the volume form is not necessarily exact so the notation can be a little misleading.

Even though manifolds are not necessarily oriented or even orientable it is still possible to define this volume form locally. The easiest way of doing so is to locally select an *orthonormal frame* $E_1, \ldots, E_n$ and declare it to be positive. A *frame* is a collection of vector fields defined on a common domain $U \subset M$ such that they form a basis for the tangent spaces T_pM for all $p \in U$. The volume form is then defined on vectors and vector fields by

$$\mathrm{vol}(X_1, \ldots, X_n) = \det\left[g\left(X_i, E_j\right)\right].$$

This formula quickly establishes the simplest version of the "height×base" principle if we replace E_i by a general vector X since

$$\mathrm{vol}(E_1, \ldots, X, \ldots, E_n) = g(X, E_i)$$

is the projection of X onto E_i and this describes the height in the ith coordinate direction.

On oriented manifolds it is possible to integrate n-forms. On oriented Riemannian manifolds we can then integrate functions f by integrating the form $f \cdot \mathrm{vol}$. In fact any manifold contains an open dense set $O \subset M$ where $TO = O \times \mathbb{R}^n$ is trivial. In particular, O is orientable and we can choose an orthonormal frame on all of O. This shows that we can integrate functions over M by integrating them over O. Thus we can integrate on all Riemannian manifolds.

1.3 Groups and Riemannian Manifolds

We shall study groups of Riemannian isometries on Riemannian manifolds and see how they can be used to construct new Riemannian manifolds.

1.3.1 Isometry Groups

For a Riemannian manifold (M, g) we use $\mathrm{Iso}(M, g)$ or $\mathrm{Iso}(M)$ to denote the group of Riemannian isometries $F : (M, g) \to (M, g)$ and $\mathrm{Iso}_p(M, g)$ the *isotropy* or *stabilizer (sub)group* at p, i.e., those $F \in \mathrm{Iso}(M, g)$ with $F(p) = p$. 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 an $F \in \mathrm{Iso}(M, g)$ such that $F(p) = q$.

Example 1.3.1. The isometry group of Euclidean space is given by

$$\begin{aligned}\mathrm{Iso}\left(\mathbb{R}^n, g_{\mathbb{R}^n}\right) &= \mathbb{R}^n \rtimes \mathrm{O}\left(n\right)\\ &= \left\{F : \mathbb{R}^n \to \mathbb{R}^n \mid F(x) = v + Ox,\ v \in \mathbb{R}^n \text{ and } O \in \mathrm{O}(n)\right\}.\end{aligned}$$

(Here H $\rtimes$ G is the semi direct product, with G acting on H.) The translational part v and rotational part O are uniquely determined. It is clear that these maps are isometries. To see the converse first observe that $G(x) = F(x) - F(0)$ is also a Riemannian isometry. Using this, we observe that at $x = 0$ the differential $DG_0 \in \mathrm{O}\left(n\right)$. Thus, G and DG_0 are Riemannian isometries on Euclidean space that both preserve the origin and have the same differential there. It is then a general uniqueness result for Riemannian isometries that $G = DG_0$ (see proposition 5.6.2). In exercise 2.5.12 there is a more elementary version for Euclidean space.

The isotropy Iso_p is always isomorphic to $\mathrm{O}(n)$ and $\mathbb{R}^n \simeq \mathrm{Iso}/\mathrm{Iso}_p$ for any $p \in \mathbb{R}^n$. In fact any homogenous space can always be written as the quotient $M = \mathrm{Iso}/\mathrm{Iso}_p$.

Example 1.3.2. We claim that spheres have

$$\mathrm{Iso}\left(S^n(R), g_{S^n(R)}\right) = \mathrm{O}(n+1) = \mathrm{Iso}_0\left(\mathbb{R}^{n+1}, g_{\mathbb{R}^{n+1}}\right).$$

Clearly $\mathrm{O}(n+1) \subset \mathrm{Iso}\left(S^n(R), g_{S^n(R)}\right)$. Conversely, when $F \in \mathrm{Iso}\left(S^n(R), g_{S^n(R)}\right)$, consider the linear map given by the $n+1$ columns vectors:

$$O = \left[\tfrac{1}{R}F\left(Re_1\right)\ DF|_{e_1}\left(e_2\right) \cdots DF|_{e_1}\left(e_{n+1}\right)\right]$$

The first vector is unit since $F\left(Re_1\right) \in S^n\left(R\right)$. Moreover, the first column is orthogonal to the others as $DF|_{Re_1}\left(e_i\right) \in T_{F(Re_1)}S^n\left(R\right) = F\left(Re_1\right)^{\perp}, i = 2, \ldots, n+1$. Finally, the last n columns form an orthonormal basis since DF is assumed to be a linear isometry. This shows that $O \in \mathrm{O}\left(n+1\right)$ and that O agrees with F and DF at Re_1. Proposition 5.6.2 can then be invoked again to show that $F = O$.

The isotropy groups are again isomorphic to $\mathrm{O}(n)$, that is, those elements of $\mathrm{O}(n+1)$ fixing a 1-dimensional linear subspace of $\mathbb{R}^{n+1}$. In particular, we have $S^n \simeq \mathrm{O}\left(n+1\right)/\mathrm{O}\left(n\right)$.

Example 1.3.3. Recall our definition of the hyperbolic spaces from example 1.1.7. The isometry group $\mathrm{Iso}(H^n(R))$ comes from the linear isometries of $\mathbb{R}^{n,1}$

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

One can, as in the case of the sphere, see that these are isometries on $H^n(R)$ as long as they preserve the condition $x^{n+1} > 0$. The group of those isometries is denoted $\mathrm{O}^+\left(n,1\right)$. As in the case of Euclidean space and the spheres we can construct an element in $\mathrm{O}^+\left(n,1\right)$ that agrees with any isometry at Re_{n+1} and such that their differentials at that point agree on the basis $e_1, \ldots, e_n$ for $T_{Re_{n+1}}H^n(R)$. Specifically, if $F \in \mathrm{Iso}(H^n(R))$ we can use:

$$O = \left[DF|_{e_{n+1}}(e_1)\; DF|_{e_{n+1}}(e_2) \cdots DF|_{e_{n+1}}(e_n)\; \tfrac{1}{R}F(Re_{n+1}) \right].$$

The isotropy group that preserves Re_{n+1} can be identified with $\mathrm{O}(n)$ (isometries we get from the metric being rotationally symmetric). One can also easily check that $\mathrm{O}^+(n,1)$ acts transitively on $H^n(R)$.

1.3.2 Lie Groups

If instead we start with a Lie group G, then it is possible to make it a group of isometries in several ways. The tangent space can be trivialized

$$T\mathrm{G} \simeq \mathrm{G} \times T_e\mathrm{G}$$

by using left- (or right-) translations on G. Therefore, any inner product on $T_e\mathrm{G}$ 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 where all left-translations are Riemannian isometries is of this form. In contrast to $\mathbb{R}^n$, not all of these Riemannian metrics need be isometric to each other. Thus a Lie group might not come with a canonical metric.

It can be shown that the left coset space $\mathrm{G}/\mathrm{H} = \{g\mathrm{H} \mid g \in \mathrm{G}\}$ is a manifold provided $\mathrm{H} \subset \mathrm{G}$ is a compact subgroup. If we endow G with a general Riemannian metric such that *right-translations* by elements in H act by isometries, then there is a unique Riemannian metric on G/H making the projection $\mathrm{G} \to \mathrm{G}/\mathrm{H}$ into a Riemannian submersion (see also section 4.5.2). When in addition the metric is also left-invariant, then G acts by isometries on G/H (on the left) thus making G/H into a homogeneous space. Proofs of all this are given in theorem 5.6.21 and remark 5.6.22.

The next two examples will be studied further in sections 1.4.6, 4.4.3, and 4.5.3. In sections 4.5.2 the general set-up is discussed and the fact that quotients are Riemannian manifolds is also discussed in section 5.6.4 and theorem 5.6.21.

Example 1.3.4. The idea of taking the quotient of a Lie group by a subgroup can be generalized. Consider $S^{2n+1}(1) \subset \mathbb{C}^{n+1}$. Then $S^1 = \{\lambda \in \mathbb{C} \mid |\lambda| = 1\}$ acts by complex scalar multiplication on both S^{2n+1} and $\mathbb{C}^{n+1}$; furthermore, this action is by isometries. We know that the quotient $S^{2n+1}/S^1 = \mathbb{CP}^n$, and since the action of S^1 is by isometries, we obtain a metric on $\mathbb{CP}^n$ such that $S^{2n+1} \to \mathbb{CP}^n$ is a Riemannian submersion. This metric is called the Fubini-Study metric. When $n = 1$, this becomes the Hopf fibration $S^3(1) \to \mathbb{CP}^1 = S^2(1/2)$.

Example 1.3.5. One of the most important nontrivial Lie groups is $\mathrm{SU}(2)$, which is defined as

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

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

$$\mathfrak{su}(2) = \left\{\begin{bmatrix} \mathrm{i}\alpha & \beta + \mathrm{i}c \\ -\beta + \mathrm{i}c & -\mathrm{i}\alpha\end{bmatrix} \mid \alpha, \beta, c \in \mathbb{R}\right\}$$

and can be spanned by

$$X_1 = \begin{bmatrix} \mathrm{i} & 0 \\ 0 & -\mathrm{i}\end{bmatrix}, \ X_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0\end{bmatrix}, \ X_3 = \begin{bmatrix} 0 & \mathrm{i} \\ \mathrm{i} & 0\end{bmatrix}.$$

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 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 $\mathrm{SU}(2) = S^3$. 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

$$\begin{bmatrix} e^{\mathrm{i}\theta} & 0 \\ 0 & e^{-\mathrm{i}\theta}\end{bmatrix}$$

since

$$\begin{bmatrix} e^{\mathrm{i}\theta} & 0 \\ 0 & e^{-\mathrm{i}\theta}\end{bmatrix}\begin{bmatrix} z & w \\ -\bar{w} & \bar{z}\end{bmatrix} = \begin{bmatrix} e^{\mathrm{i}\theta}z & e^{\mathrm{i}\theta}w \\ -e^{-\mathrm{i}\theta}\bar{w} & e^{-\mathrm{i}\theta}\bar{z}\end{bmatrix}.$$

Thus X_1 is tangent to the orbits of the Hopf circle action. The Berger spheres are then obtained from the canonical metric by multiplying the metric along the Hopf fiber by ε^2.

1.3.3 Covering Maps

Discrete groups are also common in geometry, often through *deck transformations* or *covering transformations*. Suppose that $F : M \to N$ is a covering map. Then F is, in particular, both an immersion and a submersion. Thus, any Riemannian metric on N induces a Riemannian metric on M. This makes F into an isometric immersion, also called a *Riemannian covering*. Since $\dim M = \dim N$, F 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 $F|_U : U \to F(U)$ 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 $\{U_p\}_{p \in F^{-1}(q)}$, then all the sets V and U_p are isometric to each other. In fact, if F is a normal covering, i.e., there is a group Γ of deck transformations acting on M such that:

$$F^{-1}(p) = \{g(q) \mid F(q) = p \text{ and } g \in \Gamma\},$$

then Γ acts by isometries on the pullback metric. This construction can easily be reversed. 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 a local isometry.

Example 1.3.6. If we fix a basis v_1, v_2 for $\mathbb{R}^2$, then $\mathbb{Z}^2$ acts by isometries through the translations

$$(n, m) \mapsto (x \mapsto x + nv_1 + mv_2).$$

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. These metrics will depend on $|v_1|$, $|v_2|$ and $\angle(v_1, v_2)$, so they need not be isometric to each other.

Example 1.3.7. The involution $-I$ on $S^n(1) \subset \mathbb{R}^{n+1}$ is an isometry and induces a Riemannian covering $S^n \to \mathbb{RP}^n$.

1.4 Local Representations of Metrics

1.4.1 Einstein Summation Convention

We shall often use the index and summation convention introduced by Einstein. Given a vector space V, such as the tangent space of a manifold, we use subscripts for vectors in V. Thus a basis of V is denoted by $e_1, \dots, e_n$. Given a vector $v \in V$ we can then write it as a linear combination of these basis vectors as follows

$$v = \sum_i v^i e_i = v^i e_i = \begin{bmatrix} e_1 \cdots e_n \end{bmatrix} \begin{bmatrix} v^1 \\ \vdots \\ v^n \end{bmatrix}.$$

Here we use superscripts on the coefficients and then automatically sum over indices that are repeated as both subscripts and superscripts. If we define a dual basis e^i for the dual space $V^* = \mathrm{Hom}(V, \mathbb{R})$ as follows: $e^i\left(e_j\right) = \delta^i_j$, then the coefficients can be computed as $v^i = e^i(v)$. Thus we decide to use superscripts for dual bases in V^*. The matrix representation $\left[L^j_i\right]$ of a linear map $L : V \to V$ is found by solving

$$L(e_i) = L^j_i e_j,$$

$$\left[L(e_1) \cdots L(e_n)\right] = \left[e_1 \cdots e_n\right] \begin{bmatrix} L^1_1 & \cdots & L^1_n \\ \vdots & \ddots & \vdots \\ L^n_1 & \cdots & L^n_n \end{bmatrix}$$

In other words

$$L^j_i = e^j\left(L(e_i)\right).$$

As already indicated, subscripts refer to the column number and superscripts to the row number.

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 $\left(x^i\right)$ as they are coordinate coefficients; coordinate vector fields then look like

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

and consequently have subscripts. This is natural, as they form a basis for the tangent space. The dual 1-forms dx^i satisfy $dx^j(\partial_i) = \delta^j_i$ and consequently 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

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

if e_i is the standard basis for $\mathbb{R}^n$. The velocity field is defined as the vector $\dot{c}(t) = \left(\dot{x}^i(t)\right)$. Using the coordinate vector fields this can also be written as

$$\dot{c}(t) = \frac{dx^i}{dt}\frac{\partial}{\partial x^i} = \dot{x}^i(t)\,\partial_i.$$

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. In this case we must show that it gives the same answer in different coordinates. This is simply because the chain rule tells us that

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

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

When offering coordinate dependent definitions we shall be careful that they are given in a form where they obviously conform to this philosophy and are consequently easily seen to be invariantly defined.

1.4.2 Coordinate Representations

On a manifold M we can multiply 1-forms to get bilinear forms:

$$\theta_1 \cdot \theta_2(v, w) = \theta_1(v) \cdot \theta_2(w).$$

Note that $\theta_1 \cdot \theta_2 \neq \theta_2 \cdot \theta_1$. This multiplication is actually a tensor product $\theta_1 \cdot \theta_2 = \theta_1 \otimes \theta_2$. 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 bit optimistic. 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 abstract calculations.

Example 1.4.1. The canonical metric on $\mathbb{R}^n$ in the identity chart is

$$g = \delta_{ij} dx^i dx^j = \sum_{i=1}^{n} \left(dx^i\right)^2.$$

Example 1.4.2. 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,

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

To see this recall that

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

Thus,

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

which gives

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

1.4.3 Frame Representations

A 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

$$g = g_{ij}\sigma^i\sigma^j = g(X_i, X_j)\,\sigma^i\sigma^j.$$

Example 1.4.3. Any left-invariant metric on a Lie group G can be written as

$$g = (\sigma^1)^2 + \cdots + (\sigma^n)^2$$

using a coframe dual to left-invariant vector fields $X_1, \ldots, X_n$ forming an orthonormal basis for T_eG. If instead we just begin with a frame of left-invariant vector fields $X_1, \ldots, X_n$ and dual coframe $\sigma^1, \ldots, \sigma^n$, then a left-invariant metric g depends only on its values on T_eG and can be written as $g = g_{ij}\sigma^i\sigma^j$, 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 1.4.4. A *surface of revolution* consists of a profile curve

$$c(t) = (r(t), 0, z(t)) : I \to \mathbb{R}^3,$$

where $I \subset \mathbb{R}$ is open and $r(t) > 0$ for all t. By rotating this curve around the z-axis, we get a surface that can be represented as

$$(t, \theta) \mapsto f(t, \theta) = (r(t)\cos\theta, r(t)\sin\theta, z(t)).$$

This is a cylindrical coordinate representation, and we have a natural frame $\partial_t, \partial_\theta$ on the surface with dual coframe $dt, d\theta$. We wish to calculate the induced metric on this surface from the Euclidean metric $dx^2 + dy^2 + dz^2$ on $\mathbb{R}^3$ with respect to this frame. Observe that

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

so

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

Thus

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

If the curve is parametrized by arc length, then we obtain the simpler formula:

$$g = dt^2 + r^2 d\theta^2.$$

Fig. 1.4 Surfaces of revolution

This is reminiscent of our polar coordinate description of $\mathbb{R}^2$. In figure 1.4 there are two pictures of surfaces of revolution. In the first, r starts out being zero, but the metric appears smooth as r has vertical tangent to begin with. The second shows that when $r = 0$ the metric looks conical and therefore collapses the manifold.

On the abstract manifold $I \times S^1$ we can use the frame $\partial_t, \partial_\theta$ with coframe $dt, d\theta$ to define metrics

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

These are called *rotationally symmetric* metrics since η and ρ do not depend on the rotational parameter 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 $dt^2 + r^2 d\theta^2$ is a surface of revolution, then $\dot{z}^2 + \dot{r}^2 = 1$ and, in particular, $|\dot{r}| \leq 1$.

Example 1.4.5. The round sphere $S^2(R) \subset \mathbb{R}^3$ can be thought of as a surface of revolution by revolving

$$t \mapsto R\left(\sin\left(\tfrac{t}{R}\right), 0, \cos\left(\tfrac{t}{R}\right)\right)$$

around the z-axis. The metric looks like

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

Note that $R \sin\left(\frac{t}{R}\right) \to t$ as $R \to \infty$, so very large spheres look like Euclidean space.

By formally changing R to $\mathrm{i}R$, we arrive at a different family of rotationally symmetric metrics:

$$dt^2 + R^2 \sinh^2\left(\tfrac{t}{R}\right) d\theta^2.$$

This metric coincides with the metric defined in example 1.1.7 by observing that it comes from the induced metric in $\mathbb{R}^{2,1}$ after having rotated the curve

$$t \mapsto R\left(\sinh\left(\tfrac{t}{R}\right), 0, \cosh\left(\tfrac{t}{R}\right)\right)$$

around the z-axis.

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 obtain 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 $k > 0$, it is $S^2\,(1/\sqrt{k})$; and when $k < 0$ the hyperbolic space $H^2\,(1/\sqrt{-k})$.

Corresponding to sn_k we also have cs_k defined as the solution to

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

The functions are related by

$$\begin{aligned} \frac{d\,\mathrm{sn}_k}{dt}(t) &= \mathrm{cs}_k\,(t)\,, \\ \frac{d\,\mathrm{cs}_k}{dt}(t) &= -k\,\mathrm{sn}_k\,(t)\,, \\ 1 &= \mathrm{cs}_k^2\,(t) + k\,\mathrm{sn}_k^2\,(t)\,. \end{aligned}$$

1.4.4 Polar Versus Cartesian Coordinates

In the rotationally symmetric examples we haven't discussed what happens when $\rho(t) = 0$. In the revolution case, the profile curve clearly needs to have a horizontal tangent in order to look smooth. To be specific, consider $dt^2 + \rho^2(t)d\theta^2$, where $\rho : [0,b) \to [0,\infty)$ with $\rho(0) = 0$ and $\rho(t) > 0$ for $t > 0$. All other situations can be translated or reflected into this position.

More generally, we wish to consider metrics on $I \times S^{n-1}$ of the type $dt^2 + \rho^2(t)ds_{n-1}^2$, where ds_{n-1}^2 is the canonical metric on $S^{n-1}(1) \subset \mathbb{R}^n$. These are also called *rotationally symmetric metrics* and are a special class of *warped products* (see also section 4.3). If we assume that $\rho\,(0) = 0$ and $\rho\,(t) > 0$ for $t > 0$, then we want to check that the metric extends smoothly near $t = 0$ to give a smooth metric near the origin in $\mathbb{R}^n$. There is also a discussion of how to approach this smoothness question in section 4.3.4.

The natural coordinate change to make is $x = ts$ where $x \in \mathbb{R}^n$, $t > 0$, and $s \in S^{n-1}(1) \subset \mathbb{R}^n$. Thus

$$ds_{n-1}^2 = \sum_{i=1}^{n} \left(ds^i\right)^2.$$

Keep in mind that the constraint $\sum \left(s^i\right)^2 = 1$ implies the relationship $\sum s^i ds^i = 0$ between the restriction of the differentials to $S^{n-1}(1)$.

The standard metric on $\mathbb{R}^n$ now becomes

$$
\begin{aligned}
\sum \left(dx^i\right)^2 &= \sum \left(s^i dt + t ds^i\right)^2 \\
&= \sum \left(s^i\right)^2 dt^2 + t^2 \left(ds^i\right)^2 + (tdt)\left(s^i ds^i\right) + \left(s^i ds^i\right)(tdt) \\
&= dt^2 + t^2 ds_{n-1}^2
\end{aligned}
$$

when switching to polar coordinates.

In the general situation we have to do this calculation in reverse and check that the expression becomes smooth at the origin corresponding to $x^i = 0$. Thus we have to calculate dt and ds^i in terms of dx^i. First observe that

$$
\begin{aligned}
2tdt &= 2 \sum x^i dx^i, \\
dt &= \frac{1}{t} \sum x^i dx^i,
\end{aligned}
$$

and then from $\sum \left(dx^i\right)^2 = dt^2 + t^2 ds_{n-1}^2$ that

$$
ds_{n-1}^2 = \frac{\sum \left(dx^i\right)^2 - dt^2}{t^2}.
$$

This implies

$$
\begin{aligned}
dt^2 + \rho^2(t) ds_{n-1}^2 &= dt^2 + \rho^2(t) \frac{\sum \left(dx^i\right)^2 - dt^2}{t^2} \\
&= \left(1 - \frac{\rho^2(t)}{t^2}\right) dt^2 + \frac{\rho^2(t)}{t^2} \sum \left(dx^i\right)^2 \\
&= \left(\frac{1}{t^2} - \frac{\rho^2(t)}{t^4}\right) \left(\sum x^i dx^i\right)^2 + \frac{\rho^2(t)}{t^2} \sum \left(dx^i\right)^2.
\end{aligned}
$$

Thus we have to ensure that the functions

$$
\frac{\rho^2(t)}{t^2} \text{ and } \left(\frac{1}{t^2} - \frac{\rho^2(t)}{t^4}\right)
$$

are smooth, keeping in mind that $t = \sqrt{\sum \left(x^i\right)^2}$ is not differentiable at the origin. The condition $\rho(0) = 0$ is necessary for the first function to be continuous at $t = 0$, while we have to additionally assume that $\dot{\rho}(0) = 1$ for the second function to be continuous.

The general condition for ensuring that both functions are smooth is that $\rho(0) = 0$, $\dot{\rho}(0) = 1$, and that all even derivatives vanish: $\rho^{(\text{even})}(0) = 0$. This implies that for each $l = 1, 2, 3, \ldots$

$$\rho(t) = t + \sum_{k=1}^{l} a_k t^{2k+1} + o\left(t^{2l+3}\right)$$

as all the even derivatives up to $2l + 2$ vanish. Note that

$$\begin{aligned}\frac{\rho^2(t)}{t^2} &= \left(1 + \sum_{k=1}^{l} a_k t^{2k} + o\left(t^{2l+2}\right)\right)^2 \\ &= 1 + \sum_{k=1}^{l} b_k t^{2k} + o\left(t^{2l+2}\right),\end{aligned}$$

where $b_k = \sum_{i=1}^{k} a_i a_{k-i}$. Similarly for the other function

$$\begin{aligned}\frac{1}{t^2} - \frac{\rho^2(t)}{t^4} &= \frac{1}{t^2}\left(1 - \frac{\rho^2(t)}{t^2}\right) \\ &= \frac{1}{t^2}\left(-\sum_{k=1}^{l} b_k t^{2k} + o\left(t^{2l+2}\right)\right) \\ &= -\sum_{k=1}^{l} b_k t^{2k-2} + o\left(t^{2l}\right).\end{aligned}$$

This shows that both functions can be approximated to any order by polynomials that are smooth as functions of x^i at $t = 0$. Thus the functions themselves are smooth.

Example 1.4.6. These conditions hold for all of the metrics $dt^2 + \mathrm{sn}_k^2(t) ds_{n-1}^2$, where $t \in [0, \infty)$ when $k \leq 0$, and $t \in [0, \pi/\sqrt{k}]$ when $k > 0$. The corresponding Riemannian manifolds are denoted S_k^n and are called *space forms* of dimension n with curvature k. As in example 1.4.5 we can show that these spaces coincide with $H^n(R)$, $\mathbb{R}^n$, or $S^n(R)$. When $k = 0$ we clearly get $(\mathbb{R}^n, g_{\mathbb{R}^n})$. When $k = 1/R^2$ we get $S^n(R)$. To see this, observe that there is a map

$$\begin{aligned}F : \mathbb{R}^n \times (0, R\pi) &\to \mathbb{R}^n \times \mathbb{R}, \\ F(s, r) &= (x, t) = R\left(s \cdot \sin\left(\tfrac{r}{R}\right), \cos\left(\tfrac{r}{R}\right)\right),\end{aligned}$$

that restricts to

$$\begin{aligned}G : S^{n-1} \times (0, R\pi) &\to \mathbb{R}^n \times \mathbb{R}, \\ G(s, r) &= R\left(s \cdot \sin\left(\tfrac{r}{R}\right), \cos\left(\tfrac{r}{R}\right)\right).\end{aligned}$$

Thus, G really maps into the R-sphere in $\mathbb{R}^{n+1}$. To check that G is a Riemannian isometry we just compute the canonical metric on $\mathbb{R}^n \times \mathbb{R}$ using the coordinates $R\left(s \cdot \sin\left(\frac{r}{R}\right), \cos\left(\frac{r}{R}\right)\right)$. To do the calculation keep in mind that $\sum \left(s^i\right)^2 = 1$ and $\sum s^i ds^i = 0$.

$$\begin{aligned}
&dt^2 + \sum \delta_{ij} dx^i dx^j \\
&= \left(dR\cos\left(\tfrac{r}{R}\right)\right)^2 + \sum \delta_{ij} d\left(R\sin\left(\tfrac{r}{R}\right)s^i\right) d\left(R\sin\left(\tfrac{r}{R}\right)s^j\right) \\
&= \sin^2\left(\tfrac{r}{R}\right) dr^2 \\
&\quad + \sum \delta_{ij}\left(s^i\cos\left(\tfrac{r}{R}\right)dr + R\sin\left(\tfrac{r}{R}\right)ds^i\right)\left(s^j\cos\left(\tfrac{r}{R}\right)dr + R\sin\left(\tfrac{r}{R}\right)ds^j\right) \\
&= \sin^2\left(\tfrac{r}{R}\right) dr^2 + \sum \delta_{ij} s^i s^j \cos^2\left(\tfrac{r}{R}\right) dr^2 + \sum \delta_{ij} R^2 \sin^2\left(\tfrac{r}{R}\right) ds^i ds^j \\
&\quad + \sum \delta_{ij} s^j R\cos\left(\tfrac{r}{R}\right)\sin(r)\, ds^i dr + \sum \delta_{ij} s^i R\cos\left(\tfrac{r}{R}\right)\sin\left(\tfrac{r}{R}\right) dr ds^j \\
&= \sin^2\left(\tfrac{r}{R}\right) dr^2 + \cos^2\left(\tfrac{r}{R}\right) dr^2 \sum \delta_{ij} s^i s^j + R^2\sin^2\left(\tfrac{r}{R}\right)\sum \delta_{ij} ds^i ds^j \\
&\quad + R\cos\left(\tfrac{r}{R}\right)\sin\left(\tfrac{r}{R}\right) dr \sum s^i ds^i + R\cos\left(\tfrac{r}{R}\right)\sin\left(\tfrac{r}{R}\right)\left(\sum s^i ds^i\right) dr \\
&= dr^2 + R^2 \sin^2\left(\tfrac{r}{R}\right) ds_{n-1}^2 .
\end{aligned}$$

Hyperbolic space $H^n(R) \subset \mathbb{R}^{n,1}$ is similarly realized as a rotationally symmetric metric using the map

$$\begin{aligned}
S^{n-1} \times (0,\infty) &\to \mathbb{R}^{n,1} \\
(s,r) &\mapsto (x,t) = R\left(s \cdot \sinh\left(\tfrac{r}{R}\right), \cosh\left(\tfrac{r}{R}\right)\right).
\end{aligned}$$

As with spheres this defines a Riemannian isometry from $dr^2 + R^2\sinh^2\left(\frac{r}{R}\right) ds_{n-1}^2$ to the induced metric on $H^n(R) \subset \mathbb{R}^{n,1}$. For the calculation note that the metric is induced by $g_{\mathbb{R}^{n,1}} = \delta_{ij} dx^i dx^j - dt^2$ and that $\sum (s^i)^2 = 1$ and $\sum s^i ds^i = 0$.

$$\begin{aligned}
&-dt^2 + \sum \delta_{ij} dx^i dx^j \\
&= -\left(d\left(R\cosh\left(\tfrac{r}{R}\right)\right)\right)^2 + \sum \delta_{ij} d\left(R\sinh\left(\tfrac{r}{R}\right)s^i\right) d\left(R\sinh\left(\tfrac{r}{R}\right)s^j\right) \\
&= -\sinh^2\left(\tfrac{r}{R}\right) dr^2 \\
&\quad + \sum \delta_{ij}\left(s^i\cosh\left(\tfrac{r}{R}\right)dr + R\sinh\left(\tfrac{r}{R}\right)ds^i\right)\left(s^j\cosh\left(\tfrac{r}{R}\right)dr + R\sinh\left(\tfrac{r}{R}\right)ds^j\right) \\
&= -\sinh^2\left(\tfrac{r}{R}\right) dr^2 + \sum \delta_{ij} s^i s^j \cosh^2\left(\tfrac{r}{R}\right) dr^2 + \sum \delta_{ij} R^2 \sinh^2\left(\tfrac{r}{R}\right) ds^i ds^j \\
&= dr^2 + R^2 \sinh^2\left(\tfrac{r}{R}\right) ds_{n-1}^2 .
\end{aligned}$$

1.4.5 Doubly Warped Products

We can more generally consider metrics of the type:

$$dt^2 + \rho^2(t)ds_p^2 + \phi^2(t)ds_q^2$$

on $I \times S^p \times S^q$. These are a special class of *doubly warped products*. When $\rho(t) = 0$ we can use the calculations for rotationally symmetric metrics (see 1.4.4) to check for smoothness. Note, however, that nondegeneracy of the metric implies that ρ and ϕ cannot both be zero at the same time. The following propositions explain the various possible situations:

Proposition 1.4.7. *If* $\rho : (0, b) \to (0, \infty)$ *is smooth and* $\rho(0) = 0$, *then we get a smooth metric at* $t = 0$ *if and only if*

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

and

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

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

Proposition 1.4.8. *If* $\rho : (0, b) \to (0, \infty)$ *is smooth and* $\rho(b) = 0$, *then we get a smooth metric at* $t = b$ *if and only if*

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

and

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

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

By adjusting and possibly changing the roles of these functions we obtain three different types of topologies.

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

Example 1.4.9. We exhibit spheres as doubly warped products. The claim is that the metrics

$$dt^2 + \sin^2(t)ds_p^2 + \cos^2(t)ds_q^2,\ t \in [0, {}^{\pi}/_{2}]\,,$$

are $\left(S^{p+q+1}(1), g_{S^{p+q+1}}\right)$. Since $S^p \subset \mathbb{R}^{p+1}$ and $S^q \subset \mathbb{R}^{q+1}$ we can map

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

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 calculations in example 1.4.6.

1.4.6 Hopf Fibrations

We use several of the above constructions to understand the Hopf fibration. This includes the higher dimensional analogues and other metric variations of these examples.

Example 1.4.10. First we revisit the Hopf fibration $S^3(1) \to S^2\left({}^1/_2\right)$ (see also example 1.1.5). On $S^3(1)$, write the metric as

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

and use complex coordinates

$$\left(t, e^{i\theta_1}, e^{i\theta_2}\right) \mapsto \left(\sin(t)e^{i\theta_1}, \cos(t)e^{i\theta_2}\right)$$

to describe the isometric embedding

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

Since the Hopf fibers come from complex scalar multiplication, we see that they are of the form

$$\theta \mapsto \left(t, e^{i(\theta_1+\theta)}, e^{i(\theta_2+\theta)}\right).$$

On $S^2\left({}^1/_2\right)$ use the metric

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

with coordinates

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

The Hopf fibration in these coordinates looks like

$$\left(t, e^{\mathrm{i}\,\theta_1}, e^{\mathrm{i}\,\theta_2}\right) \mapsto \left(t, e^{\mathrm{i}(\theta_1-\theta_2)}\right).$$

This conforms with Wilhelm's map defined in example 1.1.5 if we observe that

$$\left(\sin(t)e^{\mathrm{i}\,\theta_1}, \cos(t)e^{\mathrm{i}\,\theta_2}\right)$$

is supposed to be mapped to

$$\left(\frac{1}{2}\left(\cos^2 t - \sin^2 t\right), \sin(t)\cos(t)\, e^{\mathrm{i}(\theta_1-\theta_2)}\right) = \left(\frac{1}{2}\cos(2t), \frac{1}{2}\sin(2t)\, e^{\mathrm{i}(\theta_1-\theta_2)}\right).$$

On $S^3(1)$ there is an orthogonal frame

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

where the first vector is tangent to the Hopf fiber and the two other vectors have unit length. On $S^2\left({}^1/_2\right)$

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

is an orthonormal frame. The Hopf map clearly maps

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

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

Note also that the map

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

gives us the promised isometry from $S^3(1)$ to $\mathrm{SU}(2)$, where $\mathrm{SU}(2)$ has the left-invariant metric described in example 1.3.5.

Example 1.4.11. More generally, the map

$$I \times S^1 \times S^1 \to I \times S^1$$
$$\left(t, e^{\mathrm{i}\,\theta_1}, e^{\mathrm{i}\,\theta_2}\right) \mapsto \left(t, e^{\mathrm{i}(\theta_1-\theta_2)}\right)$$

is always a Riemannian submersion when the domain is endowed with the doubly warped product metric

$$dt^2 + \rho^2(t)d\theta_1^2 + \phi^2(t)d\theta_2^2$$

and the target has the rotationally symmetric metric

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

Example 1.4.12. This submersion can also be generalized to higher dimensions as follows: On $I \times S^{2n+1} \times S^1$ consider the doubly warped product metric

$$dt^2 + \rho^2(t)ds_{2n+1}^2 + \phi^2(t)d\theta^2.$$

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 an appropriate metric on the quotient space. To find this 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

$$h(v, w) = ds_{2n+1}^2(pr(v), pr(w))$$

and

$$g(v, w) = ds_{2n+1}^2(v - pr(v), w - pr(w)).$$

We can then rewrite

$$dt^2 + \rho^2(t)ds_{2n+1}^2 + \phi^2(t)d\theta^2 = dt^2 + \rho^2(t)g + \rho^2(t)h + \phi^2(t)d\theta^2.$$

Observe 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, the submersion metric on $I \times S^{2n+1}$ can be written

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

Example 1.4.13. 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

$$\begin{aligned} ds_3^2 &= (\sigma^1)^2 + (\sigma^2)^2 + (\sigma^3)^2, \\ h &= (\sigma^1)^2, \\ g &= (\sigma^2)^2 + (\sigma^3)^2, \end{aligned}$$

where $\{\sigma^1, \sigma^2, \sigma^3\}$ is the coframe coming from the identification $S^3 \simeq \mathrm{SU}(2)$ (see example 1.3.5). The Riemannian submersion in this case can then be written

$$\begin{gathered} \left(I \times S^3 \times S^1,\ dt^2 + \rho^2(t)\left((\sigma^1)^2 + (\sigma^2)^2 + (\sigma^3)^2\right) + \phi^2(t)d\theta^2\right) \\ \downarrow \\ \left(I \times S^3,\ dt^2 + \rho^2(t)\left((\sigma^2)^2 + (\sigma^3)^2\right) + \frac{(\rho(t)\cdot\phi(t))^2}{\rho^2(t)+\phi^2(t)}(\sigma^1)^2\right). \end{gathered}$$

Example 1.4.14. If we let $\rho = \sin(t)$, $\phi = \cos(t)$, and $t \in I = [0, \pi/2]$, then we obtain the generalized Hopf fibration $S^{2n+3} \to \mathbb{CP}^{n+1}$ defined in example 1.3.4. The map

$$(0, \pi/2) \times \left(S^{2n+1} \times S^1\right) \to (0, \pi/2) \times \left(\left(S^{2n+1} \times S^1\right)/S^1\right)$$

is a Riemannian submersion, and the Fubini-Study metric on $\mathbb{CP}^{n+1}$ can be represented as

$$dt^2 + \sin^2(t)(g + \cos^2(t)h).$$

1.5 Some Tensor Concepts

In this section we shall collect together some notational baggage and more general inner products of tensors that will be needed from time to time.

1.5.1 Type Change

The inner product structures on the tangent spaces to a Riemannian manifold allow us to view tensors in different ways. We shall use this for the Hessian of a function and the Ricci tensor. These are naturally bilinear tensors, but can also be viewed as endomorphisms of the tangent bundle. Specifically, if we have a metric g and an endomorphism S on a vector space, then $b(v, w) = g(S(v), w)$ is the corresponding bilinear form. Given g, this correspondence is an isomorphism. When generalizing to the pseudo-Riemannian setting it is occasionally necessary to change the formulas we develop (see also exercise 1.6.10).

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

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

Given a Riemannian metric g on M, we can make T 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 $(w \mapsto g(v,w)) \in T^*M$. Using this isomorphism we can then 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 in TM and covectors in T^*M can be written as

$$v = v^i E_i = \sigma^i(v) E_i,$$
$$\omega = \omega_j \sigma^j = \omega\left(E_j\right) \sigma^j.$$

The tensor T can then be written as

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

Using indices and simply writing $T^{i_1 \cdots i_s}_{j_1 \cdots j_t}$ is often called tensor notation.

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 $w \mapsto g(E_i, w)$, 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

$$g(v, E_i) = \delta^j_i.$$

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

$$g_{ki} v^k = \delta^j_i.$$

Letting g^{ij} denote the ijth entry in the inverse of $\left(g_{ij}\right)$, we obtain

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

Thus,

$$\begin{aligned} E_i &\mapsto g_{ij}\sigma^j, \\ \sigma^j &\mapsto g^{ij} E_i. \end{aligned}$$

Note that using Einstein notation 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 Es with σs 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: For now this is simply an abstract $(1,1)$-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

$$\mathrm{Ric} = \mathrm{Ric}_{jk} \cdot \sigma^j \otimes \sigma^k = g_{ji}\,\mathrm{Ric}_k^i \cdot \sigma^j \otimes \sigma^k,$$

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

$$\mathrm{Ric} = \mathrm{Ric}^{ik} \cdot E_i \otimes E_k = g^{ij}\,\mathrm{Ric}_j^k \cdot E_i \otimes E_k.$$

The curvature tensor: We consider a $(1,3)$-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

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

Note that we have elected to place l at the end of the $(0,4)$ version. In many texts it is placed first. Our choice appears natural given how we write these tensors in invariant notation in chapter 3. As a $(2,2)$-tensor we have:

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

Here we must be careful as there are several different possibilities for raising and lowering indices. We chose 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 what we will call the curvature operator.

1.5.2 Contractions

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

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

An instructive example comes from considering the rank 1 tensor $X \otimes \omega$ where X is a vector field and ω a 1-form. In this case contraction is simply evaluation $C(X \otimes \omega) = \omega(X)$. Conversely, contraction is a sum of such evaluations.

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 fact the Ricci tensor is a contraction of the curvature tensor:

$$\begin{aligned} \mathrm{Ric} &= \mathrm{Ric}^i_j \cdot E_i \otimes \sigma^j \\ &= R^{kj}_{ik} \cdot E_i \otimes \sigma^j \\ &= R^j_{iks} g^{sk} \cdot E_i \otimes \sigma^j, \end{aligned}$$

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 is defined as a contraction of the Ricci tensor:

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

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

1.5.3 Inner Products of Tensors

There are several conventions 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 normed spaces as

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

The Euclidean norm is given by

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

where $L^* : W \to V$ is the adjoint. These norms 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 has the advantage of actually coming from an inner product:

$$\langle L_1, L_2 \rangle = \mathrm{tr}\left(L_1 \circ L_2^*\right) = \mathrm{tr}\left(L_2 \circ L_1^*\right).$$

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

It is worthwhile to check 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 $T = T^i_j \cdot E_i \otimes \sigma^j$. 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

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

which after type change becomes

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

Finally,

$$|T|^2 = T^i_j T^k_l g^{lj} g_{ki}.$$

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

$$|T|^2 = T^i_j T^j_i.$$

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

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

In general, we can define the inner product of two tensors of the same type, by declaring that if E_i is an orthonormal frame with dual coframe σ^i then the (s,t)-tensors

$$E_{i_1} \otimes \cdots \otimes E_{i_s} \otimes \sigma^{j_1} \otimes \cdots \otimes \sigma^{j_t}$$

form an orthonormal basis for (s,t)-tensors.

The inner product just defined is what we shall call the point-wise inner product of tensors, just as $g(X,Y)$ is the point-wise inner product of two vector fields. The point-wise inner product of two compactly supported tensors of the same type can be integrated to yield an inner product structure on the space of tensors:

$$(T_1, T_2) = \int_M g(T_1, T_2) \operatorname{vol}.$$

1.5.4 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 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

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

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

$$|T|^2 = T^i_{\ j} T^{\ j}_i.$$

Nice as this notation is, it is not used consistently in the literature. 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.

1.6 Exercises

EXERCISE 1.6.1. On $M \times N$ one has the Cartesian product metrics $g = g_M + g_N$, where g_M, g_N are metrics on M, N respectively.

(1) Show that $(\mathbb{R}^n, g_{\mathbb{R}^n}) = (\mathbb{R}, dt^2) \times \cdots \times (\mathbb{R}, dt^2)$.
(2) Show that the flat square torus

$$T^2 = \mathbb{R}^2/\mathbb{Z}^2 = \left(S^1, \left(\frac{1}{2\pi}\right)^2 d\theta^2\right) \times \left(S^1, \left(\frac{1}{2\pi}\right)^2 d\theta^2\right).$$

(3) Show that

$$F(\theta_1,\theta_2)=\frac{1}{2\pi}(\cos\theta_1,\sin\theta_1,\cos\theta_2,\sin\theta_2)$$

is a Riemannian embedding: $T^2\to\mathbb{R}^4$.

EXERCISE 1.6.2. 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.

EXERCISE 1.6.3. Let $M\to N$ be a Riemannian k-fold covering map. Show, $\operatorname{vol} M=k\cdot\operatorname{vol} N$.

EXERCISE 1.6.4. Show that the volume form for a metric $dr^2+\rho^2(r)\,g_N$ on a product $I\times N$ is given by $\rho^{n-1}dr\wedge\operatorname{vol}_N$, where vol_N is the volume form on (N,g_N).

EXERCISE 1.6.5. Show that if $E_1,\dots,E_n$ is an orthonormal frame, then the dual frame is given by $\sigma^i(X)=g(E_i,X)$ and the volume form by $\operatorname{vol}=\pm\sigma^1\wedge\cdots\wedge\sigma^n$.

EXERCISE 1.6.6. Show that in local coordinates $x^1,\dots,x^n$ the volume form is given by $\operatorname{vol}=\pm\sqrt{\det[g_{ij}]}dx^1\wedge\cdots\wedge dx^n$. In the literature one often sees the simplified notation $g=\sqrt{\det[g_{ij}]}$.

EXERCISE 1.6.7. Construct paper models of the warped products $dt^2+a^2t^2d\theta^2$. If $a=1$, this is of course the Euclidean plane, and when $a<1$, they look like cones. What do they look like when $a>1$?

EXERCISE 1.6.8. Consider a rotationally symmetric metric $dr^2+\rho^2(r)\,g_{S^{n-1}(R)}$, where $S^{n-1}(R)\subset\mathbb{R}^n$ is given the induced metric. Show that if $\rho(0)=0$, then we need $\dot\rho(0)=1/R$ and $\rho^{(2k)}(0)=0$ to get a smooth metric near $r=0$.

EXERCISE 1.6.9. Show that if we think of $\mathbb{R}^n$ as any of the hyperplanes $x^{n+1}=R$ in $\mathbb{R}^{n+1}$, then $\operatorname{Iso}(\mathbb{R}^n)$ can be identified with the group of $(n+1)\times(n+1)$ matrices

$$\begin{bmatrix} O & v\\ 0 & 1\end{bmatrix},$$

where $v\in\mathbb{R}^n$ and $O\in\mathrm{O}(n)$. Further, show that these are precisely the linear maps that preserve $x^{n+1}=R$ and the degenerate bilinear form $x^1y^1+\cdots+x^ny^n$.

EXERCISE 1.6.10. Let V be an n-dimensional vector space with a symmetric nondegenerate bilinear form g of index p.

(1) Show that there exists a basis $e_1,\dots,e_n$ such that $g(e_i,e_j)=0$ if $i\neq j$, $g(e_i,e_i)=1$ if $i=1,\dots,n-p$ and $g(e_i,e_i)=-1$ if $i=n-p+1,\dots,n$. Thus V is isometric to $\mathbb{R}^{p,q}$.
(2) Show that for any v we have the expansion

$$
\begin{aligned}
v &= \sum_{i=1}^{n} \tfrac{g(v,e_i)}{g(e_i,e_i)} e_i \\
&= \sum_{i=1}^{n-p} g(v, e_i)\, e_i - \sum_{i=n-p+1}^{n} g(v, e_i)\, e_i.
\end{aligned}
$$

(3) Let $L : V \to V$ be a linear operator. Show that

$$
\operatorname{tr}(L) = \sum_{i=1}^{n} \tfrac{g(L(e_i),e_i)}{g(e_i,e_i)}.
$$

EXERCISE 1.6.11. Let g^{-1} denote the $(2, 0)$-tensor that is the inner product on the dual tangent space T^*M. Show that type change can be described as a contraction of a tensor product with g or g^{-1}.

EXERCISE 1.6.12. For a $(1, 1)$-tensor T on a Riemannian manifold, show that if E_i is an orthonormal basis, then

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

EXERCISE 1.6.13. Given $(1, 1)$-tensor tensors S, T show that if S is symmetric and T skew-symmetric, then $g(S, T) = 0$.

EXERCISE 1.6.14. Show that the inner product of two tensors of the same type can be described as (possibly several) type change(s) to one of the tensors followed by (possibly several) contraction(s).

EXERCISE 1.6.15. Consider $F : \mathbb{F}^{n+1} - \{0\} \to \mathbb{FP}^n$ defined by $F(x) = \operatorname{span}_{\mathbb{F}} \{x\}$, where $\mathbb{F} = \mathbb{R}, \mathbb{C}$ and assume that $\mathbb{FP}^n$ comes with the metric that makes the restriction of F to the unit sphere a Riemannian submersion.

(1) Show that F is a submersion.
(2) Show that F is not a Riemannian submersion with respect to the standard metric on $\mathbb{F}^{n+1} - \{0\}$.
(3) Is it possible to choose a metric on $\mathbb{F}^{n+1} - \{0\}$ so that F becomes a Riemannian submersion?

EXERCISE 1.6.16. The *arc length* of a curve $c(t) : [a, b] \to (M, g)$ is defined by

$$
L(c) = \int_{[a,b]} |\dot{c}|\, dt
$$

(1) Show that the arc length does not depend on the parametrization of c.
(2) Show that any curve with nowhere vanishing speed can be reparametrized to have unit speed.

(3) Show that it is possible to define the arclength of an absolutely continuous curve. You should, in particular, show that the concept of being absolutely continuous is well-defined for curves in manifolds.

EXERCISE 1.6.17. Show that the arclength of curves is preserved by Riemannian immersions.

EXERCISE 1.6.18. Let $F : (M, g_M) \to (N, g_N)$ be a Riemannian submersion and $c(t) : [a,b] \to (M, g_M)$ a curve. Show that $L(F \circ c) \leq L(c)$ with equality holding if and only if $\dot{c}(t) \perp \ker DF_{c(t)}$ for all $t \in [a,b]$.

EXERCISE 1.6.19. Show directly that any curve between two points in Euclidean space is longer than the Euclidean distance between the points. Moreover, if the length agrees with the distance, then the curve lies on the straight line between those points. Hint: If v is an appropriate unit vector, then calculate the length of $v \cdot c(t)$ and compare it to the length of c.

EXERCISE 1.6.20. Let $S^n \subset \mathbb{R}^{n+1}$ be the standard unit sphere and $p, q \in S^n$ and $v \in T_pS^n$ a unit vector. We think of p, q and v as unit vectors in $\mathbb{R}^{n+1}$.

(1) Show that the great circle $p \cos t + v \sin t$ is a unit speed curve on S^n that starts at p and has initial velocity v.
(2) Consider the map $F(r, v) = p \cos r + v \sin r$ for $r \in [0, \pi]$ and $v \perp p$, $|v| = 1$. Show that this map defines a diffeomorphism $(0, \pi) \times S^{n-1} \to S^n - \{\pm p\}$.
(3) Define $\partial_r = F_*(\partial_r)$ on $S^n - \{\pm p\}$. Show that if $q = F(r_0, v_0)$, then

$$\partial_r|_q = \frac{-p + (p \cdot q) q}{\sqrt{1 - (p \cdot q)^2}} = -p \sin r_0 + v_0 \cos r_0.$$

(4) Show that any curve from p to q is longer than r_0, where $q = F(r_0, v_0)$, unless it is part of the great circle. Hint: Compare the length of $c(t)$ to the integral $\int \dot{c} \cdot \partial_r dt$ and show that $\dot{c} \cdot \partial_r = \frac{dr}{dt}$, where $c(t) = F(r(t), v(t))$.
(5) Show that there is no Riemannian immersion from an open subset $U \subset \mathbb{R}^n$ into S^n. Hint: Any such map would map small equilateral triangles to triangles on S^n whose side lengths and angles are the same. Show that this is impossible by showing that the spherical triangles have sides that are part of great circles and that when such triangles are equilateral the angles are always $> \frac{\pi}{3}$.

EXERCISE 1.6.21. Let $H^n \subset \mathbb{R}^{n,1}$ be hyperbolic space: $p, q \in H^n$; and $v \in T_pH^n$ a unit vector. Thus $|p|^2 = |q|^2 = -1$, $|v|^2 = 1$, and $p \cdot v = 0$.

(1) Show that the hyperbola $p \cosh t + v \sinh t$ is a unit speed curve on H^n that starts at p and has initial velocity v.
(2) Consider $F(r, v) = p \cosh r + v \sinh r$, for $r \geq 0$ and $v \cdot p = 0$, $|v|^2 = 1$. Show that this map defines a diffeomorphism $(0, \infty) \times S^{n-1} \to H^n - \{p\}$.
(3) Define the radial field $\partial_r = F_*(\partial_r)$ on $H^n - \{p\}$. Show that if $q = F(r_0, v_0)$, then

$$\partial_r|_q = \frac{-p - (q \cdot p)\, q}{\sqrt{-1 + (q \cdot p)^2}} = p \sinh r_0 + v_0 \cosh r_0.$$

(4) Show that any curve from p to q is longer than r_0, where $q = F(r_0, v_0)$, unless it is part of the hyperbola. Hint: For a curve $c(t)$ compare the length of c to the integral $\int \dot{c} \cdot \partial_r dt$ and show that $\dot{c} \cdot \partial_r = \frac{dr}{dt}$, where $c(t) = F(r(t), v(t))$.
(5) Show that there is no Riemannian immersion from an open subset $U \subset \mathbb{R}^n$ into H^n. Hint: Any such map would map small equilateral triangles to triangles on H^n whose side lengths and angles are the same. Show that this is impossible by showing that the hyperbolic triangles have sides that are part of hyperbolas and that when such triangles are equilateral the angles are always $< \frac{\pi}{3}$.

EXERCISE 1.6.22 (F. WILHELM). The Hopf fibration from example 1.1.5 can be generalized using quaternions. Quaternions can be denoted $q = a + b\,\mathrm{i} + c\,\mathrm{j} + d\,\mathrm{k} = z + w\,\mathrm{j}$, where $z = a + b\,\mathrm{i}$, $w = c + d\,\mathrm{i}$ are complex numbers and

$$\begin{aligned} \mathrm{i}^2 = \mathrm{j}^2 = \mathrm{k}^2 &= -1, \\ \mathrm{i}\,\mathrm{j} = \mathrm{k} &= -\mathrm{j}\,\mathrm{i}, \\ \mathrm{j}\,\mathrm{k} = \mathrm{i} &= -\mathrm{k}\,\mathrm{j}, \\ \mathrm{k}\,\mathrm{i} = \mathrm{j} &= -\mathrm{i}\,\mathrm{k}. \end{aligned}$$

The set of quaternions form a 4-dimensional real vector space $\mathbb{H}$ with a product structure that is $\mathbb{R}$-bilinear and associative.

(1) Show the quaternions can be realized as a matrix algebra

$$q = \begin{pmatrix} z & w \\ -\bar{w} & \bar{z} \end{pmatrix}$$

where

$$\mathrm{i} = \begin{pmatrix} \sqrt{-1} & 0 \\ 0 & -\sqrt{-1} \end{pmatrix},$$

$$\mathrm{j} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

$$\mathrm{k} = \begin{pmatrix} 0 & \sqrt{-1} \\ \sqrt{-1} & 0 \end{pmatrix}.$$

This in particular ensures that the product structure is $\mathbb{R}$-bilinear and associative.
(2) Show that if

$$\bar{q} = a - b\,\mathrm{i} - c\,\mathrm{j} - d\,\mathrm{k},$$

then the following identities hold:

$$\begin{aligned} a^2+b^2+c^2+d^2 &= |q|^2 \\ &= q\bar{q} \\ &= \bar{q}q \\ &= |z|^2+|w|^2 \\ &= \det\begin{pmatrix} z & w \\ -\bar{w} & \bar{z} \end{pmatrix}, \end{aligned}$$

$$|pq| = |p|\,|q|\,,$$

and

$$\overline{pq} = \bar{q}\bar{p}.$$

(3) Define two maps $\mathbb{H}^2 \to \mathbb{R}\oplus\mathbb{H}$

$$H^l(p,q) = \left(\frac{1}{2}\left(|p|^2-|q|^2\right), \bar{p}q\right)$$

$$H^r(p,q) = \left(\frac{1}{2}\left(|p|^2-|q|^2\right), p\bar{q}\right)$$

Show that they both map $S^7(1)\subset\mathbb{H}^2$ to $S^4(1/2)\subset\mathbb{R}\oplus\mathbb{H}$.

(4) Show that the pre-images of $H^l : S^7(1)\to S^4(1/2)$ correspond to the orbits from left multiplication by unit quaternions on $\mathbb{H}^2$.

(5) Show that the pre-images of $H^r : S^7(1)\to S^4(1/2)$ correspond to the orbits from right multiplication by unit quaternions on $\mathbb{H}^2$.

(6) Show that both H^l and H^r are Riemannian submersions as maps $S^7(1)\to S^4(1/2)$.

EXERCISE 1.6.23. Suppose ρ and ϕ are positive on $(0,\infty)$ and consider the Riemannian submersion

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

Define $f=\rho$ and $h=\frac{(\rho(t)\cdot\phi(t))^2}{\rho^2(t)+\phi^2(t)}$ and assume that

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

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 $\mathbb{CP}^2 - \{\text{point}\}$.

EXERCISE 1.6.24. Let G be a compact Lie group.

(1) Show that G admits a biinvariant metric, i.e., both right- and left-translations are isometries. Hint: Fix a left-invariant metric g_L and a volume form $\text{vol} = \sigma^1 \wedge \cdots \wedge \sigma^1$ where σ^i are orthonormal left-invariant 1-forms. Then define g as the average over right-translations:

$$g(v, w) = \frac{1}{\int_G \text{vol}} \int_G g_L(DR_x(v), DR_x(w)) \, \text{vol}.$$

(2) Show that conjugation $\text{Ad}_h(x) = hxh^{-1}$ is a Riemannian isometry for any biinvariant metric. Conclude that its differential at $x = e$ denoted by the same letters

$$\text{Ad}_h : \mathfrak{g} \to \mathfrak{g}$$

is a linear isometry with respect to g.

(3) Use this to show that the adjoint action

$$\begin{aligned} \text{ad}_U : \mathfrak{g} &\to \mathfrak{g}, \\ \text{ad}_U X &= [U, X] \end{aligned}$$

is skew-symmetric, i.e.,

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

Hint: It is shown in section 2.1.4 that $U \mapsto \text{ad}_U$ is the differential of $h \mapsto \text{Ad}_h$.

EXERCISE 1.6.25. Let G be a Lie group with Lie algebra $\mathfrak{g}$. Show that a nondegenerate, bilinear, symmetric form (X, Y) on $\mathfrak{g}$ defines a biinvariant pseudo-Riemannian metric if and only if $(X, Y) = (\text{Ad}_h X, \text{Ad}_h Y)$ for all $h \in G$.

EXERCISE 1.6.26. Let G be a compact group acting on a Riemannian manifold. Show that M admits a Riemannian metric such that G acts by isometries. Hint: You first have to show that any manifold admits a Riemannian metric (partition of unity) and then average the metric to make it G-invariant.

EXERCISE 1.6.27. Let G be a Lie group. Define the *Killing form* on $\mathfrak{g}$ by

$$B(X,Y)=\operatorname{tr}(\operatorname{ad}_X\circ\operatorname{ad}_Y).$$

(1) Show that B is symmetric and bilinear.
(2) When G admits a biinvariant metric show that $B(X,X)\leq 0$. Hint: Use part (3) of exercise 1.6.24.
(3) Show that $B(\operatorname{ad}_Z X,Y)=-B(X,\operatorname{ad}_Z Y)$.
(4) Show that $B(\operatorname{Ad}_h X,\operatorname{Ad}_h Y)=B(X,Y)$, when G is connected. Hint: Show that

$$t\mapsto B\left(\operatorname{Ad}_{\exp(tZ)}X,\operatorname{Ad}_{\exp(tZ)}Y\right)$$

is constant, where $\exp(0)=e$ and $\frac{d}{dt}\exp(tZ)=Z$.

Note B looks like a biinvariant metric on G. When $\mathfrak{g}$ is semisimple the Killing form is nondegenerate (this can in fact be taken as the definition of semisimplicity) and thus can be used as a pseudo-Riemannian biinvariant metric. It is traditional to use $-B$ instead so as to obtain a Riemannian metric when G is also compact.

EXERCISE 1.6.28. Consider the Lie group of real $n\times n$-matrices with determinant 1, $SL(n,\mathbb{R})$. The Lie algebra $\mathfrak{sl}(n,\mathbb{R})$ consists of real $n\times n$-matrices with trace 0. Show that the symmetric bilinear form $(X,Y)=\operatorname{tr}(XY)$ on $\mathfrak{sl}(n,\mathbb{R})$ defines a biinvariant pseudo-Riemannian metric on $SL(n,\mathbb{R})$. Hint: Show that it is nondegenerate and invariant under Ad_h.

EXERCISE 1.6.29. Show that the matrices

$$\begin{bmatrix} a^{-1} & 0 & 0\\ 0 & a & b\\ 0 & 0 & 1\end{bmatrix},\ a>0,\ b\in\mathbb{R}$$

define a two-dimensional Lie group that does not admit a biinvariant pseudo-Riemannian metric.

Chapter 2
Derivatives

This chapter introduces several important notions of derivatives of tensors. In chapters 5 and 6 we also introduce partial derivatives of functions into Riemannian manifolds.

The main goal is the construction of the connection and its use as covariant differentiation. We give a motivation of this concept that depends on exterior and Lie derivatives. Covariant differentiation, in turn, allows for nice formulas for exterior derivatives, Lie derivatives, divergence and much more. It is also crucial in the development of curvature which is the central construction in Riemannian geometry.

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 a Taylor expansion of a Riemannian metric at a point with respect to a suitably chosen coordinate system. Lipschitz, Killing, and Christoffel introduced the connection in various ways as an intermediate step in computing the curvature. After this early work by the above-mentioned German mathematicians, an Italian school around Levi-Civita, Ricci, Bianchi 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 physical phenomena such as stress, torque, and divergence. At the beginning of the twentieth century Minkowski started developing the geometry of space-time as a mathematical model 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.

Much of what we do in this chapter carries over to the pseudo-Riemannian setting as long as we keep in mind how to calculate traces in this context.

P. Petersen, *Riemannian Geometry*, Graduate Texts in Mathematics 171,
DOI 10.1007/978-3-319-26654-1_2

2.1 Lie Derivatives

2.1.1 Directional Derivatives

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 Y on M we will use the following ways of writing the directional derivative of f in the direction of Y

$$\nabla_Y f = D_Y f = L_Y f = df(Y) = Y(f).$$

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 $\mathrm{grad} f = \nabla f$, defined as 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} (see also section 1.5.1). 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 use polar coordinates on $\mathbb{R}^2$, say, then

$$\nabla f = \partial_x (f) \partial_x + \partial_y (f) \partial_y \neq \partial_r (f) \partial_r + \partial_\theta (f) \partial_\theta,$$

One rule of thumb for items that are invariantly defined is that they should satisfy the Einstein summation convention. 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 Lie Derivatives

Let X be a vector field and F^t the corresponding locally defined flow on a smooth manifold M. Thus $F^t(p)$ is defined for small t and the curve $t \mapsto F^t(p)$ is the integral curve for X that goes through p at $t = 0$. The *Lie derivative* of a tensor in the direction of X is defined as the first-order term in a suitable Taylor expansion of the tensor when it is moved by the flow of X. The precise formula, however, depends on what type of tensor we use.

If $f : M \to \mathbb{R}$ is a function, then

$$f\left(F^t(p)\right) = f(p) + t(L_X f)(p) + o(t),$$

or

$$(L_X f)(p) = \lim_{t\to 0} \frac{f(F^t(p)) - f(p)}{t}.$$

Thus the Lie derivative $L_X f$ is simply the directional derivative $D_X f = df(X)$. Without specifying p we can also write

$$f \circ F^t = f + tL_X f + o(t) \text{ and } L_X f = D_X f = df(X).$$

When we have a vector field Y things get a little more complicated as $Y|_{F^t}$ can't be compared directly to Y since the vectors live in different tangent spaces. Thus we consider the curve $t \mapsto DF^{-t}\left(Y|_{F^t(p)}\right)$ that lies in T_pM. When this is expanded in t near 0 we obtain an expression

$$DF^{-t}\left(Y|_{F^t(p)}\right) = Y|_p + t(L_X Y)|_p + o(t)$$

for some vector $(L_X Y)|_p \in T_pM$. In other words we define

$$(L_X Y)|_p = \lim_{t\to 0} \frac{DF^{-t}\left(Y|_{F^t(p)}\right) - Y|_p}{t}.$$

This Lie derivative turns out to be the Lie bracket.

Proposition 2.1.1. *If X, Y are vector fields on M, then $L_X Y = [X, Y]$.*

Proof. While Lie derivatives are defined as a limit of suitable difference quotients it is generally far more convenient to work with their implicit definition through the first-order Taylor expansion.

The Lie derivative comes from

$$DF^{-t}(Y|_{F^t}) = Y + tL_X Y + o(t)$$

or equivalently

$$Y|_{F^t} - DF^t(Y) = tDF^t(L_X Y) + o(t).$$

Consider the directional derivative of a function f in the direction of $Y|_{F^t} - DF^t(Y)$

$$\begin{aligned} D_{Y|_{F^t} - DF^t(Y)} f &= D_{Y|_{F^t}} f - D_{DF^t(Y)} f \\ &= (D_Y f) \circ F^t - D_Y\left(f \circ F^t\right) \end{aligned}$$

$$
\begin{aligned}
&= D_Y f + tD_X D_Y f + o(t) \\
&\quad - D_Y (f + tD_X f + o(t)) \\
&= t(D_X D_Y f - D_Y D_X f) + o(t) \\
&= tD_{[X,Y]} f + o(t).
\end{aligned}
$$

This shows that

$$
\begin{aligned}
L_X Y &= \lim_{t \to 0} \frac{Y|_{F^t} - DF^t(Y)}{t} \\
&= [X, Y].
\end{aligned}
$$

□

We are now ready to define the Lie derivative of a $(0,k)$-tensor T and also give an algebraic formula for this derivative. Define

$$
\left(F^t\right)^* T = T + t(L_X T) + o(t)
$$

or with variables included

$$
\begin{aligned}
\left(\left(F^t\right)^* T\right)(Y_1, \dots, Y_k) &= T\left(DF^t(Y_1), \dots, DF^t(Y_k)\right) \\
&= T(Y_1, \dots, Y_k) + t(L_X T)(Y_1, \dots, Y_k) + o(t).
\end{aligned}
$$

As a difference quotient this means

$$
(L_X T)(Y_1, \dots, Y_k) = \lim_{t \to 0} \frac{(F^t)^* T - T}{t}.
$$

Proposition 2.1.2. *If X is a vector field and T a $(0,k)$-tensor on M, then*

$$
(L_X T)(Y_1, \dots, Y_k) = D_X(T(Y_1, \dots, Y_k)) - \sum_{i=1}^{k} T(Y_1, \dots, L_X Y_i, \dots, Y_k).
$$

Proof. We restrict attention to the case where $k = 1$. The general case is similar but requires more notation. Using that

$$
Y|_{F^t} = DF^t(Y) + tDF^t(L_X Y) + o(t)
$$

we get

$$
\left(\left(F^t\right)^* T\right)(Y) = T\left(DF^t(Y)\right)
$$

$$
\begin{aligned}
&= T\left(Y|_{F^t} - tDF^t\left(L_X Y\right)\right) + o\left(t\right) \\
&= T\left(Y\right) \circ F^t - tT\left(DF^t\left(L_X Y\right)\right) + o\left(t\right) \\
&= T\left(Y\right) + tD_X\left(T\left(Y\right)\right) - tT\left(DF^t\left(L_X Y\right)\right) + o\left(t\right).
\end{aligned}
$$

Thus

$$
\begin{aligned}
\left(L_X T\right)\left(Y\right) &= \lim_{t\to 0} \frac{\left(\left(F^t\right)^* T\right)\left(Y\right) - T\left(Y\right)}{t} \\
&= \lim_{t\to 0} \left(D_X\left(T\left(Y\right)\right) - T\left(DF^t\left(L_X Y\right)\right)\right) \\
&= D_X\left(T\left(Y\right)\right) - T\left(L_X Y\right).
\end{aligned}
$$

□

Finally, we have that Lie derivatives satisfy all possible product rules, i.e., they are *derivations*. From the above propositions this is already obvious when multiplying functions with vector fields or $(0, k)$-tensors.

Proposition 2.1.3. *If T_1 and T_2 be $(0, k_i)$-tensors, then*

$$
L_X\left(T_1 \cdot T_2\right) = \left(L_X T_1\right) \cdot T_2 + T_1 \cdot \left(L_X T_2\right).
$$

Proof. Recall that for 1-forms and more general $(0, k)$-tensors we define the product as

$$
T_1 \cdot T_2\left(X_1, \ldots, X_{k_1}, Y_1, \ldots, Y_{k_2}\right) = T_1\left(X_1, \ldots, X_{k_1}\right) \cdot T_2\left(Y_1, \ldots, Y_{k_2}\right).
$$

The proposition is then a simple consequence of the previous proposition and the product rule for derivatives of functions. □

Proposition 2.1.4. *If T is a $(0, k)$-tensor and $f : M \to \mathbb{R}$ a function, then*

$$
L_{fX} T\left(Y_1, \ldots, Y_k\right) = f L_X T\left(Y_1, \ldots, Y_k\right) + \sum_{i=1}^{k} \left(L_{Y_i} f\right) T\left(Y_1, \ldots, X, \ldots, Y_k\right).
$$

Proof. We have that

$$
\begin{aligned}
L_{fX} T\left(Y_1, \ldots, Y_k\right) &= D_{fX}\left(T\left(Y_1, \ldots, Y_k\right)\right) - \sum_{i=1}^{k} T\left(Y_1, \ldots, L_{fX} Y_i, \ldots, Y_k\right) \\
&= f D_X\left(T\left(Y_1, \ldots, Y_p\right)\right) - \sum_{i=1}^{k} T\left(Y_1, \ldots, \left[fX, Y_i\right], \ldots, Y_k\right)
\end{aligned}
$$

$$= fD_X\left(T\left(Y_1,\ldots,Y_p\right)\right) - f\sum_{i=1}^{k} T\left(Y_1,\ldots,[X,Y_i],\ldots,Y_k\right)$$
$$+\sum_{i=1}^{k}\left(L_{Y_i}f\right)T\left(Y_1,\ldots,X,\ldots,Y_k\right).$$

□

The case where $X|_p = 0$ is of special interest when computing Lie derivatives. We note that $F^t(p) = p$ for all t. Thus $DF^t : T_pM \to T_pM$ and

$$\begin{aligned} L_X Y|_p &= \lim_{t\to 0}\frac{DF^{-t}\left(Y|_p\right) - Y|_p}{t} \\ &= \frac{d}{dt}\left(DF^{-t}\right)|_{t=0}\left(Y|_p\right). \end{aligned}$$

This shows that $L_X = \frac{d}{dt}\left(DF^{-t}\right)|_{t=0}$ when $X|_p = 0$. From this we see that if θ is a 1-form then $L_X\theta = -\theta \circ L_X$ at points p where $X|_p = 0$. This is a general phenomenon.

Lemma 2.1.5. *If a vector field X vanishes at p, then the Lie derivative $L_X T$ at p depends only on the value of T at p.*

Proof. We have that

$$\left(L_X T\right)\left(Y_1,\ldots,Y_k\right) = D_X\left(T\left(Y_1,\ldots,Y_k\right)\right) - \sum_{i=1}^{k} T\left(Y_1,\ldots,L_X Y_i,\ldots,Y_k\right).$$

So if X vanishes at p, then

$$\left(L_X T\right)\left(Y_1,\ldots,Y_k\right)|_p = -\sum_{i=1}^{k} T\left(Y_1,\ldots,L_X Y_i,\ldots,Y_k\right)|_p.$$

□

It is also possible to define Lie derivatives of more general tensors and even multilinear maps on vector fields. An important instance of this is the Lie derivative of the Lie bracket $[Y,Z]$ or even the Lie derivative of the Lie derivative $L_Y T$. This is algebraically defined as

$$\begin{aligned} \left(L_X L\right)_Y T &= L_X\left(L_Y T\right) - L_{L_X Y}T - L_Y\left(L_X T\right) \\ &= \left[L_X, L_Y\right]T - L_{[X,Y]}T. \end{aligned}$$

Proposition 2.1.6 (The Generalized Jacobi Identity). *For all vector fields X, Y and tensors T*

$$(L_X L)_Y T = 0.$$

Proof. When T is a function this follows from the definition of the Lie bracket:

$$\begin{aligned}(L_X L)_Y f &= [L_X, L_Y] f - L_{[X,Y]} f \\ &= [D_X, D_Y] f - D_{[X,Y]} f \\ &= 0.\end{aligned}$$

When $T = Z$ is a vector field it is the usual Jacobi identity:

$$\begin{aligned}(L_X L)_Y Z &= [L_X, L_Y] Z - L_{[X,Y]} Z \\ &= [X, [Y, Z]] - [Y, [X, Z]] - [[X, Y], Z] \\ &= [X, [Y, Z]] + [Z, [X, Y]] + [Y, [Z, X]] \\ &= 0.\end{aligned}$$

When $T = \omega$ is a one-form it follows automatically from those two observations provided we know that

$$([L_X, L_Y] \omega) (Z) = [L_X, L_Y] (\omega (Z)) - \omega ([L_X, L_Y] Z)$$

since we then have

$$\begin{aligned}((L_X L)_Y \omega) (Z) &= ([L_X, L_Y] \omega) (Z) - \left(L_{[X,Y]} \omega\right) (Z) \\ &= [L_X, L_Y] (\omega (Z)) - \omega ([L_X, L_Y] Z) \\ &\quad - L_{[X,Y]} (\omega (Z)) + \omega \left(L_{[X,Y]} Z\right) \\ &= 0.\end{aligned}$$

A few cancellations must occur for the first identity to hold. Note that

$$([L_X, L_Y] \omega) (Z) = (L_X (L_Y \omega)) (Z) - (L_Y (L_X \omega)) (Z),$$

$$\begin{aligned}(L_X (L_Y \omega)) (Z) &= L_X ((L_Y \omega) (Z)) - (L_Y \omega) (L_X Z) \\ &= L_X (L_Y (\omega (Z))) - L_X (\omega (L_Y Z)) \\ &\quad - L_Y (\omega (L_X Z)) + \omega (L_Y L_X Z),\end{aligned}$$

and similarly

$$(L_Y (L_X \omega)) (Z) = L_Y (L_X (\omega (Z))) - L_Y (\omega (L_X Z)) - L_X (\omega (L_Y Z)) + \omega (L_X L_Y Z) .$$

This shows that

$$([L_X, L_Y] \omega) (Z) = [L_X, L_Y] (\omega (Z)) - \omega ([L_X, L_Y] Z) .$$

The proof for general tensors now follows by observing that these are tensor products of the above three simple types of tensors and that Lie derivatives act as derivations. □

The Lie derivative can also be used to give a formula for the exterior derivative of a k-form

$$\begin{aligned} d\omega (X_0, X_1, \dots, X_k) &= \frac{1}{2} \sum_{i=0}^{k} (-1)^i (L_{X_i} \omega) \left(X_0, \dots, \widehat{X}_i, \dots, X_k \right) . \\ &+ \frac{1}{2} \sum_{i=0}^{k} (-1)^i L_{X_i} \left(\omega \left(X_0, \dots, \widehat{X}_i, \dots, X_k \right) \right) \end{aligned}$$

For a 1-form this gives us the usual definition

$$d\omega (X, Y) = D_X (\omega (Y)) - D_Y (\omega (X)) - \omega ([X, Y]) .$$

2.1.3 Lie Derivatives and the Metric

The Lie derivative allows us to define the *Hessian* of a function on a Riemannian manifold as a $(0, 2)$-tensor:

$$\operatorname{Hess} f (X, Y) = \frac{1}{2} \left(L_{\nabla f} g \right) (X, Y) .$$

At a critical point for f this gives the expected answer. To see this, select coordinates x^i around p such that the metric coefficients satisfy $g_{ij}|_p = \delta_{ij}$. If $df|_p = 0$, then $\nabla f|_p = 0$ and it follows that

$$\begin{aligned} L_{\nabla f} \left(g_{ij} dx^i dx^j \right) |_p &= L_{\nabla f} \left(g_{ij} \right) |_p + \delta_{ij} L_{\nabla f} \left(dx^i \right) dx^j + \delta_{ij} dx^i L_{\nabla f} \left(dx^j \right) \\ &= \delta_{ij} L_{\nabla f} \left(dx^i \right) dx^j + \delta_{ij} dx^i L_{\nabla f} \left(dx^j \right) \\ &= L_{\nabla f} \left(\delta_{ij} dx^i dx^j \right) |_p . \end{aligned}$$

Thus $\operatorname{Hess} f|_p$ is the same if we compute it using g and the Euclidean metric in the fixed coordinate system.

It is perhaps still not clear why the Lie derivative formula for the Hessian is reasonable. The idea is that the Hessian measures how the metric changes as we flow along the gradient field. To justify this better let us define the *divergence* of a vector field X as the function $\operatorname{div} X$ that measures how the volume form changes along the flow for X:

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

Note that the form $L_X \operatorname{vol}$ is always exact as

$$L_X \operatorname{vol} = d i_X \operatorname{vol},$$

where $i_X T$ evaluates T on X in the first variable.

The *Laplacian* of a function is defined as in vector calculus by

$$\Delta f = \operatorname{div} \nabla f$$

and we claim that it is also given as the trace of the Hessian. To see this select a positively oriented orthonormal frame E_i and note that

$$\begin{aligned}
\operatorname{div} X &= (L_X \operatorname{vol})(E_1, \ldots, E_n) \\
&= L_X (\operatorname{vol}(E_1, \ldots, E_n)) \\
&\quad - \sum \operatorname{vol}(E_1, \ldots, L_X E_i, \ldots, E_n) \\
&= -\sum g(L_X E_i, E_i) \\
&= \frac{1}{2} \sum (L_X (g(E_i, E_i)) - g(L_X E_i, E_i) - g(E_i, L_X E_i)) \\
&= \sum \frac{1}{2} (L_X g)(E_i, E_i).
\end{aligned}$$

We can also show that the Hessian defined in this way gives us back the usual Hessian of a function $f : \mathbb{R}^n \to \mathbb{R}$ with the canonical metric on Euclidean space:

$$\begin{aligned}
L_{\nabla f} \left(\delta_{ij} dx^i dx^j \right) &= L_{\sum \partial_j f \partial_j} \sum dx^i dx^i \\
&= \sum L_{\partial_j f \partial_j} dx^i dx^i \\
&= \sum \left(L_{\partial_j f \partial_j} dx^i \right) dx^i + \sum dx^i \left(L_{\partial_j f \partial_j} dx^i \right) \\
&= \sum \partial_j f \left(L_{\partial_j} dx^i \right) dx^i + \sum \partial_j f dx^i \left(L_{\partial_j} dx^i \right)
\end{aligned}$$

$$
\begin{aligned}
&= +\sum d\left(\partial_j f\right) dx^i \left(\partial_j\right) dx^i + \sum d\left(\partial_j f\right) dx^i dx^i \left(\partial_j\right) \\
&= 2\sum d\left(\partial_i f\right) dx^i \\
&= 2\sum \partial_{ji} f dx^j dx^i \\
&= 2\,\mathrm{Hess} f.
\end{aligned}
$$

2.1.4 Lie Groups

Lie derivatives as might be expected also come in handy when working with Lie groups. For a Lie group G we have the inner automorphism $\mathrm{Ad}_h : x \mapsto hxh^{-1}$and its differential at $x = e$ denoted by the same letters $\mathrm{Ad}_h : \mathfrak{g} \to \mathfrak{g}$.

Lemma 2.1.7. *The differential of $h \mapsto \mathrm{Ad}_h$ is given by $U \mapsto \mathrm{ad}_U(X) = [U, X]$.*

Proof. If we write $\mathrm{Ad}_h(x) = R_{h^{-1}} L_h(x)$, then its differential at $x = e$ is given by $\mathrm{Ad}_h = DR_{h^{-1}} DL_h$. Now let F^t be the flow for U. Then $F^t(x) = xF^t(e) = L_x(F^t(e))$ as both curves go through x at $t = 0$ and have U as tangent everywhere since U is a left-invariant vector field. This also shows that $DF^t = DR_{F^t(e)}$. Thus

$$
\begin{aligned}
\mathrm{ad}_U(X)|_e &= \frac{d}{dt} DR_{F^{-t}(e)} DL_{F^t(e)} \left(X|_e\right)|_{t=0} \\
&= \frac{d}{dt} DR_{F^{-t}(e)} \left(X|_{F^t(e)}\right)|_{t=0} \\
&= \frac{d}{dt} DF^{-t} \left(X|_{F^t(e)}\right)|_{t=0} \\
&= L_U X = [U, X].
\end{aligned}
$$

□

This is used in the next lemma.

Lemma 2.1.8. *Let* $\mathrm{G} = \mathrm{GL}(V)$ *be the Lie group of invertible matrices on* V. *The Lie bracket structure on the Lie algebra* $\mathfrak{gl}(V)$ *of left-invariant vector fields on* $\mathrm{GL}(V)$ *is given by commutation of linear maps. i.e., if* $X, Y \in T_I \mathrm{GL}(V)$, *then*

$$[X, Y]|_I = XY - YX.$$

Proof. Since $x \mapsto hxh^{-1}$ is a linear map on the space $\mathrm{Hom}(V, V)$ we see that $\mathrm{Ad}_h(X) = hXh^{-1}$. The flow of U is given by $F^t(g) = g(I + tU + o(t))$ so we have

$$
\begin{aligned}
[U,X] &= \frac{d}{dt}\left(F^t(I)\,XF^{-t}(I)\right)|_{t=0} \\
&= \frac{d}{dt}\left((I+tU+o(t))\,X\,(I-tU+o(t))\right)|_{t=0} \\
&= \frac{d}{dt}\left(X+tUX-tXU+o(t)\right)|_{t=0} \\
&= UX-XU.
\end{aligned}
$$

□

2.2 Connections

2.2.1 Covariant Differentiation

We now come to the question of attaching a meaning to the change of a vector field. The Lie derivative is one possibility, but it is not a strong enough concept as it doesn't characterize the Cartesian coordinate fields in $\mathbb{R}^n$ as having zero derivative. A better strategy for $\mathbb{R}^n$ is to write $X = X^i\partial_i$, where ∂_i are the Cartesian coordinate fields. If we want the coordinate vector fields to have zero derivative, then it is natural to define the *covariant derivative* of X in the direction of Y as

$$\nabla_Y X = \left(\nabla_Y X^i\right)\partial_i = d\left(X^i\right)(Y)\,\partial_i.$$

Thus we measure the change in X by measuring how the coefficients change. Therefore, a vector field with constant coefficients does not change. This formula clearly depends on the fact that we used Cartesian coordinates and is not invariant under change of coordinates. If we take the coordinate vector fields

$$\partial_r = \frac{1}{r}\left(x\partial_x + y\partial_y\right),\ \partial_\theta = -y\partial_x + x\partial_y$$

that come from polar coordinates in $\mathbb{R}^2$, then we see that they are not constant.

In order to better understand such derivatives we need to find a coordinate independent definition. This is done most easily by splitting the problem of defining the change in a vector field X into two problems.

First, we can measure the change in X by asking whether or not X is a gradient field. If $i_X g = \theta_X$ is the 1-form dual to X, i.e., $(i_X g)(Y) = g(X,Y)$, then we know that X is locally the gradient of a function if and only if $d\theta_X = 0$. In general, the 2-form $d\theta_X$ then measures the extent to which X is a gradient field.

Second, we can measure how a vector field X changes the metric via the Lie derivative $L_X g$. This is a symmetric $(0,2)$-tensor as opposed to the skew-symmetric $(0,2)$-tensor $d\theta_X$. If F^t is the local flow for X, then we see that $L_X g = 0$ if and only if F^t are isometries (see also section 8.1). When this happens we say that X is a *Killing field*.

In case $X = \nabla f$ is a gradient field we saw that the expression $\frac{1}{2}L_{\nabla f}g$ is the Hessian of f. From that calculation we can also quickly see what the Killing fields on $\mathbb{R}^n$ should be: If $X = X^i\partial_i$, then X is a Killing field if and only if $\partial_k X^i + \partial_i X^k = 0$. This implies that

$$\begin{aligned}\partial_j\partial_k X^i &= -\partial_j\partial_i X^k \\ &= -\partial_i\partial_j X^k \\ &= \partial_i\partial_k X^j \\ &= \partial_k\partial_i X^j \\ &= -\partial_k\partial_j X^i \\ &= -\partial_j\partial_k X^i.\end{aligned}$$

Thus we have $\partial_j\partial_k X^i = 0$ and hence

$$X^i = \alpha^i_j x^j + \beta^i$$

with the extra conditions that

$$\alpha^i_j = \partial_j X^i = -\partial_i X^j = -\alpha^j_i.$$

In particular, the angular field ∂_θ is a Killing field. This also follows from the fact that the corresponding flow is matrix multiplication by the orthogonal matrix

$$\begin{bmatrix}\cos(t) & -\sin(t)\\ \sin(t) & \cos(t)\end{bmatrix}.$$

More generally, one can show that the flow of the Killing field X is

$$F^t(x) = \exp(At)\,x + t\beta,\ A = \left[\alpha^i_j\right],\ \beta = \left[\beta^i\right].$$

In this way we see that a vector field on $\mathbb{R}^n$ is constant if and only if it is both a Killing field and a gradient field.

Finally we make the important observation.

Proposition 2.2.1. *The covariant derivative in $\mathbb{R}^n$ is given by the implicit formula:*

$$2g(\nabla_Y X, Z) = (L_X g)(Y,Z) + (d\theta_X)(Y,Z).$$

Proof. Since both sides are tensorial in Y and Z it suffices to check the formula on the Cartesian coordinate vector fields. Write $X = a^i\partial_i$ and calculate the right-hand side

$$
\begin{aligned}
(L_X g)(\partial_k, \partial_l) + (d\theta_X)(\partial_k, \partial_l) &= D_X \delta_{kl} - g(L_X \partial_k, \partial_l) - g(\partial_k, L_X \partial_l) \\
&\quad + \partial_k g(X, \partial_l) - \partial_l g(X, \partial_k) - g(X, [\partial_k, \partial_l]) \\
&= -g\left(L_{a^i \partial_i} \partial_k, \partial_l\right) - g\left(\partial_k, L_{a^j \partial_j} \partial_l\right) \\
&\quad + \partial_k a^l - \partial_l a^k \\
&= -g\left(-\left(\partial_k a^i\right) \partial_i, \partial_l\right) - g\left(\partial_k, -\left(\partial_l a^j\right) \partial_j\right) \\
&\quad + \partial_k a^l - \partial_l a^k \\
&= +\partial_k a^l + \partial_l a^k + \partial_k a^l - \partial_l a^k \\
&= 2\partial_k a^l \\
&= 2g\left(\left(\partial_k a^i\right) \partial_i, \partial_l\right) \\
&= 2g\left(\nabla_{\partial_k} X, \partial_l\right).
\end{aligned}
$$

□

Since the right-hand side in the formula for $\nabla_Y X$ makes sense on any Riemannian manifold we can use this to give an implicit definition of the *covariant derivative* of X in the direction of Y. This covariant derivative turns out to be uniquely determined by the following properties.

Theorem 2.2.2 (The Fundamental Theorem of Riemannian Geometry). *The assignment $X \mapsto \nabla X$ on (M, g) is uniquely defined by the following properties:*

(1) *$Y \mapsto \nabla_Y X$ is a $(1,1)$-tensor, i.e., it is well-defined for tangent vectors and linear*

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

(2) *$X \mapsto \nabla_Y X$ is a* derivation:

$$
\begin{aligned}
\nabla_Y (X_1 + X_2) &= \nabla_Y X_1 + \nabla_Y X_2, \\
\nabla_Y (fX) &= (D_Y f) X + f \nabla_Y X
\end{aligned}
$$

for functions $f : M \to \mathbb{R}$.

(3) *Covariant differentiation is* torsion free:

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

(4) *Covariant differentiation is* metric:

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

Proof. We have already established (1) by using that

$$(L_X g)(Y,Z) + (d\theta_X)(Y,Z)$$

is tensorial in Y and Z. This also shows that the expression is linear in X. To check the derivation rule we observe that

$$\begin{aligned}
L_{fX} g + d\theta_{fX} &= fL_X g + df \cdot \theta_X + \theta_X \cdot df + d(f\theta_X) \\
&= fL_X g + df \cdot \theta_X + \theta_X \cdot df + df \wedge \theta_X + fd\theta_X \\
&= f(L_X g + d\theta_X) + df \cdot \theta_X + \theta_X \cdot df + df \cdot \theta_X - \theta_X \cdot df \\
&= f(L_X g + d\theta_X) + 2df \cdot \theta_X.
\end{aligned}$$

Thus

$$\begin{aligned}
2g(\nabla_Y(fX), Z) &= f2g(\nabla_Y X, Z) + 2df(Y)\, g(X,Z) \\
&= 2g(f\nabla_Y X + df(Y)\, X, Z) \\
&= 2g(f\nabla_Y X + (D_Y f)\, X, Z).
\end{aligned}$$

To establish the next two claims it is convenient to create the following expansion also known as *Koszul's formula*.

$$\begin{aligned}
2g(\nabla_Y X, Z) &= (L_X g)(Y,Z) + (d\theta_X)(Y,Z) \\
&= D_X g(Y,Z) - g([X,Y],Z) - g(Y,[X,Z]) \\
&\quad + D_Y\theta_X(Z) - D_Z\theta_X(Y) - \theta_X([Y,Z]) \\
&= D_X g(Y,Z) - g([X,Y],Z) - g(Y,[X,Z]) \\
&\quad + D_Y g(X,Z) - D_Z g(X,Y) - g(X,[Y,Z]) \\
&= D_X g(Y,Z) + D_Y g(Z,X) - D_Z g(X,Y) \\
&\quad - g([X,Y],Z) - g([Y,Z],X) + g([Z,X],Y).
\end{aligned}$$

We then see that (3) follows from

$$\begin{aligned}
2g(\nabla_X Y - \nabla_Y X, Z) &= D_Y g(X,Z) + D_X g(Z,Y) - D_Z g(Y,X) \\
&\quad - g([Y,X],Z) - g([X,Z],Y) + g([Z,Y],X) \\
&\quad - D_X g(Y,Z) - D_Y g(Z,X) + D_Z g(X,Y) \\
&\quad + g([X,Y],Z) + g([Y,Z],X) - g([Z,X],Y) \\
&= 2g([X,Y],Z).
\end{aligned}$$

And (4) from

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

Conversely, if we have a covariant derivative $\bar{\nabla}_Y X$ with these four properties, then

$$\begin{aligned}
2g\left(\nabla_Y X, Z\right) &= \left(L_X g\right)\left(Y, Z\right) + \left(d\theta_X\right)\left(Y, Z\right) \\
&= D_X g\left(Y, Z\right) + D_Y g\left(Z, X\right) - D_Z g\left(X, Y\right) \\
&\quad -g\left(\left[X, Y\right], Z\right) - g\left(\left[Y, Z\right], X\right) + g\left(\left[Z, X\right], Y\right) \\
&= g\left(\bar{\nabla}_X Y, Z\right) + g\left(Y, \bar{\nabla}_X Z\right) + g\left(\bar{\nabla}_Y Z, X\right) + g\left(Z, \bar{\nabla}_Y X\right) \\
&\quad -g\left(\bar{\nabla}_Z X, Y\right) - g\left(X, \bar{\nabla}_Z Y\right) + g\left(\bar{\nabla}_Z X, Y\right) - g\left(\bar{\nabla}_X Z, Y\right) \\
&\quad -g\left(\bar{\nabla}_X Y, Z\right) + g\left(\bar{\nabla}_Y X, Z\right) - g\left(\bar{\nabla}_Y Z, X\right) + g\left(\bar{\nabla}_Z Y, X\right) \\
&= 2g\left(\bar{\nabla}_Y X, Z\right)
\end{aligned}$$

showing that $\nabla_Y X = \bar{\nabla}_Y X$. □

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 that in addition also satisfies (3) and (4), then we call it a *Riemannian connection*. As we just saw, this connection is uniquely defined by these four properties and is given implicitly through the formula

$$\begin{aligned}
2g\left(\nabla_Y X, Z\right) &= \left(L_X g\right)\left(Y, Z\right) + \left(d\theta_X\right)\left(Y, Z\right) \\
&= D_X g\left(Y, Z\right) + D_Y g\left(Z, X\right) - D_Z g\left(X, Y\right) \\
&\quad -g\left(\left[X, Y\right], Z\right) - g\left(\left[Y, Z\right], X\right) + g\left(\left[Z, X\right], Y\right).
\end{aligned}$$

Before proceeding we need to discuss how $\nabla_Y X$ depends on X and Y. Since $\nabla_Y X$ is tensorial in Y, we see that the value of $\nabla_Y X$ at $p \in M$ depends only on $Y|_p$. But in what way does it depend on X? Since $X \mapsto \nabla_Y X$ is a derivation, it is definitely not tensorial in X. Therefore, we cannot expect $\left(\nabla_Y X\right)|_p$ to depend only on $X|_p$ and $Y|_p$. The next two lemmas explore how $\left(\nabla_Y X\right)|_p$ depends on X.

Lemma 2.2.3. *Let M be a manifold and ∇ an affine 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 $\lambda : M \to \mathbb{R}$ such that $\lambda \equiv 0$ on $M - U$ and $\lambda \equiv 1$ in a neighborhood of p. Then $\lambda X = \lambda Y$ on M. Thus at p

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

since $d\lambda|_p = 0$ and $\lambda(p) = 1$. In particular,

$$\nabla_v X = \nabla_v \lambda X = \nabla_v \lambda Y = \nabla_v Y.$$

$\square$

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

Lemma 2.2.4. *Let M be a manifold and ∇ an affine connection on M. If X is a vector field on M and $c : I \to M$ a smooth curve with $\dot{c}(0) = v \in T_pM$, then $\nabla_v X$ depends only on the values of X along c, i.e., if $X \circ c = Y \circ c$, then $\nabla_{\dot{c}} X = \nabla_{\dot{c}} Y$.*

Proof. Choose a frame $E_1, \ldots, E_n$ in a neighborhood of p and write $Y = \sum Y^i E_i$, $X = \sum X^i E_i$ on this neighborhood. From the assumption that $X \circ c = Y \circ c$ we get that $X^i \circ c = Y^i \circ c$. Thus,

$$\begin{aligned}
\nabla_v Y &= \nabla_v \left(Y^i E_i\right) \\
&= Y^i(p)\nabla_v E_i + E_i(p) dY^i(v) \\
&= X^i(p)\nabla_v E_i + E_i(p) dX^i(v) \\
&= \nabla_v X.
\end{aligned}$$

$\square$

This shows that $\nabla_v X$ makes sense as long as X is prescribed along some curve (or submanifold) that has v as a tangent.

It will occasionally be convenient to use coordinates or orthonormal frames with certain nice properties. We say that a coordinate system is *normal* at p if $g_{ij}|_p = \delta_{ij}$ and $\partial_k g_{ij}|_p = 0$. An orthonormal frame E_i is *normal* at $p \in M$ if $\nabla_v E_i(p) = 0$ for all $i = 1, \ldots, n$ and $v \in T_pM$. It is not hard to show that such coordinates and frames always exist (see exercises 2.5.20 and 2.5.19).

2.2.2 Covariant Derivatives of Tensors

The connection, as we shall see, is also useful in generalizing many of the well-known concepts (such as Hessian, Laplacian, divergence) from multivariable calculus to the Riemannian setting (see also section 2.1.3).

If S is a (s,t)-tensor field, then we can define a *covariant derivative* ∇S that we interpret as an $(s,t+1)$-tensor field. Recall 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 for a $(1,1)$-tensor S we should have

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

Therefore, it seems reasonable to define ∇S as

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

In other words

$$\nabla_X S = [\nabla_X, S].$$

It is easily checked that $\nabla_X S$ is still tensorial in Y.

More generally, when $s = 0, 1$ we obtain

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

Here ∇_X is interpreted as the directional derivative when applied to a function and covariant differentiation on vector fields. This also makes sense when $s \geq 2$, if we make sense of defining covariant derivatives of, say, tensor products of vector fields. This can also be done using the product rule:

$$\nabla_X (X_1 \otimes X_2) = (\nabla_X X_1) \otimes X_2 + X_1 \otimes (\nabla_X X_2).$$

A tensor is said to be *parallel* if $\nabla S \equiv 0$. In Euclidean space one can easily show that if a tensor is written in Cartesian coordinates, then it is parallel if and only if it has constant coefficients. Thus $\nabla X \equiv 0$ for constant vector fields. On a Riemannian manifold (M, g) the metric and volume forms are always parallel.

Proposition 2.2.5. *On a Riemannian n-manifold* (M, g)

$$\nabla g = 0,$$
$$\nabla \operatorname{vol} = 0.$$

Proof. The metric is parallel due to property (4):

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

To check that the volume form is parallel we evaluate the covariant derivative on an orthonormal frame $E_1, \ldots, E_n$:

$$\begin{aligned}
(\nabla_X \operatorname{vol})(E_1, \ldots, E_n) &= \nabla_X \operatorname{vol}(E_1, \ldots, E_n) \\
&\quad - \sum \operatorname{vol}(E_1, \ldots, \nabla_X E_i, \ldots, E_n) \\
&= -\sum g(E_i, \nabla_X E_i) \\
&= -\frac{1}{2} \sum D_X (g(E_i, E_i)) \\
&= 0.
\end{aligned}$$

□

The covariant derivative gives us a different way of calculating the Hessian of a function.

Proposition 2.2.6. *If* $f : (M, g) \to \mathbb{R}$, *then*

$$(\nabla_X df)(Y) = g(\nabla_X \nabla f, Y) = \operatorname{Hess} f(X, Y).$$

Proof. First observe that

$$\begin{aligned}
(\nabla df)(X, Y) &= (\nabla_X df)(Y) \\
&= D_X D_Y f - df(\nabla_X Y) \\
&= D_X D_Y f - D_{\nabla_X Y} f.
\end{aligned}$$

This shows that

$$(\nabla_X df)(Y) - (\nabla_Y df)(X) = [D_X, D_Y] f - D_{[X,Y]} f = 0.$$

Thus $(\nabla_X df)(Y)$ is symmetric. This can be used to establish the formulas

$$\begin{aligned}
(\nabla df)(X,Y) &= (\nabla_X df)(Y) \\
&= D_X g(\nabla f, Y) - g(\nabla f, \nabla_X Y) \\
&= g(\nabla_X \nabla f, Y) \\
&= \frac{1}{2} g(\nabla_X \nabla f, Y) + \frac{1}{2} g(X, \nabla_Y \nabla f) \\
&= \frac{1}{2}\left(\nabla_{\nabla f} g\right)(X,Y) + \frac{1}{2} g(\nabla_X \nabla f, Y) + \frac{1}{2} g(X, \nabla_Y \nabla f) \\
&= \frac{1}{2} D_{\nabla f} g(X,Y) - \frac{1}{2} g([\nabla f, X], Y) - \frac{1}{2} g(X, [\nabla f, Y]) \\
&= \frac{1}{2}\left(L_{\nabla f} g\right)(X,Y).
\end{aligned}$$

□

2.2.2.1 The Adjoint of the Covariant Derivative

The *adjoint* to the covariant derivative on (s,t)-tensors with $t > 0$ is defined as

$$\left(\nabla^* S\right)(X_2,\ldots,X_r) = -\sum (\nabla_{E_i} S)(E_i, X_2, \ldots, X_r),$$

where $E_1,\ldots,E_n$ is an orthonormal frame. This means that while the covariant derivative adds a variable, the adjoint eliminates one. The adjoint is related to the divergence of a vector field (see section 2.1.3) by

Proposition 2.2.7. *If X is a vector field and θ_X the corresponding 1-form, then*

$$\operatorname{div} X = -\nabla^* \theta_X.$$

Proof. See section 2.1.3 for the definition of divergence. Select an orthonormal frame E_i, then

$$\begin{aligned}
-\nabla^* \theta_X &= \sum (\nabla_{E_i} \theta_X)(E_i) \\
&= \sum D_{E_i} g(X, E_i) - \sum g(X, \nabla_{E_i} E_i) \\
&= \sum g(\nabla_{E_i} X, E_i) \\
&= \sum \frac{1}{2}(L_X g)(E_i, E_i) \\
&= \operatorname{div} X.
\end{aligned}$$

□

The adjoint really is the adjoint of the covariant derivative with respect to the integrated inner product.

Proposition 2.2.8. *If S is a compactly supported (s,t)-tensor and T a compactly supported $(s,t+1)$-tensor, then*

$$\int g(\nabla S, T)\,\mathrm{vol} = \int g\left(S, \nabla^* T\right)\mathrm{vol}.$$

Proof. Define a 1-form by $\omega(X) = g(i_X T, S)$. To calculate its divergence more easily, select an orthonormal frame E_i such that $\nabla_v E_i = 0$ for all $v \in T_p M$. To further simplify things a bit assume that $s = t = 1$, then

$$\begin{aligned}
-\nabla^*\omega &= (\nabla_{E_i}\omega)(E_i) \\
&= \nabla_{E_i} g\left(T\left(E_i, E_j\right), S\left(E_j\right)\right) \\
&= g\left(\nabla_{E_i} T\left(E_i, E_j\right), S\left(E_j\right)\right) + g\left(T\left(E_i, E_j\right), \nabla_{E_i} S\left(E_j\right)\right) \\
&= -g\left(\nabla^* T, S\right) + g(T, \nabla S).
\end{aligned}$$

So the result follows by the divergence theorem or Stokes' theorem:

$$\int \operatorname{div} X\,\mathrm{vol} = \int d i_X\,\mathrm{vol} = 0,$$

where X is any compactly supported vector field. □

2.2.2.2 Exterior Derivatives

The covariant derivative gives us a very nice formula for exterior derivatives of forms as the skew-symmetrized covariant derivative:

$$(d\omega)(X_0,\dots,X_k) = \sum (-1)^i (\nabla_{X_i}\omega)\left(X_0,\dots,\hat{X}_i,\dots,X_k\right).$$

While the covariant derivative clearly depends on the metric this formula shows that for forms we can still obtain derivatives that do not depend on the metric. It will also allow us to define exterior derivatives of more complicated tensors. Suppose we have a $(1,k)$-tensor T that is skew-symmetric in the k variables. Then we can define the $(1,k+1)$-tensor

$$\left(d^\nabla T\right)(X_0,\dots,X_k) = \sum (-1)^i (\nabla_{X_i} T)\left(X_0,\dots,\hat{X}_i,\dots,X_k\right).$$

In case $k = 0$ the tensor $T = Y$ is a vector field and we obtain the $(1,1)$-tensor:

$$\left(d^\nabla Y\right)(X) = \nabla_X Y.$$

When $k = 1$ we have a $(1,1)$-tensor and obtain the $(1,2)$-tensor:

$$\begin{aligned}\left(d^{\nabla}T\right)(X,Y) &= (\nabla_X T)(Y) - (\nabla_Y T)(X)\\ &= \nabla_X (T(Y)) - \nabla_Y (T(X)) - T[X,Y].\end{aligned}$$

2.2.2.3 The Second Covariant Derivative

For a (s,t)-tensor field S we define the *second covariant derivative* $\nabla^2 S$ as the $(s, t+2)$-tensor field

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

With this we obtain another definition for the $(0,2)$ version of the Hessian of a function:

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

The second covariant derivative on functions is symmetric in X and Y. For more general tensors, 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 Riemannian geometries.

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

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

2.2.2.4 The Lie Derivative of the Covariant Derivative

We can define the Lie derivative of the connection in a way similar to the Lie derivative of the Lie bracket

$$\begin{aligned}(L_X \nabla)_U V &= (L_X \nabla)(U,V)\\ &= L_X(\nabla_U V) - \nabla_{L_X U} V - \nabla_U L_X V\\ &= [X, \nabla_U V] - \nabla_{[X,U]} V - \nabla_U [X,V].\end{aligned}$$

Since $[U, V] = \nabla_U V - \nabla_V U$ it follows that

$$(L_X \nabla)(U, V) - (L_X \nabla)(V, U) = L_X L_U V = 0.$$

Moreover as $\nabla_U V$ is tensorial in U the Lie derivative $(L_X \nabla)_U V$ will also be tensorial in U. The fact that it is also symmetric shows that it is tensorial in both variables.

2.2.2.5 The Covariant Derivative of the Covariant Derivative

We can also define the covariant derivative of the covariant derivative

$$(\nabla_X \nabla)_Y T = \nabla_X (\nabla_Y T) - \nabla_{\nabla_X Y} T - \nabla_Y (\nabla_X T).$$

Note however, that this is not tensorial in X!

It is related to the *second covariant derivative* of T by

$$\nabla^2_{X,Y} T = (\nabla_X \nabla)_Y T + \nabla_Y (\nabla_X T).$$

2.3 Natural Derivations

We've seen that there are many natural derivations on tensors coming from various combinations of derivatives. We shall attempt to tie these together in a natural and completely algebraic fashion by using that all $(1, 1)$-tensors naturally act as derivations on tensors.

For clarity we define a derivation on tensors as map $T \mapsto DT$ that preserves the type of the tensor T; is linear; commutes with contractions; and satisfies the product rule

$$D(T_1 \otimes T_2) = (DT_1) \otimes T_2 + T_1 \otimes DT_2.$$

2.3.1 Endomorphisms as Derivations

The goal is to show that $(1, 1)$-tensors naturally act as derivations on the space of all tensors.

We use the natural homomorphism

$$\mathrm{GL}(V) \to \mathrm{GL}(T(V)),$$

where $T(V)$ is the space of all tensors over the vector space V. This respects the natural grading of tensors: The subspace of (s, t)-tensors is spanned by

$$v_1 \otimes \cdots \otimes v_s \otimes \phi_1 \otimes \cdots \otimes \phi_t$$

where $v_1, \dots, v_s \in V$ and $\phi_1, \dots, \phi_t : V \to \mathbb{R}$ are linear functions. The natural homomorphism acts as follows: for $\alpha \in \mathbb{R}$ we have $g \cdot \alpha = 0$; for $v \in V$ we have $g \cdot v = g(v)$; for $\phi \in V^*$ we have $g \cdot \phi = \phi \circ g^{-1}$; and on general tensors

$$\begin{aligned} &g \cdot (v_1 \otimes \cdots \otimes v_s \otimes \phi_1 \otimes \cdots \otimes \phi_t) \\ &= g(v_1) \otimes \cdots \otimes g(v_s) \otimes \left(\phi_1 \circ g^{-1}\right) \otimes \cdots \otimes \left(\phi_t \circ g^{-1}\right). \end{aligned}$$

The derivative of this action yields a linear map

$$\mathrm{End}(V) \to \mathrm{End}(T(V)),$$

which for each $L \in \mathrm{End}(V)$ induces a derivation on $T(V)$. Specifically, if $L \in \mathrm{End}(V)$, then $Lv = L(v)$ on vectors; on 1-forms $L\phi = -\phi \circ L$; and on general tensors

$$\begin{aligned} &L(v_1 \otimes \cdots \otimes v_s \otimes \phi_1 \otimes \cdots \otimes \phi_t) \\ &= L(v_1) \otimes \cdots \otimes v_s \otimes \phi_1 \otimes \cdots \otimes \phi_t \\ &\quad + \cdots \\ &\quad + v_1 \otimes \cdots \otimes L(v_s) \otimes \phi_1 \otimes \cdots \otimes \phi_t \\ &\quad - v_1 \otimes \cdots \otimes v_s \otimes (\phi_1 \circ L) \otimes \cdots \otimes \phi_t \\ &\quad - \cdots \\ &\quad - v_1 \otimes \cdots \otimes v_s \otimes \phi_1 \otimes \cdots \otimes (\phi_t \circ L). \end{aligned}$$

As the natural derivation comes from an action that preserves symmetries of tensors we immediately obtain.

Proposition 2.3.1. *The linear map*

$$\begin{aligned} \mathrm{End}(V) &\to \mathrm{End}(T(V)) \\ L &\mapsto LT \end{aligned}$$

is a Lie algebra homomorphism that preserves symmetries of tensors.

We also need to show that it is a derivation.

Proposition 2.3.2. *Any* $(1,1)$*-tensor* L *defines a derivation on tensors.*

Proof. It is easy to see from the definition that it is linear and satisfies the product rule. So it remains to show that it commutes with contractions. Consider a $(1,1)$-tensor T and in a local frame X_i with associated coframe σ^i write it as $T = T^i_j X_i \otimes \sigma^j$. The contraction of T is scalar valued and simply the trace of T so we know that $L(\mathrm{tr}\, T) = 0$. On the other hand we have

$$\begin{aligned}
L(T) &= T^i_j L(X_i) \otimes \sigma^j - T^i_j X_i \otimes \sigma^j \circ L \\
&= T^i_j L^k_i X_k \otimes \sigma^j - T^i_j L^j_l X_i \otimes \sigma^l \\
&= T^i_l L^k_i X_k \otimes \sigma^l - T^k_j L^j_l X_k \otimes \sigma^l \\
&= \left(T^i_l L^k_i - T^k_j L^j_l \right) X_k \otimes \sigma^l
\end{aligned}$$

so

$$\operatorname{tr}(L(T)) = T^i_k L^k_i - T^k_j L^j_k = 0.$$

A similar strategy can be used for general tensors $T^{i_1 \cdots i_k}_{j_1 \cdots j_l}$ where we trace or contract over a fixed superscript and subscript. □

We also need to know how this derivation interacts with an inner product. The inner product on $T(V)$ is given by declaring

$$e_{i_1} \otimes \cdots \otimes e_{i_p} \otimes e^{j_1} \otimes \cdots \otimes e^{j_q}$$

an orthonormal basis when $e_1, \ldots, e_n$ is an orthonormal basis for V and $e^1, \ldots, e^n$ the dual basis for V^*.

Proposition 2.3.3. *Assume V has an inner product:*

(1) *The adjoint of $L : V \to V$ extends to become the adjoint for $L : T(V) \to T(V)$.*
(2) *If $L \in \mathfrak{so}(V)$, i.e., L is skew-adjoint, then L commutes with type change of tensors.*

2.3.2 Derivatives

One can easily show that both the Lie derivative L_U and the covariant derivative ∇_U act as derivations on tensors (see exercises 2.5.9 and 2.5.10). However, these operations are nontrivial on functions. Therefore, they are not of the type we just introduced above.

Proposition 2.3.4. *If we think of ∇U as the $(1,1)$-tensor $X \mapsto \nabla_X U$, then*

$$L_U = \nabla_U - (\nabla U).$$

Proof. It suffices to check that this identity holds on vector fields and functions. On functions it reduces to the definition of directional derivatives, on vectors from the definition of Lie brackets and the torsion free property of the connection. □

This proposition indicates that one can make sense of the expression $\nabla_T U$ where T is a tensor and U a vector field. It has in other places been named $A_X T$, but as that now generally has been accepted as the A-tensor for a Riemannian submersion we have not adopted this notation.

2.4 The Connection in Tensor Notation

In a local coordinate system the metric is written as $g = g_{ij}dx^i dx^j$. So if $X = X^i \partial_i$ and $Y = Y^j \partial_j$ are vector fields, then

$$g(X, Y) = g_{ij} X^i Y^j.$$

We can also compute the dual 1-form θ_X to X by:

$$\begin{aligned}\theta_X &= g(X, \cdot) \\ &= g_{ij} dx^i (X)\, dx^j (\cdot) \\ &= g_{ij} X^i dx^j.\end{aligned}$$

The inverse of the matrix $\left[g_{ij}\right]$ is denoted $\left[g^{ij}\right]$. Thus we have

$$\delta^i_j = g^{ik} g_{kj}.$$

The vector field X dual to a 1-form $\omega = \omega_i dx^i$ is defined implicitly by

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

In other words we have

$$\theta_X = g_{ij} X^i dx^j = \omega_j dx^j = \omega.$$

This shows that

$$g_{ij} X^i = \omega_j.$$

In order to isolate X^i we have to multiply by g^{kj} on both sides and also use the symmetry of g_{ij}

$$\begin{aligned}g^{kj}\omega_j &= g^{kj} g_{ij} X^i \\ &= g^{kj} g_{ji} X^i \\ &= \delta^k_i X^i \\ &= X^k.\end{aligned}$$

Therefore,

$$\begin{aligned}X &= X^i \partial_i \\ &= g^{ij} \omega_j \partial_i.\end{aligned}$$

The gradient field of a function is a particularly important example of this construction

$$\begin{aligned} \nabla f &= g^{ij}\partial_j f \partial_i, \\ df &= \partial_j f dx^j. \end{aligned}$$

We proceed to find a formula for $\nabla_Y X$ in local coordinates

$$\begin{aligned} \nabla_Y X &= \nabla_{Y^i\partial_i} X^j \partial_j \\ &= Y^i \nabla_{\partial_i} X^j \partial_j \\ &= Y^i \left(\partial_i X^j\right)\partial_j + Y^i X^j \nabla_{\partial_i}\partial_j \\ &= Y^i \left(\partial_i X^j\right)\partial_j + Y^i X^j \Gamma_{ij}^k \partial_k, \end{aligned}$$

where we simply expanded the term $\nabla_{\partial_i}\partial_j$ in local coordinates. The first part of this formula is what we expect to get when using Cartesian coordinates in $\mathbb{R}^n$. The second part is the correction term coming from having a more general coordinate system and also a non-Euclidean metric. Our next goal is to find a formula for Γ_{ij}^k in terms of the metric. To this end we can simply use our defining implicit formula for the connection keeping in mind that there are no Lie bracket terms. On the left-hand side we have

$$\begin{aligned} 2g\left(\nabla_{\partial_i}\partial_j, \partial_l\right) &= 2g\left(\Gamma_{ij}^k\partial_k, \partial_l\right) \\ &= 2\Gamma_{ij}^k g_{kl}, \end{aligned}$$

and on the right-hand side

$$\begin{aligned} \left(L_{\partial_j} g\right)(\partial_i, \partial_l) + d\theta_{\partial_j}(\partial_i, \partial_l) &= \partial_j g_{il} + \partial_i\left(\theta_{\partial_j}(\partial_l)\right) - \partial_l\left(\theta_{\partial_j}(\partial_i)\right) \\ &= \partial_j g_{il} + \partial_i g_{jl} - \partial_l g_{ji}. \end{aligned}$$

Multiplying by g^{lm} on both sides then yields

$$\begin{aligned} 2\Gamma_{ij}^m &= 2\Gamma_{ij}^k \delta_k^m \\ &= 2\Gamma_{ij}^k g_{kl} g^{lm} \\ &= \left(\partial_j g_{il} + \partial_i g_{jl} - \partial_l g_{ji}\right) g^{lm}. \end{aligned}$$

Thus we have the formula

$$\begin{aligned} \Gamma_{ij}^k &= \frac{1}{2} g^{lk}\left(\partial_j g_{il} + \partial_i g_{jl} - \partial_l g_{ji}\right) \\ &= \frac{1}{2} g^{kl}\left(\partial_j g_{il} + \partial_i g_{jl} - \partial_l g_{ji}\right) \\ &= \frac{1}{2} g^{kl}\Gamma_{ij,l}. \end{aligned}$$

The symbols

$$\begin{aligned}\Gamma_{ij,k} &= \frac{1}{2}\left(\partial_j g_{ik} + \partial_i g_{jk} - \partial_k g_{ji}\right)\\ &= g\left(\nabla_{\partial_i}\partial_j, \partial_k\right)\end{aligned}$$

are called the *Christoffel symbols of the first kind*, while Γ_{ij}^k are the *Christoffel symbols of the second kind*. Classically the following notation has also been used

$$\begin{aligned}\left\{\begin{matrix} k \\ i, j\end{matrix}\right\} &= \Gamma_{ij}^k,\\ [ij, k] &= \Gamma_{ij,k}\end{aligned}$$

so as not to think that these things define a tensor. The reason why they are not tensorial comes from the fact that they may be zero in one coordinate system but not zero in another. A good example of this comes from the plane where the Christoffel symbols vanish in Cartesian coordinates, but not in polar coordinates:

$$\begin{aligned}\Gamma_{\theta\theta,r} &= \frac{1}{2}\left(\partial_\theta g_{\theta r} + \partial_\theta g_{\theta r} - \partial_r g_{\theta\theta}\right)\\ &= -\frac{1}{2}\partial_r\left(r^2\right)\\ &= -r.\end{aligned}$$

In fact, as is shown in exercise 2.5.20 it is always possible to find coordinates around a point $p \in M$ such that

$$\begin{aligned}g_{ij}|_p &= \delta_{ij},\\ \partial_k g_{ij}|_p &= 0.\end{aligned}$$

In particular,

$$\begin{aligned}g_{ij}|_p &= \delta_{ij},\\ \Gamma_{ij}^k|_p &= 0.\end{aligned}$$

In such coordinates the covariant derivative is computed exactly as in Euclidean space

$$\begin{aligned}\nabla_Y X|_p &= \left(\nabla_{Y^i\partial_i} X^j\partial_j\right)|_p\\ &= Y^i(p)\left(\partial_i X^j\right)|_p\partial_j|_p.\end{aligned}$$

The torsion free property of the connection is equivalent to saying that the Christoffel symbols are symmetric in ij as

$$\begin{aligned}\Gamma_{ij}^{k}\partial_{k} &= \nabla_{\partial_i}\partial_j\\ &= \nabla_{\partial_j}\partial_i\\ &= \Gamma_{ji}^{k}\partial_{k}.\end{aligned}$$

The metric property of the connection becomes

$$\begin{aligned}\partial_k g_{ij} &= g\left(\nabla_{\partial_k}\partial_i, \partial_j\right) + g\left(\partial_i, \nabla_{\partial_k}\partial_j\right)\\ &= \Gamma_{ki,j} + \Gamma_{kj,i}.\end{aligned}$$

This shows that the Christoffel symbols completely determine the derivatives of the metric.

Just as the metric could be used to give a formula for the gradient in local coordinates we can use the Christoffel symbols to get a local coordinate formula for the Hessian of a function. This is done as follows

$$\begin{aligned}2\,\mathrm{Hess}f\left(\partial_i,\partial_j\right) &= \left(L_{\nabla f}g\right)\left(\partial_i,\partial_j\right)\\ &= D_{\nabla f}g_{ij} - g\left(L_{\nabla f}\partial_i,\partial_j\right) - g\left(\partial_i, L_{\nabla f}\partial_j\right)\\ &= g^{kl}\left(\partial_k f\right)\left(\partial_l g_{ij}\right)\\ &\quad + g\left(L_{\partial_i}\left(g^{kl}\left(\partial_k f\right)\partial_l\right),\partial_j\right)\\ &\quad + g\left(\partial_i, L_{\partial_j}\left(g^{kl}\left(\partial_k f\right)\partial_l\right)\right)\\ &= \left(\partial_k f\right)g^{kl}\left(\partial_l g_{ij}\right)\\ &\quad + \partial_i\left(g^{kl}\left(\partial_k f\right)\right)g_{lj}\\ &\quad + \partial_j\left(g^{kl}\left(\partial_k f\right)\right)g_{il}\\ &= \left(\partial_k f\right)g^{kl}\left(\partial_l g_{ij}\right)\\ &\quad + \left(\partial_i\partial_k f\right)g^{kl}g_{lj} + \left(\partial_j\partial_k f\right)g^{kl}g_{il}\\ &\quad + \left(\partial_i g^{kl}\right)\left(\partial_k f\right)g_{lj} + \left(\partial_j g^{kl}\right)\left(\partial_k f\right)g_{il}\\ &= 2\partial_i\partial_j f\\ &\quad + \left(\partial_k f\right)\left(\left(\partial_i g^{kl}\right)g_{lj} + \left(\partial_j g^{kl}\right)g_{il} + g^{kl}\left(\partial_l g_{ij}\right)\right).\end{aligned}$$

To compute $\partial_i g^{jk}$ we note that

$$\begin{aligned}
0 &= \partial_i \delta_l^j \\
&= \partial_i \left(g^{jk} g_{kl} \right) \\
&= \left(\partial_i g^{jk} \right) g_{kl} + g^{jk} \left(\partial_i g_{kl} \right).
\end{aligned}$$

Thus we have

$$\begin{aligned}
2 \operatorname{Hess} f \left(\partial_i, \partial_j \right) &= 2 \partial_i \partial_j f \\
&\quad + \left(\partial_k f \right) \left(\left(\partial_i g^{kl} \right) g_{lj} + \left(\partial_j g^{kl} \right) g_{il} + g^{kl} \left(\partial_l g_{ij} \right) \right) \\
&= 2 \partial_i \partial_j f \\
&\quad + \left(\partial_k f \right) \left(-g^{kl} \partial_i g_{lj} - g^{kl} \partial_j g_{li} + g^{kl} \left(\partial_l g_{ij} \right) \right) \\
&= 2 \partial_i \partial_j f - g^{kl} \left(\partial_i g_{lj} + \partial_j g_{li} - \partial_l g_{ij} \right) \partial_k f \\
&= 2 \left(\partial_i \partial_j f - \Gamma_{ij}^k \partial_k f \right).
\end{aligned}$$

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

This notation, however, is at odds with the idea that the covariant derivative variable should come first as the notation forces its index to be last. A better index notation. often used in physics, is to write

$$\nabla_{j_0} S = \nabla_{E_{j_0}} S$$

and let

$$\nabla_{j_0} S^i_{j_1 \cdots j_k} = (\nabla S)^i_{j_0 \cdots j_k}.$$

This notation is also explored in exercise 2.5.34. This will also be our convention when using indices for the curvature tensor.

2.5 Exercises

EXERCISE 2.5.1. Show that the connection on Euclidean space is the only affine connection such that $\nabla X = 0$ for all constant vector fields X.

EXERCISE 2.5.2. Show that the skew-symmetry property $[X, Y] = -[Y, X]$ does not necessarily hold for C^1 vector fields. Show that the Jacobi identity holds for C^2 vector fields.

EXERCISE 2.5.3. Let ∇ be an affine connection on a manifold. Show that the torsion tensor

$$T(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y]$$

defines a $(2, 1)$-tensor.

EXERCISE 2.5.4. Show that if $c : I \to M$ has nonzero speed at $t_0 \in I$, then there is a vector X such that $X|_{c(t)} = \dot{c}(t)$ for t near t_0.

EXERCISE 2.5.5. Let (M, g) be a Riemannian manifold, f, h functions on M, and X a vector field on M. Show that

$$\begin{aligned} \operatorname{div}(fX) &= D_X f + f \operatorname{div} X, \\ \Delta(fh) &= h\Delta f + f\Delta h + 2g(\nabla f, \nabla h), \\ \operatorname{Hess}(fh) &= h \operatorname{Hess} f + f \operatorname{Hess} h + df dh + dh df. \end{aligned}$$

EXERCISE 2.5.6. Let (M, g) be a Riemannian manifold, f a function on M, and ϕ a function on $\mathbb{R}$. Show that

$$\Delta (\phi (f)) = \dot{\phi} (f) \Delta f + \ddot{\phi} (f) |df|^2 ,$$
$$\text{Hess} (\phi (f)) = \dot{\phi} (f) \text{Hess} f + \ddot{\phi} (f) df^2.$$

EXERCISE 2.5.7. Let (M, g) be a Riemannian manifold, X a vector field on M, and θ_X the dual 1-form. Show that $d\theta_X (Y, Z) = g (\nabla_Y X, Z) - g (Y, \nabla_Z X)$.

EXERCISE 2.5.8. The metric in coordinates satisfies:

(1) $\partial_s g^{ij} = g^{ik} \partial_s g_{kl} g^{lj}$.
(2) $\partial_s g^{ij} = -g^{il} \Gamma^j_{sl} - g^{jl} \Gamma^i_{sl}$.

EXERCISE 2.5.9. Let X be a vector field.

(1) Show that for any $(1, 1)$-tensor S

$$\text{tr} (\nabla_X S) = \nabla_X \text{tr} S.$$

(2) Let $T (Y, Z) = g (S (Y) , Z)$. Show that

$$(\nabla_X T) (Y, Z) = g ((\nabla_X S) (Y) , Z) .$$

(3) Show more generally that contraction and covariant differentiation commute.
(4) Finally show that type change and covariant differentiation commute.

EXERCISE 2.5.10. Let X be a vector field.

(1) Show that for any $(1, 1)$-tensor S

$$\text{tr} (L_X S) = L_X \text{tr} S.$$

(2) Let $T (Y, Z) = g (S (Y) , Z)$. Show that

$$(L_X T) (Y, Z) = (L_X g) (S (Y) , Z) + g ((L_X S) (Y, Z)) .$$

(3) Show that contraction and Lie differentiation commute.

EXERCISE 2.5.11. Show that a vector field X on a Riemannian manifold is locally a gradient field if and only if $Z \mapsto \nabla_Z X$ is self-adjoint.

EXERCISE 2.5.12. If $F : M \to M$ is a diffeomorphism, then the push-forward of a vector field is defined as

$$(F_* X) |_p = DF \left(X|_{F^{-1}(p)} \right) .$$

Let F be an isometry on (M, g).

(1) Show that $F_* (\nabla_X Y) = \nabla_{F_* X} F_* Y$ for all vector fields.
(2) Use this to show that isometries on $(\mathbb{R}^n, g_{\mathbb{R}^n})$ are of the form $F(x) = Ox + b$, where $O \in \mathrm{O}(n)$ and $b \in \mathbb{R}^n$. Hint: Show that F maps constant vector fields to constant vector fields.

EXERCISE 2.5.13. A vector field X is said to be *affine* if $L_X \nabla = 0$.

(1) Show that Killing fields are affine. Hint: The flow of X preserves the metric.
(2) Give an example of an affine field on $\mathbb{R}^n$ which is not a Killing field.

EXERCISE 2.5.14. Let G be a Lie group. Show that there is a unique affine connection such that $\nabla X = 0$ for all left-invariant vector fields. Show that this connection is torsion free if and only if the Lie algebra is Abelian.

EXERCISE 2.5.15. Show that the Hessian of a composition $\phi(f)$ is given by

$$\mathrm{Hess}\,\phi(f) = \phi'' df^2 + \phi' \,\mathrm{Hess} f.$$

EXERCISE 2.5.16. Consider a vector field X and a $(1,1)$-tensor L.

(1) Show that $L_X + L$ defines a derivation on tensors.
(2) Show that all derivations are of this form and that X is unique.
(3) Show that derivations are uniquely determined by how they act on functions and vector fields.
(4) Show that $L_{fX} = fL_X - X \otimes df$, where $X \otimes df$ is the rank 1 $(1,1)$-tensor $Y \mapsto X df(Y)$.

EXERCISE 2.5.17. Show that if X is a vector field of constant length on a Riemannian manifold, then $\nabla_v X$ is always perpendicular to X.

EXERCISE 2.5.18. Show that if we have a tensor field T on a Riemannian manifold (M, g) that vanishes at $p \in M$, then for any vector field X we have $L_X T = \nabla_X T$ at p. Conclude that the $(1,1)$ version of 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?

EXERCISE 2.5.19. For any $p \in (M, g)$ and orthonormal basis $e_1, \dots, e_n$ for $T_p M$, show that there is an orthonormal frame $E_1, \dots, E_n$ in a neighborhood of p such that $E_i = e_i$ and $(\nabla E_i)|_p = 0$. Hint: Fix an orthonormal frame $\bar{E}_i$ near $p \in M$ with $\bar{E}_i(p) = e_i$. If we define $E_i = \alpha_i^j \bar{E}_j$, where $\left[\alpha_i^j(x)\right] \in \mathrm{SO}(n)$ and $\alpha_i^j(p) = \delta_i^j$, then this will yield the desired frame provided that the directional derivatives $D_{e_k} \alpha_i^j$ are appropriately prescribed at p.

EXERCISE 2.5.20. Show that there are coordinates $x^1, \dots, x^n$ 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. Hint: Given a general set of coordinates y^i around p with $y^i(p) = 0$, let $x^i = \alpha_j^i(y) y^j$, adjust

$\alpha_j^i(0)$ to make the fields orthonormal at p, and adjust $\frac{\partial \alpha_j^i}{\partial y^k}(0)$ to make the covariant derivatives vanish at p.

EXERCISE 2.5.21. Consider coordinates x^i and $\bar{x}^s$ around $p \in M$. Show that the Christoffel symbols of a metric g in these two charts are related by

$$\bar{\Gamma}_{ij}^k = \frac{\partial^2 x^s}{\partial \bar{x}^i \partial \bar{x}^j} \frac{\partial \bar{x}^k}{\partial x^s} + \frac{\partial x^s}{\partial \bar{x}^i} \frac{\partial x^t}{\partial \bar{x}^j} \frac{\partial \bar{x}^k}{\partial x^l} \Gamma_{st}^l,$$

$$\frac{\partial^2 x^r}{\partial \bar{x}^i \partial \bar{x}^j} = \bar{\Gamma}_{ij}^k \frac{\partial x^r}{\partial \bar{x}^k} - \frac{\partial x^s}{\partial \bar{x}^i} \frac{\partial x^t}{\partial \bar{x}^j} \Gamma_{st}^r,$$

and

$$\bar{\Gamma}_{ij,k} = \frac{\partial^2 x^s}{\partial \bar{x}^i \partial \bar{x}^j} \frac{\partial x^t}{\partial \bar{x}^k} g_{st} + \frac{\partial x^s}{\partial \bar{x}^i} \frac{\partial x^t}{\partial \bar{x}^j} \frac{\partial x^l}{\partial \bar{x}^k} \Gamma_{st,l}.$$

EXERCISE 2.5.22. Let M be an n-dimensional submanifold of $\mathbb{R}^{n+m}$ with the induced metric. Further assume that we have a local coordinate system given by a parametrization $u^s\left(x^1, \ldots, x^n\right)$, $s = 1, \ldots, n+m$. Show that in these coordinates:

(1)

$$g_{ij} = \sum_{s=1}^{n+m} \frac{\partial u^s}{\partial x^i} \frac{\partial u^s}{\partial x^j}.$$

(2)

$$\Gamma_{ij,k} = \sum_{s=1}^{n+m} \frac{\partial u^s}{\partial x^k} \frac{\partial^2 u^s}{\partial x^i \partial x^j}.$$

EXERCISE 2.5.23. Let (M, g) be an oriented manifold.

(1) Show that if $v_1, \ldots, v_n$ is positively oriented, then

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

(2) Show that in positively oriented coordinates,

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

(3) Conclude that the Laplacian has the formula

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

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

$$\Delta f\left(p\right) = \sum_{i=1}^{n}\partial_i^2 f.$$

EXERCISE 2.5.24. Show that if a $(0,2)$-tensor T is given by T_{kl}, then ∇T is given by

$$(\nabla T)_{jkl} = \frac{\partial T_{kl}}{\partial x^j} - \Gamma_{jk}^i T_{il} - \Gamma_{jl}^i T_{ki}.$$

Similarly, when a $(1,1)$-tensor T is given by T_l^k, then ∇T is given by

$$(\nabla T)_{jl}^k = \frac{\partial T_l^k}{\partial x^j} - \Gamma_{jl}^i T_i^k + \Gamma_{ji}^k T_l^i.$$

EXERCISE 2.5.25. Let $F : (M, g_M) \looparrowright \left(\bar{M}, g_{\bar{M}}\right)$ be an isometric immersion. For two vector fields X, Y tangent to M we can compute both $\nabla_X^M Y$ and $\nabla_X^{\bar{M}} Y$. Show that the component of $\nabla_X^{\bar{M}} Y$ that is tangent to M is $\nabla_X^M Y$. Show that the normal component

$$\nabla_X^{\bar{M}} Y - \nabla_X^M Y = T_X Y$$

is symmetric in X, Y and use that to show that it is tensorial.

EXERCISE 2.5.26. Let $F : (M, g_M) \looparrowright \left(\bar{M}, g_{\bar{M}}\right)$ be an isometric immersion and

$$T^{\perp}M = \left\{v \in T_p\bar{M} \mid p \in M \text{ and } v \perp T_pM\right\}$$

the normal bundle. A vector field $V : M \to T\bar{M}$ such that $V_p \in T_p^{\perp}M$ is called a normal field along M. For a vector field X and normal field V show that

(1) The covariant derivative $\nabla_X^{\bar{M}} V$ can be defined.
(2) Decompose $\nabla_X^{\bar{M}} V$ into normal $\nabla_X^{\perp} V$ and tangential $T_X V$ components:

$$\nabla_X^{\bar{M}} V = \nabla_X^{\perp} V + T_X V.$$

$\nabla_X^{\perp} V$ is called the normal derivative of V along M. Show that

$$g_{\bar{M}}\left(T_X Y, V\right) = -g_M\left(Y, T_X V\right).$$

(3) Show that $\nabla_X^{\perp} V$ is linear and a derivation in the V variable and tensorial in the X variable.

EXERCISE 2.5.27. Let (M, g) be a oriented Riemannian manifold.

(1) If f has compact support, then

$$\int_M \Delta f \cdot \mathrm{vol} = 0.$$

(2) Show that

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

(3) Show that

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

(4) Establish Green's formula for functions with compact support:

$$\int_M f_1 \cdot \Delta f_2 \cdot d\,\mathrm{vol} = -\int_M g(\nabla f_1, \nabla f_2)\,\mathrm{vol}.$$

(5) Conclude that if f is subharmonic 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*.

EXERCISE 2.5.28. A vector field and its corresponding flow is said to be *incompressible* if $\operatorname{div} X = 0$.

(1) Show that X is incompressible if and only if the local flows it generates are volume preserving (i.e., leave the Riemannian volume form invariant).
(2) Let X be a unit vector field on $\mathbb{R}^2$. Show that $\nabla X = 0$ if X is incompressible.
(3) Find a unit vector field X on $\mathbb{R}^3$ that is incompressible but where $\nabla X \neq 0$.

EXERCISE 2.5.29. Let X be a unit vector field on (M, g) such that $\nabla_X X = 0$.

(1) Show that X is locally the gradient of a function if and only if the orthogonal distribution is integrable.
(2) Show that the orthogonal distribution is integrable in a neighborhood of $p \in M$ if it has an integral submanifold through p. Hint: It might help to show that $L_X \theta_X = 0$.
(3) Find X with the given conditions so that it is not a gradient field. Hint: Consider S^3.

EXERCISE 2.5.30. Suppose we have two distributions E and F on (M, g), that 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.

(1) Show that the distributions are integrable.
(2) Show that around any point $p \in M$ there is a product neighborhood $U = V_E \times V_F$ such that $(U, g) = (V_E \times V_F, g|_{V_E} + g|_{V_F})$, where V_E and V_F are the integral submanifolds through p.

EXERCISE 2.5.31. 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 $(U, g) = \left(V \times I, g|_V + dt^2\right)$, where V is a submanifold perpendicular to X.

EXERCISE 2.5.32. If we have two tensors S, T of the same type show that

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

EXERCISE 2.5.33. Recall that complex manifolds have complex tangent spaces. Thus we can multiply vectors by i. As a generalization of this we can define an *almost complex* structure. This is a $(1, 1)$-tensor J such that $J^2 = -I$. A *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 $\omega\,(X, Y) = g\,(J\,(X), Y)$.

(1) Show that the *Nijenhuis tensor*:

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

is a tensor.
(2) Show that if J comes from a complex structure, then $N = 0$. The converse is the famous theorem of Newlander and Nirenberg.
(3) Show that ω is a 2-form.
(4) Show that $d\omega = 0$ if $\nabla J = 0$.
(5) Conversely show that if $d\omega = 0$ and J is a complex structure, then $\nabla J = 0$. In this case we call the metric a Kähler metric.

EXERCISE 2.5.34. Define $\nabla_i T$ as the covariant derivative in the direction of the i^{th} coordinate vector field and $\nabla^i T = g^{ij}\nabla_j T$ as the corresponding type changed tensor.

(1) For a function f show that $df = \nabla_i f dx^i$ and $\nabla f = \nabla^i f \partial_i$.
(2) For a vector field X show that $(\nabla_i X)^i = \operatorname{div} X$.
(3) For a $(0, 2)$-tensor T show that $\left(\nabla^i T\right)_{ij} = -\left(\nabla^* T\right)_j$.

Chapter 3
Curvature

The idea of a Riemannian metric having curvature, while intuitively appealing and natural, is also often the stumbling block for further progress into the realm of geometry. The most elementary way of defining curvature is to set it up as an integrability condition. This indicates that when it vanishes it should be possible to solve certain differential equations, e.g., that the metric is Euclidean. This was in fact one of Riemann's key insights.

As we shall observe here and later in sections 5.1 and 6.1.2 one can often take two derivatives (such as in the Hessian) and have them commute in a suitable sense, but taking more derivatives becomes somewhat more difficult to understand. This is what is behind the abstract definitions below and is also related to integrability conditions.

We shall also try to justify curvature on more geometric grounds. The idea is to create 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 formulas hold all the information that is needed for computing curvatures in many examples and also for studying how curvature influences the metric.

Much of what we do in this chapter carries over to the pseudo-Riemannian setting. The connection and curvature tensor are generalized without changes. But formulas that involve contractions do need modification (see exercise 1.6.10).

3.1 Curvature

We introduced in the previous chapter the idea of covariant derivatives of tensors and explained their relation to the classical concepts of gradient, Hessian, and Laplacian. However, the Riemannian metric is parallel and consequently has no meaningful

P. Petersen, *Riemannian Geometry*, Graduate Texts in Mathematics 171,
DOI 10.1007/978-3-319-26654-1_3

derivatives. 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 affine connection on a manifold gives rise to a *curvature tensor*. 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. On a (pseudo-)Riemannian manifold it is also possible to take traces of this curvature operator to obtain various averaged curvatures.

3.1.1 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 the $(1, 3)$-tensor defined by

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

on vector fields X, Y, Z. The first line in the definition is also called the *Ricci identity* and is often written as

$$R_{X,Y}Z = \nabla^2_{X,Y}Z - \nabla^2_{Y,X}Z.$$

This also allows us to define the curvature of tensors

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

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:

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

Observe that X, Y appear skew-symmetrically in $R(X, Y)Z = R_{X,Y}Z$, while Z plays its own role on top of the line, hence the unusual notation.

In relation to derivations as explained in section 2.3 note that $R_{X,Y}$ acts as a derivation on tensors. Moreover, as the Hessian of a function is symmetric $\nabla^2_{X,Y}f = \nabla^2_{Y,X}f$ it follows that $R_{X,Y}$ acts trivially on functions. This is the content of the Ricci identity

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

where on the right-hand side we think of $R(X, Y)$ as a $(1, 1)$-tensor acting on tensors. As an example note that when T is a $(0, k)$-tensor then

$$\begin{aligned}(R_{X,Y}T)(X_1, \dots, X_k) &= (R(X, Y)T)(X_1, \dots, X_k) \\ &= -T(R(X, Y)X_1, \dots, X_k) \\ &\quad \vdots \\ &\quad -T(X_1, \dots, R(X, Y)X_k).\end{aligned}$$

Using the metric g we can change R to a $(0, 4)$-tensor as follows:

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

We justify next why the variables are treated on a more equal footing in this formula by showing several important symmetry properties.

Proposition 3.1.1. *The Riemannian curvature tensor $R(X, Y, Z, W)$ satisfies the following properties:*

(1) *R is skew-symmetric 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:

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

or

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

Proof. The first part of (1) has already been established. For part two of (1) use that $[X, Y]$ is the vector field defined implicitly 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:

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

Now (1) follows by *polarizing* the identity $R(X, Y, Z, Z) = 0$ in Z:

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

Part (3) relies on the torsion free property and the definitions from section 2.2.2.4 to first show that

$$\begin{aligned}
(L_X \nabla)_Y Z &= L_X (\nabla_Y Z) - \nabla_{L_X Y} Z - \nabla_Y L_X Z \\
&= \nabla_X \nabla_Y Z - \nabla_{\nabla_X Y} Z - \nabla_Y \nabla_X Z \\
&\quad - \nabla_{\nabla_Y Z} X + \nabla_{\nabla_Y X} Z + \nabla_Y \nabla_Z X \\
&= R_{X,Y} Z + \nabla^2_{Y,Z} X.
\end{aligned}$$

The Jacobi identity (see proposition 2.1.6) followed by the torsion free property and the Ricci identity then show that

$$\begin{aligned}
0 &= (L_X L)_Y Z \\
&= (L_X \nabla)_Y Z - (L_X \nabla)_Z Y
\end{aligned}$$

$$\begin{aligned}
&= R_{X,Y}Z + \nabla^2_{Y,Z}X - R_{X,Z}Y - \nabla^2_{Z,Y}X \\
&= R_{X,Y}Z + R_{Z,X}Y + R_{Y,Z}X.
\end{aligned}$$

Part (2) is a direct combinatorial 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 $2R(X,Y,Z,W) = 2R(Z,W,X,Y)$.

Part (4) follows from the claim that

$$(\nabla_X R)_{Y,Z} W = \nabla^3_{X,Y,Z}W - \nabla^3_{X,Z,Y}W - \nabla^3_{Y,Z,X}W + \nabla^3_{Z,Y,X}W + \nabla_{R_{Y,Z}X}W.$$

To see this simply add over the cyclic permutations of X, Y, Z:

$$\begin{aligned}
&(\nabla_X R)_{Y,Z} W + (\nabla_Z R)_{X,Y} W + (\nabla_Y R)_{Z,X} W \\
&= \nabla^3_{X,Y,Z}W - \nabla^3_{X,Z,Y}W - \nabla^3_{Y,Z,X}W + \nabla^3_{Z,Y,X}W + \nabla_{R_{Y,Z}X}W \\
&+ \nabla^3_{Z,X,Y}W - \nabla^3_{Z,Y,X}W - \nabla^3_{X,Y,Z}W + \nabla^3_{Y,X,Z}W + \nabla_{R_{X,Y}Z}W \\
&+ \nabla^3_{Y,Z,X}W - \nabla^3_{Y,X,Z}W - \nabla^3_{Z,X,Y}W + \nabla^3_{X,Z,Y}W + \nabla_{R_{Z,X}Y}W \\
&= \nabla_{R_{X,Y}Z + R_{Z,X}Y + R_{Y,Z}X}W \\
&= 0.
\end{aligned}$$

The claim can be proven directly but also follows from the two different iterated Ricci identities for taking three derivatives:

$$\nabla^3_{X,Y,Z}W - \nabla^3_{Y,X,Z}W = R_{X,Y}\nabla_Z W - \nabla_{R_{X,Y}Z}W$$

and

$$\nabla^3_{X,Y,Z}W - \nabla^3_{X,Z,Y}W = (\nabla_X R)_{Y,Z} W + R_{Y,Z}\nabla_X W.$$

These follow from the various ways one can iterate covariant derivatives (see sections 2.2.2.3 and 2.2.2.5):

$$\nabla^3_{X,Y,Z}W = \nabla^2_{X,Y}(\nabla_Z W) - \nabla_{\nabla^2_{X,Y}Z}W$$

and

$$\nabla^3_{X,Y,Z}W = \nabla_X\left(\nabla^2\right)_{Y,Z}W + \nabla^2_{Y,Z}(\nabla_X W)$$

and then using the Ricci identity. □

Example 3.1.2. $(\mathbb{R}^n, g_{\mathbb{R}^n})$ has $R \equiv 0$ since $\nabla_{\partial_i}\partial_j = 0$ for the standard Cartesian coordinates.

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

3.1.2 The Curvature Operator

First recall that we have the space $\Lambda^2 TM$ of bivectors. A decomposable bivector $v \wedge w$ can be thought of as the oriented parallelogram spanned by v, w. If e_i is an orthonormal basis for T_pM, then the inner product on $\Lambda^2 T_pM$ is such that the bivectors $e_i \wedge e_j$, $i < j$ will form an orthonormal basis. The inner product that $\Lambda^2 TM$ inherits in this way is also denoted by g. Note that this inner product on $\Lambda^2 T_pM$ has the property that

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

It is also useful to interpret bivectors as skew symmetric maps. This is done via the formula:

$$(x \wedge y)(v) = g(x, v)\, y - g(y, v)\, x.$$

This represents a skew-symmetric transformation in span $\{v, w\}$ which is a counterclockwise 90° rotation when v, w are orthonormal. (We could have used a clockwise rotation as that will in fact work more naturally with our version of the curvature tensor.) Note that

$$g(x \wedge y, v \wedge w) = g(x, v)\, g(y, w) - g(x, w)\, g(y, v) = g((x \wedge y)(v), w).$$

These operators satisfy a Jacobi-Bianchi type identity:

$$(x \wedge y)(z) + (y \wedge z)(x) + (z \wedge x)(y) = 0.$$

From the symmetry properties of the curvature tensor it follows that R defines a symmetric bilinear map

$$
\begin{aligned}
R : \Lambda^2 TM \times \Lambda^2 TM \to \mathbb{R} \\
R\left(\sum X_i \wedge Y_i, \sum V_j \wedge W_j\right) = \sum R\left(X_i, Y_i, W_j, V_j\right).
\end{aligned}
$$

Note the reversal of V and W! The relation

$$
g\left(\mathfrak{R}\left(\sum X_i \wedge Y_i\right), \sum V_j \wedge W_j\right) = \sum R\left(X_i, Y_i, W_j, V_j\right)
$$

consequently defines a self-adjoint operator $\mathfrak{R} : \Lambda^2 TM \to \Lambda^2 TM$. This operator is called the *curvature operator*. It is evidently just a different manifestation of the curvature tensor. The switch between V and W is related to our definition of the next curvature concept.

3.1.3 Sectional Curvature

For any $v \in T_pM$ let

$$
R_v(w) = R(w, v)v : T_pM \to T_pM
$$

be the *directional curvature operator*. This operator is also known as the *tidal force operator*. The latter name describes in physical (general relativity) terms the meaning of the tensor. As we shall see, this is the part of the curvature tensor that directly relates to the metric. The above symmetry properties of R imply that this operator is self-adjoint and that v is always a zero-eigenvector. The normalized biquadratic form

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

is called the *sectional curvature* of (v, w). Since the denominator is the square of the area of the parallelogram $\{tv + sw \mid 0 \leq t, s \leq 1\}$ it is easy to check that $\sec(v, w)$ depends only on the plane $\pi = \operatorname{span}\{v, w\}$. One of the important relationships between directional and sectional curvature is the following algebraic result by Riemann.

Proposition 3.1.3 (Riemann, 1854). *The following properties are equivalent:*

(1) $\sec(\pi) = k$ *for all 2-planes in* T_pM.
(2) $R(v_1, v_2)v_3 = -k\,(v_1 \wedge v_2)\,(v_3)$ *for all* $v_1, v_2, v_3 \in T_pM$.
(3) $R_v(w) = k \cdot (w - g(w, v)v) = k \cdot pr_{v^\perp}(w)$ *for all* $w \in T_pM$ *and* $|v| = 1$.
(4) $\mathfrak{R}\,(\omega) = k \cdot \omega$ *for all* $\omega \in \Lambda^2 T_pM$.

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

$$\begin{aligned} R_k(v_1, v_2)v_3 &= -k\,(v_1 \wedge v_2)\,(v_3)\,, \\ R_k\,(v_1, v_2, v_3, v_4) &= -kg\,((v_1 \wedge v_2)\,(v_3)\,, v_4) \\ &= kg\,(v_1 \wedge v_2, v_4 \wedge v_3)\,. \end{aligned}$$

The first observation is that these maps behave exactly like the curvature tensor in that they satisfy properties (1), (2), and (3) of proposition 3.1.1. Now consider the difference between the curvature tensor and this curvature-like tensor

$$D\,(v_1, v_2, v_3, v_4) = R\,(v_1, v_2, v_3, v_4) - R_k\,(v_1, v_2, v_3, v_4)\,.$$

Properties (1), (2), and (3) from proposition 3.1.1 carry over to this difference tensor. Moreover, the assumption that $\sec = k$ implies

$$D\,(v, w, w, v) = 0$$

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

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

Using properties (1) and (2) from proposition 3.1.1 it follows that D is alternating in all four variables. That, however, is in violation of Bianchi's first identity (property (3) from proposition 3.1.1) unless $D = 0$. This finishes the implication (see also exercise 3.4.29 for two other strategies.)

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 $\Lambda^2 T_pM$. Using (2) it follows 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

$$k = g\left(\mathfrak{R}\left(v \wedge w\right), v \wedge w\right) = \sec\left(v, w\right).$$

□

A Riemannian manifold (M, g) that satisfies either of these 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 $(\mathbb{R}^n, g_{\mathbb{R}^n})$ has curvature zero. In sections 4.2.1 and 4.2.3 we shall prove that the space forms S_k^n as described in example 1.4.6 have constant curvature k.

3.1.4 Ricci Curvature

Our next curvature is the Ricci curvature, which can be thought of as the Laplacian of g.

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

$$\begin{aligned}
\operatorname{Ric}(v, w) &= \operatorname{tr}\left(x \mapsto R\left(x, v\right) w\right) \\
&= \sum_{i=1}^{n} g\left(R\left(e_i, v\right) w, e_i\right) \\
&= \sum_{i=1}^{n} g\left(R\left(v, e_i\right) e_i, w\right) \\
&= \sum_{i=1}^{n} g\left(R\left(e_i, w\right) v, e_i\right).
\end{aligned}$$

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

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

We adopt the language that $\operatorname{Ric} \geq k$ if all eigenvalues of $\operatorname{Ric}(v)$ are $\geq k$. In $(0,2)$ language this means that $\operatorname{Ric}\left(v, v\right) \geq kg\left(v, v\right)$ for all v. When (M, g) satisfies $\operatorname{Ric}(v) = k \cdot v$, or equivalently $\operatorname{Ric}(v, w) = k \cdot g(v, w)$, 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 $(n-1)k$.

In chapter 4 we shall exhibit several interesting Einstein metrics that do not have constant curvature. Three basic types are

(1) The product metric $S^n(1) \times S^n(1)$ with Einstein constant $n-1$ (see section 4.2.2).
(2) The Fubini-Study metric on $\mathbb{CP}^n$ with Einstein constant $2n+2$ (see section 4.5.3).
(3) The generalized Schwarzschild metric on $\mathbb{R}^2 \times S^{n-2}$, $n \geq 4$, which is a doubly warped product metric: $dr^2 + \phi^2(r)d\theta^2 + \rho^2(r)ds_{n-2}^2$ with Einstein constant 0 (see section 4.2.5).

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

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

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, then

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

In other words:

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

As the matrix has det $= 2$ any sectional curvature can be computed from Ric. In particular, we see that (M^3, g) is Einstein if and only if (M^3, g) has constant sectional curvature. Therefore, the search for Einstein metrics that do not have constant curvature naturally begins in dimension 4.

3.1.5 Scalar Curvature

The last curvature quantity we define here is the *scalar curvature*:

$$\mathrm{scal} = \mathrm{tr}(\mathrm{Ric}) = 2 \cdot \mathrm{tr}\mathfrak{R}.$$

Notice that scal depends only on $p \in M$, so we obtain a function scal : $M \to \mathbb{R}$. In an orthonormal basis $e_1, \ldots, e_n$ for T_pM it can be calculated from the curvature tensor in several ways:

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

When $n = 2$ it follows that $\mathrm{scal}(p) = 2 \cdot \sec(T_pM)$. In section 4.2.3 we exhibit examples of scalar flat metrics that are not Ricci flat when $n \geq 3$. There is also another interesting phenomenon in dimensions ≥ 3 related to scalar curvature.

Lemma 3.1.4 (Schur, 1886). *Suppose that a Riemannian manifold (M, g) of dimension $n \geq 3$ satisfies either one of the following two conditions for a function $f : M \to \mathbb{R}$*

(1) $\sec(\pi) = f(p)$ *for all 2-planes* $\pi \subset T_pM,\ p \in M$,
(2) $\mathrm{Ric}(v) = (n-1) \cdot f(p) \cdot v$ *for all* $v \in T_pM,\ p \in M$.

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

Proof. It suffices to show (2), as the conditions for (1) imply that (2) holds. To show (2) we need an important identity relating derivatives of the scalar curvature and the $(0,2)$-version of the Ricci tensor:

$$d\mathrm{scal} = -2\nabla^* \mathrm{Ric}\,.$$

Let us see how this implies (2). First note that

$$\begin{aligned}
d\mathrm{scal} &= d\mathrm{tr}\,(\mathrm{Ric}) \\
&= d\,(n \cdot (n-1) \cdot f) \\
&= n \cdot (n-1) \cdot df.
\end{aligned}$$

On the other hand using the definition of the adjoint from section 2.2.2; the product rule; and $\nabla g = 0$ we obtain

$$\begin{aligned}
-\nabla^* \operatorname{Ric}(X) &= (n-1)\left(\nabla_{E_i}(fg)\right)(E_i, X) \\
&= (n-1)\left(\nabla_{E_i} f\right) g(E_i, X) + (n-1) f\left(\nabla_{E_i} g\right)(E_i, X) \\
&= (n-1)\, df\left(g(E_i, X) E_i\right) \\
&= (n-1)\, df(X).
\end{aligned}$$

This shows that $n \cdot df = 2 \cdot df$ and consequently: $n = 2$ or $df \equiv 0$ (i.e., f is constant). □

Proposition 3.1.5 (The Contracted Bianchi Identify). *On any Riemannian manifold the scalar and Ricci curvature are related by*

$$d\operatorname{tr}(\operatorname{Ric}) = d\operatorname{scal} = -2\nabla^* \operatorname{Ric}.$$

Proof. The identity is proved by a calculation that relies the second Bianchi identity (property (4) from proposition 3.1.1). Using that contractions and covariant differentiation commute (see exercise 2.5.9) we obtain

$$\begin{aligned}
d\operatorname{scal}(W)|_p &= D_W \operatorname{scal} \\
&= \sum (\nabla_W R)\left(E_i, E_j, E_j, E_i\right) \\
&= -\sum \left(\nabla_{E_j} R\right)\left(W, E_i, E_j, E_i\right) \\
&\quad -\sum \left(\nabla_{E_i} R\right)\left(E_j, W, E_j, E_i\right) \\
&= 2\sum \left(\nabla_{E_j} R\right)\left(E_i, W, E_j, E_i\right) \\
&= 2\sum \left(\nabla_{E_j} R\right)\left(E_j, E_i, E_i, W\right) \\
&= 2\sum g\left(\left(\nabla_{E_j} \operatorname{Ric}\right)\left(E_j, W\right)\right) \\
&= -2\left(\nabla^* \operatorname{Ric}\right)(W)(p).
\end{aligned}$$

□

Corollary 3.1.6. *An $n\,(>2)$-dimensional Riemannian manifold (M, g) is Einstein if and only if*

$$\operatorname{Ric} = \frac{\operatorname{scal}}{n} g.$$

3.1.6 Curvature in Local Coordinates

As with the connection it is sometimes convenient to know what the curvature tensor looks like in local coordinates. We first observe that when $X = X^i \partial_i$, $Y = Y^j \partial_j$, $Z = Z^k \partial_k$, then

$$R(X,Y)Z = X^i Y^j Z^k R^l_{ijk} \partial_l,$$
$$R^l_{ijk} \partial_l = R\left(\partial_i, \partial_j\right) \partial_k.$$

Using the definition of R we can calculate R^l_{ijk} in terms of the Christoffel symbols (see section 2.4)

$$\begin{aligned}
R^l_{ijk} \partial_l &= R\left(\partial_i, \partial_j\right) \partial_k \\
&= \nabla_{\partial_i} \nabla_{\partial_j} \partial_k - \nabla_{\partial_j} \nabla_{\partial_i} \partial_k \\
&= \nabla_{\partial_i} \left(\Gamma^s_{jk} \partial_s\right) - \nabla_{\partial_j} \left(\Gamma^t_{ik} \partial_t\right) \\
&= \partial_i \left(\Gamma^s_{jk}\right) \partial_s + \Gamma^s_{jk} \nabla_{\partial_i} \partial_s \\
&\quad - \partial_j \left(\Gamma^t_{ik}\right) \partial_t - \Gamma^t_{ik} \nabla_{\partial_j} \partial_t \\
&= \partial_i \left(\Gamma^l_{jk}\right) \partial_l - \partial_j \left(\Gamma^l_{ik}\right) \partial_l \\
&\quad + \Gamma^s_{jk} \Gamma^l_{is} \partial_l - \Gamma^t_{ik} \Gamma^l_{jt} \partial_l \\
&= \left(\partial_i \Gamma^l_{jk} - \partial_j \Gamma^l_{ik} + \Gamma^s_{jk} \Gamma^l_{is} - \Gamma^s_{ik} \Gamma^l_{js}\right) \partial_l.
\end{aligned}$$

So

$$R^l_{ijk} = \partial_i \Gamma^l_{jk} - \partial_j \Gamma^l_{ik} + \Gamma^s_{jk} \Gamma^l_{is} - \Gamma^s_{ik} \Gamma^l_{js}.$$

Similarly we also have

$$R_{ijkl} = \partial_i \Gamma_{jk,l} - \partial_j \Gamma_{ik,l} + g^{st} \Gamma_{ik,s} \Gamma_{jl,t} - g^{st} \Gamma_{jk,s} \Gamma_{il,t}.$$

These coordinate expression can also be used, in conjunction with the properties of the Christoffel symbols (see section 2.4), to prove all of the symmetry properties of the curvature tensor.

The formula clearly simplifies if we are at a point p where $\Gamma^k_{ij}|_p = 0$

$$R^l_{ijk}|_p = \partial_i \Gamma^l_{jk}|_p - \partial_j \Gamma^l_{ik}|_p.$$

If we use the formulas for the Christoffel symbols in terms of the metric we can create an expression for R^l_{ijk} that depends on the metric g_{ij} and its first two derivatives.

Remark 3.1.7. One often sees the following index notation for Ricci and scalar curvature in the literature

$$\mathrm{Ric}_{ij} = R_{ij} = R^k_{ijk} = g^{kl} R_{kijl},$$

$$\mathrm{scal} = R = g^{ij} R_{ij}.$$

The idea behind this notation is that these tensors are gotten by contracting indices in the curvature tensor. In this case the full curvature tensor R is denoted Rm so that it isn't confused with scalar curvature.

Remark 3.1.8. Due to how we wrote the $(0, 4)$ version of R we write

$$R_{ijkl} = g_{sl} R^s_{ijk} = R\left(\partial_i, \partial_j, \partial_k, \partial_l\right).$$

Other conventions such as

$$R_{lijk} = g_{sl} R^s_{ijk}$$

are also used in the literature.

3.2 The Equations of Riemannian Geometry

In this section we will see that curvature comes up naturally in the investigation of certain types of functions. This will lead us to a collection of formulas that will facilitate the calculation of the curvature tensor of rotationally symmetric and doubly warped product metrics (see section 4.2).

3.2.1 Curvature Equations

We start with the goal of calculating the curvatures on a Riemannian manifold using various geometric concepts that relate to a specific smooth function $f : M \to \mathbb{R}$. Often this function will only be smooth on an open subset $O \subset M$ in which case we just confine our attention to what happens on that subset.

The function has a gradient ∇f and a Hessian $\mathrm{Hess} f$. We shall also use $S(X) = \nabla_X \nabla f$ for the $(1, 1)$-tensor that corresponds to $\mathrm{Hess} f$ and $\mathrm{Hess}^2 f$ for the $(0, 2)$-tensor that corresponds to $S^2 = S \circ S$.

The *second fundamental form* of a hypersurface $H^{n-1} \subset M^n$ with a fixed unit normal vector field $N : H \to T^\perp H = \left\{v \in T_p M \mid p \in H,\ v \perp T_p H\right\}$ is defined as the $(0, 2)$-tensor $\mathrm{II}(X, Y) = g(\nabla_X N, Y)$ on H. Since $X, Y, [X, Y] \in TH$ are perpendicular to N we have

$$\begin{aligned} g\left(\nabla_X N, Y\right) &= D_X g\left(N, Y\right) - g\left(N, \nabla_X Y\right) \\ &= -g\left(N, \nabla_X Y\right) \\ &= -g\left(N, \nabla_Y X\right) \\ &= g\left(\nabla_Y N, X\right). \end{aligned}$$

This shows that II is symmetric. Note also that

$$g\left(\nabla_X N, N\right) = \tfrac{1}{2} D_X \left|N\right|^2 = 0.$$

So we can also define $\mathrm{II}\left(X, Y\right) = g\left(\nabla_X N, Y\right)$ when $X \in TH$, as $\nabla_X N$ has no normal component.

For the remainder of this section assume that f is given and that $H \subset f^{-1}\left(a\right)$ is open and consists entirely of regular points for f. In this case H is clearly a hypersurface. We start by relating the second fundamental form of H to f.

Proposition 3.2.1. *The following properties hold:*

(1) $N = \frac{\nabla f}{|\nabla f|}$ *is a unit normal to* H*,*
(2) $\mathrm{II}\left(X, Y\right) = \frac{1}{|\nabla f|} \mathrm{Hess} f\left(X, Y\right)$ *for all* $X, Y \in TH$*, and*
(3) $\mathrm{Hess} f\left(\nabla f, X\right) = \frac{1}{2} D_X \left|\nabla f\right|^2$ *for all* $X \in TM$*.*

Proof. (1) Clearly $N = \frac{\nabla f}{|\nabla f|}$ has unit length. It is perpendicular to H since $D_X f = 0$ for any vector field tangent to H.

(2) Using that choice of a normal vector tells us that when $X, Y \in TH$:

$$\begin{aligned} \mathrm{II}\left(X, Y\right) &= g\left(\nabla_X \frac{\nabla f}{|\nabla f|}, Y\right) \\ &= g\left(\frac{1}{|\nabla f|} \nabla_X \nabla f, Y\right) + g\left(D_X\left(\frac{1}{|\nabla f|}\right)\nabla f, Y\right) \\ &= \frac{1}{|\nabla f|} \mathrm{Hess} f\left(X, Y\right). \end{aligned}$$

(3) Finally the symmetry of $\mathrm{Hess} f$ implies:

$$\mathrm{Hess} f\left(\nabla f, X\right) = g\left(\nabla_{\nabla f} \nabla f, X\right) = g\left(\nabla_X \nabla f, \nabla f\right) = \tfrac{1}{2} D_X \left|\nabla f\right|^2.$$

□

Our first fundamental equation is the calculation of what's called the radial curvatures.

Theorem 3.2.2 (The Radial Curvature Equation). *When $H \subset f^{-1}(a)$ consists of regular points for f we have:*

$$\left(\nabla_{\nabla f} S\right)(X) + S^2(X) - \nabla_X(S(\nabla f)) = -R(X, \nabla f)\nabla f,$$

$$\nabla_{\nabla f} \operatorname{Hess} f + \operatorname{Hess}^2 f - \operatorname{Hess}\left(\frac{1}{2}|\nabla f|^2\right) = -R(\cdot, \nabla f, \nabla f, \cdot),$$

and

$$L_{\nabla f} \operatorname{Hess} f - \operatorname{Hess}^2 f - \operatorname{Hess}\left(\frac{1}{2}|\nabla f|^2\right) = -R(\cdot, \nabla f, \nabla f, \cdot).$$

Proof. The first formula is a straightforward computation.

$$\begin{aligned}
-R_{\nabla f}(X) &= -R(X, \nabla f)\nabla f \\
&= -\nabla^2_{X,\nabla f}\nabla f + \nabla^2_{\nabla f, X}\nabla f \\
&= -(\nabla_X S)(\nabla f) + \left(\nabla_{\nabla f} S\right)(X) \\
&= -\nabla_X(S(\nabla f)) + \nabla_{\nabla_X \nabla f}\nabla f + \left(\nabla_{\nabla f} S\right)(X) \\
&= -\nabla_X(S(\nabla f)) + S^2(X) + \left(\nabla_{\nabla f} S\right)(X).
\end{aligned}$$

The second formula follows by the definition of $\operatorname{Hess}^2 f$; observing that the gradient of $\frac{1}{2}|\nabla f|^2$ is $\nabla_{\nabla f}\nabla f$; and that covariant differentiation commutes with type change (see exercise 2.5.9):

$$(\nabla_N \operatorname{Hess} f)(X, Y) = g((\nabla_N S)(X), Y).$$

The final formula is a consequence of

$$\begin{aligned}
\left(L_{\nabla f} \operatorname{Hess} f\right)(X, Y) &= \left(\nabla_{\nabla f} \operatorname{Hess} f\right)(X, Y) \\
&\quad + \operatorname{Hess} f(\nabla_X \nabla f, Y) + \operatorname{Hess} f(X, \nabla_Y \nabla f) \\
&= \left(\nabla_{\nabla f} \operatorname{Hess} f\right)(X, Y) + 2\operatorname{Hess}^2 f(X, Y).
\end{aligned}$$

□

Remark 3.2.3. The last formula is particularly interesting as it shows how suitable curvatures can be calculated using only gradients of functions and Lie derivatives, i.e., covariant derivatives are not necessary.

The following two fundamental equations are also known as the *Gauss equations* and *Peterson-Codazzi-Mainardi equations,* respectively. They will be proved simultaneously but stated separately. For a vector we use the notation

$$\begin{aligned}
X &= X^\top + X^\perp \\
&= X - g(X, N)N + g(X, N)N
\end{aligned}$$

for decomposing it into components that are tangential and normal to H. We use the notation that g_H is the metric g restricted to H and that the curvature on H is R^H

Theorem 3.2.4 (The Tangential Curvature Equation).

$$g\left(R(X,Y)Z,W\right) = g_H\left(R^H(X,Y)Z,W\right) - \mathrm{II}\left(X,W\right)\mathrm{II}(Y,Z) + \mathrm{II}(X,Z)\,\mathrm{II}\left(Y,W\right),$$

where X,Y,Z,W are tangent to H.

Theorem 3.2.5 (The Normal or Mixed Curvature Equation).

$$g\left(R(X,Y)Z,N\right) = -\left(\nabla_X \mathrm{II}\right)(Y,Z) + \left(\nabla_Y \mathrm{II}\right)(X,Z),$$

where X,Y,Z are tangent to H.

Proof. The proofs hinge on the important fact that if X,Y are vector fields that are tangent to H, then:

$$\begin{aligned}\nabla^H_X Y &= \left(\nabla_X Y\right)^\top\\ &= \nabla_X Y - g\left(\nabla_X Y, N\right)N\\ &= \nabla_X Y + \mathrm{II}(X,Y)N.\end{aligned}$$

Here the first equality is a consequence of the uniqueness of the Riemannian connection on H. One can check either that $\left(\nabla_X Y\right)^\top$ satisfies properties (1)–(4) of a Riemannian connection (see theorem 2.2.2) or alternatively that it satisfies the Koszul formula. The latter task is almost immediate. The other equalities are immediate from our definitions.

The curvature equations that involve the second fundamental form are verified by calculating $R(X,Y)Z$ using $\nabla_X Y = \nabla^H_X Y - \mathrm{II}(X,Y)N$.

$$\begin{aligned}R(X,Y)Z &= \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z\\ &= \nabla_X(\nabla^H_Y Z - \mathrm{II}(Y,Z)N) - \nabla_Y(\nabla^H_X Z - \mathrm{II}(X,Z)N)\\ &\quad -\nabla^H_{[X,Y]}Z + \mathrm{II}([X,Y],Z)N\\ &= \nabla_X\nabla^H_Y Z - \nabla_Y\nabla^H_X Z - \nabla^H_{[X,Y]}Z\\ &\quad -\nabla_X\left(\mathrm{II}(Y,Z)N\right) + \nabla_Y\left(\mathrm{II}(X,Z)N\right) + \mathrm{II}([X,Y],Z)N\\ &= R^H(X,Y)Z - \mathrm{II}\left(X,\nabla^H_Y Z\right)N + \mathrm{II}(Y,\nabla^H_X Z)N\\ &\quad -\left(D_X\,\mathrm{II}(Y,Z)\right)N - \mathrm{II}(Y,Z)\nabla_X N + \left(D_Y\,\mathrm{II}(X,Z)\right)N + \mathrm{II}(X,Z)\nabla_Y N\\ &\quad + \mathrm{II}\left(\nabla_X Y,Z\right)N - \mathrm{II}\left(\nabla_Y X,Z\right)N\\ &= R^H(X,Y)Z - \mathrm{II}\left(X,\nabla_Y Z\right)N + \mathrm{II}(Y,\nabla_X Z)N\\ &\quad -\left(D_X\,\mathrm{II}(Y,Z)\right)N - \mathrm{II}(Y,Z)\nabla_X N + \left(D_Y\,\mathrm{II}(X,Z)\right)N + \mathrm{II}(X,Z)\nabla_Y N\end{aligned}$$

$$
\begin{aligned}
&+\mathrm{II}\left(\nabla_X Y, Z\right) N-\mathrm{II}\left(\nabla_Y X, Z\right) N \\
=\; & R^H(X, Y) Z-\mathrm{II}(Y, Z) \nabla_X N+\mathrm{II}(X, Z) \nabla_Y N \\
&+\left(-\left(\nabla_X \mathrm{II}\right)(Y, Z)+\left(\nabla_Y \mathrm{II}\right)(X, Z)\right) N .
\end{aligned}
$$

To finish we just need to recall the definition of II in terms of ∇N. □

These three fundamental equations give us a way of computing curvature tensors by induction on dimension. More precisely, if we know how to do computations on H and also how to compute S, then we can compute any curvature in M at a point in H. We shall clarify and exploit this philosophy in subsequent chapters.

Here we confine ourselves to some low dimensional observations. Recall that the three curvature quantities sec, Ric, and scal obeyed some special relationships in dimensions 2 and 3 (see sections 3.1.4 and 3.1.5). Curiously enough this also manifests itself in our three fundamental equations.

If M has dimension 1, then $\dim H = 0$. This is related to the fact that $R \equiv 0$ on all 1 dimensional spaces.

If M has dimension 2, then $\dim H = 1$. Thus $R^H \equiv 0$ and the three vectors X, Y, and Z are proportional. Thus only the radial curvature equation is relevant. The curvature is also calculated in example 3.2.12.

When M has dimension 3, then $\dim H = 2$. The radial curvature equation is not simplified, but in the other two equations 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 frame, 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 $M^3 = \mathbb{R}^3$ we have $R = 0$. The tangential curvature equation is particularly interesting as it becomes the classical Gauss equation. If we assume that E_1, E_2 is an orthonormal basis for T_pM, then

$$
\begin{aligned}
\sec(T_pH) &= R^H\left(E_1, E_2, E_2, E_1\right) \\
&= \mathrm{II}(E_1, E_1)\,\mathrm{II}(E_2, E_2) - \mathrm{II}(E_1, E_2)\,\mathrm{II}(E_1, E_2) \\
&= \det\left[\mathrm{II}\right].
\end{aligned}
$$

This was *Gauss's wonderful observation*! Namely, that the extrinsic quantity $\det[\mathrm{II}]$ for H is actually the intrinsic quantity, $\sec(T_pH)$. The two mixed curvature equations are the classical Peterson-Codazzi-Mainardi equations.

Finally, in dimension 4 everything reaches its most general level. We can start with an orthonormal frame $\{X, Y, Z, N\}$ and there are potentially twenty different curvature quantities to compute.

3.2.2 Distance Functions

The formulas in the previous section become simpler and more significant if we start by making assumptions about the function. The geometrically defined functions we shall study are *distance functions*. As we don't have a concept of distance yet, we define $r : O \to \mathbb{R}$, where $O \subset (M, g)$ is open, to be a *distance function* if $|\nabla r| \equiv 1$ on O. Distance functions are then simply solutions to the *Hamilton-Jacobi equation* or *eikonal equation* $|\nabla r|^2 = 1$. This is a nonlinear first-order PDE and can be solved by the method of characteristics (see e.g. [6]). For now we shall assume that solutions exist and investigate their properties. Later, after we have developed the theory of geodesics, we establish the existence of such functions on general Riemannian manifolds and also justify their name.

Example 3.2.6. On $(\mathbb{R}^n, g_{\mathbb{R}^n})$ define $r(x) = |x - y| = |xy|$. Then r is smooth on $\mathbb{R}^n - \{y\}$ and has $|\nabla r| \equiv 1$. If we have two different points $\{y, z\}$, then

$$r(x) = |x\{y, z\}| = \min\{|x - y|, |x - z|\}$$

is smooth away from $\{y, z\}$ and the hyperplane $\{x \in \mathbb{R}^n \mid |x - y| = |x - z|\}$ equidistant from y and z.

Example 3.2.7. If $H \subset \mathbb{R}^n$ is a submanifold, then it can be shown that

$$r(x) = |xH| = \inf\{|xy| = |x - y| \mid y \in H\}$$

is a distance function on some open set $O \subset \mathbb{R}^n$. When H is an orientable hypersurface this can justified as follows. Since H is orientable, it is possible to choose a unit normal vector field N on H. Now coordinatize $\mathbb{R}^n$ using $x = tN + y$, where $t \in \mathbb{R}$, $y \in H$. In some neighborhood O of H these coordinates are actually well-defined. In other words, there is a function $\varepsilon(y) : H \to (0, \infty)$ such that any point in

$$O = \{tN + y \mid y \in H, \ |t| < \varepsilon(y)\}$$

has unique coordinates (t, y). We can define $r(x) = t$ on O or $f(x) = |xH| = |t|$ on $O - H$. Both functions will then define distance functions on their respective domains. Here r is usually referred to as the *signed distance* to H, while f is just the regular distance.

On $I \times H$, 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 $\{r\} \times H$ that depends on r. In this case the projection $I \times H \to I$ is a distance function. Special cases of this situation are rotationally symmetric metrics, doubly warped products, and our submersion metrics on $I \times S^{2n-1}$.

Lemma 3.2.8. *Let $r : O \to I \subset \mathbb{R}$, where O is an open set in Riemannian manifold. The function r is a distance function if and only if it is a Riemannian submersion.*

Proof. In general, we have $dr(v) = g(\nabla r, v)$, so $Dr(v) = dr(v)\,\partial_t = 0$ if and only if $v \perp \nabla r$. Thus, v is perpendicular to the kernel of Dr if and only if it is proportional to ∇r. For such $v = \alpha \nabla r$ the differential is

$$Dr(v) = \alpha Dr(\nabla r) = \alpha g(\nabla r, \nabla r)\,\partial_t.$$

Now ∂_t has length 1 in I, so

$$\begin{aligned} |v| &= |\alpha|\,|\nabla r|\,, \\ |Dr(v)| &= |\alpha|\,|\nabla r|^2\,. \end{aligned}$$

Thus, r is a Riemannian submersion if and only if $|\nabla r| = 1$. □

Before continuing we introduce some simplifying notation. A distance function $r : O \to \mathbb{R}$ is fixed on an open subset $O \subset (M, g)$ of a Riemannian manifold. The gradient ∇r will usually be denoted by $\partial_r = \nabla r$. The ∂_r notation comes from our warped product metrics $dr^2 + g_r$. The level sets for r are denoted $O_r = \{x \in O \mid r(x) = r\}$, and the induced metric on O_r is g_r. In this spirit ∇^r, $R^r$are the Riemannian connection and curvature on (O_r, g_r). Since $|\nabla r| = 1$ we have that $\operatorname{Hess} r = \mathrm{II}$ and S is the $(1,1)$-tensor corresponding to both $\operatorname{Hess} r$ and II. Here S can stand for second derivative or *shape operator* or *second fundamental form*, depending on the situation. The last two terms are more or less synonymous and refer to the shape of (O_r, g_r) in $(O, g) \subset (M, g)$. The idea is that $S = \nabla \partial_r$ measures how the induced metric on O_r changes by computing how the unit normal to O_r changes.

Example 3.2.9. Let $H \subset \mathbb{R}^n$ be an orientable hypersurface, N the unit normal, and S the shape operator defined by $S(v) = \nabla_v N$ for $v \in TH$. If $S \equiv 0$ on H then N must be a constant vector field on H, and hence H is an open subset of the hyperplane

$$\left\{x + p \in \mathbb{R}^n \mid x \cdot N_p = 0\right\},$$

where $p \in H$ is fixed. As an explicit example of this, recall our isometric immersion or embedding $\left(\mathbb{R}^{n-1}, g_{\mathbb{R}^{n-1}}\right) \to (\mathbb{R}^n, g_{\mathbb{R}^n})$ from example 1.1.3 defined by

$$\left(x^1, \dots, x^{n-1}\right) \to \left(c\left(x^1\right), x^2, \dots, x^{n-1}\right),$$

where c is a unit speed curve $c : \mathbb{R} \to \mathbb{R}^2$. In this case,

$$N = \left(-\dot{c}^2\left(x^1\right), \dot{c}^1\left(x^1\right), 0, \dots, 0\right)$$

is a unit normal in Cartesian coordinates. So

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

Thus, $S \equiv 0$ if and only if $\ddot{c}^1 = \ddot{c}^2 = 0$ if and only if c is a straight line if and only if H is an open subset of a hyperplane. Thus the shape operator really does capture the idea that the hypersurface bends in $\mathbb{R}^n$, even though $\mathbb{R}^{n-1}$ 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. Extrinsic geometry is the study of how an isometric immersion $(M, g) \to (\bar{M}, g_{\bar{M}})$ bends (M, g) inside $(\bar{M}, g_{\bar{M}})$. For example, the curvature tensor on (M, g) measures how the space bends intrinsically, while the shape operator measures extrinsic bending.

3.2.3 The Curvature Equations for Distance Functions

We start by reformulating the radial curvature equation from theorem 3.2.2.

Corollary 3.2.10. *When $r : O \to \mathbb{R}$ is a distance function, then*

$$\nabla_{\partial_r}\partial_r = 0$$

and

$$\nabla_{\partial_r} S + S^2 = -R_{\partial_r}.$$

Proof. The first fact follows from part (3) of proposition 3.2.1 and the second from theorem 3.2.2. □

We conclude that:

Proposition 3.2.11. *If we have a smooth distance function $r : (O, g) \to \mathbb{R}$ and denote $\nabla r = \partial_r$, then*

(1) $L_{\partial_r} g = 2\,\mathrm{Hess}\, r$,
(2) $\left(\nabla_{\partial_r}\mathrm{Hess}\, r\right)(X, Y) + \mathrm{Hess}^2 r\,(X, Y) = -R\,(X, \partial_r, \partial_r, Y)$,
(3) $\left(L_{\partial_r}\mathrm{Hess}\, r\right)(X, Y) - \mathrm{Hess}^2 r\,(X, Y) = -R\,(X, \partial_r, \partial_r, Y)$.

Proof. (1) is simply the definition of the Hessian. (2) and (3) follow directly from theorem 3.2.2 after noting that $|\nabla r| = 1$. □

The first equation shows how the Hessian controls the metric. The second and third equations give us control over the Hessian when we have information about the curvature. These two equations are different in a very subtle way. The third equation is at the moment the easiest to work with as it only uses Lie derivatives and hence can be put in a nice form in an appropriate coordinate system. The second equation is equally useful, but requires that we find a way of making it easier to interpret.

Next we show how appropriate choices for vector fields can give us a better understanding of these fundamental equations.

3.2.4 Jacobi Fields

A *Jacobi field* for a smooth distance function r is a smooth vector field J that does not depend on r, i.e., it satisfies the *Jacobi equation*

$$L_{\partial_r} J = 0.$$

This is a first-order linear PDE, which can be solved by the method of characteristics. To see how this is done we locally select a coordinate system $\left(r, x^2, \ldots, x^n\right)$ where r is the first coordinate. Then $J = J^r \partial_r + J^i \partial_i$ and the Jacobi equation becomes:

$$\begin{aligned} 0 &= L_{\partial_r} J \\ &= L_{\partial_r} \left(J^r \partial_r + J^i \partial_i\right) \\ &= \partial_r \left(J^r\right) \partial_r + \partial_r \left(J^i\right) \partial_i. \end{aligned}$$

Thus the coefficients J^r, J^i have to be independent of r as already indicated. What is more, we can construct such Jacobi fields knowing the values on a hypersurface $H \subset M$ where $\left(x^2, \ldots, x^n\right)|_H$ is a coordinate system. In this case ∂_r is transverse to H and so we can solve the equations by declaring that J^r, J^i are constant along the integral curves for ∂_r. Note that the coordinate vector fields are themselves Jacobi fields.

The equation $L_{\partial_r} J = 0$ is equivalent to the linear equation

$$\nabla_{\partial_r} J = S\left(J\right).$$

This tells us that

$$\operatorname{Hess} r\left(J, J\right) = g\left(\nabla_{\partial_r} J, J\right) = \frac{1}{2} \partial_r g\left(J, J\right).$$

Jacobi fields also satisfy a more general second-order equation, also known as the *Jacobi Equation*:

$$\nabla_{\partial_r}\nabla_{\partial_r}J = -R(J,\partial_r)\,\partial_r,$$

as

$$-R(J,\partial_r)\,\partial_r = R(\partial_r,J)\,\partial_r = \nabla_{\partial_r}S(J)\,.$$

This is a second-order equation and has more solutions than the above first-order equation. This equation will be studied further in section 6.1.5 for general Riemannian manifolds.

Equations (1) and (3) from proposition 3.2.11 when evaluated on Jacobi fields become:

(1) $\quad \partial_r g(J_1,J_2) = 2\,\mathrm{Hess}\,r(J_1,J_2),$
(3) $\quad \partial_r \,\mathrm{Hess}\,r(J_1,J_2) - \mathrm{Hess}^2\,r(J_1,J_2) = -R(J_1,\partial_r,\partial_r,J_2).$

As we only have directional derivatives this is a much simpler version of the fundamental equations. Therefore, there is a much better chance of predicting how g and $\mathrm{Hess}\,r$ change depending on our knowledge of $\mathrm{Hess}\,r$ and R respectively.

This can be reduced a bit further if we take a product neighborhood $\Omega = (a,b)\times H \subset M$ such that $r(t,z) = t$. On this product the metric has the form

$$g = dr^2 + g_r,$$

where g_r is a one parameter family of metrics on H. If J is a vector field on H, then there is a unique extension to a Jacobi field on $\Omega = (a,b)\times H$. First observe that

$$\mathrm{Hess}\,r(\partial_r,J) = g(\nabla_{\partial_r}\partial_r,J) = 0,$$
$$g_r(\partial_r,J) = 0.$$

Thus we only need to consider the restrictions of g and $\mathrm{Hess}\,r$ to H. By doing this we obtain

$$\partial_r g = \partial_r g_r = 2\,\mathrm{Hess}\,r.$$

The fundamental equations can then be written as

(1) $\quad \partial_r g_r = 2\,\mathrm{Hess}\,r,$
(3) $\quad \partial_r\,\mathrm{Hess}\,r - \mathrm{Hess}^2\,r = -R(\cdot,\partial_r,\partial_r,\cdot).$

There is a sticky point hidden in (3). Namely, how is it possible to extract information from R and pass it on to the Hessian without referring to g_r. If we focus on sectional curvature this becomes a little more transparent as

$$\begin{aligned} R(X,\partial_r,\partial_r,X) &= \sec(X,\partial_r)\left(g(X,X)\,g(\partial_r,\partial_r)-(g(X,\partial_r))^2\right)\\ &= \sec(X,\partial_r)\,g(X-g(X,\partial_r)\,\partial_r, X-g(X,\partial_r)\,\partial_r)\\ &= \sec(X,\partial_r)\,g_r(X,X)\,. \end{aligned}$$

So if we evaluate (3) on a Jacobi field J we obtain

$$\partial_r(\operatorname{Hess} r(J,J)) - \operatorname{Hess}^2 r(J,J) = -\sec(J,\partial_r)\,g_r(J,J)\,.$$

This means that (1) and (3) are coupled as we have not eliminated the metric from (3). The next subsection shows how we can deal with this by evaluating on different vector fields.

Nevertheless, we have reduced (1) and (3) to a set of ODEs where r is the independent variable along the integral curve for ∂_r through p.

Example 3.2.12. In the special case where $\dim M = 2$ we can more explicitly write the metric as $g = dr^2 + \rho^2(r,\theta)\,d\theta^2$, where θ denotes a function that locally coordinatizes the level sets of r. In this case ∂_θ is a Jacobi field of length ρ and we obtain the formula

$$2\operatorname{Hess} r(\partial_\theta,\partial_\theta) = \partial_r\rho^2 = 2\rho\partial_r\rho.$$

Since $[\partial_r,\partial_\theta] = 0$ we further have

$$\operatorname{Hess} r(\partial_\theta,\partial_\theta) = g\left(\nabla_{\partial_\theta}\partial_r,\partial_\theta\right) = g(\nabla_{\partial_r}\partial_\theta,\partial_\theta) = \frac{1}{2}\partial_r\rho^2 = \rho\partial_r\rho.$$

As S is self-adjoint and $S(\partial_r) = 0$ this implies

$$S(\partial_\theta) = \frac{\partial_r\rho}{\rho}\partial_\theta.$$

This in turn tells us that

$$\begin{aligned} -\sec(\partial_\theta,\partial_r)\,\rho^2 &= \partial_r(\operatorname{Hess} r(\partial_\theta,\partial_\theta)) - \operatorname{Hess}^2 r(\partial_\theta,\partial_\theta)\\ &= \partial_r(\rho\partial_r\rho) - (\partial_r\rho)^2\\ &= \rho\partial_r^2\rho \end{aligned}$$

and gives us the simple formula for the curvature

$$\sec\left(T_pM\right) = -\frac{\partial_r^2\rho}{\rho}.$$

3.2.5 Parallel Fields

A *parallel field* for a smooth distance function r is a vector field X such that:

$$\nabla_{\partial_r} X = 0.$$

This is, like the Jacobi equation, a first-order linear PDE and can be solved in a similar manner. There is, however, one crucial difference: Parallel fields are almost never Jacobi fields.

If we evaluate g on a pair of parallel fields we see that

$$\partial_r g(X, Y) = g(\nabla_{\partial_r} X, Y) + g(X, \nabla_{\partial_r} Y) = 0.$$

This means that (1) from proposition 3.2.11 is not simplified by using parallel fields. The second equation, on the other hand, becomes

$$\partial_r (\mathrm{Hess}\, r(X, Y)) + \mathrm{Hess}^2 r(X, Y) = -R(X, \partial_r, \partial_r, Y).$$

If this is rewritten in terms of sectional curvature, then we obtain as in section 3.2.4

$$\partial_r (\mathrm{Hess}\, r(X, X)) + \mathrm{Hess}^2 r(X, X) = -\sec(X, \partial_r)\, g_r(X, X).$$

But this time we know that $g_r(X, X)$ is constant in r as X is parallel. We can even assume that $g(X, \partial_r) = 0$ and $g(X, X) = 1$ by first projecting X onto H and then scaling it. Therefore, (2) takes the form

$$\partial_r (\mathrm{Hess}\, r(X, X)) + \mathrm{Hess}^2 r(X, X) = -\sec(X, \partial_r)$$

on unit parallel fields that are orthogonal to ∂_r. In this way we really have decoupled the equation for the Hessian from the metric. This allows us to glean information about the Hessian from information about sectional curvature. Equation (1), when rewritten using Jacobi fields, then gives us information about the metric from the information we just obtained about the Hessian using parallel fields.

3.2.6 Conjugate Points

In general, we might think of the directional curvatures R_{∂_r} as being given or having some specific properties. We then wish to investigate how the curvatures influence the metric according to the equations from proposition 3.2.11 and their simplifications on Jacobi fields or parallel fields from sections 3.2.4 and 3.2.5. Equation (1) is linear. Thus the metric can't degenerate in finite time unless the

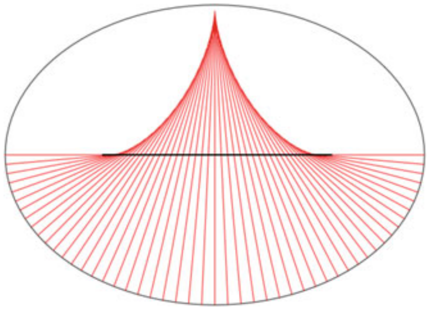


Fig. 3.1 Focal points for an ellipse and its bottom half

Hessian also degenerates. However, if we assume that the curvature is bounded, then equation (2) tells us that, if the Hessian blows up, then it must be decreasing as r increases, hence it can only go to $-\infty$. Going back to (1), we then conclude that the only degeneration which can occur along an integral curve for ∂_r, is that the metric stops being positive definite. We say that the distance function r develops a *conjugate or focal point* along this integral curve. Below we have some pictures of how focal points can develop. Note that as the metric itself is Euclidean, these singularities are relative to the coordinates. There is a subtle difference between conjugate points and focal points. A conjugate point occurs when the Hessian of r becomes undefined as we solve the differential equation for it. A focal point occurs when integral curves for ∇r meet at a point. It is not unusual for both situations to happen at the same point, but it is possible to construct metrics where there are conjugate points that are not focal points.

Figure 3.1 shows that conjugate points for the lower part of the ellipse occur along the evolute of the lower part of the ellipse. However, when we consider the entire ellipse, then the focal set is the line between the focal points of the ellipse as the normal lines from the top and bottom of the ellipse intersect along this line.

It is worthwhile investigating equations (2) and (3) a little further. If we rewrite them as

$$(2) \qquad (\nabla_{\partial_r} \operatorname{Hess} r)(X,X) = -R(X,\partial_r,\partial_r,X) - \operatorname{Hess}^2 r(X,X),$$
$$(3) \qquad (L_{\partial_r} \operatorname{Hess} r)(X,X) = -R(X,\partial_r,\partial_r,X) + \operatorname{Hess}^2 r(X,X),$$

then we can think of the curvatures as representing fixed *external forces,* while $\operatorname{Hess}^2 r$ describes an *internal reaction (or interaction)*. The reaction term is always of a fixed sign, and it will try to force $\operatorname{Hess} r$ to blow up or collapse in finite time. If, for instance $\sec \leq 0$, then $L_{\partial_r} \operatorname{Hess} r$ is positive. Therefore, if $\operatorname{Hess} r$ is positive at some point, then it will stay positive. On the other hand, if $\sec \geq 0$, then $\nabla_{\partial_r} \operatorname{Hess} r$ is negative, forcing $\operatorname{Hess} r$ to stay nonpositive if it is nonpositive at a point.

We shall study and exploit this in much greater detail throughout the book.

3.3 Further Study

In the upcoming chapters we shall mention several other books on geometry that the reader might wish to consult. A classic that is considered old fashioned by some is [40]. It offers a fairly complete treatment of the tensorial aspects of both Riemannian and pseudo-Riemannian geometry. I would certainly recommend this book to anyone who is interested in learning Riemannian geometry. There is also the authoritative guide [70]. Every differential geometer should have a copy of these tomes especially volume 2. Volume 1 contains a lot of foundational material and is probably best as a reference guide.

3.4 Exercises

EXERCISE 3.4.1. Let M be an n-dimensional submanifold of $\mathbb{R}^{n+m}$ with the induced metric. Further assume that we have a local coordinate system given by a parametrization $u^s\left(x^1,\dots,x^n\right)$, $s=1,\dots,n+m$. Show that in these coordinates R_{ijkl} depends only on the first and second partials of u^s. Hint: Look at exercise 2.5.22.

EXERCISE 3.4.2. Consider the following conditions for a smooth function $f : (M,g)\to\mathbb{R}$ on a connected Riemannian manifold:

(1) $|\nabla f|$ is constant.
(2) $\nabla_{\nabla f}\nabla f=0$.
(3) $|\nabla f|$ is constant on the level sets of f.

Show that (1) $\Leftrightarrow$ (2) $\Rightarrow$ (3) and give an example to show that the last implication is not a bi-implication.

EXERCISE 3.4.3. Let f be a function and $S(X)=\nabla_X\nabla f$ the $(1,1)$ version of its Hessian. Show that

$$L_{\nabla f}S=\nabla_{\nabla f}S,$$
$$L_{\nabla f}S+S^2-\nabla_X\left(S\left(\nabla f\right)\right)=-R_{\nabla f}.$$

How do you reconcile this with what happens in theorem 3.2.2 for the $(0,2)$-version of the Hessian?

EXERCISE 3.4.4. Show that if $r=f : M\to\mathbb{R}$ is a distance function, then the tangential and mixed curvature equations from theorems 3.2.4 and 3.2.5 can be written as

$$(R(X,Y)Z)^{\top}=R_H(X,Y)Z-\left(S\left(X\right)\wedge S\left(Y\right)\right)(Z),$$

$$g\left(R(X,Y)Z,N\right)=-g\left(\left(\nabla_XS\right)(Y),Z\right)+g\left(\left(\nabla_YS\right)(X),Z\right),$$

and

$$R(X,Y)N = \left(d^{\nabla}S\right)(X,Y)\,.$$

EXERCISE 3.4.5. Prove the two Bianchi identities at a point $p \in M$ by using a coordinate system where $\nabla_{\partial_i}\partial_j = 0$ at p.

EXERCISE 3.4.6. Show that a Riemannian manifold with constant curvature has parallel curvature tensor.

EXERCISE 3.4.7. Show that a Riemannian manifold with parallel Ricci tensor has constant scalar curvature. In section 4.2.3 it will be shown that the converse is not true, and in section 4.2.2 that a metric with parallel curvature tensor doesn't have to be Einstein.

EXERCISE 3.4.8. Show in analogy with proposition 3.1.5 that if R is the $(0,4)$-curvature tensor and Ric the $(0,2)$-Ricci tensor, then

$$\left(\nabla^{*}R\right)(Z,X,Y) = \left(\nabla_X \operatorname{Ric}\right)(Y,Z) - \left(\nabla_Y \operatorname{Ric}\right)(X,Z)\,.$$

Conclude that $\nabla^{*}R = 0$ if $\nabla \operatorname{Ric} = 0$. Then show that $\nabla^{*}R = 0$ if and only if the $(1,1)$ Ricci tensor satisfies:

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

EXERCISE 3.4.9. Suppose we have two Riemannian manifolds (M, g_M) and (N, g_N). Then the product has a natural product metric $(M \times N, g_M + g_N)$. Let X be 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 vanish.

EXERCISE 3.4.10. Show that a Riemannian manifold has constant curvature at $p \in M$ if and only if $R(v,w)z = 0$ for all orthogonal $v, w, z \in T_pM$. Hint: Start by showing: if a symmetric bilinear form $B(v,w)$ on an inner product space has the property that $B(v,w) = 0$ when $v \perp w$, then B is a multiple of the inner product.

EXERCISE 3.4.11. Use exercises 2.5.25 and 2.5.26 to show that if X, Y, Z are tangent to M, then

$$R^{\bar{M}}(X,Y)Z = R^{M}(X,Y)Z + T_X T_Y Z - T_Y T_X Z + \left(\nabla^{\perp}_X T\right)_Y Z - \left(\nabla^{\perp}_Y T\right)_X Z$$

where

$$\left(\nabla^{\perp}_X T\right)_Y Z = \nabla^{\perp}_X \left(T_Y Z\right) - T_{\nabla^M_X Y} Z - T_Y \nabla^M_X Z.$$

The tangential parts on both sides of this curvature relation form the Gauss equations and the normal parts the Peterson-Codazzi-Mainardi equations.

EXERCISE 3.4.12. Let $H^{n-1} \subset \mathbb{R}^n$ be a hypersurface. Show that $\mathrm{Ric}^H = \operatorname{tr} \mathrm{II} \cdot \mathrm{II} - \mathrm{II}^2$

EXERCISE 3.4.13. A hypersurface of a Riemannian manifold is called totally geodesic if its second fundamental form vanishes.

(1) Show that the spaces S^n_k have the property that any tangent vector is normal to a totally geodesic hypersurface.
(2) Show a Riemannian n-manifold, $n > 2$, with the property that any tangent vector is a normal vector to a totally geodesic hypersurface has constant curvature. Hint: Start by showing that $R(X,Y)Z = 0$ when the three vectors are orthogonal to each other and use exercise 3.4.10.

EXERCISE 3.4.14. Use exercise 2.5.26 to define the normal curvature

$$R^{\perp}(X, Y, V, W)$$

for tangent fields X, Y and normal fields V, W.

(1) Show that $R^{\perp}$ is tensorial and skew-symmetric in X, Y as well as V, W.
(2) Show that

$$R^{\bar{M}}(X, Y, V, W) = R^{\perp}(X, Y, V, W) + g_M(T_X V, T_Y W) - g_M(T_Y V, T_X W)$$

These are also known as the *Ricci equations*.

EXERCISE 3.4.15. For 3-dimensional manifolds, show that if the curvature operator in diagonal form is given by

$$\begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \gamma \end{pmatrix},$$

then the Ricci curvature has a diagonal given by

$$\begin{pmatrix} \alpha+\beta & 0 & 0 \\ 0 & \beta+\gamma & 0 \\ 0 & 0 & \alpha+\gamma \end{pmatrix}.$$

Moreover, the numbers α, β, γ must be sectional curvatures.

EXERCISE 3.4.16. Consider the $(0,2)$-tensor

$$T = \mathrm{Ric} + b \operatorname{scal} g + cg$$

where $b, c \in \mathbb{R}$.

(1) Show that $\nabla^* T = 0$ if $b = -\frac{1}{2}$. The tensor

$$G = \mathrm{Ric} - \frac{\mathrm{scal}}{2} g + cg.$$

is known as the *Einstein tensor* and c as the *cosmological constant*.

(2) Show that if $c = 0$, then $G = 0$ in dimension 2.

(3) When $n > 2$ show that if $G = 0$, then the metric is an Einstein metric.

(4) When $n > 2$ show that if $G = 0$ and $c = 0$, then the metric is Ricci flat.

EXERCISE 3.4.17. Let $T_{\mathbb{C}}M = TM \otimes \mathbb{C}$ be the complexified tangent bundle to a manifold. A vector $v \in T_{\mathbb{C}}M$ looks like $v = v_1 + \mathrm{i}\, v_2$, where $v_1, v_2 \in TM$, and can be conjugated $\bar{v} = v_1 - \mathrm{i}\, v_2$. Any tensorial object on TM can be complexified. For example, if S is a $(1,1)$-tensor, then its complexification is given by

$$S_{\mathbb{C}}(v) = S_{\mathbb{C}}(v_1 + \mathrm{i}\, v_2) = S(v_1) + \mathrm{i}\, S(v_2).$$

A Riemannian structure g on TM gives a natural Hermitian structure on $T_{\mathbb{C}}M$ by

$$\begin{aligned} g(v, w) &= g_{\mathbb{C}}(v, \bar{w}) \\ &= g_{\mathbb{C}}(v_1 + \mathrm{i}\, v_2, w_1 - \mathrm{i}\, w_2) \\ &= g(v_1, w_1) + g(v_2, w_2) + \mathrm{i}\,(g(v_2, w_1) - g(v_1, w_2)). \end{aligned}$$

A vector is called *isotropic* if it is Hermitian orthogonal to its conjugate

$$\begin{aligned} 0 &= g(v, \bar{v}) \\ &= g_{\mathbb{C}}(v, v) \\ &= g_{\mathbb{C}}(v_1 + \mathrm{i}\, v_2, v_1 + \mathrm{i}\, v_2) \\ &= g(v_1, v_1) - g(v_2, v_2) + \mathrm{i}\,(g(v_2, v_1) + g(v_1, v_2)). \end{aligned}$$

More generally, isotropic subspaces are defined as subspaces on which $g_{\mathbb{C}}$ vanishes. The *complex sectional curvature* spanned by Hermitian orthonormal vectors v, w is given by the expression

$$R_{\mathbb{C}}(v, w, \bar{w}, \bar{v}).$$

It is called *isotropic sectional curvature* when v, w span an isotropic plane.

(1) Show that a vector $v = v_1 + \mathrm{i}\, v_2$ is isotropic if v_1, v_2 are orthogonal and have the same length.

(2) An isotropic plane can be spanned by two Hermitian orthonormal vectors v, w that are isotropic. Show that if $v = v_1 + \mathrm{i}\, v_2$ and $w = w_1 + \mathrm{i}\, w_2$, then v_1, v_2, w_1, w_2 are orthonormal.

(3) Show that $R_{\mathbb{C}}(v, w, \bar{w}, \bar{v})$ is always a real number.
(4) Show that if the original metric is strictly quarter pinched, i.e., all sectional curvatures lie in an open interval of the form $\left(\frac{1}{4}k, k\right)$ with $k > 0$, then the complex sectional curvatures are positive.
(5) Show that the complex sectional curvatures are nonnegative (resp. positive) if the curvature operator is nonnegative (resp. positive). Hint: Calculate

$$g(\mathfrak{R}(x \wedge u - y \wedge v), x \wedge u - y \wedge v) + g(\mathfrak{R}(x \wedge v + y \wedge u), x \wedge v + y \wedge u)$$

and compare it to a suitable complex curvature.

EXERCISE 3.4.18. Consider a Riemannian metric (M, g) and *scale* the metric by multiplying it by a number λ^2. This creates a new Riemannian manifold $\left(M, \lambda^2 g\right)$.

(1) Show that the new connection and $(1, 3)$-curvature tensor remain the same.
(2) Show that sec, scal, and $\mathfrak{R}$ all get multiplied by λ^{-2}.
(3) Show that Ric as a $(1, 1)$-tensor is multiplied by λ^{-2}.
(4) Show that Ric as a $(0, 2)$-tensor is unchanged.

EXERCISE 3.4.19. We say that X is an *affine vector field* if $L_X \nabla = 0$. Show that such a field satisfies the equation: $\nabla^2_{U,V} X = -R(X, U) V$.

EXERCISE 3.4.20 (INTEGRABILITY FOR PDEs). For given functions $P^i_k(x, u)$, where $x = \left(x^1, \ldots, x^n\right)$, $u = \left(u^1, \ldots, u^m\right)$, $i = 1, \ldots, m$, and $k = 1, \ldots, n$, consider the initial value problems for a system of first-order PDEs

$$\begin{aligned}\frac{\partial u^i}{\partial x^k} &= P^i_k(x, u(x)),\\ u(x_0) &= u_0.\end{aligned}$$

(1) Show that

$$\frac{\partial^2 u^i}{\partial x^k \partial x^l} = \frac{\partial P^i_l}{\partial x^k} + \frac{\partial P^i_l}{\partial u^j} P^j_k$$

and conclude that all such initial value problems can only be solved when the *integrability conditions*

$$\frac{\partial P^i_l}{\partial x^k} + \frac{\partial P^i_l}{\partial u^j} P^j_k = \frac{\partial P^i_k}{\partial x^l} + \frac{\partial P^i_k}{\partial u^j} P^j_l$$

hold.
(2) Conversely show that all such initial value problems can be solved if the integrability conditions hold. Hint: This is equivalent to the Frobenius integrability theorem but can be established directly (see also [97, vol. 1]). When P does not depend on u, this result goes back to Clairaut. The general case appears to have been a folklore result that predates what we call the Frobenius theorem about integrability of distributions.

(3) Using coordinates x^i on a Riemannian n-manifold form the system

$$\frac{\partial U^i_j}{\partial x^k} = \Gamma^s_{kj} U^i_s,\ i,j = 1,\dots,n$$

and show that its integrability conditions are equivalent to $R^s_{klj} = 0$.

(4) Show that a flat Riemannian manifold admits Cartesian coordinates. Hint: Denote the potential Cartesian coordinates by u^i and consider the system:

$$\frac{\partial u^i}{\partial x^k} = U^i_k$$

with appropriate initial values. Make sure you check that u^i really form a Cartesian coordinate system. This way of locally characterizing Euclidean space is very close in spirit to Riemann's original approach. Hint: Consider the derivative of

$$g^{kl} \frac{\partial u^i}{\partial x^k} \frac{\partial u^j}{\partial x^l},$$

where g_{kl} denotes the metric with respect to x and use 2.5.8.

EXERCISE 3.4.21 (FUNDAMENTAL THEOREM OF (HYPER-)SURFACE THEORY). Consider a Riemannian immersion $F : M^n \looparrowright \mathbb{R}^{n+1}$. In coordinates on M it can be written as

$$\left(u^1(x), \dots, u^{n+1}(x)\right) = F(x) = F\left(x^1, \dots, x^n\right)$$

and we define

$$U^i_k = \frac{\partial u^i}{\partial x^k}.$$

(1) Show that

$$\frac{\partial U^i_j}{\partial x^k} = \Gamma^s_{kj} U^i_s - \mathrm{II}_{jk} N^i,$$

where $N = N^i \frac{\partial}{\partial u^i}$ is a choice of unit normal and the second fundamental form is $\mathrm{II}_{jk} = \mathrm{II}\left(\partial_j, \partial_k\right) = g\left(\nabla_{\partial_j} N, \partial_k\right)$.

(2) Show that the integrability conditions for this system are equivalent to the Gauss (tangential) and Codazzi (mixed) curvature equations:

$$R_{iklj} = \mathrm{II}_{ij}\,\mathrm{II}_{kl} - \mathrm{II}_{ik}\,\mathrm{II}_{jl}$$

$$\frac{\partial \, \mathrm{II}_{jk}}{\partial x^l} - \frac{\partial \, \mathrm{II}_{jl}}{\partial x^k} = \Gamma_{lj}^s \, \mathrm{II}_{sk} - \Gamma_{kj}^s \, \mathrm{II}_{sl}$$

(3) Given metric coefficients g_{ij} and a symmetric tensor II_{ij} that is related to the metric coefficients through the Gauss and Codazzi equations, show that locally there exists a Riemannian immersion such that the second fundamental form is given by II_{ij}.
(4) We can now give a local characterization of spaces with constant positive curvature. Given a metric of constant curvature $R^{-2} > 0$, show that there is a Riemannian immersion into $\mathbb{R}^{n+1}$whose image lies in a sphere of radius R. Hint: Guess what the second fundamental form should look like and show that the constant curvature condition gives the Gauss and Codazzi equations. Note that for $S^n(R)$ the unit normal is $N = \pm R^{-1} F$.

EXERCISE 3.4.22. Repeat the previous exercise with a Riemannian immersion $F : M^n \looparrowright \mathbb{R}^{n,1}$ where M is a Riemannian manifold and the normal N satisfies $|N|^2 = -1$. This time we obtain a local characterization of the hyperbolic spaces $H^n(R)$ from example 1.1.7 as the local model for spaces of constant curvature $-R^{-2}$. Note that for $H^n(R)$ the unit normal is $N = \pm R^{-1} F$.

EXERCISE 3.4.23. For two symmetric $(0, 2)$-tensors h, k define the *Kulkarni-Nomizu product* as the $(0, 4)$-tensor

$$\begin{aligned} h \circ k (v_1, v_2, v_3, v_4) = {} & \frac{1}{2} (h(v_1, v_4) \cdot k(v_2, v_3) + h(v_2, v_3) \cdot k(v_1, v_4)) \\ & - \frac{1}{2} (h(v_1, v_3) \cdot k(v_2, v_4) + h(v_2, v_4) \cdot k(v_1, v_3)). \end{aligned}$$

The factor $\frac{1}{2}$ is not used consistently in the literature, but is convenient when $h = k$. Part (6) of this exercise explains our choice.

(1) Show that $h \circ k = k \circ h$.
(2) Show that $h \circ h = 0$ if h has rank 1.
(3) Show that if $n > 2$; k is nondegenerate; and $h \circ k = 0$, then $h = 0$. Hint: Let v_i be “eigenvectors” for k and $v_2 = v_3$.
(4) Show that $h \circ k$ satisfies the first 3 properties of proposition 3.1.1.
(5) Show that $\nabla_X (h \circ k) = (\nabla_X h) \circ k + h \circ (\nabla_X k)$.
(6) Show that (M, g) has constant curvature c if and only if the $(0, 4)$-curvature tensor satisfies $R = c \cdot (g \circ g)$.

EXERCISE 3.4.24. Define the *Schouten tensor*

$$P = \frac{2}{n-2} \mathrm{Ric} - \frac{\mathrm{scal}}{(n-1)(n-2)} \cdot g$$

for Riemannian manifolds of dimension $n > 2$.

(1) Show that if P vanishes on M, then $\mathrm{Ric} = 0$.
(2) Show that the decomposition

$$P = \frac{\mathrm{scal}}{n(n-1)} g + \frac{2}{n-2}\left(\mathrm{Ric} - \frac{\mathrm{scal}}{n}\cdot g\right)$$

of the Schouten tensor is orthogonal.
(3) Show that when $n = 2$, then

$$R = \frac{\mathrm{scal}}{2} g \circ g.$$

(4) Show that when $n = 3$, then

$$R = \frac{\mathrm{scal}}{6} g \circ g + 2\left(\mathrm{Ric} - \frac{\mathrm{scal}}{3}\cdot g\right) \circ g = P \circ g.$$

(5) Show that (M, g) has constant curvature when $n > 2$ if and only if

$$R = P \circ g \text{ and } \mathrm{Ric} = \frac{\mathrm{scal}}{n} g.$$

(6) Show that

$$\mathrm{Ric}(X, Y) = \sum_{i=1}^{n} (P \circ g)(X, E_i, E_i, Y)$$

for any orthonormal frame E_i.

EXERCISE 3.4.25. The *Weyl tensor* W is defined implicitly through

$$\begin{aligned} R &= \frac{\mathrm{scal}}{n(n-1)} g \circ g + \frac{2}{n-2}\left(\mathrm{Ric} - \frac{\mathrm{scal}}{n}\cdot g\right) \circ g + W \\ &= P \circ g + W, \end{aligned}$$

where P was defined in the previous exercise.

(1) Show that if $n = 3$, then $W = 0$.
(2) Show that

$$\sum_{i=1}^{n} W(X, E_i, E_i, Y) = 0$$

for any orthonormal frame E_i. Hint: Use (6) from exercise 3.4.24.
(3) Show that the decomposition $R = P \circ g + W$ is orthogonal. Hint: This is similar to showing that homotheties and traceless matrices are perpendicular.

EXERCISE 3.4.26. Show that

$$\nabla^* P = -\frac{1}{n-1} d\,\mathrm{scal}$$

and

$$\nabla^* W(Z, X, Y) = \frac{n-3}{2}\left(\left(\nabla_X P\right)(Y, Z) - \left(\nabla_Y P\right)(X, Z)\right).$$

Hint: Use the definitions of W and P from the previous two exercises, exercise 3.4.8, and proposition 3.1.5.

EXERCISE 3.4.27. Given an orthonormal frame $E_1, \ldots, E_n$ on (M, g), define the *structure constants* c_{ij}^k by $\left[E_i, E_j\right] = c_{ij}^k E_k$, note that each c_{ij}^k is a function on M, so it is not constant! Define the Γs and Rs by

$$\nabla_{E_i} E_j = \Gamma_{ij}^k E_k,$$
$$R\left(E_i, E_j\right) E_k = R_{ijk}^l E_l$$

and compute them in terms of the structure constants. Notice that on Lie groups with left-invariant metrics the structure constants can be assumed to be constant. In this case, computations simplify considerably.

EXERCISE 3.4.28 (CARTAN FORMALISM). There is yet another 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 called the *connection forms*. Thus,

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

Suppose additionally that the frame is orthonormal and let ω^i be the dual coframe, i.e., $\omega^i\left(E_j\right) = \delta_j^i$.

(1) Show that the connection forms satisfy

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

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.

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

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

Show that they satisfy

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

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

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

EXERCISE 3.4.29. This exercise will give you a way of finding the curvature tensor from the sectional curvatures. Assume that $R(X,Y,Z,W)$ is an algebraic curvature tensor, i.e., satisfies (1), (2), and (3) of proposition 3.1.1.

(1) Show that

$$\begin{aligned} 6R(X,Y,V,W) = & \left.\frac{\partial^2 R(X+sW, Y+tV, Y+tV, X+sW)}{\partial s \partial t}\right|_{s=t=0} \\ & - \left.\frac{\partial^2 R(X+sV, Y+tW, Y+tW, X+sV)}{\partial s \partial t}\right|_{s=t=0}. \end{aligned}$$

(2) Show that

$$\begin{aligned} 6R(X,Y,V,W) = R&(X+W, Y+V, Y+V, X+W) \\ &- R(X, Y+V, Y+V, X) - R(W, Y+V, Y+V, W) \\ &- R(X+W, V, V, X+W) - R(X+W, Y, Y, X+W) \\ &+ R(X,V,V,X) + R(W,V,V,W) \\ &+ R(X,Y,Y,X) + R(W,Y,Y,W) \end{aligned}$$

$$
\begin{aligned}
&-R(X+V, Y+W, Y+W, X+V) \\
&+R(X, Y+W, Y+W, X)+R(V, Y+W, Y+W, V) \\
&+R(X+V, Y, Y, X+V)+R(X+V, W, W, X+V) \\
&-R(X, Y, Y, X)-R(V, Y, Y, V) \\
&-R(X, W, W, X)-R(V, W, W, V) .
\end{aligned}
$$

Note that 4 of the terms on the right-hand side are redundant.

EXERCISE 3.4.30. Using the previous exercise show that the norm of the curvature operator on $\Lambda^2 T_pM$ is bounded by

$$
\left|\mathfrak{R}|_p\right| \leq c(n)\,|\sec|_p
$$

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_pM.

EXERCISE 3.4.31. Let G be a Lie group with a left-invariant metric $(\cdot,\cdot)$ on $\mathfrak{g}$ (it need not be positive definite just nondegenerate). For $X \in \mathfrak{g}$ denote by $\operatorname{ad}_X^* : \mathfrak{g} \to \mathfrak{g}$ the adjoint of $\operatorname{ad}_X Y = [X, Y]$ with respect to $(\cdot,\cdot)$. Show that:

(1) $\nabla_X Y = \frac{1}{2}\left([X, Y] + \operatorname{ad}_X^* Y - \operatorname{ad}_Y^* X\right)$. Conclude that if $X, Y \in \mathfrak{g}$, then $\nabla_X Y \in \mathfrak{g}$.
(2) $R(X, Y, Z, W) = -(\nabla_Y Z, \nabla_X W) + (\nabla_X Z, \nabla_Y W) - \left(\nabla_{[X,Y]} Z, W\right)$.
(3)

$$
\begin{aligned}
R(X, Y, Y, X) = {} & \frac{1}{4}\left|\operatorname{ad}_X^* Y + \operatorname{ad}_Y^* X\right|^2 \\
& - \left(\operatorname{ad}_X^* X, \operatorname{ad}_Y^* Y\right) - \frac{3}{4}\left|[X, Y]\right|^2 \\
& - \frac{1}{2}\left([[X, Y], Y], X\right) - \frac{1}{2}\left([[Y, X], X], Y\right).
\end{aligned}
$$

EXERCISE 3.4.32. Let G be a Lie group with a biinvariant metric $(\cdot,\cdot)$ on $\mathfrak{g}$ (it need not be positive definite just nondegenerate). Using left-invariant fields establish the following formulas. Hint: First go back to the exercise 1.6.24 and take a peek at section 4.4.1 where some of these things are proved. Show that:

(1) $\nabla_X Y = \frac{1}{2}[X, Y]$.
(2) $R(X, Y) Z = \frac{1}{4}[Z, [X, Y]]$.
(3) $R(X, Y, Z, W) = -\frac{1}{4}([X, Y], [Z, W])$. Conclude that the sectional curvatures are nonnegative when $(\cdot,\cdot)$ is positive definite.
(4) Show that the curvature operator is also nonnegative when $(\cdot,\cdot)$ is positive definite by showing that:

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

(5) Assume again that $(\cdot,\cdot)$ is positive definite. Show that $\mathrm{Ric}\,(X,X) = 0$ if and only if X commutes with all other left-invariant vector fields. Thus G has positive Ricci curvature if the center of G is discrete.

EXERCISE 3.4.33. Consider a Lie group where the Killing form B is nondegenerate and use $-B$ as the left-invariant metric (see exercise 1.6.27).

(1) Show that this metric is biinvariant.
(2) Show that $\mathrm{Ric} = -\frac{1}{4}B$.

EXERCISE 3.4.34. It is illustrative to use the Cartan formalism in the previous exercise and compute all quantities in terms of the structure constants for the Lie algebra. Given that the metric is biinvariant, it follows that with respect to an orthonormal basis they satisfy

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

The first equality is skew-symmetry of the Lie bracket, and the second is biinvariance of the metric.

Chapter 4
Examples

We are now ready to compute the curvature tensors on all of the examples constructed in chapter 1. After a few more general computations, we will exhibit Riemannian manifolds with constant sectional, Ricci, and scalar curvature. In particular, we shall look at the space forms S_k^n, products of spheres, and the Riemannian version of the Schwarzschild metric. We also offer a local characterization of certain warped products and rotationally symmetric constant curvature metrics in terms of the Hessian of certain modified distance functions.

The examples we present here include a selection of important techniques such as: Conformal change, left-invariant metrics, warped products, Riemannian submersion constructions etc. We shall not always develop the techniques in complete generality. Rather we show how they work in some basic, but important, examples. The exercises also delve into important ideas that are not needed for further developments in the text.

4.1 Computational Simplifications

Before we do more concrete calculations it will be useful to have some general results that deal with how one finds the range of the various curvatures.

Proposition 4.1.1. *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 the plane π, then we have $\sec(\pi) = g(\mathfrak{R}(v \wedge w), (v \wedge w))$, so the result is immediate. □

P. Petersen, *Riemannian Geometry*, Graduate Texts in Mathematics 171,
DOI 10.1007/978-3-319-26654-1_4

Proposition 4.1.2. *Let e_i be an orthonormal basis for T_pM. If $R\left(e_i, e_j\right) e_k = 0$, when the indices are mutually distinct, then $e_i \wedge e_j$ diagonalize the curvature operator.*

Proof. If we use

$$\begin{aligned} g\left(\mathfrak{R}\left(e_i \wedge e_j\right), \left(e_k \wedge e_l\right)\right) &= -g\left(R\left(e_i, e_j\right) e_k, e_l\right) \\ &= g\left(R\left(e_i, e_j\right) e_l, e_k\right), \end{aligned}$$

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. □

We shall see that this proposition applies to all rotationally symmetric and doubly warped products. In this case, the curvature operator can then be computed by finding the expressions $R\left(e_i, e_j, e_j, e_i\right)$. In general, however, this will definitely not work.

There is also a more general situation where we can find the range of the Ricci curvatures:

Proposition 4.1.3. *Let e_i be an orthonormal basis for T_pM. If*

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

when three of the indices are mutually distinct, then e_i diagonalize Ric.

Proof. Recall that

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

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

4.2 Warped Products

So far, all we know about curvature is that Euclidean space has $R = 0$. Using this, we determine the curvature tensor on $S^{n-1}(R)$. Armed with that information we can in turn calculate the curvatures on rotationally symmetric metrics.

4.2.1 Spheres

On $\mathbb{R}^n$ consider the distance function $r(x) = |x|$ and the polar coordinate representation:

$$g = dr^2 + g_r = dr^2 + r^2 ds_{n-1}^2,$$

where ds_{n-1}^2 is the canonical metric on $S^{n-1}(1)$. The level sets are $O_r = S^{n-1}(r)$ with the usual induced metric $g_r = r^2 ds_{n-1}^2$. The differential of r is given by $dr = \sum \frac{x^i}{r} dx^i$ and the gradient is $\partial_r = \frac{1}{r} x^i \partial_i$. Since ds_{n-1}^2 is independent of r we can compute the Hessian of r as follows:

$$\begin{aligned}
2 \operatorname{Hess} r &= L_{\partial_r} g \\
&= L_{\partial_r} \left(dr^2\right) + L_{\partial_r} \left(r^2 ds_{n-1}^2\right) \\
&= L_{\partial_r} (dr)\, dr + dr L_{\partial_r} (dr) + \partial_r \left(r^2\right) ds_{n-1}^2 + r^2 L_{\partial_r} \left(ds_{n-1}^2\right) \\
&= \partial_r \left(r^2\right) ds_{n-1}^2 \\
&= 2r ds_{n-1}^2 \\
&= 2 \tfrac{1}{r} g_r.
\end{aligned}$$

The tangential curvature equation (see theorem 3.2.4) tells us that

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

since the curvature on $\mathbb{R}^n$ is zero. In particular, if e_i is any orthonormal basis, then $R^r \left(e_i, e_j\right) e_k = 0$ when the indices are mutually distinct. Therefore, $S^{n-1}(R)$ has constant curvature R^{-2} provided $n \geq 3$. This justifies our notation that S_k^n is the rotationally symmetric metric $dr^2 + \operatorname{sn}_k^2(r) ds_{n-1}^2$ when $k \geq 0$, as these metrics have curvature k in this case. In section 4.2.3 we shall see that this is also true when $k < 0$.

4.2.2 Product Spheres

Next we compute the curvatures on the product spheres

$$S_a^n \times S_b^m = S^n \left(\frac{1}{\sqrt{a}}\right) \times S^m \left(\frac{1}{\sqrt{b}}\right).$$

The metric g_r on $S^n(r)$ is $g_r = r^2 ds_n^2$, so we can write

$$S_a^n \times S_b^m = \left(S^n \times S^m, \frac{1}{a}ds_n^2 + \frac{1}{b}ds_m^2\right).$$

Let Y be a unit vector field on S^n, V a unit vector field on S^m, and X a unit vector field on either S^n or S^m that is perpendicular to both Y and V. The Koszul formula shows that

$$\begin{aligned}2g\left(\nabla_Y X, V\right) &= g\left([Y,X], V\right) + g\left([V,Y], X\right) - g\left([X,V], Y\right)\\ &= g\left([Y,X], V\right) - g\left([X,V], Y\right)\\ &= 0,\end{aligned}$$

as $[Y,X]$ is either zero or tangent to S^n and likewise with $[X,V]$. Thus $\nabla_Y X = 0$ if X is tangent to S^m and $\nabla_Y X$ is tangent to S^n if X is tangent to S^n. This shows that $\nabla_Y X$ can be computed on S_a^n. We can then calculate R knowing the curvatures on the two spheres from section 4.2.1 and invoke proposition 4.1.2 to obtain:

$$\begin{aligned}\mathfrak{R}(X \wedge V) &= 0,\\ \mathfrak{R}(X \wedge Y) &= aX \wedge Y,\\ \mathfrak{R}(U \wedge V) &= bU \wedge V.\end{aligned}$$

In particular, proposition 4.1.1 shows that all sectional curvatures lie in the interval $[0, \max\{a,b\}]$. It also follows that

$$\begin{aligned}\mathrm{Ric}(X) &= (n-1)aX,\\ \mathrm{Ric}\,(V) &= (m-1)\,bV,\\ \mathrm{scal} &= n(n-1)a + m(m-1)b.\end{aligned}$$

Therefore, we 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 $n = m = 1$), and has constant sectional curvature only when $n = m = 1$. Note also that the curvature tensor on $S_a^n \times S_b^m$ is always parallel.

4.2.3 Rotationally Symmetric Metrics

Next we consider what happens for a general rotationally symmetric metric

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

The metric is of the form $g = dr^2 + g_r$ on $(a,b) \times S^{n-1}$, with $g_r = \rho^2 ds_{n-1}^2$. As ds_{n-1}^2 does not depend on r we have that

$$
\begin{aligned}
2\,\mathrm{Hess}\, r &= L_{\partial_r} g_r \\
&= L_{\partial_r}\left(\rho^2 ds_{n-1}^2\right) \\
&= \partial_r\left(\rho^2\right) ds_{n-1}^2 + \rho^2 L_{\partial_r}\left(ds_{n-1}^2\right) \\
&= 2\rho\left(\partial_r \rho\right) ds_{n-1}^2 \\
&= 2\frac{\partial_r \rho}{\rho} g_r .
\end{aligned}
$$

The Lie and covariant derivatives of the Hessian are computed as follows:

$$
\begin{aligned}
L_{\partial_r}\,\mathrm{Hess}\, r &= L_{\partial_r}\left(\frac{\partial_r \rho}{\rho} g_r\right) \\
&= \partial_r\left(\frac{\partial_r \rho}{\rho}\right) g_r + \frac{\partial_r \rho}{\rho} L_{\partial_r}\left(g_r\right) \\
&= \frac{\left(\partial_r^2 \rho\right)\rho - \left(\partial_r \rho\right)^2}{\rho^2} g_r + 2\left(\frac{\partial_r \rho}{\rho}\right)^2 g_r \\
&= \frac{\partial_r^2 \rho}{\rho} g_r + \left(\frac{\partial_r \rho}{\rho}\right)^2 g_r \\
&= \frac{\partial_r^2 \rho}{\rho} g_r + \mathrm{Hess}^2\, r
\end{aligned}
$$

and

$$
\begin{aligned}
\nabla_{\partial_r}\,\mathrm{Hess}\, r &= \nabla_{\partial_r}\left(\frac{\partial_r \rho}{\rho} g_r\right) \\
&= \partial_r\left(\frac{\partial_r \rho}{\rho}\right) g_r + \frac{\partial_r \rho}{\rho} \nabla_{\partial_r}\left(g_r\right) \\
&= \frac{\left(\partial_r^2 \rho\right)\rho - \left(\partial_r \rho\right)^2}{\rho^2} g_r \\
&= \frac{\partial_r^2 \rho}{\rho} g_r - \left(\frac{\partial_r \rho}{\rho}\right)^2 g_r \\
&= \frac{\partial_r^2 \rho}{\rho} g_r - \mathrm{Hess}^2\, r.
\end{aligned}
$$

The fundamental equations from proposition 3.2.11 show that when restricted to S^{n-1} we have

$$
\mathrm{Hess}\, r = \frac{\partial_r \rho}{\rho} g_r ,
$$

$$R(\cdot, \partial_r, \partial_r, \cdot) = -\frac{\partial_r^2 \rho}{\rho} g_r.$$

This implies that

$$\nabla_X \partial_r = \begin{cases} \frac{\partial_r \rho}{\rho} X & \text{if } X \text{ is tangent to } S^{n-1}, \\ 0 & \text{if } X = \partial_r. \end{cases}$$

$$R(X, \partial_r)\,\partial_r = \begin{cases} -\frac{\partial_r^2 \rho}{\rho} X & \text{if } X \text{ is tangent to } S^{n-1}, \\ 0 & \text{if } X = \partial_r. \end{cases}$$

Next we calculate the other curvatures on

$$\left(I \times S^{n-1}, dr^2 + \rho^2(r) ds_{n-1}^2\right)$$

that come from the tangential and mixed curvature equations (see theorems 3.2.4 and 3.2.5)

$$g(R(X,Y)V,W) = g_r(R^r(X,Y)V,W) - \mathrm{II}(Y,V)\,\mathrm{II}(X,W) + \mathrm{II}(X,V)\,\mathrm{II}(Y,W),$$
$$g(R(X,Y)Z,\partial_r) = -(\nabla_X \mathrm{II})(Y,Z) + (\nabla_Y \mathrm{II})(X,Z).$$

Using that g_r is the metric of curvature $\frac{1}{\rho^2}$ on the sphere, we have from section 4.2.1 that

$$g_r\left(R^r(X,Y)V,W\right) = \frac{1}{\rho^2} g_r(X \wedge Y, W \wedge V).$$

Combining this with $\mathrm{II} = \mathrm{Hess}\, r$ we obtain from the first equation that

$$g(R(X,Y)V,W) = \frac{1 - (\partial_r \rho)^2}{\rho^2} g_r(X \wedge Y, W \wedge V).$$

Finally we show that the mixed curvature vanishes as $\frac{\partial_r \rho}{\rho}$ depends only on r :

$$\begin{aligned} \nabla_X \mathrm{II} &= \nabla_X \left(\frac{\partial_r \rho}{\rho} g_r\right) \\ &= D_X\left(\frac{\partial_r \rho}{\rho}\right) g_r + \frac{\partial_r \rho}{\rho} \nabla_X g_r \\ &= 0. \end{aligned}$$

From this we can use proposition 4.1.2 to conclude

$$\mathfrak{R}(X \wedge \partial_r) = -\frac{\partial_r^2 \rho}{\rho} X \wedge \partial_r = -\frac{\ddot{\rho}}{\rho} X \wedge \partial_r,$$

$$\mathfrak{R}(X \wedge Y) = \frac{1 - (\partial_r \rho)^2}{\rho^2} X \wedge Y = \frac{1 - \dot{\rho}^2}{\rho^2} X \wedge Y$$

In particular, we have diagonalized $\mathfrak{R}$. Hence all sectional curvatures lie between the two values $-\frac{\ddot{\rho}}{\rho}$ and $\frac{1-\dot{\rho}^2}{\rho^2}$. Furthermore, if we select an orthonormal basis E_i where $E_1 = \partial_r$, then the Ricci tensor and scalar curvature are

$$\begin{aligned}
\operatorname{Ric}(X) &= \sum_{i=1}^{n} R(X, E_i) E_i \\
&= \sum_{i=1}^{n-1} R(X, E_i) E_i + R(X, \partial_r) \partial_r \\
&= \left((n-2) \frac{1 - \dot{\rho}^2}{\rho^2} - \frac{\ddot{\rho}}{\rho} \right) X, \\
\operatorname{Ric}(\partial_r) &= -(n-1) \frac{\ddot{\rho}}{\rho} \partial_r \\
\operatorname{scal} &= -(n-1) \frac{\ddot{\rho}}{\rho} + (n-1) \left((n-2) \frac{1 - \dot{\rho}^2}{\rho^2} - \frac{\ddot{\rho}}{\rho} \right) \\
&= -2(n-1) \frac{\ddot{\rho}}{\rho} + (n-1)(n-2) \frac{1 - \dot{\rho}^2}{\rho^2}.
\end{aligned}$$

When $n = 2$, it follows that $\sec = -\frac{\ddot{\rho}}{\rho}$, as there are no tangential curvatures. This makes for quite a difference between 2- and higher-dimensional rotationally symmetric metrics.

Constant curvature: First, we compute the curvature of $dr^2 + \operatorname{sn}_k^2(r) ds_{n-1}^2$ on S_k^n. Since $\rho = \operatorname{sn}_k$ solves $\ddot{\rho} + k\rho = 0$ it follows that $\sec(X, \partial_r) = k$. To compute $\frac{1-\dot{\rho}^2}{\rho^2}$ recall from section 1.4.3 that if $\rho = \operatorname{sn}_k(r)$, then

$$\begin{aligned}
\dot{\rho} &= \operatorname{cs}_k, \\
1 - \dot{\rho}^2 &= k\rho^2.
\end{aligned}$$

Thus, all sectional curvatures are equal to k, 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

$$\frac{\ddot{\rho}}{\rho} = 0,$$

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

Hence, $\ddot{\rho} \equiv 0$ and $\dot{\rho}^2 \equiv 1$, when $n > 2$. Thus, $\rho(r) = a \pm r$. In case $n = 2$ we only need $\ddot{\rho} = 0$. 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

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

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

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

This is an autonomous second-order equation and can be made into a first-order equation by using ρ as a new independent variable. If $\dot{\rho} = G(\rho)$, then $\ddot{\rho} = G'\dot{\rho} = G'G$ and the first-order equation becomes

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

Separation of variables shows that G and ρ are related by

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

which after differentiation yields:

$$\ddot{\rho} = -\frac{n-2}{2}C\rho^{1-n}.$$

We focus on solutions to this family of second-order equations. Note that they will in turn solve $\dot{\rho}^2 = 1 + C\rho^{2-n}$, when the initial values are related by $(\dot{\rho}(0))^2 = 1 + C(\rho(0))^{2-n}$.

To analyze the solutions to this equation that are positive and thus 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 both $\dot{\rho}$ and $\ddot{\rho}$ approach $\pm\infty$ at points where ρ approaches 0.

$C = 0$: In this case $\ddot{\rho} \equiv 0$ and $\dot{\rho}^2(0) = 1$. Thus, $\rho = a + r$ is the only solution and the metric is the standard Euclidean metric.

$C > 0$: ρ is concave since

$$\ddot{\rho} = -\frac{n-2}{2}C\rho^{1-n} < 0.$$

Thus, if ρ is extended to its maximal interval, then it must cross the "r-axis," but as pointed out above this means that $\ddot{\rho}$ becomes undefined. Consequently, we don't get any nice metrics this way.

$C < 0$: This time the solutions are convex. If we write $C = -\rho_0^{n-2}$, then the equation $\dot{\rho}^2 = 1 - \left(\frac{\rho_0}{\rho}\right)^{2-n}$ shows that $0 < \rho_0 \le \rho$. In case $\rho(a) = \rho_0$, it follows that $\dot{\rho}(a) = 0$ and $\ddot{\rho}(a) > 0$. Thus a is a strict minimum and the solution exists in a neighborhood of a. Furthermore, $|\dot{\rho}| \le 1$ so the solutions can't blow up in finite time. This shows that ρ is defined on all of $\mathbb{R}$. Thus, there are scalar flat rotationally symmetric metrics on $\mathbb{R} \times S^{n-1}$.

We focus on the solution with $\rho(0) = \rho_0 > 0$, which forces $\dot{\rho}(0) = 0$. Notice that ρ is even as $\rho(-r)$ solves the same initial value problem. Consequently, $(r, x) \mapsto (-r, -x)$ is an isometry on

$$\left(\mathbb{R} \times S^{n-1}, dr^2 + \rho^2(r) ds_{n-1}^2\right).$$

Thus we get a Riemannian covering map

$$\mathbb{R} \times S^{n-1} \to \tau\left(\mathbb{RP}^{n-1}\right)$$

and a scalar flat metric on $\tau\left(\mathbb{RP}^{n-1}\right)$, the tautological line bundle over $\mathbb{RP}^{n-1}$.

If we use ρ as the parameter instead of r, then

$$d\rho^2 = \dot{\rho}^2 dr^2 = \left(1 - \left(\frac{\rho_0}{\rho}\right)^{n-2}\right) dr^2.$$

When $r > 0$ it follows that $\rho > \rho_0$ and the metric has the more algebraically explicit form

$$dr^2 + \rho^2(r) ds_{n-1}^2 = \frac{1}{1 - \left(\frac{\rho_0}{\rho}\right)^{n-2}} d\rho^2 + \rho^2 ds_{n-1}^2.$$

This shows that the metric looks like the Euclidean metric $d\rho^2 + \rho^2 ds_{n-1}^2$ as $\rho \to \infty$.

In section 5.6.2 we show that $\mathbb{R} \times S^{n-1}$, $n \ge 3$, does not admit a (complete) constant curvature metric. Later in section 7.3.1 and theorem 7.3.5, 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 section 5.6.2 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.

4.2.4 Doubly Warped Products

We wish to compute the curvatures on

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

This time the Hessian looks like

$$\operatorname{Hess} r = (\partial_r \rho)\, \rho ds_p^2 + (\partial_r \phi)\, \phi ds_q^2.$$

and we see as in the rotationally symmetric case that

$$\nabla_X \mathrm{II} = 0.$$

Thus the mixed curvatures vanish. Let X, Y be tangent to S^p and V, W tangent to S^q. Using our curvature calculations from the rotationally symmetric case (see section 4.2.3) and the product sphere case (see section 4.2.2) the tangential curvature equations (see theorem 3.2.4) yield

$$\begin{aligned}
\mathfrak{R}(\partial_r \wedge X) &= -\frac{\ddot{\rho}}{\rho}\partial_r \wedge X,\\
\mathfrak{R}(\partial_r \wedge V) &= -\frac{\ddot{\phi}}{\phi}\partial_r \wedge V,\\
\mathfrak{R}(X \wedge Y) &= \frac{1-\dot{\rho}^2}{\rho^2}X \wedge Y,\\
\mathfrak{R}(U \wedge V) &= \frac{1-\dot{\phi}^2}{\phi^2}U \wedge V,\\
\mathfrak{R}(X \wedge V) &= -\frac{\dot{\rho}\dot{\phi}}{\rho\phi}X \wedge V.
\end{aligned}$$

From this it follows that all sectional curvatures are convex linear combinations of

$$-\frac{\ddot{\rho}}{\rho}, -\frac{\ddot{\phi}}{\phi}, \frac{1-\dot{\rho}^2}{\rho^2}, \frac{1-\dot{\phi}^2}{\phi^2}, -\frac{\dot{\rho}\dot{\phi}}{\rho\phi}.$$

Moreover,

$$\operatorname{Ric}(\partial_r) = \left(-p\frac{\ddot{\rho}}{\rho} - q\frac{\ddot{\phi}}{\phi}\right)\partial_r,$$

$$\mathrm{Ric}(X) = \left(\frac{-\ddot{\rho}}{\rho} + (p-1)\frac{1-\dot{\rho}^2}{\rho^2} - q \cdot \frac{\dot{\rho}\dot{\phi}}{\rho\phi} \right) X,$$

$$\mathrm{Ric}(V) = \left(\frac{-\ddot{\phi}}{\phi} + (q-1)\frac{1-\dot{\phi}^2}{\phi^2} - p \cdot \frac{\dot{\rho}\dot{\phi}}{\rho\phi} \right) V.$$

4.2.5 The Schwarzschild Metric

We wish to find a Ricci flat metric on $\mathbb{R}^2 \times S^{n-2}$. Choose $p = n-2$ and $q = 1$ in the above doubly warped product case so that the metric is on $(0, \infty) \times S^{n-2} \times S^1$. We'll see that this forces $dr^2 + \rho^2(r)\, ds^2_{n-2}$ to be scalar flat (see also exercise 4.7.16 for a more general treatment).

The equations to be solved are:

$$-(n-2)\frac{\ddot{\rho}}{\rho} - \frac{\ddot{\phi}}{\phi} = 0,$$

$$-\frac{\ddot{\rho}}{\rho} + (n-3)\frac{1-\dot{\rho}^2}{\rho^2} - \frac{\dot{\rho}\dot{\phi}}{\rho\phi} = 0,$$

$$-\frac{\ddot{\phi}}{\phi} - (n-2)\frac{\dot{\rho}\dot{\phi}}{\rho\phi} = 0.$$

Subtracting the first and last gives

$$\frac{\ddot{\rho}}{\rho} = \frac{\dot{\rho}\dot{\phi}}{\rho\phi}.$$

If we substitute this into the second equation we simply obtain the scalar flat equation for $dr^2 + \rho^2(r)\, ds^2_{n-2}$:

$$-2\frac{\ddot{\rho}}{\rho} + (n-3)\frac{1-\dot{\rho}^2}{\rho^2} = 0.$$

We use the solution $\rho(r)$ from section 4.2.3 that is even in r and satisfies:

$$\rho(0) = \rho_0,$$

$$\dot{\rho}^2 = 1 - \left(\frac{\rho_0}{\rho}\right)^{n-3}.$$

Next note that $\frac{\ddot{\rho}}{\rho} = \frac{\dot{\rho}\dot{\phi}}{\rho\phi}$ implies that $\frac{\dot{\rho}}{\phi} = c$ is constant. Thus we can define ϕ using $\dot{\rho} = c\phi$.

Since $\dot{\rho}^2 = 1 - \left(\frac{\rho_0}{\rho}\right)^{n-3}$ we obtain $c^2\dot{\phi}^2 = 1 - \left(\frac{\rho_0}{\rho}\right)^{n-3}$. This forces $\phi(0) = 0$. From $2\ddot{\rho} = (n-3)\frac{1}{\rho_0}\left(\frac{\rho_0}{\rho}\right)^{n-2}$ we get

$$2c\ddot{\phi} = (n-3)\frac{1}{\rho_0}\left(\frac{\rho_0}{\rho}\right)^{n-2}.$$

To obtain a smooth metric on $\mathbb{R}^2 \times S^{n-2}$ we need ϕ to be odd with $\dot{\phi}(0) = 1$. This forces $c = \frac{n-3}{2}\rho_0^{-1}$ and gives us $\dot{\phi} = \left(\frac{\rho_0}{\rho}\right)^{n-2}$. Since ρ is even this makes $\dot{\phi}$ even and hence ϕ odd as $\phi(0) = 0$. We also see that $\ddot{\phi} = \frac{n-3}{2}(2-n)\rho_0^{n-3}\rho^{1-n}\phi$. This shows that the first equation, and hence the other two, are satisfied:

$$-(n-2)\frac{n-3}{2}\rho_0^{n-2}\rho^{1-n} - \frac{n-3}{2}(2-n)\rho_0^{n-3}\rho^{1-n} = 0.$$

If we use ρ as a parameter instead of r as in section 4.2.3, then we obtain the more explicit algebraic form

$$\frac{1}{1-\left(\frac{\rho_0}{\rho}\right)^{n-3}}d\rho^2 + \rho^2 ds_{n-2}^2 + \rho_0^2\frac{4}{(n-3)^3}\left(1-\left(\frac{\rho_0}{\rho}\right)^{n-3}\right)d\theta^2.$$

Thus, the metric looks like $\mathbb{R}^{n-1} \times S^1$ at infinity, where the metric on S^1 is suitably scaled. Therefore, the Schwarzschild metric is a Ricci flat metric on $\mathbb{R}^2 \times S^{n-2}$ that at infinity looks approximately like the flat metric on $\mathbb{R}^{n-1} \times S^1$.

The classical Schwarzschild metric is a space-time metric and is not smooth at $\rho = \rho_0$. The parameter c above is taken to be the speed of light and is not forced to depend on ρ_0. We also replace S^1 by $\mathbb{R}$. The metric looks like:

$$\frac{1}{1-\frac{\rho_0}{\rho}}d\rho^2 + \rho^2 ds_2^2 - \frac{1}{c^2}\left(1-\frac{\rho_0}{\rho}\right)dt^2.$$

4.3 Warped Products in General

We are now ready for a slightly more general context for warped products. This will allow us to characterize the rotationally symmetric constant curvature metrics through a very simple equation for the Hessian of a modified distance function.

4.3.1 Basic Constructions

Given a Riemannian metric (H, g_H) a *warped product* (over I) is defined as a metric on $I \times H$, where $I \subset \mathbb{R}$ is an open interval, with metric

$$g = dr^2 + \rho^2(r) g_H,$$

where $\rho > 0$ on all of I. One could also more generally consider

$$\psi^2(r) dr^2 + \rho^2(r) g_H.$$

However, a change of coordinates defined by relating the differentials $d\rho = \psi(r) dr$ allows us to rewrite this as

$$d\rho^2 + \rho^2(r(\rho)) g_H.$$

Important special cases are the basic product $g = dr^2 + g_H$ and polar coordinates $dr^2 + r^2 ds_{n-1}^2$ on $(0, \infty) \times S^{n-1}$ representing the Euclidean metric.

The goal is to repackage the information that describes the warped product representation with a goal of finding a simple characterization of such metrics. Rather than using both r and ρ we will see that just one function suffices. Starting with a warped product $dr^2 + \rho^2(r) g_H$ construct the function $f = \int \rho dr$ on $M = I \times H$. Since $df = \rho dr$ it is clear that

$$dr^2 + \rho^2(r) g_H = \frac{1}{\rho^2(r)} df^2 + \rho^2(r) g_H.$$

Proposition 4.3.1. *The Hessian of f has the property*

$$\operatorname{Hess} f = \dot{\rho} g.$$

Proof. The Hessian of f is calculated from the Hessian of r. The latter is calculated as in section 4.2.3

$$\begin{aligned}
\operatorname{Hess} r &= \frac{1}{2} L_{\partial_r} g \\
&= \frac{1}{2} L_{\partial_r} \left(dr^2 + \rho^2(r) g_H \right) \\
&= \frac{1}{2} \partial_r \left(\rho^2(r) \right) g_H \\
&= \dot{\rho} \rho g_H.
\end{aligned}$$

So we obtain

$$\begin{aligned}
(\mathrm{Hess} f)(X, Y) &= (\nabla_X df)(Y) \\
&= (\nabla_X \rho dr)(Y) \\
&= \dot{\rho} dr(X)\, dr(Y) + \rho \,\mathrm{Hess}\, r(X, Y) \\
&= \dot{\rho} dr^2(X, Y) + \rho \,\mathrm{Hess}\, r(X, Y) \\
&= \dot{\rho} dr^2(X, Y) + \dot{\rho}\rho^2 g_H \\
&= \dot{\rho} g.
\end{aligned}$$

□

In other words we have shown that for a warped product it is possible to find a function f whose Hessian is conformal to the metric. In fact the relationship

$$\dot{\rho} = \frac{d\rho}{dr} = \frac{d\rho}{df}\frac{df}{dr} = \frac{d\rho}{df}\rho = \frac{1}{2}\frac{d|\nabla f|^2}{df}$$

tells us that the warped product representation depends only on f and $|\nabla f|$ since we have

$$\begin{aligned}
g &= \frac{1}{|\nabla f|^2} df^2 + |\nabla f|^2 g_H, \\
\mathrm{Hess} f &= \frac{1}{2}\frac{d|\nabla f|^2}{df} g.
\end{aligned}$$

Before turning to the general characterization let us consider how these constructions work on our standard constant curvature warped products.

Example 4.3.2. Consider the warped product given by

$$dr^2 + \mathrm{sn}_k^2(r)\, ds_{n-1}^2.$$

We select the antiderivative of $\mathrm{sn}_k(r)$ that vanishes at $r = 0$. When $k = 0$

$$\begin{aligned}
f &= \int r dr = \frac{1}{2} r^2, \\
\mathrm{Hess} f &= g.
\end{aligned}$$

When $k \neq 0$

$$\begin{aligned}
f &= \int \mathrm{sn}_k(r) = \frac{1}{k} - \frac{1}{k}\mathrm{cs}_k(r), \\
\mathrm{Hess} f &= \mathrm{cs}_k(r)\, g = (1 - kf)\, g.
\end{aligned}$$

More specifically, when $k = 1$

$$f = 1 - \cos r,$$
$$\operatorname{Hess} f = \cos r = 1 - f$$

and when $k = -1$

$$f = -1 + \cosh r,$$
$$\operatorname{Hess} f = \cosh r = 1 + f.$$

4.3.2 General Characterization

We can now state and prove our main characterization of warped products.

Theorem 4.3.3 (Brinkmann, 1925). *If there is a smooth function f whose Hessian is conformal to the metric, i.e.,* $\operatorname{Hess} f = \lambda g$, *then the Riemannian structure is locally a warped product $g = dr^2 + \rho^2(r)\, g_H$ around any point where $df \neq 0$. Moreover, if $df(p) = 0$ and $\lambda(p) \neq 0$, then $g = dr^2 + \rho^2(r)\, ds^2_{n-1}$ on some neighborhood of p.*

Proof. We first focus attention on the case where df never vanishes. Thus f can locally be considered the first coordinate in a coordinate system.

Define $\rho = |\nabla f|$ and note that

$$D_X \rho^2 = 2 \operatorname{Hess} f(\nabla f, X) = 2\lambda g(\nabla f, X),$$

i.e., $d\rho^2 = 2\lambda df$. Consequently also $d\lambda \wedge df = 0$. It follows that $d\rho$ and $d\lambda$ are both proportional to df and in particular that ρ and λ are locally constant on level sets of f. Thus we can assume that $\rho = \rho(f)$ and $\lambda = \lambda(f)$. This shows in turn that $\frac{1}{\rho} df$ is closed and locally exact. Define r by $dr = \frac{1}{\rho} df$ and use r as a new parameter. Note that r is a distance function since

$$\partial_r = \nabla r = \frac{1}{\rho(f)} \nabla f$$

is a unit vector field. We can then decompose the metric as $g = dr^2 + g_r$ on a suitable domain $I \times H \subset M$, where $H \subset \{x \in M \mid r(x) = r_0\}$. When $X \perp \partial_r$ it follows that $\nabla_X dr = \frac{1}{\rho} \nabla_X df$. Thus $\operatorname{Hess} r = \frac{\lambda}{\rho} g_r$ and $L_{\partial_r} g_r = \frac{2\lambda}{\rho} g_r$.

Observe that if g_H is defined such that $g_{r_0} = \rho^2(r_0)\, g_H$ is the restriction of g to the fixed level set $r = r_0$, then also

$$L_{\partial_r}\left(\rho^2 g_H\right) = \left(\partial_r \rho^2\right) g_H = 2\lambda\rho g_H = \frac{2\lambda}{\rho}\rho^2 g_H.$$

This shows that

$$g = dr^2 + g_r = dr^2 + \rho^2 g_H.$$

Next assume that p is a nondegenerate critical point for f. After possibly replacing f by $\alpha f + \beta$, we can assume that $\operatorname{Hess} f = \lambda g$ with $f(p) = 0$, $df|_p = 0$, and $\lambda(p) = 1$. Further assume that M is the connected component of $\{f < \epsilon\}$ that contains p and that p is the only critical point for f. Since $\operatorname{Hess} f = g$ at p there exist coordinates around p with $y^i(p) = 0$ and

$$f\left(y^1, \ldots, y^n\right) = \frac{1}{2}\left(\left(y^1\right)^2 + \cdots + (y^n)^2\right).$$

Therefore, all the regular level sets for f are spheres in this coordinates system. We can use the first part of the proof to obtain a warped product structure $dr^2 + \rho^2 g_{S^{n-1}}$ on $M - \{p\} \simeq (0, b) \times S^{n-1}$, where $g_{S^{n-1}}$ is a metric on S^{n-1} and $r \to 0$ as we approach p. When all functions are written as functions of r they are determined by λ in the following simple way:

$$\begin{aligned}
f &= f(r), \\
\frac{df}{dr} &= \rho(r), \\
\frac{d^2 f}{dr^2} &= \frac{d\rho}{dr} = \lambda, \\
f(0) &= \frac{df}{dr}(0) = \rho(0) = 0, \\
\frac{d^2 f}{dr^2}(0) &= \frac{d\rho}{dr}(0) = \lambda(0) = 1.
\end{aligned}$$

The goal is to show that $g_{S^{n-1}} = ds^2_{n-1}$. The initial conditions for ρ guarantee that the metric $dr^2 + \rho^2 ds^2_{n-1}$ is continuous at p when we switch to Cartesian coordinates as in section 1.4.4. We can use a similar analysis here. First assume that $\dim M = 2$ and $x = r\cos\theta$, $y = r\sin\theta$, where r is as above and θ coordinatizes S^1. The metric g_{S^1} on S^1 must take the form $\phi^2(\theta)\, d\theta^2$ for some function $\phi : S^1 \to (0, \infty)$. The metric is then given by $g = dr^2 + \rho^2(r)\,\phi^2(\theta)\, d\theta^2$. As the new coordinate fields are

$$\partial_x = \cos\theta\,\partial_r - \frac{1}{r}\sin\theta\,\partial_\theta,$$

$$\partial_y = \sin\theta\,\partial_r + \frac{1}{r}\cos\theta\,\partial_\theta,$$

the new metric coefficients become

$$g_{xx} = \cos^2\theta + \phi^2(\theta)\frac{\rho^2(r)}{r^2}\sin^2\theta,$$

$$g_{yy} = \sin^2\theta + \phi^2(\theta)\frac{\rho^2(r)}{r^2}\cos^2\theta.$$

As $r \to 0$ we obtain the limits

$$g_{xx}(p) = \cos^2\theta + \phi^2(\theta)\sin^2\theta,$$

$$g_{yy}(p) = \sin^2\theta + \phi^2(\theta)\cos^2\theta,$$

since $\rho(0) = 0$ and $\dot{\rho}(0) = 1$. However, these limits are independent of θ as they are the metric coefficients at p. This implies first that $\phi(\theta)$ is constant since

$$g_{xx}(p) + g_{yy}(p) = 1 + \phi^2(\theta)$$

and then that $\phi = 1$ as $g_{xx}(p)$ is independent of θ.

This case can be adapted to higher dimensions. Simply select a plane that intersects the unit sphere S^{n-1} in a great circle $c(\theta)$, where θ is the arclength parameter with respect to the standard metric. The metric g restricted to this plane can then be expressed as in the 2-dimensional case and it follows that $1 = \phi^2(\theta) = g_{S^{n-1}}\left(\frac{dc}{d\theta}, \frac{dc}{d\theta}\right)$. As $\frac{dc}{d\theta}$ can be chosen to be any unit vector on S^{n-1} it follows that $g_{S^{n-1}}$ agrees with the standard metric on the unit sphere. □

This theorem can be used to characterize the warped product constant curvature metrics from example 4.3.2.

Corollary 4.3.4. *If there is a function f on a Riemannian manifold such that*

$$f(p) = 0,$$

$$df|_p = 0,$$

and

$$\mathrm{Hess} f = (1 - kf)\,g,$$

then the metric is the warped product metric of curvature k in a neighborhood of p as described in example 4.3.2.

Proof. Note that $\lambda = 1 - kf$ is an explicit function of f. So we can find $f = f(r)$ as the solution to

$$\begin{aligned} \frac{d^2 f}{dr^2} &= 1 - kf, \\ f(0) &= 0, \\ f'(0) &= 0, \end{aligned}$$

and the warping function by

$$\rho(r) = |\nabla f| = \frac{df}{dr}.$$

The solutions are consequently given by the standard warped product representations of constant curvature metrics:

Euclidean Space

$$\begin{aligned} g &= dr^2 + r^2 ds_{n-1}^2, \\ f(r) &= \tfrac{1}{2} r^2. \end{aligned}$$

Constant curvature $k \neq 0$

$$\begin{aligned} g &= dr^2 + \mathrm{sn}_k^2(r)\, ds_{n-1}^2, \\ f(r) &= \tfrac{1}{k} - \tfrac{1}{k} \mathrm{cs}_k(r). \end{aligned}$$

In all cases $r = 0$ corresponds to the point p. □

Remark 4.3.5. A function $f : M \to \mathbb{R}$ is called *transnormal* provided $|df|^2 = \rho^2(f)$ for some smooth function ρ. We saw above that functions with conformal Hessian locally have this property. However, it is easy to construct transnormal functions that do not have conformal Hessian. A good example is the function $f = \frac{1}{2}\sin(2r)$ on the doubly warped product representation of $S^3(1)$ given by $dr^2 + \sin^2(r)\, d\theta_1^2 + \cos^2(r)\, d\theta_2^2$ on $(0, \pi/2) \times S^1 \times S^1$.

4.3.3 Conformal Representations of Warped Products

If (M, g) is a Riemannian manifold and ψ is positive on M, then we can construct a new Riemannian manifold $(M, \psi^2 g)$. Such a change in metric is called a *conformal change,* and ψ^2 is referred to as the *conformal factor*.

A warped product can be made to look like a conformal metric in two basic ways.

$$dr^2 + \rho^2(r)\, g_H = \psi^2(\rho)\left(d\rho^2 + g_H\right),$$

$$dr = \psi(\rho)\, d\rho,$$
$$\rho(r) = \psi(\rho)$$

or

$$\begin{aligned} dr^2 + \rho^2(r)\, g_H &= \psi^2(\rho)\left(d\rho^2 + \rho^2 g_H\right), \\ dr &= \psi(\rho)\, d\rho, \\ \rho(r) &= \rho\psi(\rho). \end{aligned}$$

4.3.3.1 Conformal Models of Spheres

The first of these changes has been studied since the time of Mercator. The sphere of radius R and curvature $\frac{1}{R^2}$ can be written as

$$\begin{aligned} R^2 ds_n^2 &= R^2\left(dt^2 + \sin^2(t)\, ds_{n-1}^2\right) \\ &= dr^2 + R^2 \sin^2\left(\tfrac{r}{R}\right) ds_{n-1}^2. \end{aligned}$$

The conformal change envisioned by Mercator takes the form

$$R^2 ds_n^2 = \psi^2(\rho)\left(d\rho^2 + ds_{n-1}^2\right).$$

As

$$\begin{aligned} \psi(\rho)\, d\rho &= dr, \\ \psi(\rho) &= R\sin\left(\tfrac{r}{R}\right) \end{aligned}$$

we obtain

$$\begin{aligned} d\rho &= \frac{dr}{R\sin\left(\frac{r}{R}\right)}, \\ \rho &= \frac{1}{2}\log\frac{1-\cos\left(\frac{r}{R}\right)}{1+\cos\left(\frac{r}{R}\right)}. \end{aligned}$$

Thus

$$\cos\left(\frac{r}{R}\right) = \frac{1-\exp(2\rho)}{1+\exp(2\rho)}$$

and

$$\psi^2 = R^2 \sin^2\left(\frac{r}{R}\right) = R^2 \frac{4\exp(2\rho)}{(1+\exp(2\rho))^2}$$

showing that

$$R^2 ds_n^2 = R^2 \frac{4\exp(2\rho)}{(1+\exp(2\rho))^2}\left(d\rho^2 + ds_{n-1}^2\right).$$

Switching the spherical metric to being conformal to the polar coordinate representation of Euclidean space took even longer and probably wasn't studied much until the time of Riemann. The calculations in this case require that we first solve

$$\frac{d\rho}{\rho} = \frac{dr}{R\sin\left(\frac{r}{R}\right)}.$$

This integrates to

$$\rho^2 = \frac{1-\cos\left(\frac{r}{R}\right)}{1+\cos\left(\frac{r}{R}\right)}$$

and implies

$$\cos\left(\frac{r}{R}\right) = \frac{1-\rho^2}{1+\rho^2}.$$

The relationship

$$R\sin\left(\frac{r}{R}\right) = \rho\psi(\rho)$$

then gives us

$$\psi^2(\rho) = R^2 \frac{4}{(1+\rho^2)^2}$$

and consequently

$$\begin{aligned} R^2 ds_n^2 &= R^2\psi^2(\rho)\left(d\rho^2 + \rho^2 ds_{n-1}^2\right) \\ &= R^2 \frac{4}{(1+\rho^2)^2}\left(d\rho^2 + \rho^2 ds_{n-1}^2\right) \\ &= \frac{4R^2}{(1+\rho^2)^2} g_{\mathbb{R}^n}. \end{aligned}$$

This gives us a representation of the metric on the punctured sphere that only involves algebraic functions. See also exercise 4.7.13 for a geometric construction of the representation.

4.3.3.2 Conformal Models of Hyperbolic Space

We defined hyperbolic space H^n in example 1.1.7 and exhibited it as a rotationally symmetric metric in example 1.4.6. The rotationally symmetric metric on $H^n(R)$ can be written as

$$\begin{aligned} dr^2 + \mathrm{sn}^2_{R^{-2}}(r)\, ds^2_{n-1} &= dr^2 + R^2 \sinh^2\left(\tfrac{r}{R}\right) ds^2_{n-1} \\ &= R^2\left(dt^2 + \sinh^2(t)\, ds^2_{n-1}\right). \end{aligned}$$

A construction similar to what we just saw for the sphere leads to the conformal polar coordinate representation

$$R^2\left(dt^2 + \sinh^2(t)\, ds^2_{n-1}\right) = \frac{4R^2}{\left(1-\rho^2\right)^2} g_{\mathbb{R}^n}.$$

This time, however, the metric is only defined on the unit ball. This is also known as the *Poincarè model* on the unit disc. See also exercise 4.7.13 for a geometric construction of the representation.

Consider the metric

$$\left(\frac{1}{x^n}\right)^2 \left((dx^1)^2 + \cdots + (dx^n)^2\right)$$

on the open half space $x^n > 0$. If we define $r = \log(x^n)$, then this also becomes the warped product:

$$g = dr^2 + (e^{-r})^2\left((dx^1)^2 + \cdots + (dx^{n-1})^2\right).$$

The upper half space model can be realized as the Poincaré disc using an *inversion*, i.e., a conformal transformation of Euclidean space that inverts in a suitable sphere. It'll be convenient to write $x = \left(x^1, \ldots, x^{n-1}\right)$ as the first $n-1$ coordinates and $y = x^n$. The inversion in the sphere of radius $\sqrt{2}$ centered at $(0,-1) \in \mathbb{R}^{n-1} \times \mathbb{R}$ is given by

$$\begin{aligned} F(x,y) &= (0,-1) + \tfrac{2(x,y+1)}{r^2} \\ &= \left(\tfrac{2x}{r^2}, -1 + \tfrac{2(y+1)}{r^2}\right) \\ &= \tfrac{1}{r^2}\left(2x, 1 - |x|^2 - y^2\right), \end{aligned}$$

where $r^2 = |x|^2 + (y+1)^2$. This maps H to the unit ball since

$$|F(x,y)|^2 = 1 - \frac{4y}{r^2} = \rho^2.$$

The goal is to show that F transforms the conformal unit ball model to the conformal half space model. This is a direct calculation after we write F out in coordinates:

$$\begin{aligned} F^k &= 2\frac{x^k}{r^2},\ k < n, \\ F^n &= \frac{2(y+1)}{r^2} - 1. \end{aligned}$$

This allows us to calculate the differentials so that we can check how the metric is transformed:

$$\begin{aligned}
&\frac{4}{(1-\rho^2)^2}\left((dF^n)^2 + \sum_{k<n}\left(dF^k\right)^2\right) \\
&\quad = \frac{\left(r^2\right)^2}{4y^2}\left(\frac{2dy}{r^2} - \frac{2(y+1)\,2rdr}{(r^2)^2}\right)^2 \\
&\qquad + \sum_{k<n}\frac{\left(r^2\right)^2}{4y^2}\left(\frac{2dx^k}{r^2} - \frac{2x^k 2rdr}{(r^2)^2}\right)^2 \\
&\quad = \frac{1}{y^2}\left(dy - \frac{(y+1)\,2rdr}{r^2}\right)^2 + \frac{1}{y^2}\sum_{k<n}\left(dx^k - \frac{x^k 2rdr}{r^2}\right)^2 \\
&\quad = \frac{1}{y^2}\left(dy^2 + \sum_{k<n}\left(dx^k\right)^2\right) + \frac{1}{y^2}\left(\frac{(y+1)\,2rdr}{r^2}\right)^2 + \frac{1}{y^2}\sum_{k<n}\left(\frac{x^k 2rdr}{r^2}\right)^2 \\
&\qquad - \frac{1}{y^2}dy\frac{(y+1)\,2rdr}{r^2} - \frac{1}{y^2}\sum_{k<n}dx^k\frac{x^k 2rdr}{r^2} \\
&\qquad - \frac{1}{y^2}\frac{(y+1)\,2rdr}{r^2}dy - \frac{1}{y^2}\sum_{k<n}\frac{x^k 2rdr}{r^2}dx^k \\
&\quad = \frac{1}{y^2}\left(dy^2 + g_{\mathbb{R}^{n-1}}\right) + \frac{1}{y^2}r^2\left(\frac{2rdr}{r^2}\right)^2 \\
&\qquad - \frac{1}{y^2}rdr\frac{2rdr}{r^2} \\
&\qquad - \frac{1}{y^2}\frac{2rdr}{r^2}rdr \\
&\quad = \frac{1}{y^2}\left(dy^2 + g_{\mathbb{R}^{n-1}}\right).
\end{aligned}$$

More generally, we can ask when

$$\psi^2 \cdot ((dx^1)^2 + \cdots + (dx^n)^2)$$

has constant curvature? Clearly, $\psi \cdot dx^1, \ldots, \psi \cdot dx^n$ is an orthonormal coframe, and $\frac{1}{\psi}\partial_1, \ldots, \frac{1}{\psi}\partial_n$ is an orthonormal frame. We can use the Koszul formula to compute $\nabla_{\partial_i}\partial_j$ and hence the curvature tensor. This task is done in exercise 4.7.21 or in [97, vols. II and IV]. Using

$$\psi = \left(1 + \frac{k}{4}r^2\right)^{-1}$$

gives the *Riemann model* for a metric of constant curvature k on $\mathbb{R}^n$ if $k \geq 0$ and on $B(0, \frac{2}{\sqrt{|k|}})$ if $k < 0$.

The Riemann model with $k = -1$ and the Poincaré model from above are also isometric if we use the map $F(x) = 2x$. This clearly maps the unit ball to the ball of radius 2 and the metric is changed as follows

$$\frac{1}{\left(1 - \frac{1}{4}|F|^2\right)^2}\left(\sum_{k=1}^{n}\left(dF^k\right)^2\right) = \frac{4}{\left(1 - |x|^2\right)^2}\left(\sum_{k=1}^{n}\left(dx^k\right)^2\right).$$

4.3.4 Singular Points

The polar coordinate conformal model

$$dr^2 + \varphi^2(r)\, ds_{n-1}^2 = \psi^2(\rho)\left(d\rho^2 + \rho^2 ds_{n-1}^2\right)$$

offers a different approach to the study of smoothness of the metric as we approach a point $r_0 \in \partial I$ where $\varphi(r_0) = 0$. Assume that the parametrization satisfies $\rho(r_0) = 0$. When $g_H = ds_{n-1}^2$ smoothness on the right-hand side

$$\psi^2(\rho)\left(d\rho^2 + \rho^2 ds_{n-1}^2\right)$$

depends only on $\psi^2(\rho)$ being smooth (see Section 1.4.4). Thinking of ρ as being Euclidean distance indicates that this is not entirely trivial. In fact we must assume that $\psi(0) > 0$ and $\psi^{(\text{odd})}(0) = 0$. Translating back to φ we obtain the usual conditions: $\dot{\varphi}(0) = \pm 1$ and $\varphi^{(\text{even})}(0) = 0$.

4.4 Metrics on Lie Groups

We are going to study some general features of left-invariant metrics and show how things simplify in the biinvariant situation. There are two examples of left-invariant metrics. The first represents hyperbolic space H^2, and the other is the Berger sphere (see example 1.3.5).

4.4.1 Generalities on Left-invariant Metrics

We can construct a metric on a Lie group G by fixing an inner product $(,)$ on $T_e\mathrm{G}$ and then translating it to T_gM using left-translation $L_g(x) = gx$. The metric is also denoted (X, Y) on G so as not to confuse it with elements $g \in \mathrm{G}$. With this metric, L_g becomes an isometry for all g since

$$\begin{aligned}
\left(DL_g\right)|_h &= \left(DL_{ghh^{-1}}\right)|_h \\
&= \left(D\left(L_{gh} \circ L_{h^{-1}}\right)\right)|_h \\
&= \left(DL_{gh}\right)|_e \circ \left(DL_{h^{-1}}\right)|_h \\
&= \left(DL_{gh}\right)|_e \circ \left(\left(DL_h\right)|_e\right)^{-1}
\end{aligned}$$

and we have assumed that $\left(DL_{gh}\right)|_e$ and $\left(DL_h\right)|_e$ are isometries.

Left-invariant fields X, i.e., $DL_g\left(X|_h\right) = X|_{gh}$ are completely determined by their value at the identity. This identifies T_eM with $\mathfrak{g}$, the space of left-invariant fields. Note that $\mathfrak{g}$ is in a natural way a vector space as addition of left-invariant fields is left-invariant. It is also a Lie algebra as the vector field Lie bracket of two such fields is again left-invariant. In section 1.3.2 we saw that on matrix groups the Lie bracket is simply the commutator of the matrices in T_eM representing the vector fields.

If $X \in \mathfrak{g}$, then the integral curve through $e \in \mathrm{G}$ is denoted by $\exp(tX)$. In case of a matrix group the standard matrix exponential e^{tX} is in fact the integral curve since

$$\begin{aligned}
\frac{d}{dt}|_{t=t_0}\left(e^{tX}\right) &= \frac{d}{dt}|_{s=0}\left(e^{(t_0+s)X}\right) \\
&= \frac{d}{dt}|_{s=0}\left(e^{t_0X}e^{sX}\right) \\
&= \frac{d}{dt}|_{s=0}\left(L_{e^{t_0X}}e^{sX}\right) \\
&= D\left(L_{e^{t_0X}}\right)\left(\frac{d}{dt}|_{s=0}e^{sX}\right) \\
&= D\left(L_{e^{t_0X}}\right)\left(X|_I\right) \\
&= X|_{e^{t_0X}}.
\end{aligned}$$

The key property for $t \mapsto \exp(tX)$ to be the integral curve for X is evidently that the derivative at $t = 0$ is $X|_e$ and that $t \mapsto \exp(tX)$ is a homomorphism

$$\exp((t+s)X) = \exp(tX)\exp(sX).$$

The entire flow for X can be written as follows

$$F^t(x) = x\exp(tX) = L_x \exp(tX) = R_{\exp(tX)}(x).$$

The curious thing is that the flow maps $F^t : \mathrm{G} \to \mathrm{G}$ don't act by isometries unless the metric is also invariant under right-translations, i.e., the metric is biinvariant. In particular, the elements of $\mathfrak{g}$ are not in general Killing fields. In fact, it is the right-invariant fields that are Killing fields for left-invariant metrics as their flows are generated by

$$F^t(x) = \exp(tX)x = R_x \exp(tX) = L_{\exp(tX)}(x).$$

We can give a fairly reasonable way of checking that a left-invariant metric is also biinvariant. Conjugation $x \mapsto gxg^{-1}$ is denoted $\mathrm{Ad}_g(x) = gxg^{-1}$ on Lie groups and is called the *adjoint action* of G on G. The differential of this action at $e \in \mathrm{G}$ is a linear map $\mathrm{Ad}_g : \mathfrak{g} \to \mathfrak{g}$ denoted by the same symbol, and called the *adjoint action* of G on $\mathfrak{g}$. It is in fact a Lie algebra isomorphism. These two adjoint actions are related by

$$\mathrm{Ad}_g(\exp(tX)) = \exp\left(t\,\mathrm{Ad}_g(X)\right).$$

This is quite simple to prove. It only suffices to check that $t \mapsto \mathrm{Ad}_g(\exp(tX))$ is a homomorphism with differential $\mathrm{Ad}_g(X)$ at $t = 0$. The latter follows from the definition of the differential of a map and the former by noting that it is the composition of two homomorphisms $x \mapsto \mathrm{Ad}_g(x)$ and $t \mapsto \exp(tX)$. We can now give our criterion for biinvariance.

Proposition 4.4.1. *A left-invariant metric is biinvariant if and only if the adjoint action on the Lie algebra is by isometries.*

Proof. In case the metric is biinvariant we know that both L_g and $R_{g^{-1}}$ act by isometries. Thus also $\mathrm{Ad}_g = L_g \circ R_{g^{-1}}$ acts by isometries. The differential is then a linear isometry on the Lie algebra.

Conversely, assume that $\mathrm{Ad}_g : \mathfrak{g} \to \mathfrak{g}$ is always an isometry. Using that

$$\left(DR_g\right)|_h = \left(DR_{hg}\right)|_e \circ \left(\left(DR_h\right)|_e\right)^{-1}$$

it clearly suffices to prove that $(DR_g)|_e$ is always an isometry. This follows from

$$\begin{aligned} R_g &= L_g \circ \mathrm{Ad}_{g^{-1}}, \\ (DR_g)|_e &= D(L_g)|_e \circ \mathrm{Ad}_{g^{-1}}. \end{aligned}$$

□

In sections 4.4.2 and 4.4.3 we shall see how this can be used to check whether metrics are biinvariant in some specific matrix group examples.

Before giving examples of how to compute the connection and curvatures for left-invariant metrics we present the general and simpler situation of biinvariant metrics.

Proposition 4.4.2. *Consider a Lie group* G *with a biinvariant metric* $(,)$ *and* $X, Y, Z, W \in \mathfrak{g}$. *Then*

$$\begin{aligned} \nabla_Y X &= \frac{1}{2}[Y, X], \\ R(X, Y)Z &= -\frac{1}{4}[[X, Y], Z], \\ R(X, Y, Z, W) &= \frac{1}{4}([X, Y], [W, Z]). \end{aligned}$$

In particular, the sectional curvature is always nonnegative, when $(,)$ *is positive definite.*

Proof. We first need to construct the *adjoint action* $\mathrm{ad}_X : \mathfrak{g} \to \mathfrak{g}$ of the Lie algebra on the Lie algebra. If we think of the adjoint action of the Lie group on the Lie algebra as a homomorphism $\mathrm{Ad} : \mathrm{G} \to \mathrm{Aut}(\mathfrak{g})$, then $\mathrm{ad} : \mathfrak{g} \to \mathrm{End}(\mathfrak{g})$ is simply the differential $\mathrm{ad} = D(\mathrm{Ad})|_e$. In section 2.1.4 it is shown that $\mathrm{ad}_X(Y) = [X, Y]$. The biinvariance of the metric shows that the image $\mathrm{Ad}(\mathrm{G}) \subset \mathrm{O}(\mathfrak{g})$ lies in the group of orthogonal transformations on $\mathfrak{g}$. This immediately shows that the image of ad lies in the set of skew-adjoint transformations since

$$\begin{aligned} 0 &= \frac{d}{dt}(Y, Z)|_{t=0} \\ &= \frac{d}{dt}\left(\mathrm{Ad}_{\exp(tX)}(Y), \mathrm{Ad}_{\exp(tX)}(Z)\right)|_{t=0} \\ &= (\mathrm{ad}_X Y, Z) + (Y, \mathrm{ad}_X Y). \end{aligned}$$

Keeping this skew-symmetry in mind we can use the Koszul formula on $X, Y, Z \in \mathfrak{g}$ to see that

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

As for the curvature we then have

$$\begin{aligned}
R\left(X, Y\right) Z &= \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z \\
&= \frac{1}{2}\nabla_X\left[Y, Z\right] - \frac{1}{2}\nabla_Y\left[X, Z\right] - \frac{1}{2}\left[\left[X, Y\right], Z\right] \\
&= \frac{1}{4}\left[X, \left[Y, Z\right]\right] - \frac{1}{4}\left[Y, \left[X, Z\right]\right] - \frac{1}{2}\left[\left[X, Y\right], Z\right] \\
&= \frac{1}{4}\left[X, \left[Y, Z\right]\right] + \frac{1}{4}\left[Y, \left[Z, X\right]\right] + \frac{1}{4}\left[Z, \left[X, Y\right]\right] - \frac{1}{4}\left[\left[X, Y\right], Z\right] \\
&= -\frac{1}{4}\left[\left[X, Y\right], Z\right],
\end{aligned}$$

and finally

$$\begin{aligned}
\left(R\left(X, Y\right) Z, W\right) &= -\frac{1}{4}\left(\left[\left[X, Y\right], Z\right], W\right) \\
&= \frac{1}{4}\left(\left[Z, \left[X, Y\right]\right], W\right) \\
&= -\frac{1}{4}\left(\left[Z, W\right], \left[X, Y\right]\right) \\
&= \frac{1}{4}\left(\left[X, Y\right], \left[W, Z\right]\right).
\end{aligned}$$

□

We note that Lie groups with biinvariant Riemannian metrics always have nonnegative sectional curvature and with a little more work it is also possible to show that the curvature operator is nonnegative (see exercise 3.4.32).

4.4.2 Hyperbolic Space as a Lie Group

Let G be the 2-dimensional Lie group

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

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

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

If we define

$$X = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \ Y = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},$$

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.

We can also compute Ad_g:

$$\begin{aligned} \mathrm{Ad}_{\begin{bmatrix} \alpha & \beta \\ 0 & 1 \end{bmatrix}} \begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix} &= \begin{bmatrix} \alpha & \beta \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \alpha & \beta \\ 0 & 1 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} a & -a\beta + b\alpha \\ 0 & 0 \end{bmatrix} \\ &= aX + (-a\beta + b\alpha) Y. \end{aligned}$$

The orthonormal basis

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

is then mapped to the basis

$$\begin{bmatrix} 1 & -\beta \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & \alpha \\ 0 & 0 \end{bmatrix}.$$

This, however, is not an orthonormal basis unless $\beta = 0$ and $\alpha = 1$. Therefore, the metric is not biinvariant, nor are the left-invariant fields Killing fields.

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

4.4.3 Berger Spheres

On SU(2) consider the left-invariant metric such that $\lambda_1^{-1}X_1$, $\lambda_2^{-1}X_2$, $\lambda_3^{-1}X_3$ is an orthonormal frame and $[X_i, X_{i+1}] = 2X_{i+2}$ (indices are mod 3) as in example 1.3.5. The Koszul formula is:

$$2\left(\nabla_{X_i}X_j, X_k\right) = \left(\left[X_i, X_j\right], X_k\right) + \left(\left[X_k, X_i\right], X_j\right) - \left(\left[X_j, X_k\right], X_i\right).$$

From this we can quickly see that as with a biinvariant metric we have: $\nabla_{X_i}X_i = 0$. It also follows that

$$\begin{aligned}
\nabla_{X_i}X_{i+1} &= \left(\frac{\lambda_{i+2}^2 + \lambda_{i+1}^2 - \lambda_i^2}{\lambda_{i+2}^2}\right)X_{i+2}, \\
\nabla_{X_{i+1}}X_i &= [X_{i+1}, X_i] + \nabla_{X_i}X_{i+1} \\
&= \left(\frac{-\lambda_{i+2}^2 + \lambda_{i+1}^2 - \lambda_i^2}{\lambda_{i+2}^2}\right)X_{i+2}.
\end{aligned}$$

This shows that

$$\begin{aligned}
R(X_i, X_{i+1})X_{i+2} &= \nabla_{X_i}\nabla_{X_{i+1}}X_{i+2} \\
&\quad -\nabla_{X_{i+1}}\nabla_{X_i}X_{i+2} - \nabla_{[X_i, X_{i+1}]}X_{i+2} \\
&= 0 - 0 - 0.
\end{aligned}$$

Thus all curvatures between three distinct vectors vanish.

The special case of Berger spheres occur when $\lambda_1 = \varepsilon < 1$, $\lambda_2 = \lambda_3 = 1$. In this case

$$\begin{aligned}
\nabla_{X_1}X_2 &= \left(2 - \varepsilon^2\right)X_3,\ \nabla_{X_2}X_1 = -\varepsilon^2 X_3 \\
\nabla_{X_2}X_3 &= X_1,\ \nabla_{X_3}X_2 = -X_1, \\
\nabla_{X_3}X_1 &= \varepsilon^2 X_2,\ \nabla_{X_1}X_3 = \left(\varepsilon^2 - 2\right)X_2.
\end{aligned}$$

and

$$R\left(X_1, X_2\right)X_2 = \varepsilon^2 X_1,$$

$$
\begin{aligned}
R(X_3, X_1) X_1 &= \varepsilon^4 X_3, \\
R(X_2, X_3) X_3 &= \left(4 - 3\varepsilon^2\right) X_2,
\end{aligned}
$$

$$
\begin{aligned}
\mathfrak{R}(X_1 \wedge X_2) &= \varepsilon^2 X_1 \wedge X_2, \\
\mathfrak{R}(X_3 \wedge X_1) &= \varepsilon^2 X_3 \wedge X_1, \\
\mathfrak{R}(X_2 \wedge X_3) &= \left(4 - 3\varepsilon^2\right) X_2 \wedge X_3.
\end{aligned}
$$

Thus all sectional curvatures must lie in the interval $\left[\varepsilon^2, 4 - 3\varepsilon^2\right]$. Note that as $\varepsilon \to 0$ the sectional curvature $\sec(X_2, X_3) \to 4$, which is the curvature of the base space $S^2\left(\frac{1}{2}\right)$ in the Hopf fibration.

We should also consider the adjoint action in this case. The standard orthogonal basis X_1, X_2, X_3 is mapped to

$$
\begin{aligned}
\mathrm{Ad}_{\begin{bmatrix} z & w \\ -\bar{w} & \bar{z} \end{bmatrix}} X_1 &= \left(|z|^2 - |w|^2\right) X_1 - 2\mathrm{Re}\,(wz) X_2 - 2\mathrm{Im}\,(wz) X_3, \\
\mathrm{Ad}_{\begin{bmatrix} z & w \\ -\bar{w} & \bar{z} \end{bmatrix}} X_2 &= 2\,\mathrm{i}\,\mathrm{Im}\,(z\bar{w}) X_1 + \mathrm{Re}\left(w^2 + z^2\right) X_2 + \mathrm{Im}\left(w^2 + z^2\right) X_3, \\
\mathrm{Ad}_{\begin{bmatrix} z & w \\ -\bar{w} & \bar{z} \end{bmatrix}} X_3 &= 2\mathrm{Re}\,(z\bar{w}) X_1 + \mathrm{Re}\left(\mathrm{i}\left(z^2 - w^2\right)\right) X_2 + \mathrm{Im}\left(\mathrm{i}\left(z^2 - w^2\right)\right) X_3.
\end{aligned}
$$

If the three vectors X_1, X_2, X_3 have the same length, then we see that the adjoint action is by isometries, otherwise not.

4.5 Riemannian Submersions

In this section we develop formulas for curvatures that relate to Riemannian submersions. The situation is similar to that of distance functions, which as we know are Riemannian submersions. In this case, however, we determine the curvature of the base space from information about the total space.

4.5.1 Riemannian Submersions and Curvatures

Throughout this section let $F : \left(\bar{M}, \bar{g}\right) \to (M, g)$ be a Riemannian submersion. Like with the metrics we shall use the standard "bar" notation: $\bar{p}$ and p and $\bar{X}$ and X for points and vector fields that are F-related, i.e., $F(\bar{p}) = p$ and $DF\left(\bar{X}\right) = X$.

The *vertical distribution* consists of the tangent spaces to the preimages $F^{-1}(p)$ and is given by $\mathscr{V}_{\bar{p}} = \ker DF_{\bar{p}} \subset T_{\bar{p}}\bar{M}$. The *horizontal distribution* is the orthogonal complement $\mathscr{H}_{\bar{p}} = \left(\mathscr{V}_{\bar{p}}\right)^{\perp} \subset T_{\bar{p}}\bar{M}$. The fact that F is a Riemannian submersion means that $DF : \mathscr{H}_{\bar{p}} \to T_pM$ is an isometry for all $\bar{p} \in \bar{M}$. Given a vector field X on M we can always find a unique horizontal vector field $\bar{X}$ on $\bar{M}$ that is F related to X. We say that $\bar{X}$ is a *basic horizontal* lift of X. Any vector in $\bar{M}$ can be decomposed into horizontal and vertical parts: $v = v^{\mathscr{V}} + v^{\mathscr{H}}$.

The next proposition gives some important properties for relationships between vertical and basic horizontal vector fields.

Proposition 4.5.1. *Let V be a vertical vector field on $\bar{M}$ and X, Y, Z vector fields on M with basic horizontal lifts $\bar{X}, \bar{Y}, \bar{Z}$.*

(1) $\left[V, \bar{X}\right]$ *is vertical,*
(2) $(L_V\bar{g})\left(\bar{X}, \bar{Y}\right) = D_V\bar{g}\left(\bar{X}, \bar{Y}\right) = 0,$
(3) $\bar{g}\left(\left[\bar{X}, \bar{Y}\right], V\right) = 2\bar{g}\left(\nabla_{\bar{X}}\bar{Y}, V\right) = -2\bar{g}\left(\nabla_V\bar{X}, \bar{Y}\right) = 2\bar{g}\left(\nabla_{\bar{Y}}V, \bar{X}\right),$
(4) $\nabla_{\bar{X}}\bar{Y} = \overline{\nabla_X Y} + \frac{1}{2}\left[\bar{X}, \bar{Y}\right]^{\mathscr{V}}.$

Proof. (1): $\bar{X}$ is F related to X and V is F related to the zero vector field on M. Thus

$$DF\left(\left[\bar{X}, V\right]\right) = \left[DF\left(\bar{X}\right), DF\left(V\right)\right] = [X, 0] = 0.$$

(2): We use (1) to see that

$$\begin{aligned}(L_V\bar{g})\left(\bar{X}, \bar{Y}\right) &= D_V\bar{g}\left(\bar{X}, \bar{Y}\right) - \bar{g}\left(\left[V, \bar{X}\right], \bar{Y}\right) - \bar{g}\left(\bar{X}, \left[V, \bar{Y}\right]\right) \\ &= D_V\bar{g}\left(\bar{X}, \bar{Y}\right).\end{aligned}$$

Next we use that F is a Riemannian submersion to conclude that $\bar{g}\left(\bar{X}, \bar{Y}\right) = g\left(X, Y\right)$. But this implies that the inner product is constant in the direction of the vertical distribution.

(3): Using (1) and (2) the Koszul formula in each case reduces to

$$\begin{aligned}2\bar{g}\left(\nabla_{\bar{X}}\bar{Y}, V\right) &= \bar{g}\left(\left[\bar{X}, \bar{Y}\right], V\right), \\ 2\bar{g}\left(\nabla_V\bar{X}, \bar{Y}\right) &= -\bar{g}\left(\left[\bar{X}, \bar{Y}\right], V\right), \\ 2\bar{g}\left(\nabla_{\bar{Y}}V, \bar{X}\right) &= \bar{g}\left(\left[\bar{X}, \bar{Y}\right], V\right).\end{aligned}$$

This proves the claim.

(4) We have just seen in (3) that $\frac{1}{2}\left[\bar{X}, \bar{Y}\right]^{\mathscr{V}}$ is the vertical component of $\nabla_{\bar{X}}\bar{Y}$. We know that $\overline{\nabla_X Y}$ is horizontal so it only remains to be seen that it is the horizontal component of $\nabla_{\bar{X}}\bar{Y}$. The Koszul formula together with F relatedness of the fields and the fact that inner products are the same in $\bar{M}$ and M show that

$$2\bar{g}\left(\nabla_{\bar{X}}\bar{Y},\bar{Z}\right) = 2g\left(\nabla_X Y, Z\right) = 2\bar{g}\left(\overline{\nabla_X Y},\bar{Z}\right).$$

□

Note that the map that takes horizontal vector fields X, Y on $\bar{M}$ to $[X,Y]^{\mathscr{V}}$ measures the extent to which the horizontal distribution is integrable in the sense of Frobenius. It is in fact tensorial and skew-symmetric since

$$[X, fY]^{\mathscr{V}} = f[X,Y]^{\mathscr{V}} + (D_X f)\,Y^{\mathscr{V}} = f[X,Y]^{\mathscr{V}}.$$

Therefore, it defines a map $\mathscr{H}\times\mathscr{H}\to\mathscr{V}$ called the *integrability tensor*.

Example 4.5.2. In the case of the Hopf map $S^3(1)\to S^2\left(\frac{1}{2}\right)$ we have that X_1 is vertical and X_2, X_3 are horizontal. However, X_2, X_3 are not basic. Still, we know that $[X_2, X_3] = 2X_1$ so the horizontal distribution cannot be integrable.

We are now ready to give a formula for the curvature tensor on M in terms of the curvature tensor on $\bar{M}$ and the integrability tensor.

Theorem 4.5.3 (B. O'Neill and A. Grey). *Let R be the curvature tensor on M and $\bar{R}$ the curvature tensor on $\bar{M}$. These curvature tensors are related by the formula*

$$g\left(R(X,Y)Y,X\right) = \bar{g}\left(\bar{R}\left(\bar{X},\bar{Y}\right)\bar{Y},\bar{X}\right) + \frac{3}{4}\left|\left[\bar{X},\bar{Y}\right]^{\mathscr{V}}\right|^2.$$

Proof. The proof is a direct calculation using the above properties. We calculate the full curvature tensor so let X, Y, Z, H be vector fields on M with vanishing Lie brackets. This forces the corresponding Lie brackets $\left[\bar{X},\bar{Y}\right]$, etc. in $\bar{M}$ to be vertical.

$$\begin{aligned}
\bar{g}\left(\bar{R}\left(\bar{X},\bar{Y}\right)\bar{Z},\bar{H}\right) &= \bar{g}\left(\nabla_{\bar{X}}\nabla_{\bar{Y}}\bar{Z} - \nabla_{\bar{Y}}\nabla_{\bar{X}}\bar{Z} - \nabla_{[\bar{X},\bar{Y}]}\bar{Z},\bar{H}\right)\\
&= \bar{g}\left(\nabla_{\bar{X}}\left(\overline{\nabla_Y Z} + \frac{1}{2}\left[\bar{Y},\bar{Z}\right]\right),\bar{H}\right)\\
&\quad -\bar{g}\left(\nabla_{\bar{Y}}\left(\overline{\nabla_X Z} + \frac{1}{2}\left[\bar{X},\bar{Z}\right]\right),\bar{H}\right)\\
&\quad +\bar{g}\left(\left[\bar{Z},\bar{H}\right],\left[\bar{X},\bar{Y}\right]\right)\\
&= \bar{g}\left(\overline{\nabla_X\nabla_Y Z} + \frac{1}{2}\left[\bar{X},\overline{\nabla_Y Z}\right]^{\mathscr{V}} + \frac{1}{2}\nabla_{\bar{X}}\left[\bar{Y},\bar{Z}\right],\bar{H}\right)\\
&\quad -\bar{g}\left(\overline{\nabla_Y\nabla_X Z} + \frac{1}{2}\left[\bar{Y},\overline{\nabla_X Z}\right]^{\mathscr{V}} + \frac{1}{2}\nabla_{\bar{Y}}\left[\bar{X},\bar{Z}\right],\bar{H}\right)\\
&\quad -\frac{1}{2}\bar{g}\left(\left[\bar{X},\bar{Y}\right],\left[\bar{H},\bar{Z}\right]\right)
\end{aligned}$$

$$
\begin{aligned}
&= g\left(R\left(X,Y\right)Z,H\right)\\
&\quad -\frac{1}{2}\bar{g}\left(\left[\bar{Y},\bar{Z}\right],\nabla_{\bar{X}}\bar{H}\right)+\frac{1}{2}\bar{g}\left(\left[\bar{X},\bar{Z}\right],\nabla_{\bar{Y}}\bar{H}\right)\\
&\quad -\frac{1}{2}\bar{g}\left(\left[\bar{X},\bar{Y}\right],\left[\bar{H},\bar{Z}\right]\right)\\
&= g\left(R\left(X,Y\right)Z,H\right)\\
&\quad -\frac{1}{4}\bar{g}\left(\left[\bar{Y},\bar{Z}\right],\left[\bar{X},\bar{H}\right]\right)+\frac{1}{4}\bar{g}\left(\left[\bar{X},\bar{Z}\right],\left[\bar{Y},\bar{H}\right]\right)\\
&\quad -\frac{1}{2}\bar{g}\left(\left[\bar{X},\bar{Y}\right],\left[\bar{H},\bar{Z}\right]\right)
\end{aligned}
$$

When $X = H$ and $Y = Z$ we get the above formula. □

More generally, one can find formulas for $\bar{R}$ where the variables are various combinations of basic horizontal and vertical fields.

4.5.2 Riemannian Submersions and Lie Groups

One can find many examples of manifolds with nonnegative or positive curvature using the previous theorem. In this section we shall explain the terminology in the general setting. The types of examples often come about by having $\left(\bar{M},\bar{g}\right)$ with a free compact group action G by isometries and using $M = \mathrm{G}\backslash\bar{M} = \bar{M}/\mathrm{G}$. Note we normally write such quotients on the right, but the action is generally on the left so $\mathrm{G}\backslash M$ is more appropriate. Examples are:

$$
\begin{aligned}
\mathbb{CP}^n &= S^{2n+1}/S^1,\\
TS^n &= \left(\mathrm{SO}\left(n+1\right)\times\mathbb{R}^n\right)/\mathrm{SO}\left(n\right),\\
M &= \mathrm{SU}\left(3\right)/T^2.
\end{aligned}
$$

The complex projective space will be studied further in section 4.5.3.

The most important general example of a Riemannian submersion comes about by having an isometric group action by G on $\bar{M}$ such that the quotient space is a manifold $M = \bar{M}/\mathrm{G}$ (see section 5.6.4 for conditions on the action that make this true). Such a submersion is also called *fiber homogeneous* as the group acts transitively on the fibers of the submersion. In this case we have a natural map $F : \bar{M} \to M$ that takes orbits to points, i.e., $p = \{x\cdot\bar{p} \mid x \in \mathrm{G}\}$ for $\bar{p} \in \bar{M}$. The vertical space $\mathscr{V}_{\bar{p}}$ then consists of the vectors that are tangent to the action. These directions can be found using the Killing fields generated by G. If $\mathfrak{X} \in \mathfrak{g} = T_e\mathrm{G}$, then we get a vector $X|_{\bar{p}} \in T_{\bar{p}}\bar{M}$ by the formula

$$
X|_{\bar{p}} = \frac{d}{dt}\left(\exp\left(t\mathfrak{X}\right)\cdot\bar{p}\right)|_{t=0},
$$

This means that the flow for X on $\bar{M}$ is defined by $F^t(\bar{p}) = \exp(t\mathfrak{X}) \cdot \bar{p}$. As the map $\bar{p} \mapsto x \cdot \bar{p}$ is assumed to be an isometry for all $x \in \mathrm{G}$ we get that the flow acts by isometries. This means that X is a Killing field. The next observation is that the action preserves the vertical distribution, i.e., $Dx\left(\mathscr{V}_{\bar{p}}\right) = \mathscr{V}_{x \cdot \bar{p}}$. Using the Killing fields this follows from

$$\begin{aligned}
Dx\left(X|_{\bar{p}}\right) &= Dx\left(\frac{d}{dt}\left(\exp(t\mathfrak{X}) \cdot \bar{p}\right)|_{t=0}\right) \\
&= \frac{d}{dt}\left(x \cdot \left(\exp(t\mathfrak{X}) \cdot \bar{p}\right)\right)|_{t=0} \\
&= \frac{d}{dt}\left(\left(x \exp(t\mathfrak{X})\, x^{-1}\right) \cdot x \cdot \bar{p}\right)|_{t=0} \\
&= \left(\left(\mathrm{Ad}_x\left(\exp(t\mathfrak{X})\right)\right) \cdot x \cdot \bar{p}\right)|_{t=0} \\
&= \frac{d}{dt}\left(\left(\exp\left(t\,\mathrm{Ad}_x\,\mathfrak{X}\right)\right) \cdot x \cdot \bar{p}\right)|_{t=0} \\
&= \left(\mathrm{Ad}_x\left(\mathfrak{X}\right)\right)|_{x \cdot \bar{p}}.
\end{aligned}$$

Thus $Dx\left(X|_{\bar{p}}\right)$ comes from first conjugating $\mathfrak{X}$ via the adjoint action in $T_e\mathrm{G}$ and then evaluating it at $x \cdot \bar{p}$. Since $\left(\mathrm{Ad}_x\left(\mathfrak{X}\right)\right)|_{x \cdot \bar{p}} \in \mathscr{V}_{x \cdot \bar{p}}$ we get that Dx maps vertical spaces to vertical spaces. However, it doesn't preserve the Killing fields in the way one might have hoped for. As Dx is a linear isometry it also preserves the orthogonal complements. These complements are our horizontal spaces $\mathscr{H}_{\bar{p}} = \left(\mathscr{V}_{\bar{p}}\right)^{\perp} \subset T_{\bar{p}}\bar{M}$. We know that $DF : \mathscr{H}_{\bar{p}} \to T_pM$ is an isomorphism. We have also seen that all of the spaces $\mathscr{H}_{x \cdot \bar{p}}$ are isometric to $\mathscr{H}_{\bar{p}}$ via Dx. We can then define the Riemannian metric on T_pM using the isomorphism $DF : \mathscr{H}_{\bar{p}} \to T_pM$. This means that $F : \bar{M} \to M$ defines a Riemannian submersion.

In the above discussion we did not discuss what conditions to put on the action of G on $\bar{M}$ in order to ensure that the quotient becomes a nice manifold. If G is compact and acts freely, then this will happen. The general situation is studied in section 5.6.4. In the next subsection we consider the special case of complex projective space as a quotient of a sphere. There is also a general way of getting new metrics on $\bar{M}$ it self from having a general isometric group action. This will be considered in section 4.5.4.

4.5.3 *Complex Projective Space*

Recall that $\mathbb{CP}^n = S^{2n+1}/S^1$, where S^1 acts by complex scalar multiplication on $S^{2n+1} \subset \mathbb{C}^{n+1}$. If we write the metric as

$$ds^2_{2n+1} = dr^2 + \sin^2(r)ds^2_{2n-1} + \cos^2(r)d\theta^2,$$

then we can think of the S^1 action on S^{2n+1} as acting separately on S^{2n-1} and S^1. Then

$$\mathbb{CP}^n = \left[0, \frac{\pi}{2}\right] \times \left(\left(S^{2n-1} \times S^1\right)/S^1\right),$$

and the metric can be written as discussed in section 1.4.6

$$dr^2 + \sin^2(r)\left(g + \cos^2(r)h\right).$$

If we restrict our attention to the case where $n = 2$, then the metric can be written as

$$dr^2 + \sin^2(r)\left(\cos^2(r)(\sigma^1)^2 + (\sigma^2)^2 + (\sigma^3)^2\right).$$

This is a bit different from the warped product metrics we have seen so far. It is certainly still possible to apply the general techniques of distance functions to compute the curvature tensor. Instead we use the Riemannian submersion apparatus that was developed in the previous section. We shall also consider the general case rather than $n = 2$.

The O'Neill formula from theorem 4.5.3 immediately shows that $\mathbb{CP}^n$ has sectional curvature ≥ 1. Let V be the unit vector field on S^{2n+1} that is tangent to the S^1 action. Then $\mathrm{i}\,V$ is the unit inward pointing normal vector to $S^{2n+1} \subset \mathbb{C}^{n+1}$. This shows that the horizontal distribution, which is orthogonal to V, is invariant under multiplication by i. This corresponds to the fact that $\mathbb{CP}^n$ has a complex structure. It also gives us the integrability tensor for this submersion. If we let $\bar{X}, \bar{Y}$ be basic horizontal vector fields and denote the canonical Euclidean metric on $\mathbb{C}^{n+1}$ by $\bar{g}$, then

$$\begin{aligned}
\bar{g}\left(\frac{1}{2}\left[\bar{X}, \bar{Y}\right], V\right) &= \bar{g}\left(\nabla^{S^{2n+1}}_{\bar{X}}\bar{Y}, V\right) \\
&= \bar{g}\left(\nabla^{\mathbb{C}^{n+1}}_{\bar{X}}\bar{Y}, V\right) \\
&= -\bar{g}\left(\bar{Y}, \nabla^{\mathbb{C}^{n+1}}_{\bar{X}} V\right) \\
&= \bar{g}\left(\bar{Y}, \nabla^{\mathbb{C}^{n+1}}_{\mathrm{i}\bar{X}}\,\mathrm{i}\,V\right) \\
&= \mathrm{II}^{S^{2n+1}}\left(\bar{Y}, \mathrm{i}\,\bar{X}\right) \\
&= \bar{g}\left(\bar{Y}, \mathrm{i}\,\bar{X}\right).
\end{aligned}$$

Thus

$$\frac{1}{2}\left[\bar{X}, \bar{Y}\right]^{\mathcal{V}} = \bar{g}\left(\bar{Y}, \mathrm{i}\,\bar{X}\right) V.$$

If we let X, Y be orthonormal on $\mathbb{CP}^n$, then the horizontal lifts $\bar{X}, \bar{Y}$ are also orthonormal so

$$\begin{aligned}\sec(X, Y) &= 1 + \frac{3}{4}\left|\left[\bar{X}, \bar{Y}\right]^{\mathscr{V}}\right|^2 \\ &= 1 + 3\left|\bar{g}\left(\bar{Y}, \mathrm{i}\bar{X}\right)\right|^2 \\ &\leq 4,\end{aligned}$$

with equality precisely when $\bar{Y} = \pm\,\mathrm{i}\,\bar{X}$.

The proof of theorem 4.5.3 in fact gave us a formula for the full curvature tensor. One can use that formula on an orthonormal set of vectors of the form X, $\mathrm{i}\,X$, Y, $\mathrm{i}\,Y$ to see that the curvature operator is not diagonalized on a decomposable basis of the form $E_i \wedge E_j$ as was the case in the previous examples. In fact it is diagonalized by vectors of the form

$$\begin{aligned}&X \wedge \mathrm{i}\,X \pm Y \wedge \mathrm{i}\,Y, \\ &X \wedge Y \pm \mathrm{i}\,X \wedge \mathrm{i}\,Y, \\ &X \wedge \mathrm{i}\,Y \pm Y \wedge \mathrm{i}\,X\end{aligned}$$

and has eigenvalues that lie in the interval $[0, 6]$.

We can also see that this metric on $\mathbb{CP}^n$ is Einstein with Einstein constant $2n+2$. If we fix a unit vector X and an orthonormal basis for the complement $E_0, \ldots, E_{2n-2}$ so that the lifts satisfy $\mathrm{i}\,\bar{X} = \bar{E}_0$, then we get that

$$\begin{aligned}\operatorname{Ric}(X, X) &= \sum_{i=0}^{2n-2} \sec(X, E_i) \\ &= \sec(X, E_0) + \sum_{i=1}^{2n-2} \sec(X, E_i) \\ &= 1 + 3\left|\bar{g}\left(\bar{E}_0, \mathrm{i}\,\bar{X}\right)\right|^2 + \sum_{i=1}^{2n-2}\left(1 + 3\left|\bar{g}\left(\bar{E}_i, \mathrm{i}\,\bar{X}\right)\right|^2\right) \\ &= 1 + 3\left|\bar{g}\left(\mathrm{i}\,\bar{X}, \mathrm{i}\,\bar{X}\right)\right|^2 + \sum_{i=1}^{2n-2}\left(1 + 3\,|0|^2\right) \\ &= 1 + 3 + 2n - 2 \\ &= 2n + 2.\end{aligned}$$

4.5.4 Berger-Cheeger Perturbations

The construction we do here was first considered by Cheeger and was based on a slightly different construction by Berger used to construct the Berger spheres.

Fix a Riemannian manifold (M, g) and a Lie group G with a *right-invariant metric* $(,)$. If G acts by isometries on M, then it also acts by isometries on $\mathrm{G} \times M$ with respect to the product metrics $g_\lambda = \lambda(,) + g$, $\lambda > 0$ via the action $h \cdot (x, p) \mapsto \left(xh^{-1}, hp\right)$. This action is free as G acts freely on itself. The quotient $(\mathrm{G} \times M)/\mathrm{G}$ is also denoted by $\mathrm{G} \times_{\mathrm{G}} M$. The natural map $M \to \mathrm{G} \times M \to \mathrm{G} \times_{\mathrm{G}} M$ is a bijection. Thus the quotient is in a natural way a manifold diffeomorphic to M. The quotient map $Q : \mathrm{G} \times M \to M$ is explicitly given by $Q(x, p) = xp$.

As G acts by isometries with respect to the product metrics $\lambda(,) + g$ we obtain a submersion metric g_λ on $M = \mathrm{G} \times_{\mathrm{G}} M$. We wish to study this perturbed metric's relation to the original metric g. The tangent space T_pM is naturally decomposed into the vectors $\mathscr{V}_p$ that are tangent to the action and the orthogonal complement $\mathscr{H}_p$. Unlike the case where G acts freely on M this decomposition is not necessarily a nicely defined distribution. It might happen that G fixes certain but not all points in M. For example, at points p that are fixed it follows that $\mathscr{V}_p = \{0\}$. At other points $\mathscr{V}_p \neq \{0\}$. The nomenclature is, however, not inappropriate. If $\mathfrak{X} \in T_e\mathrm{G}$, then $F^t(p) = \exp(t\mathfrak{X}) \cdot p$ defines a 1-parameter group of isometries. If $X = \frac{d}{dt}F^t(p)|_{t=0}$ is the corresponding Killing field on M, then $\left(-\mathfrak{X}, X|_p\right) \in T_e\mathrm{G} \times T_pM$ is a vertical direction for this action at $(e, p) \in \mathrm{G} \times M$. Therefore, $\mathscr{V}_p$ is simply the image of the projection of the vertical distribution to T_pM. Vectors in $\mathscr{H}_p$ are thus also horizontal for the action on $\mathrm{G} \times M$. All the other horizontal vectors in $T_e\mathrm{G} \times T_pM$ depend on the choice of λ and have a component of the form $\left(\left|X|_p\right|_g^2 \mathfrak{X}, \lambda |\mathfrak{X}|^2 X|_p\right)$. The image of such a horizontal vector under $Q : \mathrm{G} \times M \to M$ is given by

$$\begin{aligned}
DQ\left(\left|X|_p\right|_g^2 \mathfrak{X}, \lambda |\mathfrak{X}|^2 X|_p\right) &= \left|X|_p\right|_g^2 DQ(\mathfrak{X}, 0) + \lambda |\mathfrak{X}|^2 DQ\left(0, X|_p\right) \\
&= -\left|X|_p\right|_g^2 DQ\left(\frac{d}{dt}(e \cdot \exp(-t\mathfrak{X}))|_{t=0}, 0\right) \\
&\quad +\lambda |\mathfrak{X}|^2 DQ\left(0, \frac{d}{dt}(\exp(t\mathfrak{X}) \cdot p)|_{t=0}\right) \\
&= -\left|X|_p\right|_g^2 \frac{d}{dt}(Q(\exp(-t\mathfrak{X}), p))|_{t=0} \\
&\quad +\lambda |\mathfrak{X}|^2 \frac{d}{dt}(Q(e, \exp(t\mathfrak{X}) \cdot p))|_{t=0} \\
&= -\left|X|_p\right|_g^2 \frac{d}{dt}(\exp(-t\mathfrak{X}) \cdot p)|_{t=0} \\
&\quad +\lambda |\mathfrak{X}|^2 \frac{d}{dt}(\exp(t\mathfrak{X}) \cdot p)|_{t=0}
\end{aligned}$$

$$= \left|X|_p\right|_g^2 X|_p + \lambda\, |\mathfrak{X}|^2 X|_p$$
$$= \left(\lambda\, |\mathfrak{X}|^2 + \left|X|_p\right|_g^2\right) X|_p$$

The horizontal lift of $X|_p \in \mathscr{V}_p$ to $T_e\mathrm{G} \times T_pM$ is consequently given by

$$\overline{X|_p} = \left(\frac{\left|X|_p\right|_g^2}{\lambda\, |\mathfrak{X}|^2 + \left|X|_p\right|_g^2} \mathfrak{X}, \frac{\lambda\, |\mathfrak{X}|^2}{\lambda\, |\mathfrak{X}|^2 + \left|X|_p\right|_g^2} X|_p \right),$$

and its length in g_λ satisfies

$$\begin{aligned}
\left|\overline{X|_p}\right|_{g_\lambda}^2 &= \left(\frac{\left|X|_p\right|_g^2}{\lambda\, |\mathfrak{X}|^2 + \left|X|_p\right|_g^2} \right)^2 \lambda\, |\mathfrak{X}|^2 \\
&\quad + \left(\frac{\lambda\, |\mathfrak{X}|^2}{\lambda\, |\mathfrak{X}|^2 + \left|X|_p\right|_g^2} \right)^2 \left|X|_p\right|_g^2 \\
&= \frac{\lambda\, |\mathfrak{X}|^2}{\lambda\, |\mathfrak{X}|^2 + \left|X|_p\right|_g^2} \left|X|_p\right|_g^2 \\
&\leq \left|X|_p\right|_g^2 .
\end{aligned}$$

In particular, $\left|\overline{X|_p}\right|_{g_\lambda}^2$ has limit 0 as $\lambda \to 0$ and limit $\left|X|_p\right|_g^2$ as $\lambda \to \infty$. This means that the metric g_λ is gotten from g by squeezing the orbits of the action of G. However, the squeezing depends on the point according to this formula. The only case where the squeezing is uniform is when the Killing fields generated by the action have constant length on M. The Berger spheres are a special case of this.

Using that we know how to compute horizontal lifts and that the metric on $\mathrm{G} \times M$ is a product metric it is possible to compute the curvature of g_λ in terms of the curvature of g, λ, the curvature of $(,)$, and the integrability tensor. We will consider one important special case.

Let $X, Y \in \mathscr{H}_p$. In this case the vectors are already horizontal for the action on $\mathrm{G} \times M$. Thus we have that $\sec_{g_\lambda}(X, Y) \geq \sec_g(X, Y)$. There is a correction coming from the integrability tensor associated with the action on $\mathrm{G} \times M$ that possibly increases these curvatures.

4.6 Further Study

The book by O'Neill [80] gives an excellent account of Minkowski geometry and also studies in detail the Schwarzschild metric in the setting of general relativity. It appears to have been the first exact 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. The third edition of [47] also contains a good number of examples. Specifically they have a lot of material on hyperbolic space. They also have a brief account of the Schwarzschild metric in the setting of general relativity.

Another book, which contains many more advanced examples, is [12]. This is also a good reference on Riemannian geometry in general.

4.7 Exercises

Remark. It will be useful to read exercises 3.4.23, 3.4.24, and 3.4.25 before doing the exercises for this chapter.

EXERCISE 4.7.1. Show that the Schwarzschild metric does not have parallel curvature tensor.

EXERCISE 4.7.2. Show that the Berger spheres ($\varepsilon \neq 1$) do not have parallel curvature tensor.

EXERCISE 4.7.3. This exercise covers a few interesting aspects of projective spaces.

(1) Show that $\mathrm{U}(n+1)$ acts by isometries on $\mathbb{CP}^n$. Hint: Use that $\mathrm{U}(n+1)$ acts by isometries on $S^{2n+1}(1)$ and commutes with the quotient action that creates $\mathbb{CP}^n$.
(2) Show that for each $p \in \mathbb{CP}^n$ there is an isometry $A_p \in \mathrm{Iso}_p$ with $DA_p|_p = -I$.
(3) Use the fact that isometries leave ∇ and R invariant to show that $\nabla R = 0$.
(4) Repeat 1,2,3 for $\mathbb{HP}^n$ using the *symplectic group* $\mathrm{Sp}(n+1)$ of matrices with quaternionic entries satisfying $A^*A = I$, where $A^* = \bar{A}^t$. See also exercise 1.6.22 for more on quaternions.

EXERCISE 4.7.4. Assume that a Riemannian manifold (M, g) has a function f such that

$$\mathrm{Hess} f = \lambda(x)\, g + \mu(f)\, df^2,$$

where $\lambda : M \to \mathbb{R}$ and $\mu : \mathbb{R} \to \mathbb{R}$. Show that the metric is locally a warped product.

EXERCISE 4.7.5. Show that if $\mathrm{Hess} f = \lambda g$, then $\lambda = \frac{\Delta f}{\dim M}$.

EXERCISE 4.7.6. Consider a function f on a Riemannian manifold (M, g) so that $\nabla f \neq 0$ and ∇f is an eigenvector for $S(X) = \nabla_X \nabla f$. Show that if S has ≤ 2 eigenvalues, then the metric is locally a warped product metric.

EXERCISE 4.7.7 (O'NEILL). For a Riemannian submersion as in section 4.5 define the A-tensors

$$A_{\bar{X}}\bar{Y} = \left[\bar{\nabla}_{\bar{X}}\bar{Y}\right]^{\mathscr{V}},$$
$$A_{\bar{X}}V = \left[\bar{\nabla}_{\bar{X}}V\right]^{\mathscr{H}}.$$

We also have the T-tensor from exercises 2.5.26 and 2.5.25 but our notation for horizontal and vertical fields is the reverse of tangent and normal fields from those exercises. Note that both $A_{\bar{X}}$ and T_V make sense. We can extend both tensors by declaring $A_V = 0$ and $T_{\bar{X}} = 0$ and thus obtain $(1, 2)$-tensors on $\bar{M}$.

(1) Show that both A-tensors are tensorial.
(2) Show that $A_{\bar{X}}\bar{Y} = \frac{1}{2}\left[\bar{X}, \bar{Y}\right]^{\mathscr{V}}$.
(3) Show that $\bar{g}\left(A_{\bar{X}}\bar{Y}, V\right) = -\bar{g}\left(\bar{Y}, A_{\bar{X}}V\right)$.
(4) Show that $(\nabla_V A)_W = -A_{T_V W}$ and $(\nabla_X A)_W = -A_{A_X W}$.
(5) Show that $(\nabla_{\bar{X}} T)_{\bar{Y}} = -T_{A_{\bar{X}}\bar{Y}}$ and $(\nabla_V T)_{\bar{Y}} = -T_{T_V \bar{Y}}$.
(6) Show that

$$\bar{g}\left((\nabla_U A)_{\bar{X}} V, W\right) = \bar{g}\left(T_U V, A_{\bar{X}}W\right) - \bar{g}\left(T_U W, A_{\bar{X}}V\right).$$

EXERCISE 4.7.8 (O'NEILL). This exercise builds on the previous exercise. The Gauss equations explain how to calculate the curvature tensor on vectors tangent to the fibers of a submersion. Show that horizontal and "verti-zontal" curvatures can be calculated by the formulas

$$\bar{R}\left(\bar{Y}, \bar{X}, \bar{X}, \bar{Y}\right) = R\left(Y, X, X, Y\right) - 3\left|A_{\bar{X}}\bar{Y}\right|^2$$

and

$$\bar{R}\left(V, \bar{X}, \bar{X}, V\right) = \bar{g}\left((\nabla_{\bar{X}} T)_V V, \bar{X}\right) + \left|A_{\bar{X}}V\right|^2 - \left|T_V \bar{X}\right|^2.$$

Compare the last formula to the radial curvature equation.

EXERCISE 4.7.9. Let $(M, g) = (M_1 \times M_2, g_1 + g_2)$ be a Riemannian product manifold.

(1) Show that $R = R_1 + R_2$, where R_i is the curvature tensor of (M_i, g_i) pulled back to M.
(2) Assume for the remainder of this exercise that (M_i, g_i) has constant curvature c_i. Show that $R = c_1 g_1 \circ g_1 + c_2 g_2 \circ g_2$.

(3) Show that (M, g) is Einstein if and only if $(n_1 - 1)\, c_1 = (n_2 - 1)\, c_2$ where $n_i = \dim M_i$.
(4) Show that the Weyl tensor for (M, g) vanishes when either $c_1 = -c_2$, $n_1 = 1$, or $n_2 = 1$. Hint: Calculate $(g_1 - g_2) \circ (g_1 + g_2)$ and compare it to R.
(5) Show that if none of the conditions in (4) hold, then the Weyl tensor does not vanish.

EXERCISE 4.7.10. Let $(M^n, g) = \left(I \times N, dr^2 + \rho^2(r)\, g_N\right)$ be a warped product metric with constant curvature k.

(1) Show that $\left(N^{n-1}, \rho^2(r)\, g_N\right)$ has constant curvature $k + \left(\frac{\dot{\rho}}{\rho}\right)^2$ if $n > 2$.
(2) Show explicitly that hyperbolic space can be represented as a warped product over both hyperbolic space and Euclidean space.

EXERCISE 4.7.11. Consider an Einstein metric $\left(N^{n-1}, g_N\right)$ with $\mathrm{Ric} = \frac{n-2}{n-1}\lambda g_N$, $\lambda < 0$. Find a $\rho : \mathbb{R} \to (0, \infty)$ such that $(M^n, g) = \left(\mathbb{R} \times N, dr^2 + \rho^2(r)\, g_N\right)$ becomes an Einstein metric with $\mathrm{Ric} = \lambda g$.

EXERCISE 4.7.12. Let $\left(N^{n-1}, g_N\right)$ have constant curvature c with $n > 2$. Consider the warped product metric $(M, g) = \left(I \times N, dr^2 + \rho^2(r)\, g_N\right)$.

(1) Show that the curvature of g is given by

$$\begin{aligned} R &= \frac{c - \dot{\rho}^2}{\rho^2} g_r \circ g_r - 2\frac{\ddot{\rho}}{\rho} dr^2 \circ g_r \\ &= \frac{c - \dot{\rho}^2}{\rho^2} g \circ g - 2\left(\frac{\ddot{\rho}}{\rho} + \frac{c - \dot{\rho}^2}{\rho^2}\right) dr^2 \circ g. \end{aligned}$$

(2) Show that the Weyl tensor vanishes.
(3) Show directly that the Schouten tensor satisfies:

$$(\nabla_X P)(Y, Z) = (\nabla_Y P)(X, Z)\,.$$

See also exercise 3.4.26 for an indirect approach when $n > 3$.

EXERCISE 4.7.13. The stereographic projection of $x^{n+1} = 0$ to a hypersurface $M \subset \mathbb{R}^n \times \mathbb{R}$ that is transverse to the lines emanating from $-e_{n+1} = (0, \ldots, 0, -1)$ is given by $x \mapsto S(x)$ where $x \in \mathbb{R}^n$ and $S(x) = -e_{n+1} + \lambda(x)\,(e_{n+1} + (x, 0))$.

(1) When $M = S^n(1)$ show that $\lambda\left(1 + |x|^2\right) = 2$ and that S is a conformal map with the property that in these coordinates the metric on $S^n(1)$ is given by

$$\frac{4}{\left(1 + |x|^2\right)^2} g_{\mathbb{R}^n}.$$

(2) When $M = H^n(1) \in \mathbb{R}^{n,1}$ show that $\lambda\left(1-|x|^2\right) = 2$ and that S is a conformal map with the property that in these coordinates the metric on $H^n(1)$ is Poincaré disc

$$\frac{4}{\left(1-|x|^2\right)^2} g_{\mathbb{R}^n}.$$

EXERCISE 4.7.14. Let $\tilde{g} = e^{2\psi} g$ be a metric conformally equivalent to g and a $\tilde{}$ referring to metric objects in the conformally changed metric.

(1) Show that

$$\tilde{\nabla}_X Y = \nabla_X Y + (D_X\psi)\, Y + (D_Y\psi)\, X - g(X,Y)\, \nabla\psi.$$

(2) With notation as in exercise 3.4.23 show that

$$\begin{aligned} e^{-2\psi}\tilde{R} &= R - 2\left(\operatorname{Hess}\psi - (d\psi)^2\right)\circ g - |d\psi|^2\, g\circ g \\ &= R - \left(2\operatorname{Hess}\psi - 2\,(d\psi)^2 + |d\psi|^2\, g\right)\circ g. \end{aligned}$$

(3) If X, Y are orthonormal with respect to g, show that

$$\begin{aligned} e^{2\psi}\,\widetilde{\sec}(X,Y) = {} & \sec(X,Y) - \operatorname{Hess}\psi\,(X,X) - \operatorname{Hess}\psi\,(Y,Y) \\ & + (D_X\psi)^2 + (D_Y\psi)^2 - |d\psi|^2. \end{aligned}$$

(4) Show that

$$\widetilde{\operatorname{Ric}} = \operatorname{Ric} - (n-2)\left(\operatorname{Hess}\psi - d\psi^2\right) - \left(\Delta\psi + (n-2)\,|d\psi|^2\right) g.$$

(5) Show that

$$e^{2\psi}\,\widetilde{\operatorname{scal}} = \operatorname{scal} - 2\,(n-1)\,\Delta\psi - (n-1)\,(n-2)\,|d\psi|^2.$$

(6) Using exercise 3.4.25 show that

$$e^{-2\psi}\,\tilde{W} = W.$$

This is referred to as the conformal invariance of the Weyl tensor under conformal changes and was discovered by Weyl.

EXERCISE 4.7.15. Show that

$$\left(\frac{1}{4}\rho_0^{n-2}+r^{2-n}\right)^{\frac{4}{n-2}}g_{\mathbb{R}^n}=\frac{1}{1-\left(\frac{\rho_0}{\rho}\right)^{n-2}}d\rho^2+\rho^2ds_{n-1}^2,$$

where the right-hand side is the scalar flat metric from section 4.2.3. Use this to rewrite the Schwarzschild metric from section 4.2.5 as

$$\left(\frac{1}{4}\rho_0^{n-3}+r^{3-n}\right)^{\frac{4}{n-3}}g_{\mathbb{R}^{n-1}}+\rho_0^2\frac{4}{(n-3)^3}\left(\frac{\frac{1}{4}\rho_0^{n-3}-r^{3-n}}{\frac{1}{4}\rho_0^{n-3}+r^{3-n}}\right)^2d\theta^2.$$

EXERCISE 4.7.16 (STATIC EINSTEIN EQUATIONS). Consider a metric of the form $(M,g)=\left(N\times\mathbb{R},g_N+w^2dt^2\right)$, where $w:N\to(0,\infty)$ and $\dim N=n-1$. Let X,Y,Z be vector fields on N. Note that they can also be considered as vector fields on M.

(1) Show that $\nabla_X^NY=\nabla_X^MY$ and $R^N(X,Y)Z=R^M(X,Y)Z$. Conclude that $\mathrm{Ric}^M(X,\partial_t)=0$.
(2) Show the vector field ∂_t satisfies $|\partial_t|^2=w^2$ in (M,g).
(3) Show that

$$\nabla^M_{\partial_t}\partial_t=-w\nabla w\text{ and }\nabla^M_X\partial_t=\nabla^M_{\partial_t}X=\frac{1}{w}(D_Xw)\,\partial_t.$$

Hint: Show that $g\left(\nabla^M_{\partial_t}\partial_t,\partial_t\right)=0$ and calculate $D_X|\partial_t|^2$.
(4) Show that

$$R^M(X,\partial_t)\,\partial_t=-w\nabla_X\nabla w,$$

and

$$\mathrm{Ric}^M(\partial_t,\partial_t)=-w\Delta w,$$
$$\mathrm{Ric}^M(X,X)=\mathrm{Ric}^N(X,X)-\frac{1}{w}\mathrm{Hess}\,(X,X).$$

(5) Show that $\mathrm{Ric}^M=\lambda g$, $\lambda\in\mathbb{R}$, if and only if

$$\mathrm{Ric}^N-\frac{1}{w}\mathrm{Hess}\,w=\lambda g_N,$$
$$w\Delta w+\lambda w^2=0,$$

if and only if

$$\begin{aligned} \mathrm{Ric}^N - \frac{1}{w}\operatorname{Hess} w &= \lambda g_N, \\ \mathrm{scal}^N &= (n-2)\,\lambda . \end{aligned}$$

EXERCISE 4.7.17. A Riemannian manifold (M,g) is said to be *locally conformally flat* if every $p \in M$ lies in a coordinate neighborhood U where

$$g = e^{-2\psi}\left(\left(dx^1\right)^2 + \cdots + (dx^n)^2\right).$$

(1) Show that the space forms S_k^n with metrics $dr^2 + \mathrm{sn}_k^2(r)\,ds_{n-1}^2$ are locally conformally flat.
(2) Show that if an Einstein metric is locally conformally flat, then it has constant curvature.
(3) When $n = 2$ Gauss showed that such coordinates always exit. They are called *isothermal coordinates*. Assume that $\dim M = 2$.
   (a) Show that if $du \neq 0$ on some open subset $O \subset M$, then up to sign there is a unique 1-form $\omega = i_{\nabla u}\,\mathrm{vol}_g$ that satisfies: $|du| = |\omega|$ and $g(du,\omega) = 0$.
   (b) Show that $d\omega = \left(\Delta_g u\right)\mathrm{vol}_g$.
   (c) Show that isothermal coordinates exit provided that for each $p \in M$ it is possible to find u on a neighborhood of p so that $\Delta_g u = 0$ and $du|_p \neq 0$.

EXERCISE 4.7.18 (SCHOUTEN 1921). Let (M,g) be a Riemannian manifold of dimension $n > 2$.

(1) Show that g is locally conformally flat if and only if $W = 0$ and locally there is a function ψ so that $P = 2\operatorname{Hess}\psi - 2\,(d\psi)^2 + |d\psi|^2\,g$. Note that the condition $W = 0$ is redundant when $n = 3$. Hint: You have to use the curvature characterization of being locally Euclidean (see exercise 3.4.20 or theorem 5.5.8).
(2) Show that if g is locally conformally flat then

$$(\nabla_X P)(Y,Z) = (\nabla_Y P)(X,Z).$$

Hint: When $n > 3$, this follows from exercise 3.4.26. When $n \geq 3$, use that $R = P \circ g$, the specific form of P from (1), and show that

$$(\nabla_X \operatorname{Hess}\psi)(Y,Z) - (\nabla_Y \operatorname{Hess}\psi)(X,Z) = R(X,Y,\nabla\psi,Z).$$

EXERCISE 4.7.19 (SCHOUTEN 1921). In this exercise assume that we have a Riemannian manifold of dimension $n > 2$ such that $W = 0$ and $(\nabla_X P)(Y,Z) = (\nabla_Y P)(X,Z)$.

(1) Show that if there is a 1-form ω such that

$$\nabla\omega = \frac{1}{2}P + \omega^2 - \frac{1}{2}|\omega|^2 g,$$

then locally $\omega = d\psi$ and $P = 2\,\mathrm{Hess}\,\psi - 2(d\psi)^2 + |\nabla\psi|^2 g$.

(2) The integrability condition for finding such an ω in the sense of exercise 3.4.20 can be stated using only covariant derivatives. On the left-hand side we take one more derivative $\nabla^2_{X,Y}\omega$ and use the Ricci formula for commuting covariant derivatives as an alternative to Clairaut's theorem on partial derivatives:

$$\nabla^2_{X,Y}\omega - \nabla^2_{Y,X}\omega = R_{X,Y}\omega.$$

Show that if $\nabla\omega = \frac{1}{2}P + \omega^2 - \frac{1}{2}|\omega|^2 g$, then

$$\begin{aligned}\left(\nabla^2_{X,Y}\omega\right)(Z) &= \frac{1}{2}(\nabla_X P)(Y,Z)\\ &\quad + (\nabla_X\omega)(Y)\,\omega(Z) + \omega(Y)(\nabla_X\omega)(Z)\\ &\quad - g(\nabla_X\omega,\omega)\,g(Y,Z).\end{aligned}$$

(3) Use $\nabla\omega = \frac{1}{2}P + \omega^2 - \frac{1}{2}|\omega|^2 g$ again to show that

$$\begin{aligned}\nabla^2_{X,Y}\omega - \nabla^2_{Y,X}\omega &= \frac{1}{2}P(X,Z)\,\omega(Y) - \frac{1}{2}P(X,V)\,g(Y,Z)\\ &\quad - \frac{1}{2}P(Y,Z)\,\omega(X) + \frac{1}{2}P(Y,V)\,g(X,Z)\\ &= (P\circ g)(X,Y,V,Z),\end{aligned}$$

where V is the vector field dual to ω.

(4) Now use $R = P\circ g$ to show that

$$(R_{X,Y}\omega)(Z) = (P\circ g)(X,Y,V,Z).$$

(5) Finally, show that this implies that the integrability conditions for solving for ω are satisfied and conclude that the manifold is locally conformally flat.

EXERCISE 4.7.20. Consider a product metric $\left(N^2\times\mathbb{R}, g_N + g_{\mathbb{R}}\right)$.

(1) Show that $P_{N\times\mathbb{R}} = \frac{\mathrm{scal}_N}{2}(g_N - g_{\mathbb{R}})$.
(2) Show that this product metric is conformally flat if and only if scal_N is constant.

EXERCISE 4.7.21. Let (M^n, g) , $n > 2$ have constant curvature k.

(1) Use exercise 4.7.19 to show that the metric is locally conformally flat.
(2) Show that if $g = e^{-2\psi}\left((dx^1)^2 + \cdots + (dx^n)^2\right)$, then

$$2e^{\psi}\partial_i\partial_j e^{\psi} = \left(k + \sum\left(\partial_k e^{\psi}\right)^2\right)\delta_{ij}.$$

Hint: Use part 2 of 4.7.14.
(3) Show that

$$e^{\psi} = a + \sum b_i x^i + c\sum\left(x^i\right)^2,$$

where $k = 4ac - \sum b_i^2$.

EXERCISE 4.7.22. The *Heisenberg group* with its Lie algebra is

$$\mathrm{G} = \left\{\begin{bmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{bmatrix} \mid a, b, c \in \mathbb{R}\right\},$$

$$\mathfrak{g} = \left\{\begin{bmatrix} 0 & x & z \\ 0 & 0 & y \\ 0 & 0 & 0 \end{bmatrix} \mid a, b, c \in \mathbb{R}\right\}.$$

A basis for the Lie algebra is:

$$X = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, Y = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, Z = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

(1) 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.
(2) Show that the Ricci tensor has both negative and positive eigenvalues.
(3) Show that the scalar curvature is constant.
(4) Show that the Ricci tensor is not parallel.

EXERCISE 4.7.23. Consider metrics of the form

$$dr^2 + \rho^2(r)\left(\phi^2(r)(\sigma^1)^2 + (\sigma^2)^2 + (\sigma^3)^2\right).$$

(1) Show that if

$$\begin{aligned}\dot{\rho} &= \phi,\\ \dot{\rho}^2 &= 1 - k\rho^{-4},\\ \rho(0) &= k^{\frac{1}{4}}, \dot{\rho}(0) = 0,\\ \phi(0) &= 0, \dot{\phi}(0) = 2,\end{aligned}$$

then we obtain a family of Ricci flat metrics on TS^2.

(2) Show that $\rho(r) \sim r$, $\dot{\rho}(r) \sim 1$, $\ddot{\rho}(r) \sim 2kr^{-5}$ as $r \to \infty$. Conclude that all curvatures are of order r^{-6} as $r \to \infty$ and that the metric looks like $(0,\infty) \times \mathbb{RP}^3 = (0,\infty) \times \mathrm{SO}(3)$ 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*.

EXERCISE 4.7.24. For the general metric

$$dr^2 + \rho^2(r)\left(\phi^2(r)(\sigma^1)^2 + (\sigma^2)^2 + (\sigma^3)^2\right)$$

show that the $(1,1)$-tensor, which in the orthonormal frame looks like

$$\begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

yields a Hermitian structure.

(1) Show that this structure is Kähler, i.e., parallel, if and only if $\dot{\rho} = \phi$.
(2) Find the scalar curvature for such metrics.
(3) 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.

EXERCISE 4.7.25. Show that $\tau\left(\mathbb{RP}^{n-1}\right)$ admits rotationally symmetric metrics $dr^2 + \rho^2(r)\,ds^2_{n-1}$ such that $\rho(r) = r$ for $r > 1$ and the Ricci 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 $\tau\left(\mathbb{RP}^{n-1}\right)$.

EXERCISE 4.7.26. We say that (M, g) admits *orthogonal coordinates* around $p \in M$ if we have coordinates on some neighborhood of p, where

$$g_{ij} = 0 \text{ for } i \neq j,$$

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. To find a counterexample, you may want to show that in such coordinates the curvatures $R_{ijk}^{l} = 0$ if all indices are distinct. It can be shown that such coordinates always exist in 3 dimensions.

EXERCISE 4.7.27. Show that the Weyl tensors for the Schwarzschild metric and the Eguchi-Hanson metrics are not zero.

EXERCISE 4.7.28. 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 bivectors $\Lambda^2 TM$ 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 $*(e_1 \wedge e_2) = e_3 \wedge e_4$.

(1) Show that his 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.)
(2) 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

$$\begin{aligned} e_1 \wedge e_2 \pm e_3 \wedge e_4 &\in \Lambda^{\pm} TM, \\ e_1 \wedge e_3 \pm e_4 \wedge e_2 &\in \Lambda^{\pm} TM, \\ e_1 \wedge e_4 \pm e_2 \wedge e_3 &\in \Lambda^{\pm} TM. \end{aligned}$$

(3) Thus, any linear map $L : \Lambda^2 TM \to \Lambda^2 TM$ has a block decomposition

$$L = \begin{bmatrix} A & D \\ B & C \end{bmatrix},$$

$$\begin{aligned} A &: \Lambda^+ TM \to \Lambda^+ TM, \\ D &: \Lambda^+ TM \to \Lambda^- TM, \\ B &: \Lambda^- TM \to \Lambda^+ TM, \\ C &: \Lambda^- TM \to \Lambda^- TM. \end{aligned}$$

In particular, we can decompose the curvature operator $\mathfrak{R} : \Lambda^2 TM \to \Lambda^2 TM$:

$$\mathfrak{R} = \begin{bmatrix} A & D \\ B & C \end{bmatrix}.$$

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

$$W = \begin{bmatrix} W^+ & 0 \\ 0 & W^- \end{bmatrix}.$$

Find these decompositions for both of the doubly warped metrics:

$$I \times S^1 \times S^2, dr^2 + \rho^2 (r)\, d\theta^2 + \phi^2 (r)\, ds_2^2,$$
$$I \times S^3, dr^2 + \rho^2(r) \left(\phi^2(r)(\sigma^1)^2 + (\sigma^2)^2 + (\sigma^3)^2\right).$$

Use as basis for TM the natural frames in which we computed the curvature tensors. Now

(4) find the curvature operators for the Schwarzschild metric, the Eguchi-Hanson metric, $S^2 \times S^2$, S^4, and $\mathbb{CP}^2$.

(5) Show that (M, g) is Einstein if and only if $B = 0$ if and only if for every plane π and its orthogonal complement $\pi^\perp$ we have: $\sec(\pi) = \sec\left(\pi^\perp\right)$.

Chapter 5
Geodesics and Distance

We are now ready to move on to the local and global geometry of Riemannian manifolds. The main tool for this will be the important concept of geodesics. These curves will help us define and understand Riemannian manifolds as metric spaces. One is led quickly 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 makes it possible to define metric distance functions. We shall study when these distance functions are smooth and show the existence of the smooth distance functions introduced in chapter 3. We also classify complete simply connected manifolds of constant curvature; showing that they are the ones we have already constructed in chapters 1 and 4.

The idea of thinking of a Riemannian manifold as a metric space must be old, but it wasn't until the early 1920s that first Cartan and then later 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. Surprisingly, it wasn't until Klingenberg's work in the 1950s that one got a thorough understanding of the maximal domain on which one has geodesic polar coordinates inside complete manifolds. This work led to the introduction of the two terms *injectivity radius* and *conjugate radius*. Many of our later results will require a detailed analysis of these concepts. The metric characterization of Riemannian isometries wasn't realized until the late 1930s with the work of Myers and Steenrod showing that groups of isometries are Lie groups. Even more surprising is Berestovskii's much more recent metric characterization of Riemannian submersions.

P. Petersen, *Riemannian Geometry*, Graduate Texts in Mathematics 171,
DOI 10.1007/978-3-319-26654-1_5

Another important topic that involves geodesics is the variation of arclength and energy. In this chapter we only develop the first variation formula. This is used to show that curves that minimize length must be geodesics if they are parametrized correctly.

We are also finally getting to results where there will be a significant difference between the Riemannian setting and the pseudo-Riemannian setting. Mixed partials and geodesics easily generalize. However, as there is no norm of vectors in the pseudo-Riemannian setting we do not have arclength or distances. Nevertheless, the energy functional does make sense so we still obtain a variational characterization of geodesics as critical points for the energy functional.

5.1 Mixed Partials

So far we have only considered the calculus of functions (and tensors) on a Riemannian manifold, and have seen that defining the gradient and Hessian requires that we use the metric structure. Here we are going to study maps into Riemannian manifolds and how to define meaningful higher derivatives for such maps. The simplest example is to consider a curve $c : I \to M$ on some interval $I \subset \mathbb{R}$. We know how to define the derivative $\dot{c}$, but not how to define the acceleration in such a way that it also gives us a tangent vector to M. A similar but slightly more general problem is that of defining mixed partial derivatives

$$\frac{\partial^2 c}{\partial t^i \partial t^j}$$

for maps c with several real variables. As we shall see, covariant differentiation plays a crucial role in the definition of these concepts. In this section we only develop a method that covers second partials. In section 6.1.2 we shall explain how to calculate higher order partials as well. This involves a slightly different approach (see section 6.1.1) that is not needed for the developments in this chapter.

Let $c : \Omega \to M$, where $\Omega \subset \mathbb{R}^m$. As we usually reserve x^i for coordinates on M we shall use t^i or s, t, u as coordinates on Ω. The first partials

$$\frac{\partial c}{\partial t^i}$$

are simply defined as the velocity field of $t^i \mapsto c\left(t^1, \dots, t^i, \dots, t^m\right)$, where the remaining coordinates are fixed. We wish to define the second partials so that they also lie TM as opposed to TTM. In addition we also require the following two natural properties:

(1) Equally of mixed second partials:

$$\frac{\partial^2 c}{\partial t^i \partial t^j} = \frac{\partial^2 c}{\partial t^j \partial t^i}.$$

(2) The product rule:

$$\frac{\partial}{\partial t^k} g\left(\frac{\partial c}{\partial t^i}, \frac{\partial c}{\partial t^j}\right) = g\left(\frac{\partial^2 c}{\partial t^k \partial t^i}, \frac{\partial c}{\partial t^j}\right) + g\left(\frac{\partial c}{\partial t^i}, \frac{\partial^2 c}{\partial t^k \partial t^j}\right).$$

The first is similar to assuming that the connection is torsion free and the second to assuming that the connection is metric. As with theorem 2.2.2, were we saw that the key properties of the connection in fact also characterized the connection, we can show that these two rules also characterize how we define second partials. More precisely, if we have a way of defining second partials such that these two properties hold, then we claim that there is a Koszul type formula:

$$2g\left(\frac{\partial^2 c}{\partial t^i \partial t^j}, \frac{\partial c}{\partial t^k}\right) = \frac{\partial}{\partial t^i} g\left(\frac{\partial c}{\partial t^j}, \frac{\partial c}{\partial t^k}\right) + \frac{\partial}{\partial t^j} g\left(\frac{\partial c}{\partial t^k}, \frac{\partial c}{\partial t^i}\right) - \frac{\partial}{\partial t^k} g\left(\frac{\partial c}{\partial t^i}, \frac{\partial c}{\partial t^j}\right).$$

This formula is established in the proof of the next lemma.

Lemma 5.1.1 (Uniqueness of mixed partials). *There is at most one way of defining mixed partials so that (1) and (2) hold.*

Proof. First we show that the Koszul type formula holds if we have a way of defining mixed partials such that (1) and (2) hold:

$$\begin{aligned}
&\frac{\partial}{\partial t^i} g\left(\frac{\partial c}{\partial t^j}, \frac{\partial c}{\partial t^k}\right) + \frac{\partial}{\partial t^j} g\left(\frac{\partial c}{\partial t^k}, \frac{\partial c}{\partial t^i}\right) - \frac{\partial}{\partial t^k} g\left(\frac{\partial c}{\partial t^i}, \frac{\partial c}{\partial t^j}\right) \\
&= g\left(\frac{\partial^2 c}{\partial t^i \partial t^j}, \frac{\partial c}{\partial t^k}\right) + g\left(\frac{\partial c}{\partial t^j}, \frac{\partial^2 c}{\partial t^i \partial t^k}\right) \\
&\quad + g\left(\frac{\partial^2 c}{\partial t^j \partial t^k}, \frac{\partial c}{\partial t^i}\right) + g\left(\frac{\partial c}{\partial t^k}, \frac{\partial^2 c}{\partial t^j \partial t^i}\right) \\
&\quad - g\left(\frac{\partial^2 c}{\partial t^k \partial t^i}, \frac{\partial c}{\partial t^j}\right) - g\left(\frac{\partial c}{\partial t^i}, \frac{\partial^2 c}{\partial t^k \partial t^j}\right) \\
&= g\left(\frac{\partial^2 c}{\partial t^i \partial t^j}, \frac{\partial c}{\partial t^k}\right) + g\left(\frac{\partial c}{\partial t^k}, \frac{\partial^2 c}{\partial t^j \partial t^i}\right) \\
&\quad + g\left(\frac{\partial c}{\partial t^j}, \frac{\partial^2 c}{\partial t^i \partial t^k}\right) - g\left(\frac{\partial^2 c}{\partial t^k \partial t^i}, \frac{\partial c}{\partial t^j}\right) \\
&\quad + g\left(\frac{\partial^2 c}{\partial t^j \partial t^k}, \frac{\partial c}{\partial t^i}\right) - g\left(\frac{\partial c}{\partial t^i}, \frac{\partial^2 c}{\partial t^k \partial t^j}\right) \\
&= 2g\left(\frac{\partial^2 c}{\partial t^i \partial t^j}, \frac{\partial c}{\partial t^k}\right).
\end{aligned}$$

Next we observe that if we have a map $c : \Omega \to M$, then we can always add an extra parameter t^0 to get a map $\bar{c} : (-\varepsilon, \varepsilon) \times \Omega \to M$ with the property that

$$\frac{\partial \bar{c}}{\partial t^0}|_p = v \in T_pM,$$

where $v \in T_pM$ is any vector and p is any point in the image of c. Using $k = 0$ in the Koszul type formula at p shows that $\frac{\partial^2 c}{\partial t^i \partial t^j}$ is uniquely defined, as our extension is independent of how mixed partials are defined. □

We can now give a local and coordinate dependent definition of mixed partials. As long as the definition gives us properties (1) and (2) the above lemma shows that we have a coordinate independent definition.

Note also that if two different maps $c_1, c_2 : \Omega \to M$ agree on a neighborhood of a point in the domain, then the right-hand side of the Koszul type formula will give the same answer for these two maps. Thus there is no loss of generality in assuming that the image of c lies in a coordinate system.

Theorem 5.1.2 (Existence of mixed partials). *It is possible to define mixed partials in a coordinate system so that (1) and (2) hold.*

Proof. Assume that we have $c : \Omega \to U \subset M$ where U is a coordinate neighborhood. Furthermore, assume that the parameters in use are called s and t. This avoids introducing more indices than necessary. Finally write $c = \left(c^1, \dots, c^n\right)$ using the coordinates on U. The velocity in the s direction is given by

$$\frac{\partial c}{\partial s} = \frac{\partial c^i}{\partial s}\partial_i.$$

This suggests that

$$\begin{aligned}\frac{\partial}{\partial t}\frac{\partial c}{\partial s} &= \frac{\partial}{\partial t}\left(\frac{\partial c^i}{\partial s}\partial_i\right) \\ &= \frac{\partial}{\partial t}\frac{\partial c^i}{\partial s}\partial_i + \frac{\partial c^i}{\partial s}\frac{\partial}{\partial t}\left(\partial_i\right).\end{aligned}$$

To make sense of $\frac{\partial}{\partial t}\left(\partial_i\right)$ we define

$$\frac{\partial X}{\partial t}|_p = \nabla_{\dot{c}(t)}X,$$

where $c\left(t\right) = p$ and X is a vector field defined in a neighborhood of p. With that in mind

$$\frac{\partial}{\partial t}\frac{\partial c}{\partial s} = \frac{\partial^2 c^k}{\partial t \partial s}\partial_k + \frac{\partial c^i}{\partial s}\nabla_{\frac{\partial c}{\partial t}}\partial_i$$

$$
\begin{aligned}
&= \frac{\partial^2 c^k}{\partial t \partial s}\partial_k + \frac{\partial c^i}{\partial s}\frac{\partial c^j}{\partial t}\nabla_{\partial_j}\partial_i \\
&= \frac{\partial^2 c^k}{\partial t \partial s}\partial_k + \frac{\partial c^i}{\partial s}\frac{\partial c^j}{\partial t}\Gamma^k_{ji}\partial_k.
\end{aligned}
$$

Thus we define

$$
\begin{aligned}
\frac{\partial^2 c}{\partial t \partial s} &= \frac{\partial^2 c^k}{\partial t \partial s}\partial_k + \frac{\partial c^i}{\partial s}\frac{\partial c^j}{\partial t}\Gamma^k_{ji}\partial_k \\
&= \left(\frac{\partial^2 c^k}{\partial t \partial s} + \frac{\partial c^i}{\partial s}\frac{\partial c^j}{\partial t}\Gamma^k_{ji}\right)\partial_k.
\end{aligned}
$$

Since $\frac{\partial^2 c^l}{\partial t \partial s}$ is symmetric in s and t by the usual theorem on equality of mixed partials (Clairaut's theorem) and the Christoffel symbol Γ^k_{ji} is symmetric in i and j it follows that (1) holds.

To check the metric property (2) we use that the Christoffel symbols satisfy the metric property (see section 2.4) $\partial_k g_{ij} = \Gamma_{ki,j} + \Gamma_{kj,i}$. With that in mind we calculate

$$
\begin{aligned}
&\frac{\partial}{\partial t} g\left(\frac{\partial c}{\partial s}, \frac{\partial c}{\partial u}\right) \\
&\quad = \frac{\partial}{\partial t}\left(g_{ij}\frac{\partial c^i}{\partial s}\frac{\partial c^j}{\partial u}\right) \\
&\quad = \frac{\partial g_{ij}}{\partial t}\frac{\partial c^i}{\partial s}\frac{\partial c^j}{\partial u} + g_{ij}\frac{\partial^2 c^i}{\partial t \partial s}\frac{\partial c^j}{\partial u} + g_{ij}\frac{\partial c^i}{\partial s}\frac{\partial^2 c^j}{\partial t \partial u} \\
&\quad = g_{ij}\left(\frac{\partial^2 c^i}{\partial t \partial s} + \frac{\partial c^k}{\partial s}\frac{\partial c^l}{\partial t}\Gamma^i_{kl}\right)\frac{\partial c^j}{\partial u} + g_{ij}\frac{\partial c^i}{\partial s}\left(\frac{\partial^2 c^j}{\partial t \partial u} + \frac{\partial c^k}{\partial u}\frac{\partial c^l}{\partial t}\Gamma^j_{kl}\right) \\
&\qquad + \frac{\partial g_{ij}}{\partial t}\frac{\partial c^i}{\partial s}\frac{\partial c^j}{\partial u} - g_{ij}\frac{\partial c^k}{\partial s}\frac{\partial c^l}{\partial t}\frac{\partial c^j}{\partial u}\Gamma^i_{kl} - g_{ij}\frac{\partial c^i}{\partial s}\frac{\partial c^k}{\partial u}\frac{\partial c^l}{\partial t}\Gamma^j_{kl} \\
&\quad = g\left(\frac{\partial^2 c}{\partial t \partial s}, \frac{\partial c}{\partial u}\right) + g\left(\frac{\partial c}{\partial s}, \frac{\partial^2 c}{\partial t \partial u}\right) \\
&\qquad + \frac{\partial g_{ij}}{\partial t}\frac{\partial c^i}{\partial s}\frac{\partial c^j}{\partial u} - \frac{\partial c^k}{\partial s}\frac{\partial c^l}{\partial t}\frac{\partial c^j}{\partial u}\Gamma_{kl,j} - \frac{\partial c^i}{\partial s}\frac{\partial c^k}{\partial u}\frac{\partial c^l}{\partial t}\Gamma^j_{kl,i} \\
&\quad = g\left(\frac{\partial^2 c}{\partial t \partial s}, \frac{\partial c}{\partial u}\right) + g\left(\frac{\partial c}{\partial s}, \frac{\partial^2 c}{\partial t \partial u}\right) \\
&\qquad + \partial_k g_{ij}\frac{\partial c^k}{\partial t}\frac{\partial c^i}{\partial s}\frac{\partial c^j}{\partial u} - \frac{\partial c^i}{\partial s}\frac{\partial c^k}{\partial t}\frac{\partial c^j}{\partial u}\Gamma_{ki,j} - \frac{\partial c^i}{\partial s}\frac{\partial c^j}{\partial u}\frac{\partial c^k}{\partial t}\Gamma_{kj,i} \\
&\quad = g\left(\frac{\partial^2 c}{\partial t \partial s}, \frac{\partial c}{\partial u}\right) + g\left(\frac{\partial c}{\partial s}, \frac{\partial^2 c}{\partial t \partial u}\right).
\end{aligned}
$$

□

In case $M \subset \bar{M}$ it is often convenient to calculate the mixed partials in $\bar{M}$ first and then project them onto M. For each $v \in T_p\bar{M}$, $p \in M$ we use the notation $v = v^\top + v^\perp$ for the decomposition into tangential T_pM and normal $T_p^\perp M$ components.

Proposition 5.1.3 (Mixed partials in submanifolds). *If $c : \Omega \to M \subset \bar{M}$ and $\frac{\partial^2 c}{\partial t^i \partial t^j} \in T_p\bar{M}$ is the mixed partial in $\bar{M}$, then*

$$\left(\frac{\partial^2 c}{\partial t^i \partial t^j}\right)^\top \in T_pM$$

is the mixed partial in M.

Proof. Let $\bar{g}$ be the Riemannian metric in $\bar{M}$ and g its restriction to the submanifold M. We know that $\frac{\partial^2 c}{\partial t^j \partial t^i} \in T\bar{M}$ satisfies

$$2\bar{g}\left(\frac{\partial^2 c}{\partial t^i \partial t^j}, \frac{\partial c}{\partial t^k}\right) = \frac{\partial}{\partial t^i}\bar{g}\left(\frac{\partial c}{\partial t^j}, \frac{\partial c}{\partial t^k}\right) + \frac{\partial}{\partial t^j}\bar{g}\left(\frac{\partial c}{\partial t^k}, \frac{\partial c}{\partial t^i}\right) - \frac{\partial}{\partial t^k}\bar{g}\left(\frac{\partial c}{\partial t^i}, \frac{\partial c}{\partial t^j}\right).$$

As $\frac{\partial c}{\partial t^i}, \frac{\partial c}{\partial t^j}, \frac{\partial c}{\partial t^k} \in TM$ this shows that

$$2\bar{g}\left(\frac{\partial^2 c}{\partial t^i \partial t^j}, \frac{\partial c}{\partial t^k}\right) = \frac{\partial}{\partial t^i}g\left(\frac{\partial c}{\partial t^j}, \frac{\partial c}{\partial t^k}\right) + \frac{\partial}{\partial t^j}g\left(\frac{\partial c}{\partial t^k}, \frac{\partial c}{\partial t^i}\right) - \frac{\partial}{\partial t^k}g\left(\frac{\partial c}{\partial t^i}, \frac{\partial c}{\partial t^j}\right).$$

Next use that $\frac{\partial c}{\partial t^k} \in TM$ to alter the left-hand side to

$$2\bar{g}\left(\frac{\partial^2 c}{\partial t^i \partial t^j}, \frac{\partial c}{\partial t^k}\right) = 2g\left(\left(\frac{\partial^2 c}{\partial t^i \partial t^j}\right)^\top, \frac{\partial c}{\partial t^k}\right).$$

This shows that $\left(\frac{\partial^2 c}{\partial t^i \partial t^j}\right)^\top$ is the correct mixed partial in M. □

5.2 Geodesics

We define the *acceleration* of a curve $c : I \to M$ by the formula

$$\ddot{c} = \frac{d^2 c}{dt^2}.$$

In local coordinates this becomes

$$\ddot{c} = \frac{d^2 c^k}{dt^2}\partial_k + \frac{dc^i}{dt}\frac{dc^j}{dt}\Gamma_{ij}^k \partial_k.$$

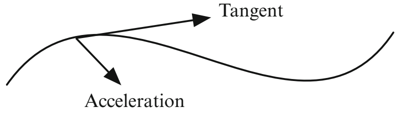


Fig. 5.1 Tangent and acceleration of a curve

A C^∞ curve $c : I \to M$ with vanishing acceleration, $\ddot{c} = 0$, is called a *geodesic* (Fig. 5.1). If c is a geodesic, then the speed $|\dot{c}| = \sqrt{g(\dot{c}, \dot{c})}$ is constant, as

$$\frac{d}{dt} g(\dot{c}, \dot{c}) = 2g(\ddot{c}, \dot{c}) = 0,$$

or phrased differently, it is parametrized proportionally to arc length. If $|\dot{c}| \equiv 1$, one says that c is parametrized by arclength.

Remark 5.2.1. If $r : U \to \mathbb{R}$ is a distance function, then $\nabla_{\partial_r} \partial_r = 0$, where $\partial_r = \nabla r$. The integral curves for $\nabla r = \partial_r$ are consequently geodesics. The theory of geodesics is developed independently of distance functions and ultimately used 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. At first sight, it is not clear that there are going to be any nonconstant geodesics to study on a general Riemannian manifold (although Riemann seems to have taken this for granted). In this section we show that every Riemannian manifold has many nonconstant geodesics. Informally speaking, there is a unique one at each point with a given tangent vector at that point. However, the question of how far it will extend from that point is subtle. To deal with the existence and uniqueness questions, we need to use some information from differential equations.

In local coordinates on $U \subset M$ the equation for a curve to be a geodesic is:

$$\begin{aligned} 0 &= \ddot{c} \\ &= \frac{d^2 c^k}{dt^2} \partial_k + \frac{dc^i}{dt} \frac{dc^j}{dt} \Gamma_{ij}^k \partial_k. \end{aligned}$$

Thus, the curve $c : I \to U$ is a geodesic if and only if the coordinate components c^k satisfy:

$$\ddot{c}^k(t) = -\dot{c}^i(t)\dot{c}^j(t)\Gamma_{ji}^k|_{c(t)}, \ k = 1, \ldots, n.$$

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 the value and first derivative, i.e.,

$$\begin{aligned} c(0) &= q, \\ \dot{c}(0) &= \dot{c}^i(0)\, \partial_i|_q. \end{aligned}$$

But because the system is nonlinear it is not clear that solutions will exist for all t.

The precise statements obtained from the theory of ordinary differential equations give us the following two theorems when we consider geodesics in a chart $U \subset M$.

Theorem 5.2.2 (Local Uniqueness). *Let I_1 and I_2 be intervals with $t_0 \in I_1 \cap I_2$. If $c_1 : I_1 \to U$ and $c_2 : I_2 \to U$ are geodesics with $c_1(t_0) = c_2(t_0)$ and $\dot{c}_1(t_0) = \dot{c}_2(t_0)$, then $c_1|_{I_1 \cap I_2} = c_2|_{I_1 \cap I_2}$.*

Theorem 5.2.3 (Existence). *For each $p \in U$ and $v \in \mathbb{R}^n$, there is a neighborhood V_1 of p, a neighborhood V_2 of v, and an $\varepsilon > 0$ such that for each $q \in V_1$ and $w \in V_2$, there is a geodesic $c_{q,w} : (-\varepsilon, \varepsilon) \to U$ with*

$$c(0) = q,$$
$$\dot{c}(0) = w^i \partial_i|_q.$$

Moreover, the mapping $(q, w, t) \mapsto c_{q,w}(t)$ is C^∞ on $V_1 \times V_2 \times (-\varepsilon, \varepsilon)$.

It is worthwhile to consider what these assertions become in informal terms. The existence statement includes not only short 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.

By relatively simple covering arguments these statements can be extended to geodesics not necessarily contained in a coordinate chart. Let us begin with the uniqueness question:

Lemma 5.2.4 (Global Uniqueness). *Let I_1 and I_2 be open intervals with $t_0 \in I_1 \cap I_2$. If $c_1 : I_1 \to M$ and $c_2 : I_2 \to M$ are geodesics with $c_1(t_0) = c_2(t_0)$ and $\dot{c}_1(t_0) = \dot{c}_2(t_0)$, then $c_1|_{I_1 \cap I_2} = c_2|_{I_1 \cap I_2}$.*

Proof. Define

$$A = \{t \in I_1 \cap I_2 \mid c_1(t) = c_2(t), \dot{c}_1(t) = \dot{c}_2(t)\}.$$

Then $t_0 \in A$. Also, A is closed in $I_1 \cap I_2$ by continuity of c_1, c_2, $\dot{c}_1$, and $\dot{c}_2$. Finally, A is open, by virtue of the local uniqueness statement for geodesics in coordinate charts: if $t_1 \in A$, then choose a coordinate chart U around $c_1(t_1) = c_2(t_1)$. Then $(t_1 - \varepsilon, t_1 + \varepsilon) \subset I_1 \cap I_2$ and $c_i|_{(t_1 - \varepsilon, t_1 + \varepsilon)}$ both have images contained in U. The coordinate uniqueness result then shows that $c_1|_{(t_1 - \varepsilon, t_1 + \varepsilon)} = c_2|_{(t_1 - \varepsilon, t_1 + \varepsilon)}$, so that $(t_1 - \varepsilon, t_1 + \varepsilon) \subset A$. $\square$

The coordinate-free global 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 $c : I \to M$, $0 \in I$, $c(0) = p$, $\dot{c}(0) = v$, 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. 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 $c : (-\varepsilon, \varepsilon) \to M$ with $c(0) = q$ and $\dot{c}(0) = v$. This is an immediate application of the local uniformity part of the differential equations existence statement together with a compactness argument.

The next point to ponder is what happens when the maximal domain of definition is not all of $\mathbb{R}$. For this, assume $c : I = (a, b) \to M$ is a maximal geodesic, where $b < \infty$. Then $c(t)$ must have a specific kind of behavior as t approaches b. If $K \subset M$ is compact, then there is a number $t_K < b$ such that; if $t_K < t < b$, then $c(t) \in M - K$. We say that *c leaves every compact set* as $t \to b$.

To see why c must leave every compact set, suppose K is a compact set it doesn't leave, i.e., there is a sequence $t_1, t_2, \ldots \in I$ with $\lim t_j = b$ and $c(t_j) \in K$ for each j. Since $|\dot{c}|$ is constant the set $\{\dot{c}(t_j) \mid j = 1, \ldots\}$ lies in a compact subset of TM, namely,

$$\widehat{K} = \{v_q \mid q \in K,\ v \in T_qM,\ |v| \le |\dot{c}|\}.$$

Thus there is an $\varepsilon > 0$ such that for each $v_q \in \widehat{K}$, there is a geodesic $c : (-\varepsilon, \varepsilon) \to M$ with $c(0) = q$, $\dot{c}(0) = v$. Now choose t_j such that $b - t_j < \varepsilon/2$. Then $c_{q,v}$ patches together with c to extend c; beginning at t_j continue c by ε, which takes us beyond b, since t_j is within $\varepsilon/2$ of b. This contradicts the maximality of I.

One important consequence of these observations is what happens when M itself is compact:

Corollary 5.2.5. *If M is a compact Riemannian manifold, then for each $p \in M$ and $v \in T_pM$, there is a geodesic $c : \mathbb{R} \to M$ with $c(0) = p$, $\dot{c}(0) = v$. In other words, geodesics exist for all time.*

A Riemannian manifold where all geodesics exist for all time is called *geodesically complete*.

A slightly trickier point is the following: Suppose $c : I \to M$ is a geodesic and $0 \in I$, where I is a bounded interval. Then we would like to say that for $q \in M$ near enough to $c(0)$ and $v \in T_qM$ near enough to $\dot{c}(0)$ there is a geodesic $c_{q,v}$ with q, v as initial position and tangent, respectively, *and* with $c_{q,v}$ defined on an interval almost as big as I. More precisely we have:

Lemma 5.2.6. *Suppose $c : [a, b] \to M$ is a geodesic on a compact interval. There is a neighborhood V in TM of $\dot{c}(0)$ such that if $v \in V$, then there is a geodesic $c_v : [a, b] \to M$ with $\dot{c}_v(a) = v$.*

Proof. A compactness argument allows us to subdivide the interval $a = b_0 < b_1 < \cdots < b_k = b$ in such a way that we have neighborhoods V_i of $\dot{c}(b_i)$ where any geodesic with initial velocity in V_i is defined on $[b_i, b_{i+1}]$. Using that the map $(t, v) \mapsto c_v(t)$ is continuous, where c_v is the geodesic with $\dot{c}_v(0) = v$, we can select a new neighborhood $U_0 \subset V_0$ of $\dot{c}(b_0)$ such that $\dot{c}_v(b_1) \in V_1$ for $v \in U_0$. Next select $U_1 \subset U_0$ so that $\dot{c}_v(b_2) \in V_2$ for $v \in U_1$ etc. In this way we get the desired neighborhood $V = U_{k-1}$ in at most k steps. □

It is easy to check that geodesics in Euclidean space are straight lines. Using this observation it is simple to give examples of the above ideas by taking M to be open subsets of $\mathbb{R}^2$ with its usual metric.

Example 5.2.7. In the punctured plane $\mathbb{R}^2 - \{(0,0)\}$ the unit speed geodesic from $(-1, 0)$ with tangent $(1, 0)$ is defined on $(-\infty, 1)$ only. But nearby geodesics from $(-1, 0)$ 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 figure 5.2.

Example 5.2.8. On the other hand, for the region $\{(x, y) \mid -1 < xy\}$, the curve $t \mapsto (t, 0)$ is a geodesic defined on all of $\mathbb{R}$ that is a limit of unit speed geodesics $t \mapsto (t, -\varepsilon)$, $\varepsilon \to 0$, each of which is defined only on a finite interval. Note that the endpoints of these intervals go to infinity as required by the above lemma. See figure 5.3.

Example 5.2.9. We think of the spheres $S^n(R) = S^n_{R^{-2}} \subset \mathbb{R}^{n+1}$. The acceleration of a curve $c : I \to S^n(R)$ can be computed as the Euclidean acceleration in $\mathbb{R}^{n+1}$ projected onto $S^n(R)$ (see proposition 5.1.3). Thus c is a geodesic if and only if $\ddot{c}$ is normal to $S^n(R)$. This means that $\ddot{c}$ and c should be proportional as vectors. Great circles $c(t) = p\cos(\alpha t) + v\sin(\alpha t)$, where $p, v \in \mathbb{R}^{n+1}$, $|p| = |v| = R$ and $p \perp v$, clearly have this property. Furthermore, since $c(0) = p \in S^n(R)$ and $\dot{c}(0) = \alpha v \in T_pS^n(R)$, we see that there is a geodesic for each initial value problem (see also exercise 1.6.20).

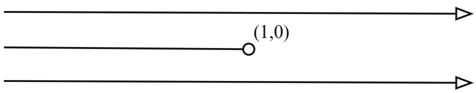


Fig. 5.2 Obstacles to continuing geodesics

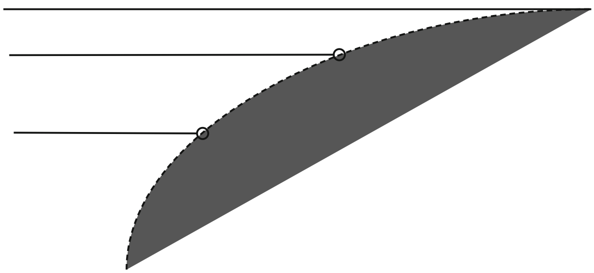


Fig. 5.3 Obstacles to continuing geodesics

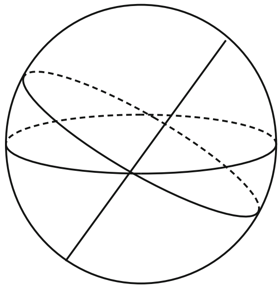


Fig. 5.4 Geodesics on the sphere

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 $2R$. Of course, this simply describes exactly the way in which we draw three-dimensional pictures on paper.

Example 5.2.10. We think of $H^n(R) = S^n_{-R^{-2}} \subset \mathbb{R}^{n,1}$ as in example 1.1.7. In this case the acceleration is also the projection of the acceleration in Minkowski space. In Minkowski space the acceleration in the usual coordinates is the same as the Euclidean acceleration. Thus we just have to find the Minkowski projection onto the hypersurface. By analogy with the sphere, one might guess that the hyperbolas $c(t) = p\cosh(\alpha t) + v\sinh(\alpha t)$, $p, v \in \mathbb{R}^{n,1}$, $|p|^2 = -R^2$, $|v|^2 = R^2$, and $p \perp v$ all in the Minkowski sense, are our geodesics. In fact the ambient acceleration is given by $\ddot{c} = \alpha^2 c$ and $T_p H^n = \{v \mid v \perp p\}$.

This time the geodesics are hyperbolas. On the space itself in Minkowski space, they are, as in the case of spheres, intersections of 2 dimensional subspaces with hyperbolic space. If we resort to the trick of projecting hyperbolic space onto the plane containing the first n coordinates, then the geodesics are hyperbolas whose asymptotes are straight lines through the origin. See also figure 5.5.

Example 5.2.11. 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. However, our Lie group model for the upper half plane does not satisfy this (see section 4.4.2). On the other hand, we did show in proposition 4.4.2 that $\nabla_X X = \frac{1}{2}[X, X] = 0$ when the metric is biinvariant and X is left-invariant. Moreover, all compact Lie groups admit biinvariant metrics (see exercise 1.6.24).

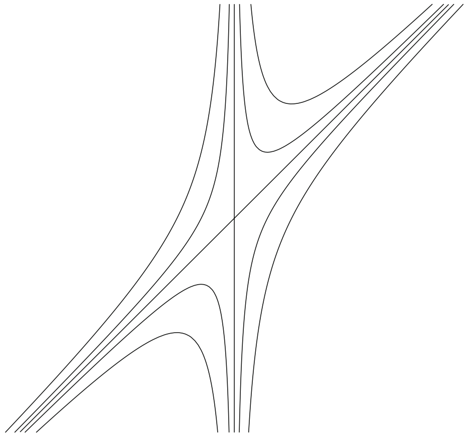


Fig. 5.5 Hyperbolas as geodesics in hyperbolic space

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 speed, i.e., length of velocity. 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 speed, 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.

Recall that a curve $c : [a, b] \to M$ is piecewise C^∞ if c is continuous and if there is a partition $a = a_1 < a_2 < \ldots < a_k = b$ of $[a, b]$ such that $c|_{[a_i, a_{i+1}]}$ is C^∞ for $i = 1, \ldots, k-1$.

Let $c : [a, b] \to M$ be a piecewise C^∞ curve in a Riemannian manifold. Then the *length* $L(c)$ is defined as follows:

$$L(c) = \int_a^b |\dot{c}(t)|\, dt = \int_a^b \sqrt{g\left(\dot{c}(t), \dot{c}(t)\right)} dt.$$

It is clear from the definition that the function $t \mapsto |\dot{c}(t)|$ is integrable in the Riemann (or Lebesgue) sense, so $L(c)$ is a well-defined finite, nonnegative number. The chain and substitution rules show that $L(c)$ is invariant under reparametrization. A curve $c : [a, b] \to M$ is said to be parametrized by arc length if $L(c|_{[a,t]}) = t - a$ for all $t \in [a,b]$, or equivalently, if $|\dot{c}(t)| = 1$ for all $t \in [a,b]$. A regular curve $c : [a,b] \to M$, i.e., the velocity never vanishes, admits a reparametrization to an arclength parametrized curve. To see this define the new parameter as

$$s = \varphi(t) = \int_a^t |\dot{c}(\tau)|\, d\tau.$$

Clearly $\varphi : [a,b] \to [0, L(c)]$ is strictly increasing and piecewise smooth. Thus the curve $c \circ \varphi^{-1} : [0, L(c)] \to M$ is piecewise smooth with unit speed everywhere.

We are now ready to introduce the idea of distance between points. For each pair of points $p, q \in M$ define the path space

$$\Omega_{p,q} = \{c : [0,1] \to M \mid c \text{ is piecewise } C^\infty \text{ and } c(0) = p,\ c(1) = q\}$$

and the distance $d(p,q) = |pq|$ between points $p, q \in M$ as

$$|pq| = \inf\left\{L(c) \mid c \in \Omega_{p,q}\right\}.$$

It follows immediately from this definition that $|pq| = |qp|$ and $|pq| \leq |pr| + |rq|$. The fact that $|pq| = 0$ only when $p = q$ will be established in the proof of theorem 5.3.8. Thus, $|\cdot\cdot|$ satisfies all the properties of a metric. When it is necessary to specify the Riemannian metric we write $|pq|_g$.

As for metric spaces, we have various metric balls

$$B(p,r) = \{x \in M \mid |px| < r\},$$
$$\bar{B}(p,r) = \{x \in M \mid |px| \leq r\}.$$

More generally, we can define the distance between subsets $A, B \subset M$ as

$$d(A,B) = |AB| = \inf\{|pq| \mid p \in A, q \in B\}.$$

Finally, we define

$$B(A,r) = \{x \in M \mid |Ax| < r\},$$
$$\bar{B}(A,r) = D(A,r) = \{x \in M \mid |Ax| \leq r\}.$$

Example 5.3.1. The infimum of curve lengths in the definition of $|pq|$ can fail to be realized. This is illustrated, for instance, by the "punctured plane" $\mathbb{R}^2 - \{(0,0)\}$ with the induced Euclidean metric. The distance $|(-1,0)(1,0)| = 2$, 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 $|pq|$ is realized.

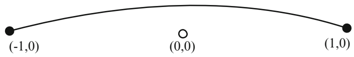


Fig. 5.6 Distance is not realized by a curve

A curve $\sigma \in \Omega_{p,q}$ is a *segment* if $L(\sigma) = |pq|$ and σ is parametrized proportionally to arc length, i.e., $|\dot{\sigma}|$ is constant. We also use the notation $\overline{pq}$ for a specific segment parameterized on $[0, |pq|]$ with $\overline{pq}(0) = p$ and $\overline{pq}\,(|pq|) = q$.

Let us relate these new concepts to our distance functions from section 3.2.2.

Lemma 5.3.2. *If $r : U \to \mathbb{R}$ is a smooth distance function and $U \subset (M, g)$ is open, then the integral curves for ∇r are segments in (U, g). Moreover, if $c \in \Omega_{p,q}(U)$ satisfies $L(c) = r(c(1)) - r(c(0))$, then $c = \sigma \circ \varphi$, where σ is the integral curve for ∇r through $c(0)$ and $\varphi(s) = \int_0^s |\dot{c}|\, dt$.*

Proof. Fix $p, q \in U$ and let $c(t) : [0, b] \to U$ be a curve from p to q. Since $dr(v) = g(\nabla r, v) \leq |v|$ it follows that

$$L(c) = \int_0^b |\dot{c}|\, dt \geq \int_0^b g(\nabla r, \dot{c})\, dt = \int_0^b dr(\dot{c})\, dt = r(q) - r(p).$$

This shows that $|pq| \geq |r(q) - r(p)|$ since $|pq| = |qp|$. If we choose c as an integral curve for ∇r, i.e., $\dot{c} = \nabla r \circ c$, then $dr(\dot{c}) = 1$. Thus $L(c) = |r(q) - r(p)|$. This shows that integral curves must be segments. Moreover, when $L(c) = r(q) - r(p)$, then it follows that $\dot{c} = |\dot{c}|\, \nabla r$. This implies the last claim.

Notice that we only considered curves in U, and thus only established the result for (U, g) and not (M, g). □

Example 5.3.3. In Euclidean space $\mathbb{R}^n$, straight line segments parametrized with constant speed, i.e. curves of the form $t \mapsto p + t \cdot v$, are in fact segments. This follows from lemma 5.3.2 if we use the smooth distance function $r(x) = v \cdot x$, where v is a unit vector. In $\mathbb{R}^n$, each pair of points p, q is joined by a segment $t \mapsto p + t(q - p)$ that is unique up to reparametrization. See also exercise 1.6.19.

Example 5.3.4. Consider $M = S^1$ and $U = S^1 - \{(1, 0)\}$. On U we have the distance function $r(\theta) = \theta$, $\theta \in (0, 2\pi)$. The previous lemma shows that any curve $c(\theta) = (\cos\theta, \sin\theta)$, $\theta \in I \subset (0, 2\pi)$ is a segment in U. If, however, the length of I is $> \pi$, then such curves can clearly not be segments in S^1.

Example 5.3.5. Next we complete our understanding of segments on $S^n(1) \subset \mathbb{R}^{n+1}$ with its standard round metric (see also the proof of theorem 5.5.4 where this is covered in greater generality and detail or exercise 1.6.20). Given two points $p, q \in S^n$ we create a warped product structure

$$ds_n^2 = dr^2 + \sin^2(r)\, ds_{n-1}^2$$

such that: when p, q are antipodal, then they correspond to $r = 0, \pi$; and otherwise $p = (a, x_0) \in (0, \pi) \times S^{n-1}$ and $q = (b, x_0) \in (0, \pi) \times S^{n-1}$. The distance function we use is r and the domain where it is smooth is $U \simeq (0, \pi) \times S^{n-1}$. When the points are antipodal they are joined by several curves of length π. A general curve between these points can always be shortened so it looks like $c : [0, b] \to S^n$, where $c(t) \in U$ for $t \in (0, b)$ and $c(0) = -c(b)$ correspond to the antipodal points where $r = 0, \pi$. Now lemma 5.3.2 shows that $L(c|_{[\epsilon, b-\epsilon]}) \geq |r \circ c(b - \epsilon) - r \circ c(\epsilon)|$. Therefore,

$$L(c) \geq \lim_{\epsilon \to 0} |r \circ c(b - \epsilon) - r \circ c(\epsilon)| = \pi.$$

When the points are not antipodal they lie on a unique integral curve for ∇r which is part of a great circle in U. This segment will again be the shortest among curves in U. However, any curve that leaves U will pass through either $r = 0$ or $r = \pi$. We can argue as with antipodal points that any such curve must have length

$$\geq \min \{r(p) + r(q), \pi - r(p) + \pi - r(q)\} \geq |r(p) - r(q)|.$$

Example 5.3.6. The same strategy can also be used to show that all geodesics in hyperbolic space are segments. See also exercise 1.6.21.

Example 5.3.7. In $\mathbb{R}^2 - \{(0, 0)\}$, as already noted, not every pair of points is joined by a segment.

In section 5.4 we show that segments are always geodesics. Conversely, we show in section 5.5.2 that geodesics are segments when they are sufficiently short. Specifically, if $c : [0, b) \to M$ is a geodesic, then $c|_{[0,\varepsilon]}$ 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 either metrically or geodesically complete. This result explains what is "wrong" with the punctured plane. It also explains why spheres 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, we consider the question of compatibility of topologies.

Theorem 5.3.8. *The metric topology obtained from the distance $|\cdot\cdot|$ on a Riemannian manifold is the same as the manifold topology.*

Proof. Fix $p \in M$ and a coordinate neighborhood U of p such that $x^i(p) = 0$. We assume in addition that $g_{ij}|_p = \delta_{ij}$. On U we have the given Riemannian metric g and also a Euclidean metric g_0 defined by $g_0(\partial_i, \partial_j) = \delta_{ij}$. Thus g_0 is constant and equal to g at p. Finally, after possibly shrinking U, we can further assume that

$$\begin{aligned} U &= B^{g_0}(p, \varepsilon) \\ &= \left\{x \in U \mid |px|_{g_0} < \varepsilon\right\} \\ &= \left\{x \in U \mid \sqrt{(x^1)^2 + \cdots + (x^n)^2} < \varepsilon\right\}. \end{aligned}$$

For $x \in U$ we can compare these two metrics as follows: There are continuous functions: $\lambda, \mu : U \to (0, \infty)$ such that if $v \in T_xM$, then

$$\lambda(x)|v|_{g_0} \leq |v|_g \leq \mu(x)|v|_{g_0}.$$

Moreover, $\lambda(x), \mu(x) \to 1$ as $x \to p$.

Now let $c : [0,1] \to M$ be a curve from p to $x \in U$.

1: If c is a straight line in the Euclidean metric, then it lies in U and

$$\begin{aligned} |px|_{g_0} &= L_{g_0}(c) \\ &= \int_0^1 |\dot{c}|_{g_0}\, dt \\ &\geq \frac{1}{\max \mu(c(t))} \int_0^1 |\dot{c}|_g\, dt \\ &= \frac{1}{\max \mu(c(t))} L_g(c) \\ &\geq \frac{1}{\max \mu(c(t))} |px|_g. \end{aligned}$$

2: If c is a general curve that lies entirely in U, then

$$\begin{aligned} L_g(c) &= \int_0^1 |\dot{c}|_g\, dt \\ &\geq (\min \lambda(c(t))) \int_0^1 |\dot{c}|_{g_0}\, dt \\ &\geq (\min \lambda(c(t))) |px|_{g_0}. \end{aligned}$$

3: If c leaves U, then there will be a smallest t_0 such that $c(t_0) \notin U$, then

$$\begin{aligned} L_g(c) &\geq \int_0^{t_0} |\dot{c}|_g\, dt \\ &\geq (\min \lambda(c(t))) \int_0^{t_0} |\dot{c}|_{g_0}\, dt \\ &\geq (\min \lambda(c(t)))\, \varepsilon \\ &\geq (\min \lambda(c(t))) |px|_{g_0}. \end{aligned}$$

By possibly shrinking U again we can guarantee that $\min \lambda \geq \lambda_0 > 0$ and $\max \mu \leq \mu_0 < \infty$. We have then proven that

$$|px|_g \leq \mu_0 |px|_{g_0}$$

and

$$\lambda_0 |px|_{g_0} \leq \inf L_g(c) = |px|_g .$$

Thus the Euclidean and Riemannian distances are comparable on a neighborhood of p. This shows that the metric topology and the manifold topology (coming from the Euclidean distance) are equivalent. It also shows that $p = q$ if $|pq| = 0$.

Finally note that

$$\lim_{x \to p} \frac{|px|_g}{|px|_{g_0}} = 1$$

since $\lambda(x), \mu(x) \to 1$ as $x \to p$. □

Just as compact Riemannian manifolds are automatically geodesically complete, this theorem also shows that such spaces are metrically complete.

Corollary 5.3.9. *If M is a compact manifold and g is a Riemannian metric on M, then $(M, |\cdot\cdot|_g)$ is a complete metric space, where $|\cdot\cdot|_g$ is the Riemannian distance function determined by g.*

The proof of theorem 5.3.8 also tells us that any curve can be replaced by a regular curve that has almost the same length.

Corollary 5.3.10. *For any $c \in \Omega_{pq}$ and $\epsilon > 0$, there exists a constant speed curve $\bar{c} \in \Omega_{pq}$ with $L(\bar{c}) \leq (1 + \epsilon) L(c)$.*

Proof. First note that it suffices to find a regular curve with the desired property. Next observe that in Euclidean space this can be accomplished by approximating a curve with a possibly shorter polygonal curve. In a Riemannian manifold we can use the same procedure in a chart to approximate a curve by a regular curve. We select a chart and Euclidean metric g_0 as above such that $\lambda_0 L_{g_0}(c) \leq L_g(c) \leq \mu_0 L_{g_0}(c)$ for any curve in the chart. We can then approximate c by a regular curve $\bar{c}$ such that

$$L_g(\bar{c}) \leq \mu_0 L_{g_0}(\bar{c}) \leq \mu_0 L_{g_0}(c) \leq \frac{\mu_0}{\lambda_0} L_g(c) .$$

By shrinking the chart we can make the ratio $\frac{\mu_0}{\lambda_0} < 1 + \epsilon$. Finally we can use compactness to cover the original curve by finitely many such charts to get the desired regular curve. □

Remark 5.3.11. It is possible to develop the theory here using other classes of curves without changing the distance concept. A natural choice would be to expand the class to all absolutely continuous curves. As corollary 5.3.10 indicates we could also have restricted attention to piecewise smooth curves with constant speed.

The *functional distance* d_F between points in a manifold is defined as

$$d_F(p,q) = \sup\{|f(p) - f(q)| \mid f : M \to \mathbb{R} \text{ has } |\nabla f| \le 1 \text{ on } M\}.$$

This distance is always smaller than the arclength distance. One can, however, show as before that it generates the standard manifold topology. In fact, after we have established the existence of smooth distance functions, it will become clear that the two distances are equal provided p and q are sufficiently close to each other.

5.4 First Variation of Energy

In this section we study the arclength functional

$$L(c) = \int_0^1 |\dot{c}|\, dt, \ c \in \Omega_{p,q}$$

in further detail. The minima, if they exist, are pre-segments. That is, they have minimal length, but are not guaranteed to have the correct parametrization. We also saw that in some cases sufficiently short geodesics minimize this functional. One issue with this functional is that it is invariant under change of parametrization. Minima, if they exist, consequently do not come with a fixed parameter. This problem can be overcome by considering the energy functional

$$E(c) = \frac{1}{2}\int_0^1 |\dot{c}|^2\, dt, \ c \in \Omega_{p,q}.$$

This functional measures the total kinetic energy of a particle traveling along the curve. Note that the energy will depend on how the curve is parametrized.

Proposition 5.4.1. *If $\sigma \in \Omega_{p,q}$ is a constant speed curve that minimizes $L : \Omega_{p,q} \to [0,\infty)$, then σ minimizes $E : \Omega_{p,q} \to [0,\infty)$. Conversely, if σ minimizes $E : \Omega_{p,q} \to [0,\infty)$, then it also minimizes $L : \Omega_{p,q} \to [0,\infty)$.*

Proof. The Cauchy-Schwarz inequality for functions tells us that

$$\begin{aligned} L(c) &= \int_0^1 |\dot{c}| \cdot 1 dt \\ &\le \sqrt{\int_0^1 |\dot{c}|^2\, dt}\sqrt{\int_0^1 1^2 dt} \\ &= \sqrt{\int_0^1 |\dot{c}|^2\, dt} \\ &= \sqrt{2E(c)}, \end{aligned}$$

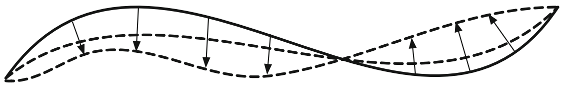


Fig. 5.7 A proper variation

with equality holding if $|\dot{c}|$ is a constant multiple of 1, i.e., c has constant speed. Conversely, when equality holds the speed is forced to be constant. Let $\sigma \in \Omega_{p,q}$ be a curve that has constant speed. If it minimizes L and $c \in \Omega_{p,q}$. Then

$$E(\sigma) = \frac{1}{2}(L(\sigma))^2 \leq \frac{1}{2}(L(c))^2 \leq E(c),$$

so σ also minimizes E.

Conversely, let $\sigma \in \Omega_{p,q}$ minimize E and $c \in \Omega_{p,q}$ be any curve. If c does not have constant speed we can use corollary 5.3.10 to find c_ϵ with constant speed and $L(c_\epsilon) \leq (1+\epsilon) L(c)$ for any $\epsilon > 0$. Then

$$L(\sigma) \leq \sqrt{2E(\sigma)} \leq \sqrt{2E(c_\epsilon)} = L(c_\epsilon) \leq (1+\epsilon) L(c).$$

As $\epsilon > 0$ is arbitrary the result follows. □

The next goal is to show that minima of E must be geodesics. To establish this we have to develop the *first variation formula of energy*. A *variation of a curve* $c : I \to M$ is a family of curves $\bar{c} : (-\varepsilon, \varepsilon) \times [a,b] \to M$, such that $\bar{c}(0,t) = c(t)$ for all $t \in [a,b]$. We say that such a variation is piecewise smooth if it is continuous and $[a,b]$ can be partitioned into intervals $[a_i, a_{i+1}]$, $i = 0, \ldots, m-1$, where $\bar{c} : (-\varepsilon, \varepsilon) \times [a_i, a_{i+1}] \to M$ is smooth. Thus the curves $t \mapsto c_s(t) = \bar{c}(s,t)$ are all piecewise smooth, while the curves $s \mapsto \bar{c}(s,t)$ are smooth. The velocity field for this variation is the field $\frac{\partial \bar{c}}{\partial t}$ which is well-defined on each interval $[a_i, a_{i+1}]$. At the break points $t = a_i$, there are two possible values for this field; a right derivative and a left derivative:

$$\frac{\partial \bar{c}}{\partial t^+}(s, a_i) = \frac{\partial \bar{c}|_{[a_i, a_{i+1}]}}{\partial t}(s, a_i),$$
$$\frac{\partial \bar{c}}{\partial t^-}(s, a_i) = \frac{\partial \bar{c}|_{[a_{i-1}, a_i]}}{\partial t}(s, a_i).$$

The *variational field* is defined as $\frac{\partial \bar{c}}{\partial s}$. This field is well-defined everywhere. It is smooth on each $(-\varepsilon, \varepsilon) \times [a_i, a_{i+1}]$ and continuous on $(-\varepsilon, \varepsilon) \times I$. The special case where $a = 0$, $b = 1$, $\bar{c}(s,0) = p$, and $\bar{c}(s,1) = q$ for all s is of special importance as all of the curves $c_s \in \Omega_{p,q}$. Such variations are called *proper variations* of c (Figure 5.7).

Lemma 5.4.2 (The First Variation Formula). *If* $\bar{c} : (-\varepsilon, \varepsilon) \times [a,b] \to M$ *is a piecewise smooth variation, then*

$$\frac{dE(c_s)}{ds} = -\int_a^b g\left(\frac{\partial^2 \bar{c}}{\partial t^2}, \frac{\partial \bar{c}}{\partial s}\right) dt + g\left(\frac{\partial \bar{c}}{\partial t^-}, \frac{\partial \bar{c}}{\partial s}\right)\bigg|_{(s,b)} - g\left(\frac{\partial \bar{c}}{\partial t^+}, \frac{\partial \bar{c}}{\partial s}\right)\bigg|_{(s,a)}$$
$$+ \sum_{i=1}^{m-1} g\left(\frac{\partial \bar{c}}{\partial t^-} - \frac{\partial \bar{c}}{\partial t^+}, \frac{\partial \bar{c}}{\partial s}\right)\bigg|_{(s,a_i)}.$$

Proof. It suffices to prove the formula for smooth variations as we can otherwise split up the integral into parts that are smooth:

$$E(c_s) = \int_a^b \left|\frac{\partial \bar{c}}{\partial t}\right|^2 dt = \sum_{i=0}^{m-1} \int_{a_i}^{a_{i+1}} \left|\frac{\partial \bar{c}}{\partial t}\right|^2 dt$$

and apply the formula to each part of the variation.

For a smooth variation $\bar{c} : (-\varepsilon, \varepsilon) \times [a, b] \to M$ we have

$$\begin{aligned}
\frac{dE(c_s)}{ds} &= \frac{d}{ds}\frac{1}{2}\int_a^b g\left(\frac{\partial \bar{c}}{\partial t}, \frac{\partial \bar{c}}{\partial t}\right) dt \\
&= \frac{1}{2}\int_a^b \frac{\partial}{\partial s} g\left(\frac{\partial \bar{c}}{\partial t}, \frac{\partial \bar{c}}{\partial t}\right) dt \\
&= \int_a^b g\left(\frac{\partial^2 \bar{c}}{\partial s \partial t}, \frac{\partial \bar{c}}{\partial t}\right) dt \\
&= \int_a^b g\left(\frac{\partial^2 \bar{c}}{\partial t \partial s}, \frac{\partial \bar{c}}{\partial t}\right) dt \\
&= \int_a^b \frac{\partial}{\partial t} g\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial \bar{c}}{\partial t}\right) dt - \int_a^b g\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial^2 \bar{c}}{\partial t^2}\right) dt \\
&= g\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial \bar{c}}{\partial t}\right)\bigg|_a^b - \int_a^b g\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial^2 \bar{c}}{\partial t^2}\right) dt \\
&= -\int_a^b g\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial^2 \bar{c}}{\partial t^2}\right) dt + g\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial \bar{c}}{\partial t}\right)\bigg|_{(s,b)} - g\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial \bar{c}}{\partial t}\right)\bigg|_{(s,a)}.
\end{aligned}$$

□

We can now completely characterize the local minima for the energy functional. The proof in fact characterizes geodesics $c \in \Omega_{p,q}$ as *stationary points* for $E : \Omega_{p,q} \to [0, \infty)$.

Theorem 5.4.3 (Characterization of local minima). *If $c \in \Omega_{p,q}$ is a local minimum for $E : \Omega_{p,q} \to [0, \infty)$, then c is a smooth geodesic.*

Proof. The assumption guarantees that c is a stationary point for the energy functional, i.e.,

$$\frac{dE(c_s)}{ds} = 0$$

for any proper variation of c. This is in fact the only property that we shall use. The trick is to find appropriate variations. If $V(t)$ is any vector field along $c(t)$, i.e., $V(t) \in T_{c(t)}M$, then there is a variation so that $V(t) = \frac{\partial c}{\partial s}|_{(0,t)}$. One such variation can be obtained by declaring the variational curves $s \mapsto c(s,t)$ to be geodesics with $\frac{\partial c}{\partial s}|_{(0,t)} = V(t)$. As geodesics are unique and vary nicely with respect to the initial data, this variation is well-defined and as smooth as V is (see theorems 5.2.2 and 5.2.3). Moreover, if $V(a) = 0$ and $V(b) = 0$, then the variation is proper.

Using such a variational field the first variation formula at $s = 0$ depends only on c itself and the variational field V

$$\begin{aligned}\frac{dE(c_s)}{ds}|_{s=0} &= -\int_a^b g(\ddot{c}, V)\,dt + g\left(\frac{dc}{dt^-}(b), V(b)\right) - g\left(\frac{dc}{dt^+}(a), V(a)\right) \\ &\quad + \sum_{i=1}^{m-1} g\left(\frac{dc}{dt^-}(a_i) - \frac{dc}{dt^+}(a_i), V(a_i)\right) \\ &= -\int_a^b g(\ddot{c}, V)\,dt + \sum_{i=1}^{m-1} g\left(\frac{dc}{dt^-}(a_i) - \frac{dc}{dt^+}(a_i), V(a_i)\right).\end{aligned}$$

We now specify V further. First select $V(t) = \lambda(t)\,\ddot{c}(t)$, where $\lambda(a_i) = 0$ at the break points a_i where c might not be smooth, $\lambda(a) = \lambda(b) = 0$, and $\lambda(t) > 0$ elsewhere. Then

$$\begin{aligned}0 &= \frac{dE(c_s)}{ds}|_{s=0} \\ &= -\int_a^b g(\ddot{c}, \lambda(t)\,\ddot{c})\,dt \\ &= -\int_a^b \lambda(t)\,|\ddot{c}|^2\,dt.\end{aligned}$$

Since $\lambda(t) > 0$ where $\ddot{c}$ is defined it must follow that $\ddot{c} = 0$ at those points. Thus c is a broken geodesic. Next select a new variational field V such that

$$\begin{aligned}V(a_i) &= \frac{dc}{dt^-}(a_i) - \frac{dc}{dt^+}(a_i), \\ V(a) &= V(b) = 0\end{aligned}$$

and otherwise arbitrary, then

$$\begin{aligned}0 &= \frac{dE(c_s)}{ds}|_{s=0} \\ &= \sum_{i=1}^{m-1} g\left(\frac{dc}{dt^-}(a_i) - \frac{dc}{dt^+}(a_i), V(a_i)\right) \\ &= \sum_{i=1}^{m-1} \left|\frac{dc}{dt^-}(a_i) - \frac{dc}{dt^+}(a_i)\right|^2.\end{aligned}$$

This forces

$$\frac{dc}{dt^-}(a_i) = \frac{dc}{dt^+}(a_i)$$

and hence the broken geodesic has the same velocity from the left and right at the places where it is potentially broken. Uniqueness of geodesics (theorem 5.2.2) then shows that c is a smooth geodesic. □

This also shows:

Corollary 5.4.4 (Characterization of segments). *Any piecewise smooth segment is a geodesic.*

While this result shows precisely what the local minima of the energy functional must be it does not guarantee that geodesics are local minima. In Euclidean space all geodesics are minimal as they are the integral curves for globally defined distance functions: $u(x) = v \cdot x$, where v is a unit vector. On the unit sphere, however, no geodesic of length $> \pi$ can be locally minimizing. Such geodesics always form part of a great circle where the complement of the geodesic in the great circle has length $< \pi$, so they can't be absolute minima. One can also easily construct a variation where the nearby curves are all shorter. We shall spend much more time on these issues in the subsequent sections as well as the next chapter. Certainly much more work has to be done before we can characterize what makes geodesics minimal.

5.5 Riemannian Coordinates

The goal of this section is to introduce a natural set of coordinates around each point in a Riemannian manifold. These coordinates will depend on the geometry and also allow us to show the existence of smooth distance functions as well as many other things. They go under the name of exponential or Riemannian normal coordinates. They are normal in the sense of exercise 2.5.20, but have further local and infinitesimal properties. Gauss first introduced such coordinates for surfaces and Riemann in the general context.

5.5.1 The Exponential Map

For a tangent vector $v \in T_pM$, let c_v be the unique geodesic with $c(0) = p$ and $\dot{c}(0) = v$, and $[0, L_v)$ the nonnegative part of the maximal interval on which c is defined. Notice that uniqueness of geodesics implies the *homogeneity property*: $c_{\alpha v}(t) = c_v(\alpha t)$ for all $\alpha > 0$ and $t < L_{\alpha v}$. In particular, $L_{\alpha v} = \alpha^{-1} L_v$. Let $O_p \subset T_pM$ be the set of vectors v such that $1 < L_v$. In other words $c_v(t)$ is defined on $[0, 1]$. The *exponential map* at p, $\exp_p : O_p \to M$, is defined by

$$\exp_p(v) = c_v(1).$$

In exercise 5.9.35 the relationship between the just defined exponential map and the Lie group exponential map is elucidated. Figure 5.8 depicts how radial lines in the tangent space are mapped to radial geodesics in M via the exponential map. The homogeneity property $c_v(t) = c_{tv}(1)$ shows that $\exp_p(tv) = c_v(t)$. Therefore, it is natural to think of $\exp_p(v)$ in a polar coordinate representation, where from p one goes "distance" $|v|$ in the direction of $\frac{v}{|v|}$. This gives the point $\exp_p(v)$, since $c_{\frac{v}{|v|}}(|v|) = c_v(1)$.

The collection of maps, $\exp_p$, can be combined to form a map $\exp : \bigcup O_p \to M$ by setting $\exp|_{O_p} = \exp_p$. This map exp is also called the *exponential map*.

Lemma 5.2.6 shows that the set $O = \bigcup O_p$ is open in TM and theorem 5.2.3 that $\exp : O \to M$ is smooth. Similarly, $O_p \subset T_pM$ is open and $\exp_p : O_p \to M$ is smooth. It is an important property that $\exp_p$ is in fact a local diffeomorphism around $0 \in T_pM$. The details of this are given next.

Proposition 5.5.1. *Let (M, g) be a Riemannian manifold.*

(1) *If $p \in M$, then*

$$D\exp_p : T_0(T_pM) \to T_pM$$

is nonsingular at the origin of T_pM. Consequently, $\exp_p$ is a local diffeomorphism.

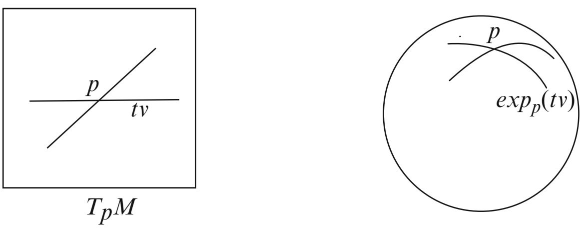


Fig. 5.8 The exponential map at p

(2) *Define* $E : O \to M \times M$ *by* $E(v) = (\pi(v), \exp v)$, *where* $\pi(v)$ *is the base point of* v, *i.e.,* $v \in T_{\pi(v)}M$. *Then for each* $p \in M$ *and* $0_p \in T_pM$,

$$DE : T_{(p,0_p)}(TM) \to T_{(p,p)}(M \times M)$$

is nonsingular. Consequently, E is a diffeomorphism from a neighborhood of the zero section of TM onto an open neighborhood of the diagonal in $M \times M$.

Proof. That the differentials are nonsingular follows from the homogeneity property of geodesics given an important identification of tangent spaces. Let $I_0 : T_pM \to T_0T_pM$ be the canonical isomorphism, i.e., $I_0(v) = \frac{d}{dt}(tv)|_{t=0}$. Recall that if $v \in O_p$, then $c_v(t) = c_{tv}(1)$ for all $t \in [0, 1]$. Thus,

$$\begin{aligned} D\exp_p(I_0(v)) &= \frac{d}{dt}\exp_p(tv)|_{t=0} \\ &= \frac{d}{dt}c_{tv}(1)|_{t=0} \\ &= \frac{d}{dt}c_v(t)|_{t=0} \\ &= \dot{c}_v(0) \\ &= v. \end{aligned}$$

In other words $D\exp_p \circ I_0$ is the identity map on T_pM. This shows that $D\exp_p$ is nonsingular. 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 $T_{(p,0_p)}(TM)$ is also naturally identified with $T_pM \times T_{0_p}(T_pM) \simeq T_pM \times T_pM$. We can think of points in TM as given by (p, v) with $v \in T_pM$. This shows that $E(p, v) = \left(p, \exp_p(v)\right)$. So varying p is just the identity map in the first coordinate, but something unpredictable in the second. While if we fix p and vary v in T_pM, then the first coordinate is fixed and we simply have $\exp_p(v)$ in the second coordinate. This explains what the differential $DE|_{(p,0_p)}$ is. If we consider it as a linear map $T_pM \times T_pM \to T_pM \times T_pM$, then it is the identity on the first factor to the first factor, identically 0 from the second factor to the first, and the identity from the second factor to the second factor as it is $D\exp_p \circ I_{0_p}$. Thus it looks like the nonsingular matrix

$$\begin{bmatrix} I & 0 \\ * & I \end{bmatrix}.$$

Now, the inverse function theorem gives (local) diffeomorphisms via E of neighborhoods of $\left(p, 0_p\right) \in TM$ onto neighborhoods of $(p, p) \in M \times M$. Since E maps the zero section of TM diffeomorphically to the diagonal in $M \times M$ and the zero section is a properly embedded submanifold of TM 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 $M \times M$. □

The largest $\epsilon > 0$ such that

$$\exp_p : B(0, \epsilon) \to M$$

is defined and a diffeomorphism onto its image is called the *injectivity radius* at p and denoted inj_p.

This formalism with the exponential maps yields some results with geometric meaning. First, we get a coordinate system around p by identifying T_pM with $\mathbb{R}^n$ via an isomorphism, and using that the exponential map $\exp_p : T_pM \to M$ is a diffeomorphism on a neighborhood of the origin. Such coordinates are called *exponential* or *Riemannian normal 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 $\mathbb{R}^n$. In section 5.5.3 we show that they are indeed normal in the sense that the Christoffel symbols vanish at p.

The second item of geometric interest is the following idea: On S^2 we know that geodesics are part of great circles. Thus any two points will be joined by both long and short geodesics. What might be hoped is that points that are close together would have a unique short geodesic connecting them. 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 $|v|$ small.

For now we show that:

Corollary 5.5.2. *Let $K \subset (M, g)$ be compact. There exists $\varepsilon > 0$ such that for every $p \in K$, the map $\exp_p : B(0, \varepsilon) \to M$ is defined and a diffeomorphism onto its image.*

Proof. This follows from compactness if we can find $\varepsilon > 0$ such that the statement holds for all p in a neighborhood of a fixed point $x \in M$. This in turn is a consequence of part (2) of proposition 5.5.1. To see this, select a neighborhood U of $(x, 0_x) \in TM$ such that $E : U \to M \times M$ is a diffeomorphism on to its image. Next select a neighborhood $x \in V \subset M$ and a diffeomorphism $F : V \times T_xM \to \pi^{-1}(V) \subset TM$ that is a linear isomorphism $\{p\} \times T_xM \to T_pM$ for each $p \in V$. We can then find $\delta > 0$ so that $F(V \times B(0_x, \delta)) \subset U$. The continuity of the metric on T_pM, $p \in V$, when pulled back to $\{p\} \times T_xM$ via F shows that there is $\varepsilon > 0$ so that $B\left(0_p, \varepsilon\right) \subset F(\{p\} \times B(0_x, \delta))$ for all p in a neighborhood $x \in W \subset V$. Finally, the restriction of E to $B\left(0_p, \varepsilon\right) \subset T_pM$ is a diffeomorphism onto its image in $\{p\} \times M$. This is exactly the map $\exp_p$ if we forget the first factor $\{p\}$. We can then invoke compactness to complete the proof. □

There is a similar construction that leads to a geometric version of the tubular neighborhood theorem from differential topology. Let N be a properly embedded submanifold of M. The normal bundle of N in M is the vector bundle over N consisting of the orthogonal complements of the tangent spaces $T_pN \subset T_pM$,

$$T^{\perp}N = \left\{v \in T_pM \mid p \in N,\ v \in (T_pN)^{\perp} \subset T_pM\right\}.$$

So for each $p \in N$, $T_pM = T_pN \oplus (T_pN)^{\perp}$ is an orthogonal direct sum. Define the *normal exponential map* $\exp^{\perp}$ by restricting exp to $O \cap TN^{\perp}$ and only recording the second factor: $\exp^{\perp} : O \cap TN^{\perp} \to M$. As in part (2) of proposition 5.5.1, one can show:

Corollary 5.5.3. *The map $D\exp^{\perp}$ is nonsingular at 0_p, for all $p \in N$ and 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 when N is a curve in $\mathbb{R}^3$ it looks like a solid tube around the curve.

5.5.2 Short Geodesics Are Segments

We just saw that points that are close together on a Riemannian manifold are connected by a short geodesic, and in fact by exactly one short geodesic. But so far, we don't have any real evidence that such short geodesics are segments. It is the goal of this section to take care of this last piece of the puzzle. 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 first result is the precise statement that we wish to prove in this section.

Theorem 5.5.4. *Let (M, g) be a Riemannian manifold, $p \in M$, and $\varepsilon > 0$ chosen such that*

$$\exp_p : B(0, \varepsilon) \to U \subset M$$

is a diffeomorphism onto its image $U \subset M$. Then $U = B(p, \varepsilon)$ and for each $v \in B(0, \varepsilon)$, the geodesic $\exp_p(tv)$, $t \in [0, 1]$ is the one and only segment with speed $|v|$ from p to $\exp_p v$ in M.

On $U = \exp_p(B(0, \varepsilon))$ we define the function $r(x) = |\exp_p^{-1}(x)|$. That is, r is simply the Euclidean distance function from the origin on $B(0, \varepsilon) \subset T_pM$ in exponential coordinates. This function can be continuously extended to $\bar{U}$ by defining $r(\partial U) = \varepsilon$. We know that $\nabla r = \partial_r = \frac{1}{r}x^i\partial_i$ in Cartesian coordinates on T_pM. In order to prove the theorem we show that this is also the gradient with respect to the general metric g.

Lemma 5.5.5 (The Gauss Lemma). *On (U, g) the function r has gradient $\nabla r = \partial_r$, where $\partial_r = D\exp_p(\partial_r)$.*

Let us see how this implies the theorem.

Proof of Theorem 5.5.4. The proof is analogous to the specific situation on the round sphere covered in example 5.3.5, where $\exp_p : B(0, \pi) \to B(p, \pi)$ is a diffeomorphism.

First observe that in $B(0, \varepsilon) - \{0\}$ the integral curves for ∂_r are the line segments $c(s) = s \cdot \frac{v}{|v|}$ of unit speed. The integral curves for ∂_r on U are then forced to be the unit speed geodesics $c(s) = \exp\left(s \cdot \frac{v}{|v|}\right)$. Thus lemma 5.5.5 implies that r is a distance function on $U - \{p\}$. First note that $U \subset B(p, \varepsilon)$ as the short geodesic that joins p to any point $q \in U$ has length $L < \varepsilon$. To see that this geodesic is the only segment in M, we must show that any other curve from p to q has length $> L$. Suppose we have a curve $c : [0, b] \to M$ from p to q. If $a \in [0, b]$ is the largest value so that $c(a) = p$, then $c|_{[a,b]}$ is a shorter curve from p to q. Next let $b_0 \in (a, b)$ be the first value for which $c(t_0) \notin U$, if such points exist, otherwise $b_0 = b$. The curve $c|_{(a,b_0)}$ lies entirely in $U - \{p\}$ and is shorter than the original curve. It's length is estimated from below as in lemma 5.3.2

$$L\left(c|_{(a,b_0)}\right) = \int_a^{b_0} |\dot{c}|\, dt \geq \int_a^{b_0} dr\,(\dot{c})\, dt = r\,(c\,(b_0))\,,$$

where we used that $r(p) = r\,(c\,(a)) = 0$. If $c\,(b_0) \in \partial U$, then c is not a segment from p to q as it has length $\geq \varepsilon > L$. If $b = b_0$, then $L\left(c|_{(a,b)}\right) \geq r\,(c\,(b)) = L$ and equality can only hold if $\dot{c}\,(t)$ is proportional to ∇r for all $t \in (a, b]$. This shows the short geodesic is a segment and that any other curve of the same length must be a reparametrization of this short geodesic.

Finally we have to show that $B(p, \varepsilon) = U$. We already have $U \subset B(p, \varepsilon)$. Conversely if $q \in B(p, \varepsilon)$ then it is joined to p by a curve of length $< \varepsilon$. The above argument then shows that this curve lies in U. Whence $B(p, \varepsilon) \subset U$. □

Proof of Lemma 5.5.5. We select an orthonormal basis for T_pM and introduce Cartesian coordinates. These coordinates are then also used on U via the exponential map. Denote these coordinates by $(x^1, \ldots, x^n)$ and the coordinate vector fields by $\partial_1, \ldots, \partial_n$. Then

$$\begin{aligned} r^2 &= (x^1)^2 + \cdots + (x^n)^2, \\ \partial_r &= \frac{1}{r} x^i \partial_i. \end{aligned}$$

To show that this is the gradient field for $r(x)$ on (M, g), we must prove that $dr(v) = g(\partial_r, v)$. We already know that

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

but have no knowledge of g, since it is just some abstract metric.

One can show that $dr(v) = g(\partial_r, v)$ by using suitable Jacobi fields for r in place of v. Let us start with $v = \partial_r$. The right-hand side is 1 as the integral curves for ∂_r

are unit speed geodesics. The left-hand side can be computed directly and is also 1. Next, take a rotational field $J = -x^i\partial_j + x^j\partial_i$, $i, j = 1, \ldots, n$, $i < j$. In dimension 2 this is simply the angular field ∂_θ. An immediate calculation shows that the left-hand side vanishes: $dr(J) = 0$. For the right-hand side we first note that J really is a Jacobi field as $L_{\partial_r} J = [\partial_r, J] = 0$. Using that $\nabla_{\partial_r}\partial_r = 0$ we obtain

$$\begin{aligned}
\partial_r g(\partial_r, J) &= g(\nabla_{\partial_r}\partial_r, J) + g(\partial_r, \nabla_{\partial_r} J) \\
&= 0 + g(\partial_r, \nabla_{\partial_r} J) \\
&= g(\partial_r, \nabla_J \partial_r) \\
&= \frac{1}{2} D_J g(\partial_r, \partial_r) \\
&= 0.
\end{aligned}$$

Thus $g(\partial_r, J)$ is constant along geodesics emanating from p. To show that it vanishes first observe that

$$\begin{aligned}
|g(\partial_r, J)| &\leq |\partial_r|\,|J| \\
&= |J| \\
&\leq \left|x^i\right| \left|\partial_j\right| + \left|x^j\right| |\partial_i| \\
&\leq r(x) \left(|\partial_i| + \left|\partial_j\right|\right).
\end{aligned}$$

Continuity of $D\exp_p$ shows that ∂_i, ∂_j are bounded near p. Thus $|g(\partial_r, J)| \to 0$ as $r \to 0$. This forces $g(\partial_r, J) = 0$. Finally, observe that any vector v is a linear combination of ∂_r and rotational fields. This proves the claim. □

The next corollary is an immediate consequence of theorem 5.5.4 and its proof.

Corollary 5.5.6. *If $p \in M$ and $\varepsilon > 0$ is such that $\exp_p : B(0, \varepsilon) \to B(p, \varepsilon)$ is defined and a diffeomorphism, then for each $\delta < \varepsilon$,*

$$\exp_p(B(0, \delta)) = B(p, \delta),$$

and

$$\exp_p(\overline{B}(0, \delta)) = \overline{B}(p, \delta).$$

5.5.3 Properties of Exponential Coordinates

Let us recapture what we have achieved in this section so far. Given $p \in (M, g)$ we found coordinates near p using the exponential map such that the distance function

$r(x) = |px|$ to p has the formula

$$r(x) = \sqrt{(x^1)^2 + \cdots + (x^n)^2}.$$

The Gauss lemma told us that $\nabla r = \partial_r$. This is equivalent to the statement that

$$\exp_p : B(0, \varepsilon) \to B(p, \varepsilon)$$

is a *radial isometry*, i.e.,

$$g\left(D\exp_p(\partial_r), D\exp_p(v)\right) = g_p(\partial_r, v).$$

To see this note that being a radial isometry can be expressed as

$$\tfrac{1}{r} g_{ij} x^i v^j = g\left(\tfrac{1}{r} x^i \partial_i, v^j \partial_j\right) = \tfrac{1}{r} \delta_{ij} x^i v^j.$$

Since $dr(v) = \frac{1}{r}\delta_{ij}x^i v^j$ this is equivalent to the assertion $\nabla r = \partial_r = \frac{1}{r}x^i \partial_i$.

We can rewrite this as the condition

$$g_{ij} x^j = \delta_{ij} x^j.$$

This relationship, as we shall see, fixes the behavior of g_{ij} around p up to first-order and shows that the coordinates are normal.

Lemma 5.5.7. *In exponential coordinates*

$$g_{ij} = \delta_{ij} + O\left(r^2\right).$$

Proof. The fact that $g_{ij}|_p = \delta_{ij}$ follows from taking one partial derivative on both sides of the formula $g_{ij}x^j = \delta_{ij}x^j$

$$\begin{aligned}
\delta_{ik} &= \delta_{ij}\partial_k x^j \\
&= \partial_k \sum_j g_{ij} x^j \\
&= \left(\partial_k g_{ij}\right) x^j + g_{ij}\partial_k x^j \\
&= \left(\partial_k g_{ij}\right) x^j + g_{ik}.
\end{aligned}$$

As $x^j(p) = 0$, the claim follows.

Taking two partial derivatives on both sides gives

$$\begin{aligned}
0 &= \partial_l\left(\left(\partial_k g_{ij}\right) x^j\right) + \partial_l g_{ik} \\
&= \left(\partial_l \partial_k g_{ij}\right) x^j + \partial_k g_{ij}\partial_l x^j + \partial_l g_{ik} \\
&= \left(\partial_l \partial_k g_{ij}\right) x^j + \partial_k g_{il} + \partial_l g_{ik}.
\end{aligned}$$

Evaluating at p we obtain

$$\partial_k g_{il}|_p + \partial_l g_{ik}|_p = 0.$$

The claim that $\partial_k g_{ij}|_p = 0$ follows from evaluating the general formula

$$2\partial_k g_{ij} = \left(\partial_k g_{ij} + \partial_j g_{ik}\right) + \left(\partial_k g_{ji} + \partial_i g_{jk}\right) - \left(\partial_i g_{kj} + \partial_j g_{ki}\right)$$

at p. □

Since r is a distance function whose level sets near p are S^{n-1} we obtain a polar coordinate representation $g = dr^2 + g_r$, where g_r is the restriction of g to S^{n-1}. The Euclidean metric looks like $\delta_{ij} = dr^2 + r^2 ds_{n-1}^2$, where ds_{n-1}^2 is the canonical metric on S^{n-1}. Since these two metrics agree up to first-order it follows that

$$\lim_{r\to 0} g_r = 0,$$
$$\lim_{r\to 0} \left(\partial_r g_r - \partial_r \left(r^2 ds_{n-1}^2\right)\right) = 0.$$

As $\partial_r g_r = 2\,\mathrm{Hess}\, r$ this implies

$$\lim_{r\to 0} \left(\mathrm{Hess}\, r - r ds_{n-1}^2\right) = \lim_{r\to 0} \left(\mathrm{Hess}\, r - \frac{1}{r} g_r\right) = 0.$$

Theorem 5.5.8 (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. We use exponential coordinates around $p \in M$ and the asymptotic behavior of g_r and $\mathrm{Hess}\, r$ near p that was just established. The constant curvature assumption implies that the radial curvature equation (see proposition 3.2.11) can be written as

$$\nabla_{\partial_r} \mathrm{Hess}\, r + \mathrm{Hess}^2\, r = -k g_r,$$
$$\lim_{r\to 0} \mathrm{Hess}\, r = 0.$$

If we think of g_r as given, then this equation has a unique solution. However, $\frac{\mathrm{sn}_k'(r)}{\mathrm{sn}_k(r)} g_r$ also solves this equation since

$$\begin{aligned}
&\nabla_{\partial_r}\left(\frac{\mathrm{sn}_k'(r)}{\mathrm{sn}_k(r)} g_r\right) + \left(\frac{\mathrm{sn}_k'(r)}{\mathrm{sn}_k(r)}\right)^2 g_r \\
&= \left(\frac{\mathrm{sn}_k'(r)}{\mathrm{sn}_k(r)}\right)' g_r + \left(\frac{\mathrm{sn}_k'(r)}{\mathrm{sn}_k(r)}\right)^2 g_r + \frac{\mathrm{sn}_k'(r)}{\mathrm{sn}_k(r)} \nabla_{\partial_r} g_r
\end{aligned}$$

$$= -kg_r - \frac{\operatorname{sn}_k'(r)}{\operatorname{sn}_k(r)} \nabla_{\partial_r} dr^2, \text{ since } 0 = \nabla_{\partial_r} g_r + \nabla_{\partial_r} dr^2,$$

$$= -kg_r - \frac{\operatorname{sn}_k'(r)}{\operatorname{sn}_k(r)} \left(\operatorname{Hess} r\,(\partial_r, \cdot)\, dr + dr \operatorname{Hess} r\,(\partial_r, \cdot) \right)$$

$$= -kg_r.$$

Using that $\operatorname{Hess} r = \frac{\operatorname{sn}_k'(r)}{\operatorname{sn}_k(r)} g_r$ together with $g = dr^2 + g_r$ implies that

$$f_k(r) = \begin{cases} \frac{1}{2} r^2, & \text{when } k = 0 \\ \frac{1}{k} - \frac{1}{k} \operatorname{cs}_k(r), & \text{when } k \neq 0 \end{cases}$$

satisfies $\operatorname{Hess} f_k = (1 - kf_k)\, g$. The result then follows from corollary 4.3.4. □

Exercises 3.4.20 and 3.4.21 explain how this theorem was proved classically and exercise 4.7.21 offers an approach focusing on conformal flatness.

Remark 5.5.9. Some remarks are in order in regards to the above proof. First note that neither of the two systems

$$L_{\partial_r} g_r = 2 \operatorname{Hess} r,$$
$$\nabla_{\partial_r} \operatorname{Hess} r + \operatorname{Hess}^2 r = -kg_r$$

or

$$L_{\partial_r} g_r = 2 \operatorname{Hess} r,$$
$$L_{\partial_r} \operatorname{Hess} r - \operatorname{Hess}^2 r = -kg_r$$

have a unique solution with the initial conditions that both g_r and $\operatorname{Hess} r$ vanish at $r = 0$. In fact there is also a trivial solution where both $g_r = 0$ and $\operatorname{Hess} r = 0$.

Moreover, it is also not clear that $\operatorname{Hess} r = \frac{\operatorname{sn}_k'(r)}{\operatorname{sn}_k(r)} g_r$ solves

$$L_{\partial_r} \operatorname{Hess} r - \operatorname{Hess}^2 r = -kg_r$$

unless we know in advance that

$$L_{\partial_r} g_r = 2 \frac{\operatorname{sn}_k'(r)}{\operatorname{sn}_k(r)} g_r.$$

Finally, note that the initial value problem

$$L_{\partial_r} g_r = 2 \frac{\operatorname{sn}_k'(r)}{\operatorname{sn}_k(r)} g_r, \ \lim_{r \to 0} g_r = 0$$

has infinitely many solutions $\lambda\,\mathrm{sn}_k^2(r)\,ds_{n-1}^2$, $\lambda \in \mathbb{R}$. Although only one of these with give a smooth metric at p.

5.6 Riemannian Isometries

We are now ready to explain the key properties of Riemannian isometries. After a general discussion of Riemannian isometries we classify all geodesically complete simply connected Riemannian manifolds with constant sectional curvature.

5.6.1 *Local Isometries*

A map $F : (M, g_M) \to (N, g_N)$ is a *local Riemannian isometry* if for each $p \in M$ the differential $DF_p : T_pM \to T_{F(p)}N$ is a linear isometry. A special and trivial example of such a map is a local coordinate system $\varphi : U \to \Omega \subset \mathbb{R}^n$ where we use the induced metric g on U and its coordinate representation $\left(\varphi^{-1}\right)^* g = g_{ij}dx^i dx^j$ on Ω.

Proposition 5.6.1. *Let $F : (M, g_M) \to (N, g_N)$ be a local Riemannian isometry.*

(1) *F maps geodesics to geodesics.*
(2) *$F \circ \exp_p(v) = \exp_{F(p)} \circ DF_p(v)$ when $\exp_p(v)$ is defined. In other words*

$$\begin{array}{ccc} T_pM \supset O_p & \xrightarrow{DF} & O_{F(p)} \subset T_{F(p)}N \\ \exp_p \downarrow & & \downarrow \exp_{F(p)} \\ M & \xrightarrow{F} & N \end{array}$$

(3) *F is distance decreasing.*
(4) *If F is also a bijection, then it is distance preserving.*

Proof. (1) The geodesic equation depends on the metric and its first derivatives in a coordinate system. A local Riemannian isometry preserves the metric and is a local diffeomorphism. So it induces coordinates on N with the same metric coefficients. In particular, it must take geodesics to geodesics.

(2) If $\exp_p(v)$ is defined, then $t \mapsto \exp_p(tv)$ is a geodesic. Thus $t \mapsto F\left(\exp_p(tv)\right)$ is also a geodesic. Since

$$\begin{aligned} \frac{d}{dt}F\left(\exp_p(tv)\right)|_{t=0} &= DF\left(\frac{d}{dt}\exp_p(tv)\,|_{t=0}\right) \\ &= DF(v), \end{aligned}$$

we have that $F\left(\exp_p(tv)\right) = \exp_{F(p)}(tDF(v))$. Setting $t = 1$ then proves the claim.

(3) This is also obvious as F must preserve the length of curves.
(4) Both F and F^{-1} are distance decreasing so they must both be distance preserving.

□

This proposition quickly yields two important results for local Riemannian isometries. The first proposition establishes the important uniqueness for Riemannian isometries and thus quickly allows us to conclude that the groups of isometries on space forms discussed in section 1.3.1 are the isometry groups.

Proposition 5.6.2 (Uniqueness of Riemannian Isometries). *Consider two local Riemannian isometries $F, G : (M, g_M) \to (N, g_N)$. If M is connected, $F(p) = G(p)$, and $DF_p = DG_p$, then $F = G$ on M.*

Proof. Let

$$A = \{x \in M \mid F(x) = G(x),\ DF_x = DG_x\}.$$

We know that $p \in A$ and that A is closed. Property (2) from the above proposition tells us that

$$\begin{aligned} F \circ \exp_x(v) &= \exp_{F(x)} \circ DF_x(v) \\ &= \exp_{G(x)} \circ DG_x(v) \\ &= G \circ \exp_x(v), \end{aligned}$$

if $x \in A$. Since $\exp_x$ maps onto a neighborhood of x it follows that some neighborhood of x also lies in A. This shows that A is open and hence all of M by connectedness. □

Proposition 5.6.3. *Let $F : (M, g_M) \to (N, g_N)$ be a Riemannian covering map. (M, g_M) is geodesically complete if and only if (N, g_N) is geodesically complete.*

Proof. Let $c : (-\varepsilon, \varepsilon) \to N$ be a geodesic with $c(0) = p$ and $\dot{c}(0) = v$. For any $\bar{p} \in F^{-1}(p)$ there is a unique lift $\bar{c} : (-\varepsilon, \varepsilon) \to M$, i.e., $F \circ \bar{c} = c$, with $\bar{c}(0) = \bar{p}$. Since F is a local isometry, the inverse is locally defined and also an isometry. Thus $\bar{c}$ is also a geodesic.

If we assume N is geodesically complete, then c and also $\bar{c}$ will exist for all time. As all geodesics in M must be of the form $\bar{c}$ this shows that all geodesics in M exist for all time.

Conversely, when M is geodesically complete, then $\bar{c}$ can be extended to be defined for all time. Then $F \circ \bar{c}$ is a geodesic defined for all time that extends c. Thus N is geodesically complete. □

Lemma 5.6.4. *Let $F : (M, g_M) \to (N, g_N)$ be a local Riemannian isometry. If M is geodesically complete and N is connected, then F is a Riemannian covering map.*

Proof. Fix $q \in N$ and assume that $\exp_q : B(0,\varepsilon) \to B(q,\varepsilon)$ is a diffeomorphism. We claim that $F^{-1}(B(q,\varepsilon))$ is evenly covered by the sets $B(p,\varepsilon)$ where $F(p) = q$. Geodesic completeness of M guarantees that $\exp_p : B(0,\varepsilon) \to B(p,\varepsilon)$ is defined and property (2) that

$$F \circ \exp_p(v) = \exp_q \circ DF_p(v)$$

for all $v \in B(0,\varepsilon) \subset T_pM$. As $\exp_q : B(0,\varepsilon) \to B(q,\varepsilon)$ and $DF_p : B(0,\varepsilon) \to B(0,\varepsilon)$ are diffeomorphisms it follows that $F \circ \exp_p : B(0,\varepsilon) \to B(q,\varepsilon)$ is a diffeomorphism. Thus each of the maps $\exp_p : B(0,\varepsilon) \to B(p,\varepsilon)$ and $F : B(p,\varepsilon) \to B(q,\varepsilon)$ are diffeomorphisms as well.

Next we need to make sure that

$$F^{-1}(B(q,\varepsilon)) = \bigcup_{F(p)=q} B(p,\varepsilon).$$

If $x \in F^{-1}(B(q,\varepsilon))$, then we can join q and $F(x)$ by a unique geodesic $c(t) = \exp_q(tv)$, $v \in B(0,\varepsilon)$. Geodesic completeness of M implies that there is a geodesic $\sigma : [0,1] \to M$ with $\sigma(1) = x$ and $DF_x(\dot{\sigma}(1)) = \dot{c}(1)$. Since $F \circ \sigma$ is a geodesic with the same initial values as c at $t = 1$ we must have $F(\sigma(t)) = c(t)$ for all t. Since $q = c(0)$ we have proven that $F(\sigma(0)) = q$ and hence that $x \in B(\sigma(0),\varepsilon)$.

Finally, we need to show that F is surjective. Clearly $F(M) \subset N$ is open. The above argument also shows that it is closed. To see this, consider a sequence $q_i \in F(M)$ that converges to $q \in N$. We can use corollary 5.5.2 to find an $\epsilon > 0$ such that $\exp_x : B(0,\epsilon) \to B(x,\epsilon)$ is a diffeomorphism for all $x \in \{q, q_1, \ldots, q_k, \ldots\}$. For k sufficiently large it follows that $q \in B(q_k,\epsilon)$. This shows that $q \in F(M)$, since we proved that $B(q_k,\epsilon) \subset F(M)$. □

If $S \subset \mathrm{Iso}(M,g)$ is a set of isometries, then the *fixed point set* of S is defined as those points in M that are fixed by all isometries in S

$$\mathrm{Fix}(S) = \{x \in M \mid F(x) = x \text{ for all } F \in S\}.$$

While the fixed point set for a general set of diffeomorphisms can be quite complicated, the situation for isometries is much more manageable. A submanifold $N \subset (M,g)$ is said to be *totally geodesic* if for each $p \in N$ a neighborhood of $0 \in T_pN \subset T_pM$ is mapped into N via the exponential map $\exp_p$ for M. This means that geodesics in N are also geodesics in M and conversely that any geodesic in M which is tangent to N at some point must lie in N for a short time.

Proposition 5.6.5. *If $S \subset \mathrm{Iso}(M,g)$ is a set of isometries, then each connected component of the fixed point set is a totally geodesic submanifold.*

Proof. Let $p \in \mathrm{Fix}(S)$ and $V \subset T_pM$ be the *Zariski tangent space*, i.e., the set of vectors fixed by the linear isometries $DF_p : T_pM \to T_pM$, where $F \in S$. Note that each such F fixes p so we know that $DF_p : T_pM \to T_pM$. If $v \in V$, then

$t \mapsto \exp_p(tv)$ must be fixed by each of the isometries in S as the initial position and velocity is fixed by these isometries. Thus $\exp_p(tv) \in \text{Fix}(S)$ as long as it is defined. This shows that $\exp_p : V \to \text{Fix}(S)$.

Next let $\varepsilon > 0$ be chosen so that $\exp_p : B(0,\varepsilon) \to B(p,\varepsilon)$ is a diffeomorphism. If $q \in \text{Fix}(S) \cap B(p,\varepsilon)$, then the unique geodesic $c : [0,1] \to B(p,\varepsilon)$ from p to q has the property that its endpoints are fixed by each $F \in S$. Now $F \circ c$ is also a geodesic from p to q which in addition lies in $B(p,\varepsilon)$ as the length is unchanged. Thus $F \circ c = c$ and hence c lies in $\text{Fix}(S) \cap B(p,\varepsilon)$.

This shows that $\exp_p : V \cap B(0,\varepsilon) \to \text{Fix}(S) \cap B(p,\varepsilon)$ is a bijection and proves the lemma. □

Remark 5.6.6. Note that if $DF_p : T_pM \to T_pM$ is orientation preserving for $p \in \text{Fix}(F)$, then the Zariski tangent space at p must have even codimension as the $+1$-eigenspace of an element in $\text{SO}(n)$ has even codimension. In particular each component of $\text{Fix}(F)$ has even codimension.

5.6.2 Constant Curvature Revisited

We just saw that isometries are uniquely determined by their differential. What about the existence question? Given any linear isometry $L : T_pM \to T_qN$, is there an isometry $F : M \to N$ such that $DF_p = L$? In case $M = N$, 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 $F : M \to M$ such that $F(\pi) = \tilde{\pi}$. But this would imply that M has constant sectional curvature. Therefore, the problem cannot be solved in general. From our knowledge of $\text{Iso}(S^n_k)$ it follows that these spaces have enough isometries so that any linear isometry $L : T_pS^n_k \to T_qS^n_k$ can be extended to a global isometry $F : S^n_k \to S^n_k$ with $DF_p = L$ (see section 1.3.1). We show below that in a suitable sense these are the only spaces with this property. However, there are other interesting results in this direction for other spaces (see section 10.1.2).

Theorem 5.6.7. *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 local Riemannian isometry called the* monodromy map *$F : M \to S^n_k$ with $DF_p = L$. Furthermore, this map is a diffeomorphism if (M,g) is geodesically complete.*

Before giving the proof, let us look at some examples.

Example 5.6.8. Suppose we have an immersion $M^n \to S^n_k$. Then F will be one of the maps described in the theorem if we use the pullback metric on M. Such maps can fold in wild ways when $n \geq 2$ and need not resemble covering maps in any way whatsoever.

Example 5.6.9. If $U \subset S_k^n$ is an open disc with smooth boundary, then one can easily construct a diffeomorphism $F : M = S_k^n - \{p\} \to S_k^n - U$. Near the missing point in M the metric will necessarily look pretty awful, although it has constant curvature.

Example 5.6.10. If $M = \mathbb{RP}^n$ or $(\mathbb{R}^n - \{0\}) /$antipodal map, then M is not simply connected and does not admit an immersion into S^n.

Example 5.6.11. If M is the universal covering of $S^2 - \{\pm p\}$, then the monodromy map is not one-to-one. In fact it must be the covering map $M \to S^2 - \{\pm p\}$.

Corollary 5.6.12. *If M is a closed simply connected manifold with constant curvature k, then $k > 0$ and $M = S^n$. Thus, $S^p \times S^q, \mathbb{CP}^n$ do not admit any constant curvature metrics.*

Corollary 5.6.13. *If M is geodesically complete and noncompact with constant curvature k, then $k \leq 0$ and the universal covering is diffeomorphic to $\mathbb{R}^n$. 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 exercise 6.7.4 when M is complete.

Proof of Theorem 5.6.7. We know from theorem 5.5.8 that given $x \in M$ sufficiently small balls $B(x,r)$ are isometric to balls $B(\bar{x},r) \subset S_k^n$. Furthermore, by composing with elements of $\mathrm{Iso}\left(S_k^n\right)$ (these are calculated in sections 1.3.1) we have: if $q \in B(x,r)$, $\bar{q} \in S_k^n$, and $L : T_qU \to T_{\bar{q}}S_k^n$ is a linear isometry, then there is a unique isometric embedding: $F : B(x,r) \to S_k^n$, where $F(q) = \bar{q}$ and $DF|_q = L$. Note that when $k \leq 0$, all metric balls in S_k^n are convex, while when $k > 0$ we need their radius to be $< \frac{\pi}{2\sqrt{k}}$ for this to be true. So for small radii metric balls in M are either disjoint or have connected intersection. For the remainder of the proof assume that all such metric balls are chosen to be isometric to convex balls in the space form.

The construction of F proceeds basically in the same way one does analytic continuation on simply connected domains. Fix base points $p \in M$, $\bar{p} \in S_k^n$ and a linear isometry $L : T_pM \to T_{\bar{p}}S_k^n$. Next, let $x \in M$ be an arbitrary point. If $c \in \Omega_{p,x}$ is a curve from p to x in M, then we can cover c by a string of balls $B(p_i,r)$, $i = 0,\ldots,m$, where $p = p_0$, $x = p_m$, and $B(p_{i-1},r) \cap B(p_i,r) \neq \varnothing$. Define $F_0 : B(p_0,r) \to S_k^n$ so that $F(p) = \bar{p}$ and $DF_0|_{p_0} = L$. Then define $F_i : B(p_i,r) \to S_k^n$ successively to make it agree with F_{i-1} on $B(p_{i-1},r) \cap B(p_i,r)$ (this just requires their values and differentials agree at one point since the intersection is connected). Define a function $G : \Omega_{p,x} \to S_k^n$ by $G(c) = F_m(x)$. We have to check that it is well-defined in the sense that it doesn't depend on our specific way of covering the curve. This is easily done by selecting a different covering and then showing that the set of values in $[0,1]$ where the two choices agree is both open and closed as in proposition 5.6.2.

If $\bar{c} \in \Omega_{p,x}$ is sufficiently close to c, then it lies in such a covering of c, but then it is clear that $G(c) = G(\bar{c})$. This implies that G is locally constant. In particular, G

has the same value on all curves in $\Omega_{p,x}$ that are homotopic to each other. Simple-connectivity then implies that G is constant on $\Omega_{p,x}$. This means that $F(x)$ becomes well-defined and a Riemannian isometry.

If M is geodesically complete we know from lemma 5.6.4 that F has to be a covering map. As S^n_k is simply connected it must be a diffeomorphism. □

We can now give the classification of complete simply connected Riemannian manifolds with constant curvature. Killing first proved the result assuming in effect that the manifold has an $\varepsilon > 0$ such that for all p the map $\exp_p : B(0,\varepsilon) \to B(p,\varepsilon)$ is a diffeomorphism, i.e., the manifold has a uniform lower bound for the injectivity radius. Hopf realized that it was sufficient to assume that the manifold was geodesically complete. Since metric completeness easily implies geodesic completeness this is clearly the best result one could have expected at the time.

Corollary 5.6.14 (Classification of Constant Curvature Spaces, Killing, 1893 and 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^n_k.*

This result shows how important the geodesic 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 hence carry incomplete metrics with zero curvature. Carrying a geodesically 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.

5.6.3 Metric Characterization of Maps

For a Riemannian manifold (M, g) we denote the corresponding metric space by $\left(M, |\cdot\cdot|_g\right)$ or simply $(M, |\cdot\cdot|)$ if only one metric is in play. It is natural to ask whether one can somehow recapture the Riemannian metric g from the distance $|\cdot\cdot|_g$. If for instance $v, w \in T_pM$, then we would like to be able to compute $g(v, w)$ from knowledge of $|\cdot\cdot|_g$. First note that it suffices to compute the length of vectors as the inner product $g(v, w)$ can be computed by polarization:

$$g(v, w) = \frac{1}{2}\left(|v + w|^2 - |v|^2 - |w|^2\right).$$

One way of computing $|v|$ from the metric is by taking a curve α such that $\dot{\alpha}(0) = v$ and observe that

$$|v| = \lim_{t\to 0}\frac{|\alpha(t)\alpha(0)|}{t}.$$

Thus, g really can be found from $|\cdot\cdot|_g$ by using 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 $\left(M, |\cdot\cdot|_g\right)$.

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 $B(q, \mathrm{inj}(q)) \supset U$ (see corollary 5.5.2 and theorem 5.5.4). Thus, for each $q \in U$ the distance function $r_q(x) = |qx|$ is smooth on $U - \{q\}$. Now choose $q_1, \ldots, q_n \in U - \{p\}$, where $n = \dim M$. If the vectors $\nabla r_{q_1}(p), \ldots, \nabla r_{q_n}(p) \in T_pM$ are linearly independent, the inverse function theorem tells us that $\varphi = \left(r_{q_1}, \ldots, r_{q_n}\right)$ 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 obtained from suitable 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.

We just saw that bijective Riemannian isometries are distance preserving. The next result shows that the converse is also true.

Theorem 5.6.15 (Myers and Steenrod, 1939). *If (M, g_M) and (N, g_N) are Riemannian manifolds and $F : M \to N$ a bijection, then F is a Riemannian isometry if F is distance preserving, i.e., $|F(p)F(q)|_{g_N} = |pq|_{g_M}$ for all $p, q \in M$.*

Proof. Let F be distance preserving. First we show that F is differentiable. Fix $p \in M$ and let $q = F(p)$. Near q introduce distance coordinates $\left(r_{q_1}, \ldots, r_{q_n}\right)$ and find p_i such that $F(p_i) = q_i$. Now observe that

$$\begin{aligned} r_{q_i} \circ F(x) &= |F(x)q_i| \\ &= |F(x)F(p_i)| \\ &= |xp_i| \,. \end{aligned}$$

Since $|pp_i| = |qq_i|$, we can assume that the q_is and p_is are chosen such that $r_{p_i}(x) = |xp_i|$ are smooth at p. Thus, $\left(r_{q_1}, \ldots, r_{q_n}\right) \circ F$ is smooth at p, showing that F must be smooth at p.

To show that F is a Riemannian isometry it suffices to check that $|DF(v)| = |v|$ for all tangent vectors $v \in TM$. For a fixed $v \in T_pM$ let $c(t) = \exp_p(tv)$. For small t we know that c is a constant speed segment. Thus, for small t, s we can conclude

$$|t - s| \cdot |v| = |c(t)c(s)|_{g_M} = |F(c(t))\, F(c(s))|_{g_N} \,,$$

implying

$$\begin{aligned}|DF(v)| &= \left|\frac{d(F\circ c)}{dt}\right|_{t=0} \\ &= \lim_{t\to 0}\frac{|F(c(t))\,F(c(0))|_{g_N}}{|t|} \\ &= \lim_{t\to 0}\frac{|c(t)c(0)|_{g_M}}{|t|} \\ &= |\dot{c}(0)| \\ &= |v|\,.\end{aligned}$$

□

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 $F : \left(\bar{M}, g_{\bar{M}}\right) \to (M, g_M)$ be a function. We call F a *submetry* if for every $\bar{p} \in \bar{M}$ there is an $r > 0$ such that $F(B(\bar{p}, \varepsilon)) = B(F(\bar{p}), \varepsilon)$ for all $\varepsilon \le r$. Submetries are locally distance nonincreasing and hence also continuous. In addition, we have that the composition of submetries (or Riemannian submersions) are again submetries (or Riemannian submersions).

Theorem 5.6.16 (Berestovskii, 1995). *If $F : \left(\bar{M}, g_{\bar{M}}\right) \to (M, g_M)$ is a surjective submetry, then F is a C^1 Riemannian submersion.*

Proof. We use the notation $\bar{p} \in F^{-1}(p)$ for points in the pre-image. The goal is to show that we have unique horizontal lifts of vectors in M that vary continuously with $\bar{p}$.

Assume that $r < \mathrm{inj}_p, \mathrm{inj}_{\bar{p}}$ in the submersion property so that all geodesic segments are unique between the end points

The submetry property shows: If $|p\,q| < r$, then for each $\bar{p} \in F^{-1}(p)$ there exists a unique $\bar{q} \in F^{-1}(q)$ with $|\bar{p}\,\bar{q}| = |p\,q|$. Moreover, the map $\bar{p} \mapsto \bar{q}$ is continuous. We can then define horizontal lifts of unit vectors by $\overrightarrow{\bar{p}\bar{q}} = \overrightarrow{pq}$. This is well defined since $\overrightarrow{pq_2} = \overrightarrow{pq_2}$ implies that q_1, q_2 lie on the same segment emanating from p and thus the same will be true for $\bar{q}_1, \bar{q}_2$.

Select distance coordinates $(r_1, \ldots, r_k)$ around p. Observe that all of the r_is are Riemannian submersions and therefore also submetries. Then the compositions $r_i \circ F$ are also submetries. Thus, F is C^1 if and only if all the maps $r_i \circ F$ are C^1. Therefore, it suffices to prove the result in the case of functions $r : U \subset M \to (a, b)$.

The observation is simply that ∇r is the horizontal lift of ∂_r on (a, b). Continuity of ∇r follows from continuity of $\bar{p} \mapsto \bar{q}$. □

Remark 5.6.17. It can be shown that submetries are $C^{1,1}$, i.e., their derivatives are locally Lipschitz. In terms of the above proof this follows from showing that the map $\bar{p} \mapsto \bar{q}$ is locally Lipschitz. It is in general not possible to improve this. Consider, e.g., $K = [0,1]^2 \subset \mathbb{R}^2$ and let $r(x) = |xK|$. Then the levels $r = r_0 > 0$ are not C^2 as they consist of a rounded square with sides parallel to the sides of K and rounded corners that are quarter circles centered at the corners of K.

5.6.4 The Slice Theorem

In this section we establish several important results about actions on manifolds. First we show that the isometry group is a Lie group and then proceed with a study of the topology near the orbits of actions by isometries.

The action by a topological group H on a manifold M is said to be *proper* if the map $\mathrm{H} \times M \to M \times M$ defined by $(h,p) \mapsto (hp,p)$ is a proper map. The orbit of H through $p \in M$ is $\mathrm{H}p = \{hp \mid h \in \mathrm{H}\}$. The topology on the quotient $\mathrm{H}\backslash M$, that consists of the space of orbits of the action, is the quotient topology (note the we are careful to divide on the left as we shall use both right and left cosets in this section). This makes $M \to \mathrm{H}\backslash M$ continuous and open. This topology is clearly second countable and also Hausdorff when the action is proper.

The *isotropy group* of an action H at $p \in M$ is $\mathrm{H}_p = \{h \in \mathrm{H} \mid hp = p\}$. Note that along an orbit the isotropy groups are always conjugate: $\mathrm{H}_{hp} = h\mathrm{H}_p h^{-1}$. When the action is proper H_p is compact. This gives us a proper action $(k,h) \mapsto hk^{-1}$ of H_p on H. The orbit space H/H_p is the natural coset space of left translates of H_p. The natural identification $\mathrm{H}/\mathrm{H}_p \to \mathrm{H}p$ is a bijection that is both continuous and proper and hence a homeomorphism. We say that H is *free or acts freely* if $\mathrm{H}_p = \{e\}$ for all $p \in M$.

The topology on $\mathrm{Iso}(M,g)$ is defined and studied in exercise 5.9.41. The key property we shall use is that $\mathrm{Iso}(M,g) \ni F \mapsto \left(F(p), DF|_p\right)$ is continuous and a homeomorphism onto its image. Note that the last fact factor is a "linear" map $T_pM \to TM$.

Example 5.6.18. The Arzela-Ascoli lemma implies that $\mathrm{Iso}(M)$ acts properly on M (see also exercise 5.9.41). However, a subgroup $\mathrm{H} \subset \mathrm{Iso}(M)$ does not necessarily act properly unless it is a closed subgroup. The action $\mathbb{R} \times S^1 \times S^1 \to S^1 \times S^1$ defined by $\theta \cdot (z_1, z_2) = \left(e^{\theta\,\mathrm{i}} z_1, e^{\alpha\theta\,\mathrm{i}} z_2\right)$ is proper if and only if α is rational.

Theorem 5.6.19 (Myers and Steenrod, 1939). *If* H *is a closed subgroup of* $\mathrm{Iso}(M,g)$*, then the orbits of the action are submanifolds. In particular, the isometry group is a Lie group.*

Proof. The proof is a streamlined version of the original proof by Myers and Steenrod. They showed that the orbits are C^1, fortunately a little trick allows us to bootstrap the construction to obtain smoothness.

Throughout the proof we work locally and use that any Riemannian manifold looks like Euclidean space via exponential coordinates both around a point as in proposition 5.5.1 and around a small tube as in corollary 5.5.3. Since we work locally all metric balls have smooth boundary.

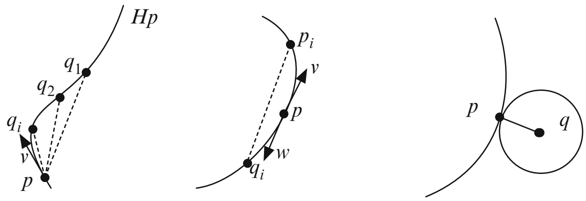


Fig. 5.9 Tangent and normal vectors to orbits

We say that $v \in T_pM$ is tangent to H$p \subset M$ if $v = \lim \dot{c}_i(0)$, where c_i are geodesic segments from p to $p_i \in$ Hp with $\lim p_i = p$ (see figure 5.9). Since H_p is compact we can always write $p_i = h_i p$ with $\lim h_i = e$. In this case $Dh_i|_p$ converges to the identity (see exercise 5.9.41). The set of all such tangent vectors at p is denoted $T_p\mathrm{H}p$. We claim that $T_p\mathrm{H}p \subset T_pM$ is a subspace. First note that this set is invariant under scaling by positive scalars as we can reparametrize the geodesic segments. Next consider $v = \lim v_i$ and $w = \lim w_i$, where v_i and w_i are initial velocities for geodesic segments from p to $p_i = h_i p$ and $q_i \in$ Hp, respectively. Using that the metric is locally Euclidean near p it follows that the velocity $\dot{c}_i(0)$ for a suitably parametrized geodesic c_i from p_i to q_i is close to $w - v$ in TM (see figure 5.9). Using the isometry h_i^{-1} to move $p_i = h_i p$ to p and $\lim h_i = e$ implies that

$$\lim \frac{d\left(h_i^{-1} \circ c_i\right)}{dt}(0) = w - v.$$

This shows that $w - v \in T_p\mathrm{H}p$.

The group structure preserves the orbits and maps tangent vectors to tangent vectors by $T_{hp}\mathrm{H}p = Dh\left(T_p\mathrm{H}p\right)$. As $Dh = \exp_{hp}^{-1} \circ h \circ \exp_p$ locally, it follows that these tangent spaces vary continuously along Hp.

We say that $v \in T_pM$ is normal to Hp if it is proportional to $\overrightarrow{pq}$ where $|qp| = |q\,\mathrm{H}p|$. Clearly $B(q, |qp|) \cap \mathrm{H}p = \varnothing$ so the angle between tangent and normal vectors must be $\geq \pi/2$. Since the tangent vectors form a subspace they must in fact be perpendicular to all normal vectors (see figure 5.9). This shows that if $O \subset M$ is an open subset with smooth boundary and $O \cap \mathrm{H}p = \varnothing$, then for any $q \in \partial O \cap \mathrm{H}p$ we have $T_q\mathrm{H}p \subset T_q\partial O$.

Let N^k be a small k-dimensional submanifold with $T_pN = T_p\mathrm{H}p$. Use the normal exponential map to introduce coordinates (x, y) on a tubular product neighborhood of N diffeomorphic to $N \times B$ with $B = B(0, \epsilon) \subset \mathbb{R}^{n-k}$, where $n = \dim M$ (see figure 5.10). Since $\{x\} \times B \subset N \times B$ is perpendicular to N at $(x, 0)$ it follows that $N \times \{y\}$ is almost perpendicular to $\{x\} \times B$ at $(x, y) \in N \times B$ as long as N and ϵ are sufficiently small. Since $T_{hp}\mathrm{H}p$ varies continuously with h it follows that it has trivial intersection with the tangent spaces to $\{x\} \times B$.

We now claim that the map $(x, y) \mapsto x$ projects a neighborhood of $p \in \mathrm{H}p$ to a neighborhood of $p \in N$. Since Hp is closed its image in N is also closed. Let the complement of the image in N be denoted N'.

If $p_k = h_k p$ and $q_k \in \mathrm{H}p$ are mapped to the same point in N, then $\overrightarrow{p_k q_k}$ is tangent to B. On the other hand, if $\lim p_k = p = \lim q_k$, then $\overrightarrow{p_k q_k}$ will (sub)converge to a

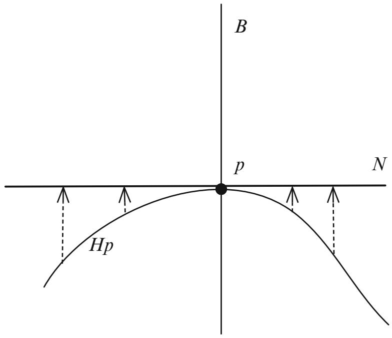


Fig. 5.10 Making an orbit a graph

vector orthogonal to $T_p\mathrm{H}p$. Then $\overrightarrow{p\,h_i^{-1}q_k}$ also (sub)converges to a vector orthogonal to $T_p\mathrm{H}p$, which is a contradiction.

Assume that p is on the boundary of N'. Then we can find a sequence of open sets $O_i' \subset N'$ with smooth boundary; $\partial O_i' \cap \partial N' \neq \varnothing$; and $p = \lim_{i\to\infty} q_i$ for any $q_i \in O_i'$. This means that if $O_i = O_i' \times B$, then $O_i \cap \mathrm{H}p = \varnothing$ and we can find $p_i = (x_i, y_i) \in \partial O_i \cap \mathrm{H}p$ that converge to $(p, 0) = p$. In particular, $T_{p_i}\mathrm{H}p \subset T_{p_i}\partial O_i$. Now $\dim T_{p_i}\mathrm{H}p = k$ and $\dim T_{p_i}(\{x_i\} \times B) = n - k$ so it follows that they have a nontrivial intersection as they are both subspaces of the $(n-1)$-dimensional space $T_{p_i}\partial O_i = T_{x_i}O_i' \oplus T_{y_i}B$. On the other hand $T_{p_i}(\{x_i\} \times B)$ converges to $T_p^{\perp}N$ and so by continuity must be almost perpendicular to $T_{p_i}\mathrm{H}p$. This contradicts that $p \in \partial N'$.

By shrinking N if necessary we can write $\mathrm{H}p \cap (N \times B)$ as a continuous graph over N. The tangent spaces to the orbits also vary continuously and are almost orthogonal to TB. Thus tangent vectors to the orbits are uniquely determined by their projection on to TN. In particular, any smooth curve in N is mapped to a curve in the orbit. Moreover, the velocity field of the curve has a unique continuous lift to the tangent space of the orbit. It is easy to see that this lifted velocity field is the velocity of the corresponding curve. Similarly we see that the graph is C^1 and consequently that the orbit is a C^1 submanifold.

To see that the isometry group of M is a C^1 Lie group first note that it acts properly on $M^{n+1} = M \times \cdots \times M$ and thus forms a closed subgroup of the isometry of this space. Moreover, this action is well-defined and free on the open subset of points $(p_0, \ldots, p_n) \in O \subset M^{n+1}$ where $\overrightarrow{p_0 p_i}$, $i = 1, \ldots, n$ are linearly independent. Thus the isometry group of M is naturally identified with a C^1 submanifold of O.

Finally, note that the formulas $T_{hp}\mathrm{H}p = Dh\left(T_p\mathrm{H}p\right)$ and $Dh = \exp_{hp}^{-1} \circ h \circ \exp_p$ show that the tangent spaces $T\mathrm{H}p$ to $\mathrm{H}p$ form a submanifold of TM that is as smooth as the group H. So if H is C^k, $k \geq 1$, then so is $T\mathrm{H}p$. But this implies that $\mathrm{H}p$ is a C^{k+1} submanifold. The above construction then shows that H itself is C^{k+1}. This finishes the proof that the isometry group is a smooth Lie group. □

There are other proofs of this theorem that also work without metric assumptions (see theorem 8.1.6 and [83] or use various profound characterizations of Lie groups as in exercise 6.7.26 and [79]).

The goal is to refine our understanding of the topology near the orbits of the action by a closed subgroup $\mathrm{H} \subset \mathrm{Iso}(M, g)$. Such groups are necessarily Lie groups and as such have a Lie group exponential map $\exp : T_e\mathrm{H} = \mathfrak{h} \to \mathrm{H}$. The Lie subalgebra of H_p is denoted $\mathfrak{h}_p$. Observe that $v \in \mathfrak{h}_p$ if and only if $\exp(tv) \in \mathrm{H}_p$ for all t.

Proposition 5.6.20. *Let* $\mathrm{O}_p(h) = hp$ *be the orbit map. Then* $\ker\left(\left(D\mathrm{O}_p\right)|_e\right) = \mathfrak{h}_p$ *and more generally* $\ker\left(\left(D\mathrm{O}_p\right)|_x\right) = DL_x\left(\mathfrak{h}_p\right)$.

Proof. Note that the last statement follows from the first by the chain rule and $(h_1h_2)\,p = h_1\,(h_2p)$. To establish the first claim we first note that

$$\left(D\mathrm{O}_p\right)|_e(v) = \frac{d}{dt}\left(\exp(tv)\cdot p\right)|_{t=0}.$$

Next observe that

$$\begin{aligned}\frac{d}{dt}\left(\exp(tv)\cdot p\right)|_{t=t_0} &= \frac{d}{ds}\left(\exp(t_0v)\exp(sv)\cdot p\right)|_{s=0}\\ &= D\left(\exp(t_0v)\right)\left(\frac{d}{ds}\left(\exp(sv)\cdot p\right)|_{s=0}\right).\end{aligned}$$

So if $\left(D\mathrm{O}_p\right)|_e(v) = 0$, then $\exp(tv)\cdot p = p$ for all t and hence $v \in \mathfrak{h}_p$. The converse is trivially true. □

When H acts freely this proposition implies that all orbits $\mathrm{H}p$ are immersed submanifolds.

Since H consists of isometries there is a natural H-invariant map $E : \mathrm{H} \times T_pM \to M$ defined by

$$(h, v) \mapsto h\exp_p(v) = \exp_{hp}\left(Dh|_p v\right),$$

i.e., $E(hx, v) = hE(x, v)$ for all $h, x \in \mathrm{H}$ and $v \in T_pM$.

Theorem 5.6.21 (The Free Slice Theorem). *If* $\mathrm{H} \subset \mathrm{Iso}(M, g)$ *is closed and acts freely, then the quotient* $\mathrm{H}\backslash M$ *can be given a smooth manifold structure and Riemannian metric so that* $M \to \mathrm{H}\backslash M$ *is a Riemannian submersion.*

Proof. We just saw that the orbits are properly embedded copies of H. If we restrict the map E to the normal bundle to $\mathrm{H}p$ at p, then we obtain a H-invariant map $\exp^{\perp} : \mathrm{H} \times T_p^{\perp}\mathrm{H}p \to M$. Note that there is a natural trivialization $\mathrm{H} \times T_p^{\perp}\mathrm{H}p \to T^{\perp}\mathrm{H}p$ defined by $(h, v) \mapsto Dh|_p(v)$, which is a linear isometry on the fibers. Moreover, $\exp^{\perp}$ is in fact the normal exponential map $\exp^{\perp} : T^{\perp}\mathrm{H}p \to M$ via this identification. We can then invoke the tubular neighborhood theorem (corollary 5.5.3) to obtain a diffeomorphism from some neighborhood of the zero section in $\mathrm{H} \times T_p^{\perp}\mathrm{H}p$ to a neighborhood of the orbit in M. However, we need a

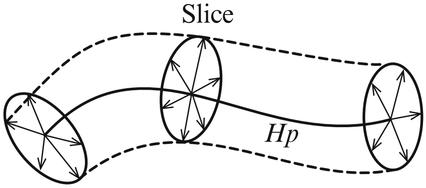


Fig. 5.11 Slices along an orbit

uniform neighborhood of the form $\exp^{\perp} : \mathrm{H} \times B(0,\epsilon) \to M$, where $B(0,\epsilon) \subset T_p^{\perp}\mathrm{H}p$. In such a uniform neighborhood the set $B(0,\epsilon)$ is called a *slice* of the action. Thus a slice is a cross section of a uniform tube (see figure 5.11).

First we find an $\epsilon > 0$ so that $\exp^{\perp} : U \times B(0,\epsilon) \to M$, $e \in U \subset \mathrm{H}$ is an embedding. Thus the usual normal exponential map is also an embedding on the ϵ-neighborhood of the zero section in $T^{\perp}Up$. We can further assume that all closed ϵ-balls centered in the image are compact and thus have compact intersection with all orbits.

We can further decrease ϵ so that if $v \in B(0,\epsilon)$ and $\left|(hp)\exp_p(v)\right| < \epsilon$, then $h \in U$. This shows that p is the unique closest point in $\mathrm{H}p$ to $\exp_p(v)$. In fact the first variation formula shows that any segment from $\exp_p(v)$ to $\mathrm{H}p$ is perpendicular to $\mathrm{H}p$. Moreover, any such a segment will end at a point hp with $h \in U$. But then v and the tangent vector to the segment from hp to $\exp_p(v)$ are normal vectors to Up that are mapped to the same point. This violates the choice of ϵ. This in turn shows that $\exp^{\perp} : \mathrm{H} \times B(0,\epsilon) \to M$ is an embedding. It is clearly nonsingular since this is true at all points (e,v), $|v| < \epsilon$, and $\exp^{\perp}(h,v) = h\exp_p(v)$, where h is a diffeomorphism on M. It is also injective since $\exp^{\perp}(h_1,v_1) = \exp^{\perp}(h_2,v_2)$ first implies that $\exp^{\perp}\left(h_2^{-1}h_1,v_1\right) = \exp^{\perp}(e,v_2)$. This shows that $h_2^{-1}h_1 \in U$ and then by choice of ϵ that $h_1 = h_2$ and $v_1 = v_2$. Finally the map is proper since $\mathrm{H}p$ is properly embedded. This shows that it is closed and an embedding.

We have shown that $M \to \mathrm{H}\backslash M$ looks like a locally trivial bundle. The manifold structure on the quotient comes from the fact that for each $p \in M$ the slice $B(0,\epsilon)$ is mapped homeomorphically to its image in $\mathrm{H}\backslash M$. These charts are easily shown to have smooth transition functions. Finally, the metric on $\mathrm{H}\backslash M$ is constructed as in section 4.5.2 by identifying the tangent space at a point $\mathrm{H}p \in \mathrm{H}\backslash M$ with one of the normal spaces $T_{hp}^{\perp}\mathrm{H}p$ and noting that $Dh|_p$ maps $T_p^{\perp}\mathrm{H}p$ isometrically to $T_{hp}^{\perp}\mathrm{H}p$. Thus all of these normal spaces are isometric to each other. This induces a natural Riemannian metric on the quotient that makes the quotient map a Riemannian submersion. □

Remark 5.6.22. Let $\mathrm{K} \subset \mathrm{H}$ be a compact subgroup of a Lie group. Consider the action $(k,x) \mapsto xk^{-1}$ of K on H. As K is compact we can average any metric on H to make it right-invariant under this action by K. Thus we obtain a free action by isometries and we can use the above to make H/K a manifold with a Riemannian submersion metric. In case the metric on H is also left-invariant we obtain an isometric action of H on H/K that makes H/K a homogeneous space.

In case $\mathrm{K} \subset \mathrm{H}$ is closed we still obtain a proper action by right multiplication. This can again be made isometric by using a right-invariant metric. However, it is not necessarily possible to also have the metric on H be left-invariant so that H acts by isometries on H/K.

Corollary 5.6.23. *Let* $\mathrm{H} \times M \to M$ *be a proper isometric action. For each* $p \in M$ *the orbits* $\mathrm{H}p$ *are properly embedded submanifolds* $\mathrm{H}/\mathrm{H}_p \to \mathrm{H}p$*.*

Proof. We already know that it is a proper injective map and that H/H_p has a manifold structure. Furthermore proposition 5.6.20 shows that the differential is also injective. This shows that it is a proper embedding. □

The *slice representation* of a proper isometric action is the linear representation $\mathrm{H} \times T_p^{\perp}\mathrm{H}p \to T_p^{\perp}\mathrm{H}p$ given by $(h, v) \mapsto Dh|_p(v)$. If we let H_p act on H on the right as above, then H_p naturally acts on $\mathrm{H} \times T_p^{\perp}\mathrm{H}p$ and corollary 5.6.23 shows that the quotient $\mathrm{H} \times_{\mathrm{H}_p} T_p^{\perp}\mathrm{H}p$ can be given a natural manifold structure.

Theorem 5.6.24 (The Slice Theorem). *Let* $\mathrm{H} \subset \mathrm{Iso}(M, g)$ *be a closed subgroup. For each* $p \in M$ *there is a map* $\exp^{\perp} : \mathrm{H} \times_{\mathrm{H}_p} T_p^{\perp}\mathrm{H}p \to M$ *that is a diffeomorphism on a uniform tubular neighborhood* $\mathrm{H} \times_{\mathrm{H}_p} B(0, \epsilon)$ *on to an* ϵ*-neighborhood of the orbit* $\mathrm{H}p$*.*

Proof. The proof is as in the free case now that we have shown that all orbits are properly embedded. For fixed $p \in M$ consider the bundle map

$$\mathrm{H} \times T_p^{\perp}\mathrm{H}p \to T^{\perp}\mathrm{H}p$$
$$(h, v) \mapsto Dh|_p(v).$$

This map is H-invariant, an isomorphism on the fibers, and $\mathrm{H} \times \{0\}$ is mapped to the zero section in $T^{\perp}\mathrm{H}p$ represented by the orbit $\mathrm{H}p$. Since $D\left(h \circ k^{-1}\right)|_p\left(Dk|_p(v)\right) = Dh|_p(v)$ for any element $k \in \mathrm{H}_p$ this gives us a natural bundle isomorphism from $\mathrm{H} \times_{\mathrm{H}_p} T_p^{\perp}\mathrm{H}p$ to $T^{\perp}\mathrm{H}p$. Now define $\exp^{\perp} : \mathrm{H} \times_{\mathrm{H}_p} T_p^{\perp}\mathrm{H}p \to M$ as the normal exponential map $T^{\perp}\mathrm{H}p \to M$ via this identification (see figure 5.12).

It is now possible to find $\epsilon > 0$ as in theorem 5.6.21 so that $\exp^{\perp} : \mathrm{H} \times_{\mathrm{H}_p} B(0, \epsilon) \to M$ becomes an H-invariant embedding. □

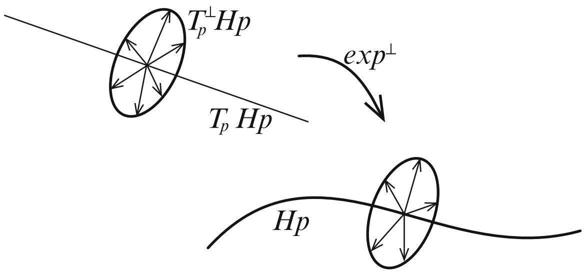


Fig. 5.12 A linear slice and a slice in the manifold

This theorem tells us exactly how H acts near an orbit and allows us to calculate the isotropy of points near a given point.

Corollary 5.6.25. *For small $v \in T_p^{\perp}\mathrm{H}p$ the isotropy at* $\exp_p(v)$ *is given by*

$$\mathrm{H}_{\exp_p(v)} = \left\{h \in \mathrm{H}_p \mid Dh|_p v = v\right\}.$$

This in turn implies.

Corollary 5.6.26. *If* $\mathrm{H} \subset \mathrm{Iso}(M, g)$ *is a closed subgroup with the property that all its isotropy groups are conjugate to each other, then the quotient space is a Riemannian manifold and the quotient map a Riemannian submersion.*

5.7 Completeness

5.7.1 The Hopf-Rinow Theorem

One of the foundational centerpieces of 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 open subsets of Euclidean space. For it seems that in these examples, geodesic and metric completeness break down in exactly the same places.

Theorem 5.7.1 (Hopf and Rinow, 1931). *The following statements are equivalent for a Riemannian manifold* (M, g)*:*

(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)⇒(2) and (3)⇒(4) are trivial.

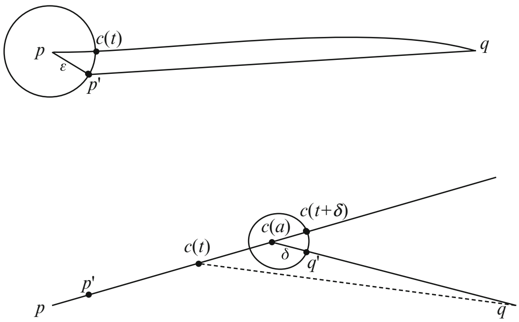
(4)⇒(1): Recall that every geodesic $c : [0, b) \to M$ defined on a maximal interval must leave every compact set if $b < \infty$. This violates metric completeness as $c(t)$, $t \to b$ is a Cauchy sequence.

(2)⇒(3): Consider $\exp_p : T_pM \to M$. It suffices to show that

$$\exp_p\left(\overline{B}(0, R)\right) = \overline{B}(p, R)$$

for all R (note that $\subset$ always holds). This will follow if we can show that any point $q \in M$ is joined to p by a segment. By corollary 5.5.6 we can find $\epsilon > 0$ such that any point in the compact set $\overline{B}(p, \epsilon) = \exp_p\left(\overline{B}(0, \epsilon)\right)$ can be joined to p by a minimal geodesic. This shows that if $p' \in \overline{B}(p, \epsilon) - B(p, \epsilon)$ is closest to q, then $|pp'| +$

Fig. 5.13 Two short cuts from p to q



$|p'q| = |pq|$. Otherwise, corollary 5.3.10 guarantees a unit speed curve $c \in \Omega_{p,q}$ with $L(c) < |pp'| + |p'q|$ (see top of figure 5.13). Choose t so that $c(t) \in \bar{B}(p,\epsilon) - B(p,\epsilon)$. Since $t + |c(t)q| \leq L(c) < |pp'| + |p'q|$ it follows that $|c(t)q| < |p'q|$ contradicting the choice of p'.

Let $c(t) : [0,\infty) \to M$ be the unit speed geodesic with $c(0) = p$ and $c(\epsilon) = p'$. We just saw that $|pq| = \epsilon + |c(\epsilon)q|$.

Consider

$$A = \{t \in [0, |pq|] \mid |pq| = t + |c(t)q|\}.$$

Clearly $0, \epsilon \in A$. Note that if $t \in A$, then

$$|pq| = t + |c(t)q| \geq |pc(t)| + |c(t)q| \geq |pq|,$$

which implies that $t = |pc(t)|$. We first claim that if $a \in A$, then $[0,a] \subset A$. When $t < a$ note that

$$\begin{aligned} |pq| &\leq |pc(t)| + |c(t)q| \\ &\leq |pc(t)| + |c(t)c(a)| + |c(a)q| \\ &\leq t + a - t + |c(a)q| \\ &\leq a + |c(a)q| \\ &= |pq|. \end{aligned}$$

This implies that $|pc(t)| + |c(t)q| = |pq|$ and $t = |pc(t)|$, showing that $t \in A$ (see also figure 5.13).

Since $t \mapsto |c(t)q|$ is continuous it follows that A is closed.

Finally, we claim that if $a \in A$, then $a + \delta \in A$ for sufficiently small $\delta > 0$. Use corollary 5.5.6 to find $\delta > 0$ so that any point in $\bar{B}(c(a), \delta)$ can be joined to $c(a)$ by a segment (see also figure 5.13). If we select $q' \in \bar{B}(c(a), \delta) - B(c(a), \delta)$ closest to q, then

$$\begin{aligned}|pq| &= a + |c(a)\,q| \\ &= a + \left|c(a)\,q'\right| + \left|q'q\right| \\ &= a + \delta + \left|q'q\right| \\ &\geq \left|pq'\right| + \left|q'q\right| \\ &\geq |pq|\,.\end{aligned}$$

It follows that $|pq'| = a + \delta$ which tells us that the piecewise smooth geodesic that goes from p to $c(a)$ and then from $c(a)$ to q' is a segment. By corollary 5.4.4 this segment is a geodesic and $q' = c(a + \delta)$. It then follows from $|pq| = a + \delta + |q'q|$ that $c(a + \delta) \in A$.

This shows that $A = [0, |pq|]$. □

From (2) ⇒ (3) we get the additional result:

Corollary 5.7.2. *If (M, g) is complete in any of the above ways, then any two points in M can be joined by a segment.*

Corollary 5.7.3. *If (M, g) admits a proper Lipschitz function $f : M \to \mathbb{R}$, then M is complete.*

Proof. We establish the Heine-Borel property. Let $C \subset M$ be bounded and closed. Since f is Lipschitz the image $f(C)$ is also bounded. Thus $f(C) \subset [a, b]$ and $C \subset f^{-1}([a, b])$. As f is proper the pre-image $f^{-1}([a, b])$ is compact. Since C is closed and a subset of a compact set it must itself be compact. □

This corollary also makes it easy to check completeness for all of our examples related to warped products. In these examples, the distance function can be extended to a proper continuous function on the entire space.

From now on, virtually all Riemannian manifolds will automatically be assumed to be connected and complete.

5.7.2 *Warped Product Characterization*

In theorem 4.3.3 we offered a local characterization of Riemannian manifolds that admit functions whose Hessian is conformal to the metric and saw that these were all locally given by warped product structures. Here we extend this to a global result for complete Riemannian manifolds.

Theorem 5.7.4 (Tashiro, 1965). *Let (M, g) be a complete Riemannian n-mani-fold that admits a nontrivial function f whose Hessian is conformal, i.e.,* $\operatorname{Hess} f = \lambda g$. *Then (M, g) is isometric to a complete warped product metric and must have one of the three forms:*

(1) $M = \mathbb{R} \times N$ *and* $g = dr^2 + \rho^2(r)\, g_N$,
(2) $M = \mathbb{R}^n$ *and* $g = dr^2 + \rho^2(r)\, ds^2_{n-1}$, $r \geq 0$,
(3) $M = S^n$ *and* $g = dr^2 + \rho^2(r)\, ds^2_{n-1}$, $r \in [a, b]$.

Proof. We start by identifying N. Recall from theorem 4.3.3 that $|\nabla f|$ is locally constant on $\{f = f_0\} \cap \{df \neq 0\}$ for each $f_0 \in \mathbb{R}$. From this it follows that the connected components of $\{f = f_0\} \cap \{df \neq 0\}$ must be closed. As f is nontrivial there will be points where the differential doesn't vanish. Define $N \subset \{f = f_0\} \cap \{df \neq 0\}$ as any nonempty connected component and note that N is a closed hypersurface in M.

For $p \in N$ the unit speed geodesic though p that is normal to N is given by:

$$c_p(t) = \exp_p\left(t \frac{\nabla f|_p}{|\nabla f|}\right).$$

For fixed numbers $a < 0 < b$ consider the set $C \subset N$ such that f is regular at $c_p(t)$ for all $p \in C$ and $t \in [a, b]$. Since the set of regular points is open it follows that $C \subset N$ is open. Theorem 4.3.3 shows that on $U = \{c_p(t) \mid p \in C,\ t \in [a, b]\}$ we have a warped product structure $g|_U = dr^2 + \rho^2(r)\, g_N$, where r is the signed distance function to N and

$$\nabla r = \frac{\nabla f}{|\nabla f|},$$

$$f(r) = \int \rho(r)\, dr.$$

For all $p \in N$

$$\frac{d^2(f \circ c_p)}{dt^2} = g(\nabla f, \ddot{c}_p) + \operatorname{Hess} f(\dot{c}_p, \dot{c}_p) = \lambda \circ c_p$$

with

$$(f \circ c_p)(0) = f_0,$$

$$\frac{d(f \circ c_p)}{dt}(0) = g(\nabla f, \dot{c}_p(0)) = |\nabla f|.$$

When we restrict attention to U we have $\lambda \circ c_p = \lambda(f \circ c_p)$. Thus $f \circ c_p$ satisfies a second-order equation with initial values that do not depend on $p \in C$. In particular, $f \circ c_p(t)$ depends only on $t \in [a, b]$ and not on $p \in C$. Similarly,

$$\frac{d(f \circ c_p)}{dt}(t) = g\left(\nabla f|_{c_p(t)}, \dot{c}_p(t)\right) = \left|\nabla f|_{c_p(t)}\right| = \rho(t)$$

depends only on $t \in [a,b]$ and not on $p \in C$. Continuity of $\left|\nabla f|_{c_p(t)}\right|$ with respect to p and t, combined with the fact that f is regular on U, shows that for fixed $t \in [a,b]$ the value $\left|\nabla f|_{c_p(t)}\right|$ cannot vanish when $p \in \partial C \subset N$. Thus we have shown that $C \subset N$ is both open and closed. Since N connected we conclude that $C = N$.

Finally, we obtain nontrivial maximal open interval $(a,b) \ni 0$ such that $c_p(t)$ is regular for all $t \in (a,b)$ and $p \in N$. Moreover, the warped product structure $dr^2 + \rho^2(r)\, g_N$ extends to hold on $(a,b) \times N$.

When $(a,b) = \mathbb{R}$ we obtain a global warped product structure.

If, say, $b < \infty$, then the level set $\{r = b\}$ consists of critical points for f. Since $\rho = |\nabla f|$ it follows that $\lim_{t \to b} \rho(t) = 0$. The warped product structure then shows that any two points in N will approach each other as $t \to b$. In other words

$$\lim_{t \to b} \left| c_p(t)\, c_q(t) \right| = 0.$$

Consequently, $\{r = b\}$ is a single critical point x. Now consider all of the unit vectors $\dot{c}_p(b) \in T_x M$. Since N is a closed submanifold this set of vectors is both closed and open and thus consists of all unit vectors at x. This shows that not only will any geodesic $c_p(t)$ approach x as $t \to b$, but after it has passed through x it must coincide with another such geodesic. Thus $\{f_0 \le f < f(x)\} \simeq [0,b) \times N$ and x is the only critical point in $\{f_0 \le f \le f(x)\}$. It also follows that $N \simeq S^{n-1}$ since the level sets for f near x are exactly distance spheres centered at x. The argument that g_N is a round metric on S^{n-1} can be completed exactly as in the proof of theorem 4.3.3.

A similar argument holds when $-\infty < a$. We finish the proof by observing that we are in case 3 when both a and b are finite and in case 2 when only one of a or b are finite. □

With more information about λ we expect a more detailed picture of what M can be. In particular, there is a global version of the local classification from corollary 4.3.4.

Theorem 5.7.5. *Let (M,g) be a complete Riemannian n-manifold that admits a nontrivial function f whose Hessian satisfies* $\operatorname{Hess} f = (\alpha f + \beta)\, g$, $\alpha, \beta \in \mathbb{R}$. *Then (M,g) falls in to one of the following three categories:*

(a) $(M,g) = \left(I \times S^{n-1}, dr^2 + \operatorname{sn}_k^2(r)\, ds_{n-1}^2\right)$, *i.e., a constant curvature space form.*
(b) $(M,g) = \left(\mathbb{R} \times N, dr^2 + g_N\right)$, *i.e., a product metric.*
(c) $(M,g) = \left(\mathbb{R} \times N, dr^2 + \left(A \exp\left(\sqrt{\alpha} r\right) + B \exp\left(-\sqrt{\alpha} r\right)\right)^2 h\right)$, $A, B \ge 0$.

Proof. From the previous theorem we already know that $g = dr^2 + \rho^2(r)\, g_N$, $r \in I$, with $f = f(r)$, $\rho(r) = f'(r)$, and $\lambda = \alpha f + \beta = f''$.

It'll be convenient to divide into various special cases.

When $\alpha = \beta = 0$ it follows that ρ is constant. This is case (b).

When $\alpha = 0$ and $\beta \neq 0$ it follows that $\rho = \beta r + \gamma$. Thus I is a half line where ρ vanishes at the boundary point. This point must correspond to a single critical point for f. The metric is the standard Euclidean metric.

When $\alpha \neq 0$ we can change f to $f + \frac{\beta}{\alpha}$. Then $\operatorname{Hess}\left(f + \frac{\beta}{\alpha}\right) = \alpha\left(f + \frac{\beta}{\alpha}\right) g$ so we can assume that $\beta = 0$. In case $\alpha < 0$, it follows that $\rho = A \sin\left(\sqrt{-\alpha} r + r_0\right)$. Thus I is a compact interval and the metric becomes a round sphere. In case $\alpha > 0$, we have that $\rho = A \exp\left(\sqrt{\alpha} r\right) + B \exp\left(-\sqrt{\alpha} r\right)$, where at least one of A or B must be positive. If they are both nonnegative we are in case (c). If they have the opposite sign we can rewrite $\rho = C \sinh\left(\sqrt{\alpha} r + r_0\right)$. Then I is a half line and the metric becomes a constant negatively curved warped product. □

Note that in case (a) the function f has at least one critical point, while in cases (b) and (c) f has no critical points. In 1961 Obata established this theorem for round spheres using the equation

$$\operatorname{Hess} f = (1 - f)\, g.$$

Remark 5.7.6. In a separate direction it is shown in [100] that transnormal functions (see remark 4.3.5) on a complete Riemannian manifold give a similar topological decomposition of the manifold. Specifically such functions can have zero, one, or two critical values. All level sets for f are smooth submanifolds, including the critical levels. Moreover, $f = \phi(r)$ where r is the signed distance to a fixed level set of f.

5.7.3 *The Segment Domain*

In this section we characterize when a geodesic is a segment and use this to find a maximal domain in T_pM on which the exponential map is an embedding. This is achieved through a systematic investigation of when distance functions to points are smooth. All Riemannian manifolds are assumed to be complete in this section, but it is possible to make generalizations to incomplete metrics by working on suitable star-shaped domains.

Fix $p \in (M, g)$ and let $r(x) = |px|$. We know that r is smooth near p and that the integral curves for ∂_r are geodesics emanating from p. Since M is complete, these integral curves can be continued indefinitely beyond the places where r is smooth. These geodesics could easily intersect after some time and consequently fail to generate a flow on M. But having the geodesics at points where r might not be smooth helps us understand the lack of smoothness. We know from section (3.2.6) that another obstruction to r being smooth is the possibility of conjugate points. It is interesting to note that while distance functions generally aren't smooth, they always have one sided directional derivatives (see exercise 5.9.28).

To clarify matters we introduce some terminology: The *segment domain* is

$$\operatorname{seg}(p) = \left\{ v \in T_pM \mid \exp_p(tv) : [0,1] \to M \text{ is a segment} \right\}.$$

The Hopf-Rinow theorem (see theorem 5.7.1) implies that $M = \exp_p(\operatorname{seg}(p))$. Clearly $\operatorname{seg}(p)$ is a closed star-shaped subset of T_pM. The star interior of $\operatorname{seg}(p)$ is

$$\operatorname{seg}^0(p) = \{ sv \mid s \in [0,1),\ v \in \operatorname{seg}(p) \}.$$

Below we show that this set is in fact the interior of $\operatorname{seg}(p)$, but this requires that we know it is open. We start by proving

Proposition 5.7.7. *If $x \in \exp_p(\operatorname{seg}^0(p))$, then it is joined to p by a unique segment. In particular, $\exp_p$ is injective on $\operatorname{seg}^0(p)$.*

Proof. To see this note that there is a segment $\sigma : [0,1) \to M$ with $\sigma(0) = p$ and $\sigma(t_0) = x$, $t_0 < 1$. Therefore, should $\hat{\sigma} : [0, t_0] \to M$ be another segment from p to x, then we could construct a nonsmooth segment

$$c(s) = \begin{cases} \hat{\sigma}(s), & s \in [0, t_0], \\ \sigma(s), & s \in [t_0, 1]. \end{cases}$$

Corollary 5.4.4 shows this is impossible. □

On the image $U_p = \exp_p(\operatorname{seg}^0(p))$ define $\partial_r = D\exp_p(\partial_r)$. We expect this to be the gradient for

$$r(x) = |px| = \left| \exp_p^{-1}(x) \right|.$$

From the proof of lemma 5.5.5 it follows that r will be smooth on U_p with gradient ∂_r if $\exp_p : \operatorname{seg}^0(p) \to U_p$ is a diffeomorphism between open sets. Since the map is injective we have to show that it is nonsingular and that $\operatorname{seg}^0(p)$ is open. The image will then automatically also be open by the inverse function theorem. We start by proving that the map is nonsingular.

Lemma 5.7.8. $\exp_p : \operatorname{seg}^0(p) \to U_p$ *is nonsingular everywhere, or, in other words, $D\exp_p$ is nonsingular at every point in* $\operatorname{seg}^0(p)$.

Proof. The standard proof of this statement uses Jacobi fields and is outlined in exercise 6.7.24, but in essence there is very little difference between the two proofs.

The proof is by contradiction. As the set of singular points is closed we can assume that $\exp_p$ is singular at $v \in \operatorname{seg}^0(p)$ and nonsingular at all points tv, $t \in [0,1)$. Since $c(t) = \exp_p(tv)$ is an embedding on $[0,1)$ we can find neighborhoods U around $[0,1)v \subset T_pM$ and V around $c([0,1)) \subset M$ such that $\exp_p : U \to V$ is a diffeomorphism. Note that $v \notin U$ and $c(1) \notin V$. If we take a tangent vector $w \in T_vT_pM$, then we can extend it to a Jacobi field J on T_pM, i.e., $[\partial_r, J] = 0$. Next J can be pushed forward via $\exp_p$ to a vector field on V, also called J, that also commutes with ∂_r. If $D\exp_p|_v w = 0$, then

$$\lim_{t\to 1} J|_{\exp(tv)} = \lim_{t\to 1} D\exp_p(J)|_{\exp(tv)} = 0.$$

In particular, we see that $D\exp_p$ is singular at v if and only if $\exp_p(v)$ is a conjugate point for r. This characterization naturally assumes that r is smooth on a region that has $\exp_p(v)$ as a accumulation point.

The fact that

$$\lim_{t\to 1} |J|^2|_{\exp(tv)} \to 0 \text{ as } t \to 1$$

implies that

$$\lim_{t\to 1} \log |J|^2|_{\exp(tv)} \to -\infty \text{ as } t \to 1.$$

Therefore, there must be a sequence of numbers $t_n \to 1$ such that

$$\frac{\partial_r |J|^2}{|J|^2}|_{\exp(t_n v)} \to -\infty \text{ as } n \to \infty.$$

Now use the first fundamental equation evaluated on the Jacobi field J (see proposition 3.2.11 and section 3.2.4) to conclude that: $\partial_r |J|^2 = 2\operatorname{Hess} r(J, J)$. This shows that:

$$\frac{\operatorname{Hess} r(J,J)}{|J|^2}|_{\exp(t_n v)} \to -\infty \text{ as } n \to \infty.$$

By assumption $c(t) = \exp_p(tv)$ is a segment on some interval $[0, 1+\varepsilon]$, $\varepsilon > 0$. Use corollary 5.5.2 to choose ε so small that $\tilde{r}(x) = |x\, c(1+\varepsilon)|$ is smooth on a ball $B(c(1+\varepsilon), 2\varepsilon)$ (see figure 5.14 for a schematic picture of J and a corresponding Jacobi field for $\tilde{r}$ that agrees with J at t_n). Then consider the function

$$e(x) = r(x) + \tilde{r}(x).$$

From the triangle inequality, we know that

$$e(x) \geq 1 + \varepsilon = |p\, c(1+\varepsilon)|\,.$$

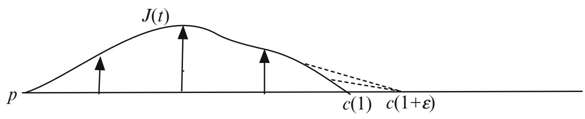


Fig. 5.14 A field that gives shorter curves from p to $c(1+\epsilon)$

Furthermore, $e(x) = 1 + \varepsilon$ whenever $x = c(t)$, $t \in [0, 1 + \varepsilon]$. Thus, e has an absolute minimum along $c(t)$ and consequently has nonnegative Hessian at all the points $c(t)$. On the other hand,

$$\frac{\operatorname{Hess} e\,(J,J)}{|J|^2}\Big|_{\exp(t_n v)} = \frac{\operatorname{Hess} r\,(J,J)}{|J|^2}\Big|_{\exp(t_n v)} + \frac{\operatorname{Hess} \tilde{r}\,(J,J)}{|J|^2}\Big|_{\exp(t_n v)} \underset{n\to\infty}{\to} -\infty$$

since $\operatorname{Hess}\tilde{r}$ is bounded in a neighborhood of $c(1)$ and the term involving $\operatorname{Hess} r$ goes to $-\infty$ as $n \to \infty$. □

We have shown that $\exp_p$ is injective and has nonsingular differential on $\operatorname{seg}^0(p)$. Before showing that $\operatorname{seg}^0(p)$ is open we characterize elements in the star "boundary" of $\operatorname{seg}^0(p)$ as points that fail to have one of these properties.

Lemma 5.7.9. *If $v \in \operatorname{seg}(p) - \operatorname{seg}^0(p)$, then either*

(1) $\exists w\,(\neq v) \in \operatorname{seg}(p)$ *such that* $\exp_p(v) = \exp_p(w)$*, or*
(2) $D\exp_p$ *is singular at* v.

Proof. Let $c(t) = \exp_p(tv)$. For $t > 1$ choose segments $\sigma_t(s)$, $s \in [0,1]$, with $\sigma_t(0) = p$, and $\sigma_t(1) = c(t)$. Since we have assumed that $c|_{[0,t]}$ is not a segment for $t > 1$ we see that $\dot{\sigma}_t(0)$ is never proportional to $\dot{c}(0)$. Now choose $t_n \to 1$ such that $\dot{\sigma}_{t_n}(0) \to w \in T_pM$. We have that

$$L(\sigma_{t_n}) = |\dot{\sigma}_{t_n}(0)| \to L(c|_{[0,1]}) = |\dot{c}(0)|,$$

so $|w| = |\dot{c}\,(0)|$. Now either $w = \dot{c}(0)$ or $w \neq \dot{c}(0)$. In the latter case w cannot be a positive multiple of $\dot{c}\,(0)$ since $|w| = |\dot{c}\,(0)|$. Therefore, we have found the promised w in (1). If the former happens, we must show that $D\exp_p$ is singular at v. If, in fact, $D\exp_p$ is nonsingular at v, then $\exp_p$ is an embedding near v. Thus, $\dot{\sigma}_{t_n}(0) \to v = \dot{c}(0)$ together with $\exp_p(\dot{\sigma}_{t_n}(0)) = \exp_p(t_n\dot{c}(0))$ implies $\dot{\sigma}_{t_n}(0) = t_n \cdot v$. This shows that c is a segment on some interval $[0, t_n]$, $t_n > 1$ which is a contradiction. □

Notice that in the first case the gradient ∂_r on M becomes undefined at $x = \exp_p(v)$, since it could be either $D\exp_p(v)$ or $D\exp_p(w)$; while in the second case the Hessian of r becomes undefined, since it is forced to go to $-\infty$ along certain fields. Finally we show

Proposition 5.7.10. $\operatorname{seg}^0(p)$ *is open.*

Proof. If we fix $v \in \operatorname{seg}^0(p)$, then there is going to be a neighborhood $V \subset T_pM$ around v on which $\exp_p$ is a diffeomorphism onto its image. If $v_i \in V$ converge to v, then $D\exp_p$ is also nonsingular at v_i. For each i choose $w_i \in \operatorname{seg}(p)$ such that $\exp_p(v_i) = \exp_p(w_i)$. When w_i has an accumulation point $w \neq v$ it follows that $v \notin \operatorname{seg}^0(p)$. Hence $w_i \to v$ and $w_i \in V$ for large i. As $\exp_p$ is a diffeomorphism on V this implies that $w_i = v_i$ and that $v_i \in \operatorname{seg}(p)$. We already know that $\exp_p$ is nonsingular at v_i. Moreover, as $w_i = v_i$ condition (1) in lemma 5.7.9 cannot hold.

It follows that $v_i \in \mathrm{seg}^0(p)$ for large i and hence that $V \cap \mathrm{seg}^0(p)$ is a neighborhood of v. $\square$

All of this implies that $r(x) = |px|$ is smooth on the open and dense subset $U_p - \{p\} \subset M$ and in addition that it is not smooth on $M - U_p$.

The set $\mathrm{seg}(p) - \mathrm{seg}^0(p)$ is called the *cut locus* of p in T_pM. Thus, being inside the cut locus means that we are on the region where r^2 is smooth. Going back to our characterization of segments, we have

Corollary 5.7.11. *Let $c : [0, \infty) \to M$ be a geodesic with $c(0) = p$. If*

$$\mathrm{cut}(\dot{c}(0)) = \sup \left\{ t \mid c|_{[0,t]} \text{ is a segment} \right\},$$

then r is smooth at $c(t)$, $t < \mathrm{cut}(\dot{c}(0))$, but not smooth at $x = c(\mathrm{cut}(\dot{c}(0)))$. Furthermore, the failure of r to be smooth at x is because $\exp_p : \mathrm{seg}(p) \to M$ either fails to be one-to-one at x or has x as a critical value.

5.7.4 The Injectivity Radius

In a complete Riemannian manifold the injectivity radius is the largest radius ε for which

$$\exp_p : B(0, \varepsilon) \to B(p, \varepsilon)$$

is a diffeomorphism. If $v \in \mathrm{seg}(p) - \mathrm{seg}^0(p)$ is the closest point to 0 in this set, then in fact $\mathrm{inj}(p) = |v|$. It turns out that such v can be characterized as follows:

Lemma 5.7.12 (Klingenberg). *Suppose $v \in \mathrm{seg}(p) - \mathrm{seg}^0(p)$ and that $|v| = \mathrm{inj}(p)$. Either*

(1) *there is precisely one other vector v' with*

$$\exp_p(v') = \exp_p(v),$$

and v' is characterized by

$$\frac{d}{dt}|_{t=1} \exp_p(tv') = -\frac{d}{dt}|_{t=1} \exp_p(tv),$$

or

(2) $x = \exp_p(v)$ *is a critical value for* $\exp_p : \mathrm{seg}(p) \to M$.

In the first case there are exactly two segments from p to $x = \exp_p(v)$, and they fit together smoothly at x to form a geodesic loop based at p.

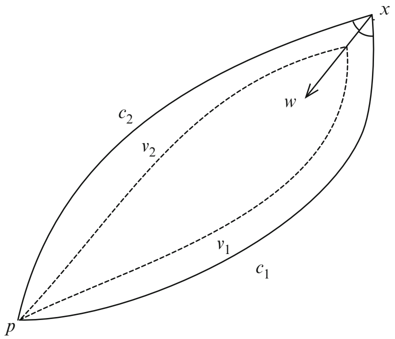


Fig. 5.15 Moving closer to p from x

Proof. Suppose x is a regular value for $\exp_p : \mathrm{seg}(p) \to M$ and that $c_1, c_2 : [0,1] \to M$ are segments from p to $x = \exp_p(v)$. If $\dot{c}_1(1) \neq -\dot{c}_2(1)$, then we can find $w \in T_xM$ such that $g(w, \dot{c}_1(1)) < 0$ and $g(w, \dot{c}_2(1)) < 0$, i.e., w forms an angle $> \frac{\pi}{2}$ with both $\dot{c}_1(1)$ and $\dot{c}_2(1)$. Next select $c(s)$ with $\dot{c}(0) = w$. As $D\exp_p$ is nonsingular at $\dot{c}_i(0)$ there are unique curves $v_i(s) \in T_pM$ with $v_i(0) = \dot{c}_i(0)$ and $D\exp_p(v_i(s)) = c(s)$ (see also figure 5.15). But the curves $t \mapsto \exp_p(tv_i(s))$ have length

$$|v_i| = |pc(s)| < |px| = |v|.$$

This implies that $\exp_p$ is not one-to-one on $\mathrm{seg}^0(p)$, a contradiction. □

5.8 Further Study

There are several textbooks on Riemannian geometry such as [23, 24, 47, 65] and [80] that treat most of the more basic material included in this chapter. All of these books, as is usual, emphasize the variational approach as being *the* basic technique used to prove every theorem. To see how the variational approach works the text [75] is also highly recommended.

5.9 Exercises

EXERCISE 5.9.1. Assume that (M, g) has the property that all unit speed geodesics exist for a fixed time $\varepsilon > 0$. Show that (M, g) is geodesically complete.

EXERCISE 5.9.2. Let $c : I \to (M, g)$ and $\phi : J \to I$, where I, J are intervals. Show that

$$\frac{d(c \circ \phi)}{dt} = \dot{c} \circ \phi \frac{d\phi}{dt},$$

$$\frac{d^2(c\circ\phi)}{dt^2} = \dot{c}\circ\phi\frac{d^2\phi}{dt^2} + \ddot{c}\circ\phi\left(\frac{d\phi}{dt}\right)^2.$$

EXERCISE 5.9.3. Show that if the coordinate vector fields in a chart are orthogonal ($g_{ij} = 0$ for $i \neq j$), then the geodesic equations can be written as

$$\frac{d}{dt}\left(g_{ii}\frac{dc^i}{dt}\right) = \frac{1}{2}\sum_j \frac{\partial g_{jj}}{\partial x^i}\left(\frac{dc^j}{dt}\right)^2.$$

EXERCISE 5.9.4. Show that a regular curve can be reparametrized to be a geodesic if and only if the acceleration is tangent to the curve.

EXERCISE 5.9.5. Let $O \subset (M, g)$ be an open subset of a Riemannian manifold. Show that if (O, g) is complete, then $O = M$.

EXERCISE 5.9.6. A Riemannian manifold is called *Misner complete* if every geodesic $c : (a, b) \to M$ with $b - a < \infty$ lies in a compact set. Show that Misner completeness implies completeness.

EXERCISE 5.9.7. Consider a curve $c \in \Omega_{p,q}$ with $L(c) = |pq|$.

(1) Show that $L\left(c|_{[a,b]}\right) = |c(a)c(b)|$ for all $a, b \in [0, 1]$.
(2) Show that there is a segment $\sigma \in \Omega_{p,q}$ and a monotone function $\varphi : [0, 1] \to [0, 1]$ such that $c = \sigma \circ \varphi$. Note that φ need not be smooth everywhere.

EXERCISE 5.9.8. Let (M, g) be a metrically complete Riemannian manifold and $\tilde{g}$ another metric on M such that $\tilde{g} \geq g$. Show that $(M, \tilde{g})$ is also metrically complete.

EXERCISE 5.9.9. A Riemannian manifold is said to be homogeneous if the isometry group acts transitively. Show that homogeneous manifolds are geodesically complete.

EXERCISE 5.9.10. Consider a Riemannian metric $(M, g) = \left(\mathbb{R} \times N, dr^2 + g_r\right)$, where (N, g_r) is complete for all $r \in \mathbb{R}$, e.g., $(M, g) = \left(\mathbb{R} \times N, dr^2 + \rho^2(r)\, g_N\right)$ where $\rho : \mathbb{R} \to (0, \infty)$ and (N, g_N) is metrically complete. Show that (M, g) is metrically complete.

EXERCISE 5.9.11. Consider metrics $(M, g) = \left((0, \infty) \times N, dr^2 + \rho^2(r)\, g_N\right)$, where $\rho : (0, \infty) \to (0, \infty)$ and (N, g_N) is complete. Give examples that are complete and examples that are not complete.

EXERCISE 5.9.12. Assume $F : (M, g) \to \left(\mathbb{R}^k, g_{\mathbb{R}^k}\right)$ is a Riemannian submersion, where (M, g) is complete. Show that if each of the components of F has zero Hessian, then $(M, g) = (N, h) \times \left(\mathbb{R}^k, g_{\mathbb{R}^k}\right)$.

EXERCISE 5.9.13. Find and fill in the gap in the proof of theorem 5.6.16.

EXERCISE 5.9.14. Show that a Riemannian manifold that is isotropic at every point is also homogeneous. Being isotropic at $p \in M$ means that Iso_p acts transitively on the unit sphere in T_pM.

EXERCISE 5.9.15. Assume that we have coordinates in a Riemannian manifold so that $g_{1i} = \delta_{1i}$. Show that x^1 is a distance function.

EXERCISE 5.9.16. Let $r : U \to \mathbb{R}$ be a distance function on an open set $U \subset (M, g)$. Define another metric $\hat{g}$ on M with the property: $\hat{g}(\nabla r, v) = g(\nabla r, v)$ for all v, where ∇r is the gradient with respect to g. Show that r is also a distance function with respect to $\hat{g}$.

EXERCISE 5.9.17. The *projective models* of $S^n(R)$ and $H^n(R)$ come from projecting the spaces along straight lines through the origin to the hyperplane $x^{n+1} = R$.

(1) Show that if $x \in \mathbb{R}^{n+1}$ and $x^{n+1} > 0$, then the projected point is

$$P(x) = R\left(\frac{x^1}{x^{n+1}}, \dots, \frac{x^n}{x^{n+1}}, 1\right).$$

(2) Show that geodesics on $S^n(R)$ and $H^n(R)$ are given by intersections with 2-dimensional subspaces.
(3) Show that the upper hemisphere of $S^n(R)$ projects to all of $x^{n+1} = R$.
(4) Show that $H^n(R)$ projects to an open disc of radius R in $x^{n+1} = R$.
(5) Show that geodesics on $S^n(R)$ and $H^n(R)$ project to straight lines in $x^{n+1} = R$.

EXERCISE 5.9.18. Show that any Riemannian manifold (M, g) admits a conformal change $(M, \lambda^2 g)$ that is complete. Hint: Choose $\lambda : M \to [1, \infty)$ to be a proper function that grows rapidly.

EXERCISE 5.9.19. On an open subset $U \subset \mathbb{R}^n$ we have the induced distance from the Riemannian metric, and also the induced distance from $\mathbb{R}^n$.

(1) Give examples where U isn't convex and the two distance concepts agree.
(2) Give examples of U, where $\bar{U}$ is convex, but the two distance concepts do not agree.

EXERCISE 5.9.20. Let $M \subset (\bar{M}, g)$ be a submanifold. Using the T-tensor introduced in 2.5.25 show that $T \equiv 0$ on M if and only if $M \subset (\bar{M}, g)$ is totally geodesic.

EXERCISE 5.9.21. Let $f : (M, g) \to \mathbb{R}$ be a smooth function on a Riemannian manifold.

(1) Let $c : (a, b) \to M$ be a geodesic. Compute the first and second derivatives of $f \circ c$.
(2) Use this to show that at a local maximum (or minimum) for f the gradient is zero and the Hessian nonpositive (or nonnegative).

(3) Show that f has everywhere nonnegative Hessian if and only if $f \circ c$ is convex for all geodesics c in (M, g).

EXERCISE 5.9.22. Assume the volume form near a point in a Riemannian manifold is written as $\lambda (r, \theta)\, dr \wedge \mathrm{vol}_{n-1}$, where vol_{n-1} denotes the standard volume form on the unit sphere. Show that $\lambda (r, \theta) = r^{n-1} + O\left(r^{n+1}\right)$.

EXERCISE 5.9.23. Let $N \subset M$ be a properly embedded submanifold of a complete Riemannian manifold (M, g).

(1) The distance from N to $x \in M$ is defined as

$$|xN| = \inf \{|xp| \mid p \in N\}.$$

A unit speed curve $\sigma : [a, b] \to M$ with $\sigma (a) \in N$, $\sigma (b) = x$, and $L (\sigma) = |xN|$ 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 $\dot{\sigma} (a)$ is perpendicular to N.

(2) Show that if N is a closed subset of M and (M, g) is complete, then any point in M can be joined to N by a segment.

(3) Show that in general there is an open neighborhood of N in M where all points are joined to N by segments.

(4) Show that $r (x) = |xN|$ is smooth on a neighborhood of N with N excluded.

(5) Show that the integral curves for ∇r are the geodesics that are perpendicular to N.

EXERCISE 5.9.24. Find the cut locus on a square torus $\mathbb{R}^2/\mathbb{Z}^2$.

EXERCISE 5.9.25. Find the cut locus on a sphere and real projective space with the constant curvature metrics.

EXERCISE 5.9.26. Show that in a Riemannian manifold,

$$\left|\exp_p (v)\ \exp_p (w)\right| = |v - w| + O\left(r^2\right),$$

where $|v|, |w| \leq r$.

EXERCISE 5.9.27. Consider a Riemannian manifold and let $r (x) = |xp|$. Introduce exponential normal coordinates x^i at p.

(1) Show that

$$\left(\mathrm{Hess}\, x^i\right)_{kl} = \Gamma^i_{kl} = O (r).$$

(2) Use that $\frac{1}{2} r^2 = \frac{1}{2} \sum \left(x^i\right)^2$ together with $g = \delta_{ij} + O\left(r^2\right)$ to show that

$$\mathrm{Hess}\, \frac{1}{2} r^2 = g + O\left(r^2\right).$$

(3) Show that

$$\operatorname{Hess} r = \tfrac{1}{r} g_r + O(r).$$

EXERCISE 5.9.28. Let (M, g) be a complete Riemannian manifold; $K \subset M$ a compact (or properly embedded) submanifold; and $r(x) = |xK|$ the distance function to K. The goal is to show that r has well-defined one sided directional derivatives at all points.

(1) Show that if r is differentiable at $x \notin K$, then $\overrightarrow{\overrightarrow{xK}}$ only contains one vector.
(2) Let $c : I \to M$ be a unit speed curve. Show that if $f = r \circ c$ is differentiable at t, then all the vectors $\overrightarrow{c(t)\,K}$ form the same angle with $\dot{c}(t)$.
(3) More generally show that

$$\overline{D}^{+} f(t_0) = \limsup_{t \to t_0^+} \frac{f(t) - f(t_0)}{t - t_0} \leq g\left(\dot{c}(t_0), -\overrightarrow{c(t_0)\,K}\right).$$

Hint: Use the first variation formula for a variation of the segment to K with initial velocity $\overrightarrow{c(t_0)\,K}$.
(4) Show that for small $h = t - t_0$

$$|c(t_0)\,c(t)| = h + O\left(h^2\right).$$

(5) Select a point q on a segment from $c(t)$ to K such that $|c(t)\,q| = h^\alpha$ where $\alpha \in (0, 1)$ and let θ be the angle between $\dot{c}(t)$ and the initial direction $\overrightarrow{c(t)\,K}$ for the segment through q. For small $h = t - t_0 > 0$ justify the following:

$$\begin{aligned}
|c(t_0)\,K| &\leq |c(t_0)\,q| + |qK| \\
&= \sqrt{h^{2\alpha} + h^2 - 2h^{1+\alpha}\cos(\pi - \theta) + O\left(h^{2+\alpha}\right)} + |qK| + O\left(h^{2\alpha}\right) \\
&\leq |c(t)\,q| + |q\,K| - h\cos(\pi - \theta) + \frac{1}{2}h^{2-\alpha} + O\left(h^2\right) + O\left(h^{2\alpha}\right) \\
&= |c(t)\,K| - h\cos(\pi - \theta) + \frac{1}{2}h^{2-\alpha} + O\left(h^2\right) + O\left(h^{2\alpha}\right).
\end{aligned}$$

Hint: Use 5.9.26 and part (4) to estimate $|c(t_0)\,q|$.
(6) Show that for suitable α

$$\underline{D}^{+} f(t_0) = \liminf_{t \to t_0^+} \frac{f(t) - f(t_0)}{t - t_0} \geq \min_{\overrightarrow{c(t_0)\,K}} g\left(\dot{c}(t_0), -\overrightarrow{c(t_0)\,K}\right).$$

(7) Conclude that the right-hand (and left-hand) derivatives of f exist everywhere.

EXERCISE 5.9.29. In a metric space $(X, |\cdot\cdot|)$ one can measure the length of continuous curves $c : [a, b] \to X$ by

$$L(c) = \sup\left\{\sum |c(t_i)\ c(t_{i+1})| \mid a = t_1 \le t_2 \le \cdots \le t_{k-1} \le t_k = b\right\}.$$

(1) Show that a curve has finite length if it is absolutely continuous. Hint: Use the characterization that $c : [a, b] \to X$ is absolutely continuous if and only if for each $\varepsilon > 0$ there is a $\delta > 0$ so that $\sum |c(s_i)\ c(s_{i+1})| \le \varepsilon$ provided $\sum |s_i - s_{i+1}| \le \delta$.
(2) Show that the Cantor step function is a counter example to the converse of (1).
(3) Show that this definition gives back our previous definition for smooth curves on Riemannian manifolds. In fact it will also give us the same length for absolutely continuous curves. Hint: If you know how to prove this in the Euclidean situation, then exercise 5.9.26 helps to approximate with the Riemannian metric.
(4) Let $c : [a, b] \to M$ be an absolutely continuous curve of length $|c(a)\,c(b)|$. Show that $c = \sigma \circ \varphi$ for some segment σ and monotone $\varphi : [0, L] \to [a, b]$.

EXERCISE 5.9.30. Assume that we have coordinates x^i around a point $p \in (M, g)$ such that $x^i(p) = 0$ and $g_{ij}x^j = \delta_{ij}x^j$. Show that these must be exponential normal coordinates. Hint: Define $r = \sqrt{\delta_{ij}x^i x^j}$; show that it is a smooth distance function away from p; and that the integral curves for the gradient are geodesics emanating from p.

EXERCISE 5.9.31. If $N_1, N_2 \subset M$ are totally geodesic submanifolds, show that each component of $N_1 \cap N_2$ is a submanifold which is totally geodesic. Hint: The potential tangent space at $p \in N_1 \cap N_2$ should be the Zariski tangent space $T_pN_1 \cap T_pN_2$.

EXERCISE 5.9.32. Let $F : (M, g) \to (M, g)$ be an isometry that fixes $p \in M$. Show that $DF|_p = -I$ on T_pM if and only if $F^2 = id_M$ and p is an isolated fixed point.

EXERCISE 5.9.33. Show that for a complete manifold the functional distance is the same as the distance.

EXERCISE 5.9.34. Let $c : [0, 1] \to M$ be a geodesic such that $\exp_{c(0)}$ is regular at all $t\dot{c}(0)$ with $t \le 1$. Show that c is a local minimum for the energy functional. Hint: Show that the lift of c via $\exp_{c(0)}$ is a minimizing geodesic in the pull-back metric.

EXERCISE 5.9.35. Consider a Lie group G with a biinvariant pseudo-Riemannian metric.

(1) Show that homomorphisms $\mathbb{R} \to \mathrm{G}$ are precisely the integral curves for left-invariant vector fields through $e \in \mathrm{G}$.
(2) Show that geodesics through the identity are exactly the homomorphisms $\mathbb{R} \to \mathrm{G}$. Conclude that the Lie group exponential map coincides with the exponential map generated by the biinvariant Riemannian metric. The Lie theoretic exponential map $\exp : T_e\mathrm{G} \to \mathrm{G}$ is precisely the map that takes $v \in T_e\mathrm{G}$ to $c(1)$, where $c : \mathbb{R} \to \mathrm{G}$ is the integral curve with $c(0) = e$ for the left-invariant field generated by v.

(3) Show that when the metric is Riemannian, then every element in $x \in \mathrm{G}$ has a square root $y \in \mathrm{G}$ with $y^2 = x$. Hint: This uses metric completeness.
(4) Show that $\mathrm{SL}(n, \mathbb{R})$ does not admit a biinvariant Riemannian metric and compare this to exercise 1.6.28.

EXERCISE 5.9.36. Show that a Riemannian submersion is a submetry.

EXERCISE 5.9.37 (HERMANN). Let $F : (M, g_M) \to (N, g_N)$ be a Riemannian submersion.

(1) Show that (N, g_N) is complete if (M, g_M) is complete.
(2) Show that F is a fibration if (M, g_M) 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_M) is not complete.

EXERCISE 5.9.38. Let S be a set of orientation preserving isometries on a Riemannian manifold (M, g). Show that if all elements in S commute with each other, then each component of $\mathrm{Fix}(S)$ has even codimension.

EXERCISE 5.9.39. A local diffeomorphism $F : (M, g_M) \to (N, g_N)$ is said to be *affine* if $F_*\left(\nabla^M_X Y\right) = \nabla^N_{F_*(X)} F_*(Y)$ for all vector fields X, Y on M.

(1) Show that affine maps take geodesics to geodesics.
(2) Show that given $p \in M$ an affine map F is uniquely determined by $F(p)$ and $DF|_p$.
(3) Give an example of an affine map $\mathbb{R}^n \to \mathbb{R}^n$ that isn't an isometry.

EXERCISE 5.9.40. Consider the real or complex projective space $\mathbb{FP}^n$.

(1) Show that $\mathrm{GL}(n+1, \mathbb{F})$ acts on $\mathbb{FP}^n$ by mapping 1-dimensional subspaces in $\mathbb{F}^{n+1}$ to 1-dimensional subspaces.
(2) Let $\mathrm{H} \subset \mathrm{GL}(n+1, \mathbb{F})$ be the transformations that act trivially. Show that $\mathrm{H} = \{\lambda I_{n+1} \mid \lambda \in \mathbb{F}\}$ and is a normal subgroup of $\mathrm{GL}(n+1, \mathbb{F})$.
(3) Define $\mathrm{PGL}(n+1, \mathbb{F}) = \mathrm{GL}(n+1, \mathbb{F})/\mathrm{H}$. Show that given $p \in \mathbb{FP}^n$ each element $F \in \mathrm{PGL}(n+1, \mathbb{F})$ is uniquely determined by $F(p)$ and $DF|_p$.
(4) Show that there is no Riemannian metric on $\mathbb{FP}^n$ such that this action is by isometries.
(5) Show that the action is by affine transformations with respect to the standard (submersion) metric on $\mathbb{FP}^n$ (see exercise 5.9.39 for the definition of affine transformations).
(6) For a subgroup $\mathrm{G} \subset \mathrm{GL}$, define $\mathrm{PG} = \mathrm{G}/\mathrm{H} \cap \mathrm{G}$. Show that the isometry group of $\mathbb{RP}^n$ is given by $\mathrm{PO}(n+1)$.
(7) Show that the isometry group of $\mathbb{CP}^n$ is given by $\mathrm{PU}(n+1)$.
(8) Show that the isometry group of $H^n(R)$ can be naturally identified with $\mathrm{PO}(n, 1)$.
(9) As in exercise 1.6.9 consider $\mathrm{Iso}(\mathbb{R}^n)$ as the matrix group

$$\mathrm{G} = \left\{ \begin{bmatrix} O & v \\ 0 & 1 \end{bmatrix} \mid O \in \mathrm{O}(n), \ v \in \mathbb{R}^n \right\} \subset \mathrm{GL}(n+1, \mathbb{R}).$$

Show that $\mathrm{PG} = \mathrm{G}$.

EXERCISE 5.9.41. Let $\text{Diff}(M)$ denote the group of diffeomorphisms on a manifold. Define $\text{Diff}(M; K, O) = \{F \in \text{Diff}(M) \mid F(K) \subset O\}$.

(1) Show that finite intersections of $\text{Diff}(M; K, O)$ where K is always compact and O open define a topology. This is the *compact-open topology*.
(2) Show that the compact-open topology is second countable.
(3) When M has a Riemannian structure, show that convergence in the compact-open topology is the same as uniform convergence on compact sets.
(4) Show that a sequence in $\text{Iso}(M, g)$ converges in the compact-open topology if and only if it converges pointwise. Hint: Use the Arzela-Ascoli lemma
(5) Show that $\text{Iso}(M, g)$ is always locally compact in the compact-open topology. Hint: Use the Arzela-Ascoli lemma.
(6) Show that $\text{Iso}_p(M, g)$ is always compact in the compact-open topology.
(7) Show that $\text{Iso}(M, g)$ defines a proper action on M.
(8) Show that for fixed p the evaluation map $F \mapsto \left(F(p), DF|_p\right)$ is continuous on $\text{Iso}(M, g)$. Note that $DF|_p : T_pM \to TM$ so that convergence of the values of the evaluation map makes sense. Hint: Start by showing that $F \mapsto (F(p), F(p_1), \ldots, F(p_n))$ is continuous.
(9) Show that the evaluation map in (8) is a homeomorphism on to its image when restricted to $\text{Iso}(M, g)$.

EXERCISE 5.9.42. Consider exponential normal coordinates around $p \in M$, i.e., $\delta_{ij}x^j = g_{ij}x^j$ and $x^i(p) = 0$. All calculations below are at p.

(1) Show that the second partials of the metric satisfy the Bianchi identity

$$\partial_l \partial_k g_{ji} + \partial_j \partial_l g_{ki} + \partial_k \partial_j g_{li} = 0.$$

Hint: Take three derivatives of the defining relation $x^i = \Sigma_s g_{is} x^s$ as in lemma 5.5.7.
(2) Use all four of these Bianchi identities with the last index being i, j, k, or l to conclude

$$\partial_i \partial_j g_{kl} = \partial_k \partial_l g_{ij}.$$

(3) Use the formula for the curvature tensor in normal coordinates from section 3.1.6 to show

$$R_{ikjl} = \partial_i \partial_j g_{kl} - \partial_i \partial_l g_{jk}.$$

(4) Use (3) and (1) to show

$$\partial_i \partial_j g_{kl} = \frac{1}{3}\left(R_{ikjl} + R_{jkil}\right).$$

(5) Show that we have a Taylor expansion

$$g_{kl} = \delta_{kl} + \frac{1}{3} R_{ikjl} x^i x^j + O\left(|x|^3\right).$$

(6) *(Riemann)* Use the symmetries of the curvature tensor to conclude

$$\begin{aligned}
g &= \sum_{i,j=1}^{n} g_{kl} dx^k dx^l \\
&= \sum_{i=1}^{n} dx^i dx^i \\
&\quad + \frac{1}{12} \sum_{i,j,k,l} R_{ikjl} \left(x^i dx^k - x^k dx^i\right) \left(x^j dx^l - x^l dx^j\right) + O\left(|x|^3\right) \\
&= \sum_{i=1}^{n} dx^i dx^i \\
&\quad + \frac{1}{3} \sum_{i<k,j<l} R_{ikjl} \left(x^i dx^k - x^k dx^i\right) \left(x^j dx^l - x^l dx^j\right) + O\left(|x|^3\right)
\end{aligned}$$

(7) *(Gauss)* Show that in dimension 2 we have

$$\begin{aligned}
g &= dx^2 + dy^2 + \frac{1}{3} R_{1212} (xdy - ydx)^2 + o\left(x^2 + y^2\right) \\
&= dx^2 + dy^2 - \frac{1}{3} \sec(p) (xdy - ydx)^2 + o\left(x^2 + y^2\right).
\end{aligned}$$

Riemann's construction of the curvature tensor proceeded as follows: Start with the normal coordinates, next use the radial isometry property to conclude that the Taylor expansion has the form

$$\begin{aligned}
g &= \sum_{i=1}^{n} dx^i dx^i \\
&\quad + \frac{1}{3} \sum_{i<k,j<l} C_{ikjl} \left(x^i dx^k - x^k dx^i\right) \left(x^j dx^l - x^l dx^j\right) + O\left(|x|^3\right)
\end{aligned}$$

for some tensor C. This tensor has some obvious symmetry properties from the form of the expansion. It is possible to calculate it from the derivatives $\partial_i \partial_j g_{kl}$ provided they satisfy $\partial_i \partial_j g_{kl} = \partial_k \partial_l g_{ij}$. Finally, one has to show that this property is equivalent to the assertion that the above expansion is possible.

EXERCISE 5.9.43. With notation as in the previous exercise show:

(1) $\sqrt{\det(g_{kl})} = 1 - \frac{1}{6}\operatorname{Ric}_{ij} x^i x^j + O\left(|x|^3\right)$.

(2) *(A. Gray)* $\operatorname{vol} B(p,r) = \omega_n r^n \left(1 - \frac{\operatorname{scal}(p)}{6(n+2)} r^2 + O\left(r^3\right)\right)$, where $\omega_n = \operatorname{vol}(B(0,1) \subset \mathbb{R}^n)$. Hint: Use (1) and expand the integral using polar coordinates.

Chapter 6
Sectional Curvature Comparison I

In the previous chapter we classified complete spaces with constant curvature. The goal of this chapter is to compare manifolds with variable curvature to spaces with constant curvature. Our first global result is the Hadamard-Cartan theorem, which says that a simply connected complete manifold with sec ≤ 0 is diffeomorphic to $\mathbb{R}^n$. There are also several interesting restrictions on the topology in positive curvature that we shall investigate, notably, the Bonnet-Myers diameter bound and Synge's theorem stating that an orientable even-dimensional manifold with positive curvature is simply connected. Finally, we also cover the classical quarter pinched sphere theorem of Rauch, Berger, and Klingenberg. In subsequent chapters we deal with some more advanced and modern topics in the theory of manifolds with lower curvature bounds.

We start by introducing the concept of differentiation of vector fields along curves. This generalizes and ties in nicely with mixed second partials from the last chapter and also allows us to define higher order partials. This is then used to define parallel fields, Jacobi fields along geodesics, and finally to establish the second variation formula of Synge.

We also establish some basic comparison estimates that are needed here and later in the text. These results are used to show how geodesics and curvature can help in estimating the injectivity, conjugate, and convexity radii.

6.1 The Connection Along Curves

Recall that in sections 3.2.4 and 3.2.5 we introduced Jacobi and parallel fields for a smooth distance function. Here we will generalize these concepts to allow for Jacobi and parallel fields along a single geodesic, rather than the whole family of geodesics associated to a distance function. This will be quite useful when we study variations.

P. Petersen, *Riemannian Geometry*, Graduate Texts in Mathematics 171,
DOI 10.1007/978-3-319-26654-1_6

6.1.1 Vector Fields Along Curves

Let $c : I \to M$ be a curve in M. A vector field V along c is by definition a map $V : I \to TM$ with $V(t) \in T_{c(t)}M$ for all $t \in I$. The goal is to define the covariant derivative

$$\dot{V}(t) = \frac{d}{dt}V(t) = \nabla_{\dot{c}}V$$

of V along c. We know that V can be thought of as the variational field for a variation $\bar{c} : (-\varepsilon, \varepsilon) \times I \to M$. So it is natural to assume that

$$\frac{d}{dt}V(t) = \frac{\partial^2 \bar{c}}{\partial t \partial s}(0, t) .$$

Doing the calculation in local coordinates (see section 5.1) gives

$$\begin{aligned} V(t) &= V^k(t)\,\partial_k \\ &= \frac{\partial \bar{c}^k}{\partial s}(0, t)\,\partial_k \end{aligned}$$

and

$$\begin{aligned} \frac{\partial^2 \bar{c}}{\partial t \partial s}(0, t) &= \frac{\partial^2 \bar{c}^k}{\partial t \partial s}(0, t)\,\partial_k + \frac{\partial \bar{c}^i}{\partial s}(0, t)\,\frac{\partial \bar{c}^j}{\partial t}(0, t)\,\Gamma_{ij}^k \partial_k \\ &= \frac{dV^k}{dt}(t)\,\partial_k + V^i(t)\,\frac{dc^j}{dt}(t)\,\Gamma_{ij}^k \partial_k . \end{aligned}$$

This shows that $\dot{V}$ does not depend on how the variation was chosen. Since the variation can be selected independently of the coordinate system we see that the local coordinate formula is independent of the coordinate system. The formula also shows that if $V(t) = X_{c(t)}$ for some vector field X defined in a neighborhood of $c(t_0)$, then this derivative is a covariant derivative

$$\dot{V}(t_0) = \nabla_{\dot{c}(t_0)}X.$$

Some caution is necessary when thinking of $\dot{V}$ in this way as it is not in general true that $\dot{V}(t_0) = 0$ when $\dot{c}(t_0) = 0$. It could, e.g., happen that c is the constant curve. In this case $V(t)$ is simply a curve in $T_{c(t_0)}M$ and as such has a well-defined velocity that doesn't have to be zero.

From the product rule for mixed partials (see section 5.1) we get the product rule:

$$\frac{d}{dt}g(V, W) = g\left(\dot{V}, W\right) + g\left(V, \dot{W}\right)$$

for vector fields V, W along c by selecting a two-parameter variation $\bar{c}(s,u,t)$ such that

$$\frac{\partial \bar{c}}{\partial s}(0,0,t) = V(t),$$
$$\frac{\partial \bar{c}}{\partial u}(0,0,t) = W(t).$$

The local coordinate formula also shows that we have:

$$\frac{d}{dt}(V(t)+W(t)) = \frac{d}{dt}V(t) + \frac{d}{dt}W(t),$$
$$\frac{d}{dt}(\lambda(t)V(t)) = \frac{d\lambda}{dt}(t)V(t) + \lambda(t)\frac{dV}{dt}(t),$$

where $\lambda : I \to \mathbb{R}$ is a function.

As with second partials, differentiation along curves can be done in a larger space and then projected on to M. Specifically, if $M \subset \bar{M}$ and $c : I \to M$ is a curve and $V : I \to TM$ a vector field along c, then we can compute $\dot{V} \in T\bar{M}$ and then project $(\dot{V})^{\top} \in TM$ to obtain the derivative of V along c in M. Example 6.1.1 shows what can go wrong if we are not careful about projecting the derivatives.

6.1.2 Third Partials

One of the uses of taking derivatives of vector fields along curves is that we can now define third and higher order partial derivatives. If we wish to compute

$$\frac{\partial^3 c}{\partial s \partial t \partial u}(s_0, t_0, u_0),$$

then consider the vector field $s \mapsto \frac{\partial^2 c}{\partial t \partial u}(s, t_0, u_0) = V(s)$ and define

$$\frac{\partial^3 c}{\partial s \partial t \partial u}(s_0, t_0, u_0) = \frac{dV}{ds}(s_0).$$

Something rather interesting happens with this definition. We expected and proved that second partials commute. This, however, does not carry over to third partials. It is true that

$$\frac{\partial^3 c}{\partial s \partial t \partial u} = \frac{\partial^3 c}{\partial s \partial u \partial t},$$

but if we switch the first two variables the derivatives might be different. One reason we are not entitled to have these derivatives commute lies in the fact that they were defined with a specific order of derivatives in mind.

Example 6.1.1. Let

$$c(t,\theta)=\begin{bmatrix}\cos(t)\\ \sin(t)\cos(\theta)\\ \sin(t)\sin(\theta)\end{bmatrix}$$

be the standard parametrization of $S^2(1)\subset\mathbb{R}^3$ as a surface of revolution around the x-axis. We can compute all derivatives in $\mathbb{R}^3$ and then project them on to $S^2(1)$ in order to find the intrinsic partial derivatives. The curves $t\mapsto c(t,\theta)$ are geodesics. We can see this by direct calculation as

$$\frac{\partial c}{\partial t}=\begin{bmatrix}-\sin(t)\\ \cos(t)\cos(\theta)\\ \cos(t)\sin(\theta)\end{bmatrix}\in TS^2(1),$$

$$\frac{\partial^2 c}{\partial t^2}=\begin{bmatrix}-\cos(t)\\ -\sin(t)\cos(\theta)\\ -\sin(t)\sin(\theta)\end{bmatrix}\in T\mathbb{R}^3.$$

Thus the Euclidean acceleration is proportional to the base point c and so has zero projection onto $S^2(1)$. Next we compute

$$\frac{\partial^2 c}{\partial\theta\partial t}=\begin{bmatrix}0\\ -\cos(t)\sin(\theta)\\ \cos(t)\cos(\theta)\end{bmatrix}\in T\mathbb{R}^3.$$

This vector is tangent to $S^2(1)$ and therefore represents the actual intrinsic mixed partial. Finally we calculate

$$\frac{\partial^3 c}{\partial t\partial\theta\partial t}=\begin{bmatrix}0\\ \sin(t)\sin(\theta)\\ -\sin(t)\cos(\theta)\end{bmatrix}\in T\mathbb{R}^3,$$

$$\frac{\partial^3 c}{\partial\theta\partial t^2}=\begin{bmatrix}0\\ \sin(t)\sin(\theta)\\ -\sin(t)\cos(\theta)\end{bmatrix}\in T\mathbb{R}^3.$$

These are equal as we would expect in $\mathbb{R}^3$. They are also both tangent to $S^2(1)$. The first term is consequently $\frac{\partial^3 c}{\partial t\partial\theta\partial t}$ as computed in $S^2(1)$. The second has no meaning in $S^2(1)$ as we are supposed to first project $\frac{\partial^2 c}{\partial t^2}$ on to $S^2(1)$ before computing $\frac{\partial}{\partial\theta}\frac{\partial^2 c}{\partial t^2}$ in $\mathbb{R}^3$ and then again project to $S^2(1)$. It follows that in $S^2(1)$ we have $\frac{\partial^3 c}{\partial\theta\partial t^2}=0$ while $\frac{\partial^3 c}{\partial t\partial\theta\partial t}\neq 0$.

In this example it is also interesting to note that the equator $t=0$ given by $\theta\mapsto c(0,\theta)$ is a geodesic and that $\frac{\partial^2 c}{\partial\theta\partial t}=0$ along this equator.

We are now ready to prove what happens when the first two partials in a third-order partial are interchanged.

Lemma 6.1.2. *The third mixed partials are related to the curvatures by the formula:*

$$\frac{\partial^3 c}{\partial u \partial s \partial t} - \frac{\partial^3 c}{\partial s \partial u \partial t} = R\left(\frac{\partial c}{\partial u}, \frac{\partial c}{\partial s}\right)\frac{\partial c}{\partial t}.$$

Proof. This result is hardly surprising if we recall the definition of curvature and think of these partial derivatives as covariant derivatives. It is, however, not so clear what happens when the derivatives are not covariant derivatives. We are consequently forced to do the calculation in local coordinates. To simplify matters assume that we are at a point $p = c(u,s,t)$, where $g_{ij}|_p = \delta_{ij}$ and $\Gamma_{ij}^k|_p = 0$. This implies that

$$\frac{\partial}{\partial u}(\partial_i)|_p = 0.$$

Thus

$$\begin{aligned}
\frac{\partial^3 c}{\partial u \partial s \partial t}|_p &= \frac{\partial}{\partial u}\left(\frac{\partial^2 c^l}{\partial s \partial t}\partial_l + \frac{\partial c^i}{\partial t}\frac{\partial c^j}{\partial s}\Gamma_{ij}^l \partial_l\right) \\
&= \frac{\partial^3 c^l}{\partial u \partial s \partial t}\partial_l + \frac{\partial c^i}{\partial t}\frac{\partial c^j}{\partial s}\frac{\partial}{\partial u}\left(\Gamma_{ij}^l\right)\partial_l \\
&= \frac{\partial^3 c^l}{\partial u \partial s \partial t}\partial_l + \frac{\partial c^i}{\partial t}\frac{\partial c^j}{\partial s}\frac{\partial c^k}{\partial u}\left(\partial_k \Gamma_{ij}^l\right)\partial_l, \\
\frac{\partial^3 c}{\partial s \partial u \partial t}|_p &= \frac{\partial^3 c^l}{\partial s \partial u \partial t}\partial_l + \frac{\partial c^i}{\partial t}\frac{\partial c^j}{\partial u}\frac{\partial c^k}{\partial s}\left(\partial_k \Gamma_{ij}^l\right)\partial_l.
\end{aligned}$$

Using our formula for R_{ijk}^l in terms of the Christoffel symbols from section 3.1.6 gives

$$\begin{aligned}
\frac{\partial^3 c}{\partial u \partial s \partial t}|_p - \frac{\partial^3 c}{\partial s \partial u \partial t}|_p &= \frac{\partial c^i}{\partial t}\frac{\partial c^j}{\partial s}\frac{\partial c^k}{\partial u}\left(\partial_k \Gamma_{ij}^l\right)\partial_l - \frac{\partial c^i}{\partial t}\frac{\partial c^j}{\partial u}\frac{\partial c^k}{\partial s}\left(\partial_k \Gamma_{ij}^l\right)\partial_l \\
&= \frac{\partial c^i}{\partial t}\frac{\partial c^j}{\partial s}\frac{\partial c^k}{\partial u}\left(\partial_k \Gamma_{ij}^l\right)\partial_l - \frac{\partial c^i}{\partial t}\frac{\partial c^k}{\partial u}\frac{\partial c^j}{\partial s}\left(\partial_j \Gamma_{ik}^l\right)\partial_l \\
&= \frac{\partial c^i}{\partial t}\frac{\partial c^j}{\partial s}\frac{\partial c^k}{\partial u}\left(\partial_k \Gamma_{ij}^l - \partial_j \Gamma_{ik}^l\right)\partial_l \\
&= \frac{\partial c^i}{\partial t}\frac{\partial c^j}{\partial s}\frac{\partial c^k}{\partial u}\left(\partial_k \Gamma_{ji}^l - \partial_j \Gamma_{ki}^l\right)\partial_l \\
&= \frac{\partial c^i}{\partial t}\frac{\partial c^j}{\partial s}\frac{\partial c^k}{\partial u}R_{kji}^l \partial_l \\
&= R\left(\frac{\partial c}{\partial u}, \frac{\partial c}{\partial s}\right)\frac{\partial c}{\partial t}.
\end{aligned}$$

□

6.1.3 Parallel Transport

A vector field V along c is said to be *parallel along* c provided $\dot{V} \equiv 0$. We know that the tangent field $\dot{c}$ along a geodesic is parallel. We also just saw in example 6.1.1 that the unit field perpendicular to a great circle in $S^2(1)$ is a parallel field.

If V, W are two parallel fields along c, then we clearly have that $g(V, W)$ is constant along c. In particular, parallel fields along a curve neither change their lengths nor their angles relative to each other; just as parallel fields in Euclidean space are of constant length and make constant angles. Based on example 6.1.1 we can pictorially describe parallel translation around certain triangles in $S^2(1)$ (see figure 6.1). Exercise 6.7.2 covers some basic features of parallel translation on surfaces to aid the reader's geometric understanding.

Theorem 6.1.3 (Existence and Uniqueness of Parallel fields). *If $t_0 \in I$ and $v \in T_{c(t_0)}M$, 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 c forming a basis for $T_{c(t)}M$ for all $t \in I$. Any vector field $V(t)$ along c can then be written $V(t) = V^i(t)E_i(t)$ for $V^i : I \to \mathbb{R}$. Thus,

$$\begin{aligned}
\dot{V} = \nabla_{\dot{c}} V &= \sum \dot{V}^i(t)E_i(t) + V^i(t)\nabla_{\dot{c}} E_i \\
&= \sum \dot{V}^j(t)E_j(t) + \sum_{i,j} V^i(t) \cdot \alpha_i^j(t)E_j(t), \text{ where } \nabla_{\dot{c}} E_i = \sum \alpha_i^j(t)E_j \\
&= \sum_j (\dot{V}^j(t) + V^i(t)\alpha_i^j(t))E_j(t).
\end{aligned}$$

Hence, V is parallel if and only if $V^1(t), \ldots, V^n(t)$ satisfy the system of first-order linear differential equations

$$\dot{V}^j(t) = -\sum_{i=1}^{n} \alpha_i^j(t)V^i(t), \ j = 1, \ldots, n.$$

Such systems have the property that for given initial values $V^1(t_0), \ldots, V^n(t_0)$, there is a unique solution defined on all of I with these initial values. □

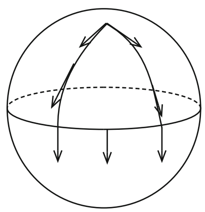


Fig. 6.1 Parallel translation along a spherical triangle

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 in showing that the solution exists on all of I. Nonlinear equations can fail to have solutions over a whole given interval as we saw with geodesics in section 5.2.

Parallel fields can be used as a substitute for Cartesian coordinates. Namely, if we choose a parallel orthonormal frame $E_1(t), \ldots, E_n(t)$ along the curve $c(t) : I \to (M, g)$, then we've seen that any vector field $V(t)$ along c has the property that

$$\begin{aligned}\frac{dV}{dt} &= \frac{d}{dt}\left(V^i(t)E_i(t)\right)\\ &= \dot{V}^i(t)E_i(t) + V^i(t)\cdot\dot{E}_i(t)\\ &= \dot{V}^i(t)E_i(t).\end{aligned}$$

So $\frac{d}{dt}V$, when represented in the coordinates of the frame, is exactly what we would expect. We could more generally choose a tensor T along $c(t)$ of type $(0,p)$ or $(1,p)$ and compute $\frac{d}{dt}T$. For the sake of simplicity, choose a $(1,1)$ tensor S. Then write $S(E_i(t)) = S_i^j(t)E_j(t)$. Thus S is represented by the matrix $\left[S_i^j(t)\right]$ along the curve. As before, we see that $\frac{d}{dt}S$ is represented by $\left[\dot{S}_i^j(t)\right]$.

This makes it possible to understand equations involving only one covariant derivative of the type ∇_X. Let F^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 $E_1, \ldots, E_n$ on H which form an orthonormal frame for the tangent space to M. Finally, construct an orthonormal frame in a neighborhood of p by parallel translating $E_1, \ldots, E_n$ along the integral curves for X. Thus, $\nabla_X E_i = 0$, $i = 1, \ldots, n$. Therefore, if we have a vector field Y near p, we can write $Y = Y^i E_i$ and $\nabla_X Y = D_X(Y^i)E_i$. Similarly, if S is a $(1,1)$-tensor, we have $S(E_i) = S_i^j E_i$, and $\nabla_X S$ is represented by $(D_X(S_i^j))$.

In this way parallel frames make covariant derivatives look like standard derivatives in the same fashion that coordinate vector fields make Lie derivatives look like standard derivatives.

6.1.4 Jacobi Fields

Another variational field that is often quite useful is the field that comes from a *geodesic variation*, i.e., $t \mapsto \bar{c}(s,t)$ is a geodesic for all s. We encountered these fields in section 3.2.4 as vector fields satisfying $L_{\partial_r}J = 0$. Here they need only be defined along a single geodesic so the Lie derivative equation no longer makes sense. The second-order Jacobi equation, however, does make sense in this context:

$$
\begin{aligned}
0 &= \frac{\partial^3 \bar{c}}{\partial s \partial^2 t} \\
&= R\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial \bar{c}}{\partial t}\right)\frac{\partial \bar{c}}{\partial t} + \frac{\partial^3 \bar{c}}{\partial t \partial s \partial t} \\
&= R\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial \bar{c}}{\partial t}\right)\frac{\partial \bar{c}}{\partial t} + \frac{\partial^3 \bar{c}}{\partial^2 t \partial s} \\
&= R\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial \bar{c}}{\partial t}\right)\frac{\partial \bar{c}}{\partial t} + \frac{\partial^2}{\partial t^2}\frac{\partial \bar{c}}{\partial s}.
\end{aligned}
$$

So if the variational field along c is $J(t) = \frac{\partial \bar{c}}{\partial s}(0,t)$, then this field solves the linear second-order *Jacobi Equation*

$$\ddot{J} + R(J, \dot{c})\,\dot{c} = 0.$$

Given $J(0)$ and $\dot{J}(0)$ there will be a unique Jacobi field with these initial conditions as the Jacobi equation is a linear second-order equation. These variational fields are called *Jacobi fields* along c. In case $J(0) = 0$, they can easily be constructed via the geodesic variation

$$\bar{c}(s,t) = \exp_p\left(t\left(\dot{c}(0) + s\dot{J}(0)\right)\right).$$

Since $\bar{c}(s,0) = p$ for all s we must have $J(0) = \frac{\partial \bar{c}}{\partial s}(0,0) = 0$. The derivative is computed as follows

$$
\begin{aligned}
\frac{\partial^2 \bar{c}}{\partial t \partial s}(0,0) &= \frac{\partial^2 \bar{c}}{\partial s \partial t}(0,0) \\
&= \frac{\partial}{\partial s}\left(\dot{c}(0) + s\dot{J}(0)\right)|_{s=0} \\
&= \dot{J}(0).
\end{aligned}
$$

What is particularly interesting about these Jacobi fields is that they control two things we are interesting in studying.

First, observe that they tie in with the differential of the exponential map since

$$
\begin{aligned}
J(t) &= \frac{\partial \bar{c}}{\partial s}(0,t) \\
&= \frac{\partial}{\partial s}\exp_p\left(t\left(\dot{c}(0) + s\dot{J}(0)\right)\right)|_{(0,t)} \\
&= D\exp_p\left(\frac{\partial}{\partial s}\left(t\left(\dot{c}(0) + s\dot{J}(0)\right)\right)|_{(0,t)}\right) \\
&= D\exp_p\left(t\dot{J}(0)\right),
\end{aligned}
$$

where we think of $t\dot{J}(0) \in T_{t\dot{c}(0)}T_pM$. This shows, in particular, that $D\exp_p$ is nonsingular at t_0v if and only if for each vector $J(t_0) \in T_{\exp_p(t_0v)}$ there is a Jacobi field along $t \mapsto \exp_p(tv)$ that vanishes at $t = 0$ and has value $J(t_0)$ at t_0.

Second, Jacobi fields can also be used to calculate the Hessian of the function $r(x) = |xp|$. Assume that $c(t)$ is a unit speed geodesic with $c(0) = p$ and $J(t)$ a Jacobi field along c with $J(0) = 0$. As long as $t\dot{c}(0) \in \text{seg}_p^0$, it follows that $\dot{c}(t) = \nabla r|_{c(t)}$ and consequently:

$$\begin{aligned}
\operatorname{Hess} r(J(t), J(t)) &= g\left(\nabla_{J(t)}\nabla r, J(t)\right) \\
&= g\left(\frac{\partial^2 \bar{c}}{\partial s \partial t}, J\right)|_{(0,t)} \\
&= g\left(\frac{\partial^2 \bar{c}}{\partial t \partial s}, J\right)|_{(0,t)} \\
&= g\left(\dot{J}(t), J(t)\right).
\end{aligned}$$

6.1.5 Second Variation of Energy

Recall from section 5.4 that all geodesics are stationary points for the energy functional. To better understand what happens near a geodesic we do exactly what we would do in calculus, namely, compute the second derivative of any variation of a geodesic.

Theorem 6.1.4 (Synge's second variation formula, 1926). *If $\bar{c} : (-\varepsilon, \varepsilon) \times [a, b]$ is a smooth variation of a geodesic $c(t) = \bar{c}(0, t)$, then*

$$\frac{d^2 E(c_s)}{ds^2}|_{s=0} = \int_a^b \left|\frac{\partial^2 \bar{c}}{\partial t \partial s}\right|^2 dt - \int_a^b g\left(R\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial \bar{c}}{\partial t}\right)\frac{\partial \bar{c}}{\partial t}, \frac{\partial \bar{c}}{\partial s}\right) dt + g\left(\frac{\partial^2 \bar{c}}{\partial s^2}, \frac{\partial \bar{c}}{\partial t}\right)\Bigg|_a^b.$$

Proof. The first variation formula (see lemma 5.4.2) tells us that

$$\frac{dE(c_s)}{ds} = -\int_a^b g\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial^2 \bar{c}}{\partial t^2}\right) dt + g\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial \bar{c}}{\partial t}\right)\Bigg|_{(s,a)}^{(s,b)}.$$

With this in mind we can calculate

$$\begin{aligned}
\frac{\partial^2 E(c_s)}{\partial s^2} &= -\frac{\partial}{\partial s}\int_a^b g\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial^2 \bar{c}}{\partial t^2}\right) dt + \frac{\partial}{\partial s} g\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial \bar{c}}{\partial t}\right)\Bigg|_{(s,a)}^{(s,b)} \\
&= -\int_a^b g\left(\frac{\partial^2 \bar{c}}{\partial s^2}, \frac{\partial^2 \bar{c}}{\partial t^2}\right) dt - \int_a^b g\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial^3 \bar{c}}{\partial s \partial t^2}\right) dt \\
&\quad + g\left(\frac{\partial^2 \bar{c}}{\partial s^2}, \frac{\partial \bar{c}}{\partial t}\right)\Bigg|_{(s,a)}^{(s,b)} + g\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial^2 \bar{c}}{\partial s \partial t}\right)\Bigg|_{(s,a)}^{(s,b)}.
\end{aligned}$$

Setting $s = 0$ and using that $c(0,t)$ is a geodesic we obtain

$$
\begin{aligned}
&\frac{\partial^2 E(c_s)}{\partial s^2}|_{s=0} \\
&= -\int_a^b g\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial^3 \bar{c}}{\partial s \partial t^2}\right) dt + g\left(\frac{\partial^2 \bar{c}}{\partial s^2}, \frac{\partial \bar{c}}{\partial t}\right)\Big|_a^b + g\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial^2 \bar{c}}{\partial s \partial t}\right)\Big|_a^b \\
&= -\int_a^b g\left(\frac{\partial \bar{c}}{\partial s}, R\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial \bar{c}}{\partial t}\right)\frac{\partial \bar{c}}{\partial t}\right) dt - \int_a^b g\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial^3 \bar{c}}{\partial t \partial s \partial t}\right) dt \\
&\quad + g\left(\frac{\partial^2 \bar{c}}{\partial s^2}, \frac{\partial \bar{c}}{\partial t}\right)\Big|_a^b + g\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial^2 \bar{c}}{\partial s \partial t}\right)\Big|_a^b \\
&= -\int_a^b g\left(\frac{\partial \bar{c}}{\partial s}, R\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial \bar{c}}{\partial t}\right)\frac{\partial \bar{c}}{\partial t}\right) dt + \int_a^b g\left(\frac{\partial^2 \bar{c}}{\partial t \partial s}, \frac{\partial^2 \bar{c}}{\partial s \partial t}\right) dt \\
&\quad - \int_a^b \frac{\partial}{\partial t} g\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial^2 \bar{c}}{\partial s \partial t}\right) dt + g\left(\frac{\partial^2 \bar{c}}{\partial s^2}, \frac{\partial \bar{c}}{\partial t}\right)\Big|_a^b + g\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial^2 \bar{c}}{\partial s \partial t}\right)\Big|_a^b \\
&= -\int_a^b g\left(\frac{\partial \bar{c}}{\partial s}, R\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial \bar{c}}{\partial t}\right)\frac{\partial \bar{c}}{\partial t}\right) dt + \int_a^b g\left(\frac{\partial^2 \bar{c}}{\partial t \partial s}, \frac{\partial^2 \bar{c}}{\partial s \partial t}\right) dt \\
&\quad - g\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial^2 \bar{c}}{\partial s \partial t}\right)\Big|_a^b + g\left(\frac{\partial^2 \bar{c}}{\partial s^2}, \frac{\partial \bar{c}}{\partial t}\right)\Big|_a^b + g\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial^2 \bar{c}}{\partial s \partial t}\right)\Big|_a^b \\
&= \int_a^b g\left(\frac{\partial^2 \bar{c}}{\partial t \partial s}, \frac{\partial^2 \bar{c}}{\partial t \partial s}\right) dt - \int_a^b g\left(R\left(\frac{\partial \bar{c}}{\partial s}, \frac{\partial \bar{c}}{\partial t}\right)\frac{\partial \bar{c}}{\partial t}, \frac{\partial \bar{c}}{\partial s}\right) dt + g\left(\frac{\partial^2 \bar{c}}{\partial s^2}, \frac{\partial \bar{c}}{\partial t}\right)\Big|_a^b
\end{aligned}
$$

□

The formula is going to be used in different ways below. First we observe that for proper variations the last term drops out and the formula depends only on the variational field $V(t) = \frac{\partial \bar{c}}{\partial s}(0,t)$ and the velocity field $\dot{c}$ of the original geodesic:

$$\frac{d^2 E(c_s)}{ds^2}|_{s=0} = \int_a^b \left|\dot{V}\right|^2 dt - \int_a^b g(R(V,\dot{c})\dot{c}, V)\, dt.$$

Another special case occurs when the variational field is parallel $\dot{V} = 0$. In this case the first term drops out:

$$\frac{d^2 E(c_s)}{ds^2}|_{s=0} = -\int_a^b g(R(V,\dot{c})\dot{c}, V)\, dt + g\left(\frac{\partial^2 \bar{c}}{\partial s^2}, \dot{c}\right)\Big|_a^b$$

but the formula still depends on the variation and not just on V. If, however, we select the variation such that $s \mapsto \bar{c}(s,t)$ are geodesics, then the last term also drops out.

6.2 Nonpositive Sectional Curvature

In this section we show that the exponential map $\exp_p : T_pM \to M$ is a covering map, provided (M, g) is complete and has nonpositive sectional curvature everywhere. This implies, in particular, that no compact simply connected manifold admits such a metric. We shall also prove some interesting results about the fundamental groups of such manifolds.

The first observation about manifolds with nonpositive curvature is that any geodesic from p to q must be a local minimum for $E : \Omega(p, q) \to [0, \infty)$ by our second variation formula. This is in sharp contrast to what we shall prove in positive curvature, where sufficiently long geodesics can never be local minima.

Recall from our discussion of the fundamental equations in section 3.2 and 3.2.4 that Jacobi fields seem particularly well-suited for the task of studying nonpositive curvature. This will be borne out here and later in section 6.4.

6.2.1 *Manifolds Without Conjugate Points*

We start with a result that gives strong restrictions on the behavior of the exponential map.

Lemma 6.2.1. *If* $\exp_p : T_pM \to M$ *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 local Riemannian isometry. We then know from lemma 5.6.4 that $\exp_p$ is a covering map provided this new metric on T_pM is complete. To see this, simply observe that the metric is geodesically complete at the origin, since straight lines through the origin are still geodesics. □

We can now prove our first big result. It was originally established by Mangoldt for surfaces. Hadamard in a survey article offered a different proof. Cartan extended the result to higher dimensions under the assumption that the manifold is metrically complete.

Theorem 6.2.2 (Mangoldt, 1881, Hadamard, 1889, and Cartan, 1925). *If* (M, g) *is complete, connected, and has* $\sec \leq 0$*, then the universal covering is diffeomorphic to* $\mathbb{R}^n$.

Proof. The goal is to show that $\left|D\exp_p(w)\right| > 0$ for all nonzero $w \in T_vT_pM$. This will imply that $\exp_p$ is nonsingular everywhere and hence a covering map.

Select a Jacobi field J along $c(t) = \exp_p(tv)$ such that $J(0) = 0$ and $\dot{J}(0) = w$ so that $\left|D\exp_p(w)\right| = |J(1)|$. Consider the function $t \mapsto \frac{1}{2}|J(t)|^2$ and its first and second derivatives:

$$\begin{aligned}
\frac{d}{dt}\left(\frac{1}{2}|J(t)|^2\right) &= g\left(\dot{J}, J\right), \\
\frac{d^2}{dt^2}\left(\frac{1}{2}|J(t)|^2\right) &= \frac{d}{dt} g\left(\dot{J}, J\right) \\
&= g\left(\ddot{J}, J\right) + g\left(\dot{J}, \dot{J}\right) \\
&= -g\left(R(J, \dot{c})\dot{c}, J\right) + \left|\dot{J}\right|^2 \\
&\geq \left|\dot{J}\right|^2.
\end{aligned}$$

The last inequality follows from the assumption that $g(R(x,y)y,x) \leq 0$ for all tangent vectors x, y. Integrating this inequality gives

$$\begin{aligned}
g\left(\dot{J}, J\right) &\geq \int_0^t \left|\dot{J}\right|^2 dt + g\left(\dot{J}(0), J(0)\right) \\
&= \int_0^t \left|\dot{J}\right|^2 dt \\
&> 0
\end{aligned}$$

unless $\dot{J}(t) = 0$ for all t, in which case $\dot{J}(0) = w = 0$. Assuming $w \neq 0$, integrating the last inequality yields

$$\frac{1}{2}|J(t)|^2 > 0,$$

which is what we wanted to prove. □

No similar theorem can hold for Riemannian manifolds with Ric ≤ 0 or scal ≤ 0, since we saw in sections 4.2.3 and 4.2.5 that there exist Ricci flat metrics on $\mathbb{R}^2 \times S^{n-2}$ and scalar flat metrics on $\mathbb{R} \times S^{n-1}$.

6.2.2 The Fundamental Group in Nonpositive Curvature

We are going to prove two results on the structure of the fundamental group for manifolds with nonpositive curvature. The interested reader is referred to the book by Eberlein [38] for further results on manifolds with nonpositive curvature.

First we need a little preparation. Let (M, g) be a complete simply connected Riemannian manifold of nonpositive curvature. The two key properties we use are that any two points in M lie on a unique geodesic, and that distance functions are everywhere smooth and convex.

We just saw that $\exp_p : T_pM \to M$ is a diffeomorphism for all $p \in M$. This shows, as in Euclidean space, that there is only one geodesic through p and $q\ (\neq p)$.

This also shows that the distance function $|xp|$ is smooth on $M-\{p\}$. The modified distance function

$$x \mapsto f_0(x) = f_{0,p}(x) = \frac{1}{2}|xp|^2 = \frac{1}{2}(r(x))^2$$

is then smooth everywhere and its Hessian is given by

$$\operatorname{Hess} f_0 = dr^2 + r \operatorname{Hess} r.$$

If $J(t)$ is a Jacobi field along a unit speed geodesic emanating from p with $J(0) = 0$, then from section 6.1.4

$$\begin{aligned} \operatorname{Hess} r(J(b), J(b)) &= g\left(\dot{J}(b), J(b)\right) \\ &\geq \int_0^b \left|\dot{J}\right|^2 dt \\ &> 0. \end{aligned}$$

Since $J(b)$ can be arbitrary we have shown that the Hessian is positive definite. If c is a geodesic, this implies that $f_0 \circ c$ is convex as

$$\begin{aligned} \frac{d}{dt} f_0 \circ c &= g(\nabla f_0, \dot{c}), \\ \frac{d^2}{dt^2} f_0 \circ c &= \frac{d}{dt} g(\nabla f_0, \dot{c}) \\ &= g(\nabla_{\dot{c}} \nabla f_0, \dot{c}) + g(\nabla f_0, \ddot{c}) \\ &= \operatorname{Hess} f_0(\dot{c}, \dot{c}) \\ &> 0. \end{aligned}$$

With this in mind we can generalize the idea of convexity slightly (see also section 7.1.3). A function is *(strictly) convex* if its restriction to all geodesics is (strictly) convex. One sees that the maximum of any collection 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 in particular consider the strictly convex function

$$x \mapsto \max\left\{f_{0,p_1}(x), \ldots, f_{0,p_k}(x)\right\}.$$

In general, any proper, nonnegative, and strictly convex function has a unique minimum. To see this, first note that there must be a minimum as the function is proper and bounded from below. If there were two minima, then the function would be strictly convex when restricted to a geodesic joining these two minima. But then the function would have smaller values on the interior of this segment than at the endpoints.

The uniquely defined minimum for

$$x \mapsto \max\left\{f_{0,p_1}(x), \ldots f_{0,p_k}(x)\right\}$$

is denoted by $\mathrm{cm}_\infty \{p_1, \ldots, p_k\}$ and called the L^∞ *center of mass* of $\{p_1, \ldots, p_k\}$. It is the center q of the smallest ball $\overline{B}(q, R) \supset \{p_1, \ldots, p_k\}$. If instead we had considered

$$x \mapsto \sum_{i=1}^{k} f_{0,p_i}(x)$$

we would have arrived at the usual center of mass also known as the L^2 center of mass.

The first theorem is concerned with fixed points of isometries.

Theorem 6.2.3 (Cartan, 1925). *If (M, g) is a complete simply connected Riemannian manifold of nonpositive curvature, then any isometry $F : M \to M$ 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. First, define the order of F as the smallest integer k such that $F^k = id$. Second, for any $p \in M$ consider the orbit $\left\{p, F(p), \ldots, F^{k-1}(p)\right\}$ of p. Then construct the center of mass

$$q = \mathrm{cm}_\infty \left\{p, F(p), \ldots, F^{k-1}(p)\right\}.$$

We claim that $F(q) = q$. This is because the function

$$x \mapsto f(x) = \max\left\{f_{0,p}(x), \ldots f_{0,F^{k-1}(p)}(x)\right\}$$

has not only q as a minimum, but also $F(q)$. To see this just observe that since F is an isometry, we have

$$\begin{aligned}
f(F(q)) &= \max\left\{f_{0,p}(F(q)), \ldots f_{0,F^{k-1}(p)}(F(q))\right\} \\
&= \frac{1}{2}\left(\max\left\{|F(q)\,p|, \ldots, \left|F(q)\,F^{k-1}(p)\right|\right\}\right)^2 \\
&= \frac{1}{2}\left(\max\left\{\left|F(q)\,F^{k}(p)\right|, \ldots, \left|F(q)\,F^{k-1}(p)\right|\right\}\right)^2 \\
&= \frac{1}{2}\left(\max\left\{\left|qF^{k-1}(p)\right|, \ldots, \left|qF^{k-2}(p)\right|\right\}\right)^2 \\
&= f(q).
\end{aligned}$$

The uniqueness of minima for strictly convex functions now implies $F(q) = q$. □

Corollary 6.2.4. *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 more preparation and a 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: $x \mapsto \frac{1}{2}r^2 = f_0(x)$ and a unit speed geodesic $c : \mathbb{R} \to M$. The Hessian estimate from above only implies that $\frac{d^2}{dt^2}(f_0 \circ c) > 0$. However, we know that this second derivative is 1 in Euclidean space. So it shouldn't be surprising that we have a much better quantitative estimate.

Lemma 6.2.5. *If (M, g) has nonpositive curvature, then any modified distance function satisfies:*

$$\operatorname{Hess} f_0 \geq g.$$

Proof. We follow the notation in the proof of theorem 6.2.2. If $r(x) = |xp|$, then

$$\operatorname{Hess} \frac{1}{2}r^2 = dr^2 + r \operatorname{Hess} r.$$

So the claim follows if we can show that

$$r \operatorname{Hess} r \geq g_r,$$

where $g = dr^2 + g_r$. This estimate in turn holds if we can prove that

$$\begin{aligned} t \cdot \operatorname{Hess} r\,(J(t), J(t)) &= t \cdot g\left(\dot{J}(t), J(t)\right) \\ &\geq g(J(t), J(t)). \end{aligned}$$

The reason behind the proof of this is slightly tricky and is known as *Jacobi field comparison*. Consider the ratio

$$\lambda(t) = \frac{|J(t)|^2}{g\left(\dot{J}(t), J(t)\right)}.$$

By l'Hospital's rule it follows that

$$\lambda(0) = \frac{2g\left(\dot{J}(0), J(0)\right)}{-g\left(R(J(0), \dot{c}(0))\,\dot{c}(0), J(0)\right) + \left|\dot{J}(0)\right|^2} = \frac{0}{\left|\dot{J}(0)\right|^2} = 0.$$

Using that the sectional curvature is nonpositive and then the Cauchy-Schwarz inequality it follows that the derivative satisfies

$$\begin{aligned}\dot{\lambda}(t) &= \frac{2\left(g\left(J,\dot{J}\right)\right)^2 - \left|\dot{J}\right|^2 |J|^2 + g\left(R\left(J,\dot{c}\right)\dot{c},J\right)|J|^2}{\left(g\left(J,\dot{J}\right)\right)^2} \\ &\leq \frac{2\left(g\left(J,\dot{J}\right)\right)^2 - \left|\dot{J}\right|^2 |J|^2}{\left(g\left(J,\dot{J}\right)\right)^2} \\ &\leq \frac{2\left(g\left(J,\dot{J}\right)\right)^2 - \left(g\left(J,\dot{J}\right)\right)^2}{\left(g\left(J,\dot{J}\right)\right)^2} \\ &= 1.\end{aligned}$$

Hence $\lambda(t) \leq t$ and $t \cdot g\left(J(t), \dot{J}(t)\right) \geq |J(t)|^2$. □

Integrating the inequality $\frac{d^2}{dt^2}\left(f_{0,p} \circ c\right) \geq 1$, where c is a unit speed geodesic, yields

$$\begin{aligned}|pc(t)|^2 &\geq |pc(0)|^2 + 2g\left(\nabla f_{0,p}, \dot{c}(0)\right) \cdot t + t^2 \\ &= |pc(0)|^2 + |c(0)c(t)|^2 \\ &\quad +2|pc(0)||c(0)c(t)|\cos\angle\left(\nabla f_{0,p}, \dot{c}(0)\right).\end{aligned}$$

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 6.2.

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 c. In particular, the angle sum in any nondegenerate quadrilateral is $< 2\pi$. This will be crucial for the proof of the next theorem

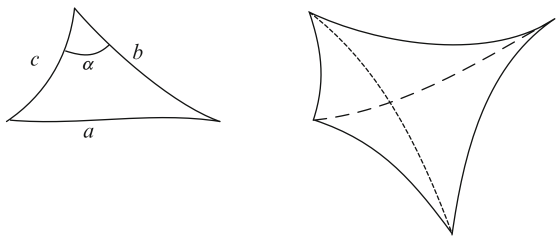


Fig. 6.2 Triangle and quadrilateral in negative curvature

Theorem 6.2.6 (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.*

The proof requires some preliminary results that can also be used in other contexts as they do not assume that the manifold has nonpositive curvature.

An *axis* for an isometry $F : M \to M$ is a geodesic $c : \mathbb{R} \to M$ such that $F(c)$ is a reparametrization of c. Since isometries map geodesics to geodesics, it must follow that

$$F \circ c(t) = c(\pm t + a).$$

Note that if $-$ occurs, then $c\left(\frac{a}{2}\right)$ is fixed by F. When $F \circ c(t) = c(t + a)$ we call a the *period* of F with respect to c. The period depends on the parametrization of c.

Given an isometry $F : M \to M$ the *displacement function* is defined as

$$x \mapsto \delta_F(x) = |xF(x)|.$$

Lemma 6.2.7. *Let $F : M \to M$ be an isometry on a complete Riemannian manifold. If the displacement function δ_F has a positive minimum, then F has an axis.*

Proof. Let δ_F have a minimum at $p \in M$ and $c : [0, 1] \to M$ be a segment from p to $F(p)$. Then $F \circ c$ is a segment from $F(p)$ to $F^2(p)$ with the same speed. We claim that these two geodesics form an angle π at $F(p)$ and thus fit together as the geodesic extension of c to $[0, 2]$. If we fix $t \in [0, 1]$, then

$$\begin{aligned}
\delta_F(p) &\leq \delta_F(c(t)) \\
&= |c(t)(F \circ c)(t)| \\
&\leq |c(t)c(1)| + |c(1)(F \circ c)(t)| \\
&= |c(t)c(1)| + |(F \circ c)(0)(F \circ c)(t)| \\
&= |c(t)c(1)| + |c(0)c(t)| \\
&= |c(0)c(1)| \\
&= |pF(p)|.
\end{aligned}$$

This means that the curve that consists of $c|_{[t,1]}$ followed by $F \circ c|_{[0,t]}$ must be a segment and thus a geodesic by corollary 5.4.4 (see also figure 6.3). This geodesic is obviously just the extension of c, so $(F \circ c)(t) = c(1 + t)$. We can repeat this argument forwards and backwards along the extension of c to $\mathbb{R}$ to show that it becomes an axis for F of period 1. □

Let $\pi : \tilde{M} \to M$ be the universal cover of M. A *deck transformation* $F : \tilde{M} \to \tilde{M}$ is a map such that $\pi \circ F = \pi$, i.e., a lift of π. As such, it is determined by the value of $F(p) \in \pi^{-1}(q)$ for a given $p \in \pi^{-1}(q)$. We can think of the fundamental group $\pi_1(M, q)$ as acting by deck transformations: Given $p \in \pi^{-1}(q)$, a loop in $[\alpha] \in \pi_1(M, q)$ yields a deck transformation with $F(p) = \tilde{\alpha}(1)$, where $\tilde{\alpha}$ is the lift of α such that $p = \tilde{\alpha}(0)$. Finally note that in the Riemannian setting deck transformations are isometries since $\pi : \tilde{M} \to M$ is a local isometry.

Lemma 6.2.8. *If $F : \tilde{M} \to \tilde{M}$ is a nontrivial deck transformation on the universal cover over a compact base M, then the dilation δ_F has a positive minimum. The axis corresponding to this minimum is mapped to a closed geodesic in M whose length is minimal in its free homotopy class. Moreover, $\delta_F(x) \geq 2\,\mathrm{inj}(M)$.*

Proof. Fix a nontrivial deck transformation $F : \tilde{M} \to \tilde{M}$. We start by characterizing the loops in M generated by F. First we show that when $x_i \in \tilde{M}$, $i = 0, 1$ are joined to $F(x_i)$ by curves $c_i : [0, 1] \to \tilde{M}$, then the loops $\pi \circ c_i$ are freely homotopic through a homotopy of loops in M. To see this choose a path $H(s, 0) : [0, 1] \to \tilde{M}$ with $H(i, 0) = x_i$, $i = 0, 1$. Then define $H(s, 1) = F(H(s, 0))$ and $H(i, t) = c_i(t)$, $i = 0, 1$. This defines H on $\partial\left([0, 1]^2\right)$. Simple connectivity of $\tilde{M}$ shows this can be extended to a map $H : [0, 1]^2 \to \tilde{M}$. Now $\pi(H(s, t))$ is the desired homotopy in M since

$$\pi(H(s, 1)) = \pi \circ F(H(s, 0)) = \pi(H(s, 0)).$$

Conversely we claim that any loop at $\pi(x_1) \in M$ that is freely homotopic through loops to $\pi \circ c_0$ must lift to a curve from x_1 to $F(x_1)$. Let $H : [0, 1]^2 \to M$ be such a homotopy, i.e., $H(0, t) = (\pi \circ c_0)(t)$, $H(s, 0) = H(s, 1)$, and $H(1, 0) = \pi(x_1)$. Let $\tilde{H}$ be the lift of H to $\tilde{M}$ such that $\tilde{H}(0, 0) = x_0$. Unique path lifting guarantees that $c_0(t) = \tilde{H}(0, t)$. Now both $\tilde{H}(s, 0)$ and $\tilde{H}(s, 1)$ are lifts of the same curve $H(s, 0)$. As F is a deck transformation $\left(F \circ \tilde{H}\right)(s, 0)$ is also a lift of $H(s, 0)$. However, $\left(F \circ \tilde{H}\right)(0, 0) = \tilde{H}(0, 1)$ so it follows that $\left(F \circ \tilde{H}\right)(s, 0) = \tilde{H}(s, 1)$. Letting $s = 1$ gives the claim.

In particular, we have shown that if F is nontrivial, then none of these loops can be homotopically trivial. This implies that $\delta_F(x) \geq 2\,\mathrm{inj}_{\pi(x)}(M)$, as otherwise the segment from x to $F(x)$ would generate a loop of length $< 2\,\mathrm{inj}_{\pi(x)}(M)$. However, such loops are contractible as they lie in $B\left(\pi(x), \mathrm{inj}_{\pi(x)}(M)\right)$.

We are now ready to minimize the dilatation. Consider a sequence $q_i \in \tilde{M}$ such that $\lim \delta_F(q_i) = \inf \delta_F \geq \mathrm{inj}\, M$ and with it a sequence of segments $\tilde{c}_i : [0, 1] \to \tilde{M}$ with $c_i(0) = q_i$ and $c_i(1) = F(q_i)$. Let $c_i = \pi \circ \tilde{c}_i$ be the corresponding loops in M. Since $|\dot{c}_i| = \delta_F(q_i)$, compactness of M implies that after possibly passing to a subsequence we can assume that $\dot{c}_i(0)$ converge to a vector $v \in T_qM$ where $q = \lim c_i(0)$ and $|v| = \inf \delta_F$. Continuity of the exponential map implies that the curves c_i converge to the geodesic $c(t) = \exp_q(tv)$. This geodesic is in turn a loop at q that is freely homotopic through loops to c_i for large i; because when

$|c_i(t)\,c(t)| < \operatorname{inj}(M)$, they can be joined by unique short geodesics resulting in a homotopy. The above characterization of loops generated by F, then shows that any lift $\tilde{c}$ of c must satisfy $F\left(c\,\tilde{(0)}\right) = \tilde{c}(1)$. All in all,

$$\delta_F\left(c\,\tilde{(0)}\right) \le L(\tilde{c}) = L(c) = |v| = \inf \delta_F.$$

It is clear that c has minimal length in its free homotopy class. A simple application of the first variation formula (see 5.4.2) then shows that it must be a closed geodesic. □

These preliminaries allow us to prove the theorem.

Proof of Theorem 6.2.6. We know that nontrivial deck transformations have axes.

To see that axes are unique in negative curvature, assume that we have two different axes c_1 and c_2 for F. If these intersect in one point, they must, by virtue of being invariant under F, intersect in at least two points. But then they must be equal. Thus they do not intersect. Select $p_1 \in c_1$ and $p_2 \in c_2$, and join these points by a segment σ. Then $F \circ \sigma$ is a segment from $F(p_1)$ to $F(p_2)$. Since F is an isometry that preserves c_1 and c_2, we see that the adjacent angles along the two axes formed by the quadrilateral $p_1, p_2, F(p_1), F(p_2)$ 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.

Finally pick a deck transformation G that commutes with F. If 1 is the period, then

$$(G \circ c)(t+1) = (G \circ F \circ c)(t) = (F \circ G \circ c)(t).$$

This implies that $G \circ c$ is an axis for F, and so must be c itself. Next consider the group H generated by F, G. Any element in this group has c as an axis. Thus we get a map $\mathrm{H} \to \mathbb{R}$ that sends an isometry to its uniquely defined period. This map is a homomorphism with trivial kernel. Consider an additive subgroup $\mathrm{A} \subset \mathbb{R}$ and let $a = \inf\{x \in \mathrm{A} \mid x > 0\}$. It is easy to check that if $a = 0$, then A is dense, while if $a > 0$, then $\mathrm{A} = \{na \mid n \in \mathbb{Z}\}$. The image of H in $\mathbb{R}$ must have the second property as no nonzero period along c can be smaller than $\frac{\operatorname{inj} M}{|\dot{c}|}$. This shows that H is cyclic. □

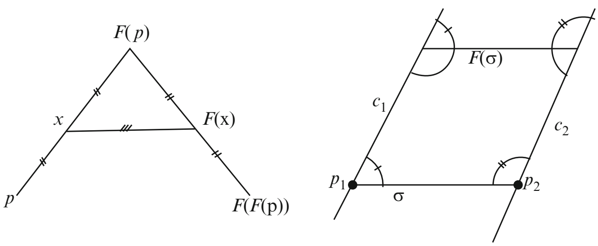


Fig. 6.3 Dilatation and axes

6.3 Positive Curvature

In this section we establish several of the classical results for manifolds with positive curvature. In contrast to 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.

In our discussion of the fundamental equations in section 3.2 we saw that using parallel fields most easily gave useful information about Hessians of distance functions when the curvature is nonnegative. This will be confirmed here through the use of suitable variational fields to find the second variation of energy. In section 6.5 below we show how more sophisticated techniques can be used in conjunction with the developments here to establish stronger results.

6.3.1 The Diameter Estimate

Our first restriction on positively curved manifolds is an estimate for how long minimal geodesics can be. It was first proven by Bonnet for surfaces and later by Synge for general Riemannian manifolds as an application of his second variation formula.

Lemma 6.3.1 (Bonnet, 1855 and Synge, 1926). *If* (M, g) *satisfies* $\sec \geq k > 0$, *then geodesics of length* $> \pi/\sqrt{k}$ *cannot be locally minimizing.*

Proof. Let $c : [0, l] \to M$ be a unit speed geodesic of length $l > \pi/\sqrt{k}$. Along c consider the variational field

$$V(t) = \sin\left(\frac{\pi}{l}t\right) E(t),$$

where E is a unit parallel field perpendicular to c. Since V vanishes at $t = 0$ and $t = l$, it corresponds to a proper variation. By theorem 6.1.4 the second derivative of this variation is

$$\begin{aligned}
\frac{d^2E}{ds^2}|_{s=0} &= \int_0^l \left|\dot{V}\right|^2 dt - \int_0^l g(R(V, \dot{c})\dot{c}, V)\, dt \\
&= \int_0^l \left|\frac{\pi}{l}\cos\left(\frac{\pi}{l}t\right) E(t)\right|^2 dt \\
&\quad - \int_0^l g\left(R\left(\sin\left(\frac{\pi}{l}t\right) E(t), \dot{c}\right)\dot{c}, \sin\left(\frac{\pi}{l}t\right) E(t)\right) dt \\
&= \left(\frac{\pi}{l}\right)^2 \int_0^l \cos^2\left(\frac{\pi}{l}t\right) dt - \int_0^l \sin^2\left(\frac{\pi}{l}t\right) \sec(E, \dot{c})\, dt
\end{aligned}$$

$$
\begin{aligned}
&\leq \left(\frac{\pi}{l}\right)^2 \int_0^l \cos^2\left(\frac{\pi}{l}t\right) dt - k \int_0^l \sin^2\left(\frac{\pi}{l}t\right) dt \\
&< k \int_0^l \cos^2\left(\frac{\pi}{l}t\right) dt - k \int_0^l \sin^2\left(\frac{\pi}{l}t\right) dt \\
&= 0.
\end{aligned}
$$

Thus all nearby curves in the variation are shorter than c. □

The next result is a very interesting and completely elementary consequence of the above result. It seems to have been pointed out first by Hopf-Rinow for surfaces in their famous paper on completeness and soon after by Myers for general Riemannian manifolds.

Corollary 6.3.2 (Hopf and Rinow, 1931 and Myers, 1932). *If* (M, g) *is complete and satisfies* $\sec \geq k > 0$*, then* M *is compact and* $\operatorname{diam}(M, g) \leq {}^{\pi}/\sqrt{k} = \operatorname{diam} S^n_k$*. In particular,* M *has finite fundamental group.*

Proof. As no geodesic of length $> {}^{\pi}/\sqrt{k}$ can realize the distance between endpoints and M is complete, the diameter cannot exceed ${}^{\pi}/\sqrt{k}$. Finally use that the universal cover has the same curvature condition to conclude that it must also be compact. Thus, the fundamental group is finite. □

These results were later extended to manifolds with positive Ricci curvature by Myers.

Theorem 6.3.3 (Myers, 1941). *If* (M, g) *is a complete Riemannian manifold with* $\operatorname{Ric} \geq (n-1)k > 0$*, then* $\operatorname{diam}(M, g) \leq {}^{\pi}/\sqrt{k}$*. Furthermore,* (M, g) *has finite fundamental group.*

Proof. It suffices to show as before that no geodesic of length $> {}^{\pi}/\sqrt{k}$ can be minimal. If $c : [0, l] \to M$ is the geodesic we can select $n-1$ variational fields

$$
V_i(t) = \sin\left(\frac{\pi}{l}t\right) E_i(t), i = 2, \ldots, n
$$

as before. This time we also assume that $\dot{c}, E_2, \ldots E_n$ form an orthonormal basis for $T_{c(t)}M$. By adding up the contributions to the second variation formula for each variational field we get

$$
\begin{aligned}
\sum_{i=2}^n \frac{d^2E}{ds^2}|_{s=0} &= \sum_{i=2}^n \int_0^l \left|\dot{V}_i\right|^2 dt - \int_0^l g\left(R\left(V_i, \dot{c}\right)\dot{c}, V_i\right) dt \\
&= (n-1)\left(\frac{\pi}{l}\right)^2 \int_0^l \cos^2\left(\frac{\pi}{l}t\right) \\
&\quad - \sum_{i=2}^n \int_0^l \sin^2\left(\frac{\pi}{l}t\right) \sec\left(E_i, \dot{c}\right) dt
\end{aligned}
$$

$$\begin{aligned}
&= (n-1)\left(\frac{\pi}{l}\right)^2 \int_0^l \cos^2\left(\frac{\pi}{l}t\right) dt - \int_0^l \sin^2\left(\frac{\pi}{l}t\right) \operatorname{Ric}(\dot{c}, \dot{c})\, dt \\
&< (n-1)\,k \int_0^l \cos^2\left(\frac{\pi}{l}t\right) dt - (n-1)\,k \int_0^l \sin^2\left(\frac{\pi}{l}t\right) dt \\
&< 0.
\end{aligned}$$

Thus the second variation is negative for at least one of the variational fields. □

Example 6.3.4. The incomplete Riemannian manifold $S^2 - \{\pm p\}$ clearly has constant curvature 1 and infinite fundamental group. To make things worse; the universal covering also has diameter π.

Example 6.3.5. The manifold $S^1 \times \mathbb{R}^3$ admits a complete doubly warped product metric

$$dr^2 + \rho^2(r)d\theta^2 + \phi^2(r)ds_2^2,$$

that has $\operatorname{Ric} > 0$ everywhere. Curvatures are calculated as in 1.4.5. If we define $\rho(t) = t^{-1/4}$ and $\phi(t) = t^{3/4}$ for $t \geq 1$, then the Ricci curvature will be positive. Next extend to $[0, \infty]$ so that the metric becomes smooth at $t = 0$; the functions are C^1 and piecewise smooth at $t = 1$; $-1 < \dot{\rho} \leq 0$; $0 < \dot{\phi} \leq 1$; $\ddot{\phi} < 0$; and on $[0, 1]$ $\ddot{\rho} \leq 0$. This will result in a C^1 metric that has positive Ricci curvature except at $t = 1$. Finally, smooth out ρ at $t = 1$ ensuring that the Ricci curvature stays positive.

6.3.2 *The Fundamental Group in Even Dimensions*

For the next result we need to study what happens when we have a closed geodesic in a Riemannian manifold of positive curvature.

Let $c : [0, l] \to M$ be a closed unit speed geodesic, i.e., $\dot{c}(0) = \dot{c}(l)$. Let $p = c(0) = c(l)$ and consider parallel translation along c. This defines a linear isometry $P : T_pM \to T_pM$. Since c is a closed geodesic we have that $P(\dot{c}(0)) = \dot{c}(l) = \dot{c}(0)$. Thus, P preserves the orthogonal complement to $\dot{c}(0)$ in T_pM. Now recall that linear isometries $L : \mathbb{R}^k \to \mathbb{R}^k$ with $\det L = (-1)^{k+1}$ have 1 as an eigenvalue, i.e., $L(v) = v$ for some$v \in \mathbb{R}^k$. We can use this to construct a closed parallel field around c in one of two ways:

(1) If M is orientable and even-dimensional, then parallel translation around a closed geodesic preserves orientation, i.e., $\det = 1$. Since the orthogonal complement to $\dot{c}(t)$ in T_pM is odd dimensional there must exist a closed parallel field around c.
(2) If M is not orientable, has odd dimension, and furthermore, c is a nonorientable loop, i.e., the orientation changes as we go around this loop, then parallel

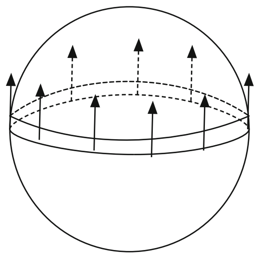


Fig. 6.4 Finding shorter curves near a closed geodesic

translation around c is orientation reversing, i.e., $\det = -1$. Now, the orthogonal complement to $\dot{c}(t)$ in T_pM is even-dimensional, and since $P(\dot{c}(0)) = \dot{c}(0)$, it follows 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.

This discussion leads to an interesting and surprising topological result for positively curved manifolds.

Theorem 6.3.6 (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. The proof goes by contradiction. So in either case assume we have a nontrivial universal covering $\pi : \tilde{M} \to M$. Let F be a nontrivial deck transformation that in the odd-dimensional case reverses orientation. From lemma 6.2.8 we obtain a unit speed geodesic (axis) $\tilde{c} : \mathbb{R} \to \tilde{M}$ that is mapped to itself by F. Moreover, $c = \pi \circ \tilde{c}$ is the shortest curve in its free homotopy class in M when restricted to an interval $[a, b]$ of length $b - a = \min \delta_F$.

In both cases our assumptions are such that the closed geodesics have closed perpendicular parallel fields. We can now use the second variation formula with this parallel field as variational field. Note that the variation isn't proper, but since the geodesic is closed the end point terms cancel each other

$$\begin{aligned}\frac{d^2E(c_s)}{ds^2}|_{s=0} &= -\int_a^b g(R(E,\dot{c})\dot{c},E)\,dt + g\left(\frac{\partial^2\bar{c}}{\partial s^2},\dot{c}\right)\Big|_a^b \\ &= -\int_a^b g(R(E,\dot{c})\dot{c},E)\,dt \\ &= -\int_a^b \sec(E,\dot{c})\,dt \\ &< 0.\end{aligned}$$

Thus all nearby curves in this variation are closed curves whose lengths are shorter than c. This contradicts our choice of c as the shortest curve in its free homotopy class. □

The first important conclusion we get from this result is that while $\mathbb{RP}^2 \times \mathbb{RP}^2$ has positive Ricci curvature, it cannot support a metric of positive sectional curvature. It is, on the other hand, completely unknown whether $S^2 \times S^2$ admits a metric of positive sectional curvature. This is known as the Hopf problem. Recall that in section 6.2.2 we showed, using fundamental group considerations, that no product manifold admits negative curvature. In this case, fundamental group considerations cannot take us as far.

6.4 Basic Comparison Estimates

In this section we lay the foundations for the comparison estimates that will be needed later in the text.

6.4.1 Riccati Comparison

We start with a general result for differential inequalities.

Proposition 6.4.1 (Riccati Comparison Principle). *If we have two smooth functions $\rho_{1,2} : (0, b) \to \mathbb{R}$ such that*

$$\dot{\rho}_1 + \rho_1^2 \leq \dot{\rho}_2 + \rho_2^2,$$

then

$$\rho_2 - \rho_1 \geq \limsup_{t \to 0} (\rho_2(t) - \rho_1(t)).$$

Proof. Let $F(t) = \int (\rho_2 + \rho_1)\, dt$ be an antiderivative for $\rho_2 + \rho_1$ on $(0, b)$. The claim follows since the function $(\rho_2 - \rho_1) e^F$ is increasing:

$$\frac{d}{dt}\left((\rho_2 - \rho_1) e^F\right) = \left(\dot{\rho}_2 - \dot{\rho}_1 + \rho_2^2 - \rho_1^2\right) e^F \geq 0.$$

□

This can be turned into more concrete estimates.

Corollary 6.4.2 (Riccati Comparison Estimate). *Consider a smooth function $\rho : (0, b) \to \mathbb{R}$ with $\rho(t) = \frac{1}{t} + O(t)$ and a real constant k.*

(1) *If* $\dot{\rho} + \rho^2 \leq -k$*, then*

$$\rho(t) \leq \frac{\mathrm{sn}'_k(t)}{\mathrm{sn}_k(t)}.$$

Moreover, $b \leq \pi/\sqrt{k}$ *when* $k > 0$.

(2) *If* $-k \leq \dot{\rho} + \rho^2$*, then*

$$\frac{\mathrm{sn}'_k(t)}{\mathrm{sn}_k(t)} \leq \rho(t)$$

for all $t < b$ *when* $k \leq 0$ *and* $t < \min\{b, \pi/\sqrt{k}\}$ *when* $k > 0$.

Proof. First note that for any k the comparison function satisfies

$$\frac{\mathrm{sn}'_k(t)}{\mathrm{sn}_k(t)} = \frac{1}{t} + O(t)$$

and solves

$$\dot{\rho} + \rho^2 = -k.$$

When $k > 0$ this function is only defined on $(0, \pi/\sqrt{k})$ and

$$\lim_{t \to \frac{\pi}{\sqrt{k}}} \frac{\mathrm{sn}'_k(t)}{\mathrm{sn}_k(t)} = -\infty.$$

In case $\dot{\rho} + \rho^2 \leq -k$ this will prevent ρ from being smooth when $b > \pi/\sqrt{k}$.

Similarly, when $-k \leq \dot{\rho} + \rho^2$ we are forced to assume that $b \leq \pi/\sqrt{k}$ in order for the comparison function to be defined. □

Let us apply these results to one of the most commonly occurring geometric situations. Suppose that on a Riemannian manifold (M, g) we have introduced exponential coordinates around a point $p \in M$ so that $g = dr^2 + g_r$ on a star shaped open set in $T_pM - \{0\} = (0, \infty) \times S^{n-1}$. Along any given geodesic from p the metric g_r is thought of as being on S^{n-1}. It is not important for the next result that M be complete as it is essentially local in nature.

Theorem 6.4.3 (Rauch Comparison). *Assume that* (M, g) *satisfies* $k \leq \sec \leq K$. *If* $g = dr^2 + g_r$ *represents the metric in the polar coordinates, then*

$$\frac{\mathrm{sn}'_K(r)}{\mathrm{sn}_K(r)} g_r \leq \mathrm{Hess}\, r \leq \frac{\mathrm{sn}'_k(r)}{\mathrm{sn}_k(r)} g_r.$$

Consequently, the modified distance functions from corollary 4.3.4 satisfy:

$$\mathrm{Hess} f_k \leq (1 - k f_k)\, g,$$

$$\mathrm{Hess} f_K \geq (1 - K f_K)\, g.$$

Proof. It'll be convenient to use slightly different techniques for lower and upper curvature bounds. Specifically, for lower curvature bounds parallel fields are the easiest to use, while Jacobi fields are better suited to upper curvature bounds.

In both cases assume that we have a unit speed geodesic $c(t)$ with $c(0) = p$ and that $t \in [0, b]$, with $c([0, b]) \subset \exp_p\left(\mathrm{seg}_p^0\right)$ so that $r(x) = |xp|$ is smooth along the entire geodesic segment.

We start with the upper curvature situation as it is quite close in spirit to lemma 6.2.5. In fact that proof can be easily adapted to the case where $\sec \leq K \leq 0$, but when $K > 0$ it runs into trouble (see exercise 6.7.11). Instead consider the reciprocal ratio

$$\rho(t) = \frac{g\left(\dot{J}, J\right)}{|J|^2} = \mathrm{Hess}\, r\left(\frac{J}{|J|}, \frac{J}{|J|}\right)$$

for a Jacobi field along c with $J(0) = 0$ and $\dot{J}(0) \perp \dot{c}(0)$. It follows that $J(t) \perp \dot{c}(t)$ for all t and

$$\begin{aligned}
\dot{\rho} &= \frac{-R(J, \dot{c}, \dot{c}, J)|J|^2 + \left|\dot{J}\right|^2 |J|^2 - 2\left(g\left(\dot{J}, J\right)\right)^2}{|J|^4} \\
&\geq -K + \frac{\left|\dot{J}\right|^2 |J|^2 - 2\left(g\left(\dot{J}, J\right)\right)^2}{|J|^4} \\
&\geq -K - \rho^2.
\end{aligned}$$

In case there is a lower curvature bound, select instead a unit parallel field E along c that is perpendicular to $\dot{c}$ and consider

$$\rho = g(S(E), E) = \mathrm{Hess}\, r(E, E).$$

From part (2) of proposition 3.2.11 we obtain

$$\begin{aligned}
\dot{\rho} &= -R(E, \dot{c}, \dot{c}, E) - g(S(E), S(E)) \\
&\leq -k - (g(S(E), E))^2 \\
&= -k - \rho^2.
\end{aligned}$$

In both cases we have the initial conditions that $\rho(t) = \frac{1}{t} + O(t)$ and so we obtain the desired inequalities for ρ and hence $\mathrm{Hess}\, r$ from corollary 6.4.2.

The Hessian estimates for the modified distance functions follow immediately.

□

Remark 6.4.4. A more traditional proof technique using the index form is discussed in exercise 6.7.25 within the context of lower curvature bounds. It can also be adapted to deal with upper curvature bounds.

6.4.2 *The Conjugate Radius*

As in the proof of theorem 6.2.2 we are going to estimate where the exponential map is nonsingular.

Example 6.4.5. Consider S_K^n, $K > 0$. If we fix $p \in S_K^n$ and use polar coordinates, then the metric looks like $dr^2 + \mathrm{sn}_K^2 ds_{n-1}^2$. At distance $\frac{\pi}{\sqrt{K}}$ from p we will hit a conjugate point no matter what direction we go in.

As a generalization of our result on no conjugate points when $\sec \leq 0$ we can show

Theorem 6.4.6. *If* (M, g) *has* $\sec \leq K,\ K > 0$*, then*

$$\exp_p : B\left(0, \frac{\pi}{\sqrt{K}}\right) \to M$$

has no critical points.

Proof. Let $c(t)$ be a unit speed geodesic and $J(t)$ a Jacobi field along c with $J(0) = 0$ and $\dot{J}(0) \perp \dot{c}(0)$. We have to show that $J(t)$ can't vanish for any $t \in (0, \pi/\sqrt{K})$. Assume that $J > 0$ on $(0, b)$ and $J(b) = 0$. From the proof of theorem 6.4.3 we obtain

$$\frac{g\left(\dot{J}(t), J(t)\right)}{|J(t)|^2} \geq \frac{\mathrm{sn}'_K(t)}{\mathrm{sn}_K(t)}$$

for $t < \min\{b, \pi/\sqrt{K}\}$. This is equivalent to saying that

$$\frac{d}{dt}\left(\frac{|J(t)|}{\mathrm{sn}_K(t)}\right) \geq 0.$$

Since $D\exp_p$ is the identity at the origin it follows that $|J(t)| = t\left|\dot{J}(0)\right| + O\left(t^2\right)$. This together with l'Hospital's rule shows that

$$\begin{aligned}
\lim_{t\to 0}\left(\frac{|J(t)|}{\mathrm{sn}_K(t)}\right) &= \lim_{t\to 0}\frac{g\left(\dot{J}(t), J(t)\right)}{|J(t)|} \\
&= \lim_{t\to 0}\frac{\mathrm{Hess}\, r\,(J(t), J(t))}{|J(t)|} \\
&= \lim_{t\to 0}\frac{t^{-1}|J(t)|^2 + O\left(t^3\right)}{|J(t)|} \\
&= \lim_{t\to 0} t^{-1}|J(t)| \\
&= \left|\dot{J}(0)\right|.
\end{aligned}$$

It follows that $J(t) \geq \left|\dot{J}(0)\right| \mathrm{sn}_K(t) > 0$ for all $t < \min\{b, \pi/\sqrt{K}\}$. This shows that we can't have $b < \pi/\sqrt{K}$ and the claim follows. □

With this information about conjugate points, we also get estimates for the injectivity radius using the characterization from lemma 5.7.12. For Riemannian manifolds with $\sec \le 0$ the injectivity radius satisfies

$$\operatorname{inj}(p) = \frac{1}{2} \cdot (\text{length of shortest geodesic loop based at } p)$$

as there are no conjugate points whatsoever. On a closed Riemannian manifold with $\sec \le 0$ we claim that

$$\operatorname{inj}(M) = \inf_{p \in M} \operatorname{inj}(p) = \frac{1}{2} \cdot (\text{length of shortest closed geodesic}).$$

Since M is closed, the infimum must be a minimum. This follows from continuity $p \mapsto \operatorname{inj}(p)$, which in turn is a consequence of $\exp : TM \to M \times M$ being smooth and the characterization of $\operatorname{inj}(p)$ from lemma 5.7.12. If $p \in M$ realizes this infimum, and $c : [0, 1] \to M$ is the geodesic loop realizing $\operatorname{inj}(p)$, then we can split c into two equal segments joining p and $c\left(\frac{1}{2}\right)$. Thus, $\operatorname{inj}\left(c\left(\frac{1}{2}\right)\right) \le \operatorname{inj}(p)$, but this means that c must also be a geodesic loop as seen from $c\left(\frac{1}{2}\right)$. In particular, it is smooth at p and forms a closed geodesic.

The same line of reasoning yields the following more general result.

Lemma 6.4.7 (Klingenberg). *Let (M, g) be a compact Riemannian manifold with* $\sec \le K$, *where* $K > 0$. *Then*

$$\operatorname{inj}(p) \ge \min\left\{\frac{\pi}{\sqrt{K}}, \frac{1}{2} \cdot (\textit{length of shortest geodesic loop based at } p)\right\},$$

and

$$\operatorname{inj}(M) \ge \frac{\pi}{\sqrt{K}} \textit{ or } \operatorname{inj}(M) = \frac{1}{2} \cdot (\textit{length of shortest closed geodesic}).$$

These estimates will be used in the next section.

Next we turn our attention to the convexity radius.

Theorem 6.4.8. *Suppose R satisfies*

(1) $R \le \frac{1}{2} \cdot \operatorname{inj}(x)$, *for $x \in B(p, R)$, and*
(2) $R \le \frac{1}{2} \cdot \frac{\pi}{\sqrt{K}}$, *where* $K = \sup\{\sec(\pi) \mid \pi \subset T_xM,\ x \in B(p, R)\}$.

Then $r(x) = |xp|$ is 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 in M, and that $r(x)$ is smooth on $B(p, 2 \cdot R) - \{p\}$. The second condition ensures that $\operatorname{Hess} r \ge 0$ on $B(p, R)$. It then remains to be shown that if $x, y \in B(p, R)$, and $c : [0, 1] \to M$ is the unique segment joining them, then

$c \subset B(p,R)$. For fixed $x \in B(p,R)$, define C_x to be the set of ys for which this holds. Certainly $x \in C_x$ and C_x is open. If $y \in \overline{B(p,R)} \cap \partial C_x$, then the segment $c : [0,1] \to M$ joining x to y must lie in $\overline{B(p,R)}$ by continuity. Now consider $\varphi(t) = r(c(t))$. By assumption

$$\varphi(0), \varphi(1) < R,$$
$$\ddot{\varphi}(t) = \operatorname{Hess} r\,(\dot{c}(t), \dot{c}(t)) \geq 0.$$

Thus, φ is convex, and consequently

$$\max \varphi(t) \leq \max\{\varphi(0), \varphi(1)\} < R,$$

showing that $c \subset B(p,R)$. □

The largest R such that $r(x)$ 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,

$$\operatorname{conv.rad}(M,g) = \inf_{p \in M} \operatorname{conv.rad}(p).$$

The previous result tell us

$$\operatorname{conv.rad}(M,g) \geq \min\left\{\frac{\operatorname{inj}(M,g)}{2}, \frac{\pi}{2\sqrt{K}}\right\}, \; K = \sup \sec(M,g).$$

In nonpositive curvature this simplifies to

$$\operatorname{conv.rad}(M,g) = \frac{\operatorname{inj}(M,g)}{2}.$$

6.5 More on Positive Curvature

In this section we shall establish some further restrictions on the topology of manifolds with positive curvature. The highlight will be the classical quarter pinched sphere theorem of Rauch, Berger, and Klingenberg. To prove this theorem requires considerable preparation. We shall elaborate further on this theorem and its generalizations in section 12.3.

6.5.1 The Injectivity Radius in Even Dimensions

Using the ideas of the proof of theorem 6.3.6 we get another interesting restriction on the geometry of positively curved manifolds.

Theorem 6.5.1 (Klingenberg, 1959). *If* (M, g) *is a compact orientable even-dimensional manifold with* $0 < \sec \leq 1$*, then* $\operatorname{inj}(M, g) \geq \pi$*. If M is not orientable, then* $\operatorname{inj}(M, g) \geq \frac{\pi}{2}$.

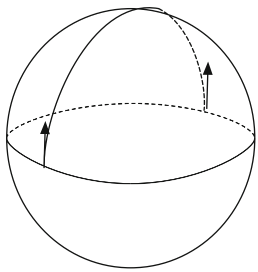
Proof. The nonorientable case follows from the orientable case, as the orientation cover will have $\operatorname{inj}(M, g) \geq \pi$.

By lemma 6.4.7 and the upper curvature bound it follows that if $\operatorname{inj} M < \pi$, then the injectivity radius is realized by a closed geodesic. So let us assume that there is a closed geodesic $c : [0, 2 \operatorname{inj} M] \to M$ parametrized by arclength, where $2 \operatorname{inj} M < 2\pi$. Since M is orientable and even dimensional, we know from section 6.3.2 and the proof of theorem 6.3.6 that for all small $\varepsilon > 0$ there are curves $c_\varepsilon : [0, 2 \operatorname{inj} M] \to M$ that converge to c as $\varepsilon \to 0$ and with $L(c_\varepsilon) < L(c) = 2 \operatorname{inj} M$. Since $c_\varepsilon \subset B(c_\varepsilon(0), \operatorname{inj} M)$ there is a unique segment from $c_\varepsilon(0)$ to $c_\varepsilon(t)$. Thus, if $c_\varepsilon(t_\varepsilon)$ is the point at maximal distance from $c_\varepsilon(0)$ on c_ε, we get a segment σ_ε joining these points that in addition is perpendicular to c_ε at $c_\varepsilon(t_\varepsilon)$. As $\varepsilon \to 0$, it follows that $t_\varepsilon \to \operatorname{inj} M$, and thus the segments σ_ε must subconverge to a segment from $c(0)$ to $c(\operatorname{inj} M)$ that is perpendicular to c at $c(\operatorname{inj} M)$. However, as the conjugate radius is $\geq \pi > \operatorname{inj} M$, and c is a geodesic loop realizing the injectivity radius at $c(0)$, we know from lemma 5.7.12 that there can only be two segments from $c(0)$ to $c(\operatorname{inj} M)$. Thus, we have a contradiction with our assumption $\pi > \operatorname{inj} M$. □

In figure 6.5 we have pictured a fake situation that gives the 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 the quotients of spheres $S^3/\mathbb{Z}_k$ for all positive integers k are all orientable. The image of the Hopf fiber via the covering map $S^3 \to S^3/\mathbb{Z}_k$ is a closed geodesic of length $\frac{2\pi}{k}$ that 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 $\left[\frac{\varepsilon^2}{4-3\varepsilon^2}, 1\right]$. When $\varepsilon < \frac{1}{\sqrt{3}}$, the Hopf fibers have length $< 2\pi$. In this case the lower curvature bound becomes smaller than $\frac{1}{9}$.

Fig. 6.5 Equator with short cut through the Northpole



A much deeper result by Klingenberg asserts that if a simply connected manifold has all its sectional curvatures in the interval $\left(\frac{1}{4}, 1\right]$, then the injectivity radius is still $\geq \pi$ (see the next section for the proof). This result has been improved first by Klingenberg-Sakai and Cheeger-Gromoll to allow for the curvatures to be in $\left[\frac{1}{4}, 1\right]$. More 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 hypothesis on the fundamental group being trivial is necessary in order to eliminate all the constant curvature spaces with small injectivity radius.

These injectivity radius estimates will be used to prove some fascinating sphere theorems.

6.5.2 Applications of Index Estimation

Some notions and results from topology are needed to explain the material here.

We say that $A \subset X$ is l-connected if the relative homotopy groups $\pi_k(X, A)$ vanish for $k \leq l$. A theorem of Hurewicz then shows that the relative homology groups $H_k(X, A)$ also vanish for $k \leq l$. The long exact sequences for the pair (X, A)

$$\pi_{k+1}(X, A) \to \pi_k(A) \to \pi_k(X) \to \pi_k(X, A)$$

and

$$H_{k+1}(X, A) \to H_k(A) \to H_k(X) \to H_k(X, A)$$

then show that $\pi_k(A) \to \pi_k(X)$ and $H_k(A) \to H_k(X)$ are isomorphisms for $k < l$ and surjective for $k = l$.

We say that a critical point $p \in M$ for a smooth function $f : M \to \mathbb{R}$ has index $\geq m$ if the Hessian of f is negative definite on a m-dimensional subspace in T_pM. Note that if $m \geq 1$, then p can't be a local minimum for f as the function must decrease in the directions where the Hessian is negative definite. The index of a critical point gives us information about how the topology of M changes as we pass through this point. In Morse theory a much more precise statement is proven, but it also requires the critical points to be nondegenerate, an assumption we do not wish make here (see [75]).

Theorem 6.5.2. *Let $f : M \to \mathbb{R}$ be a smooth proper function. If b is not a critical value for f and all critical points in $f^{-1}([a, b])$ have index $\geq m$, then*

$$f^{-1}((-\infty, a]) \subset f^{-1}((-\infty, b])$$

is $(m-1)$-connected.

Outline of Proof. If there are no critical points in $f^{-1}([a, b])$, then the gradient flow will deform $f^{-1}((-\infty, b])$ to $f^{-1}((-\infty, a])$. This is easy to prove and is explained in lemma 12.1.1. If there are critical points, then by compactness we can cover the set of critical points by finitely many open sets $U_i \approx (-a, a)^n$, $0 < a < 1$, where $\bar{U}_i \subset V_i$ and $\bar{V}_i \approx [-1,1]^n$ is a closed box coordinate chart where the first m coordinates correspond to directions where $\operatorname{Hess} f$ is negative definite.

Consider a map $\phi : N^{k-1} \to f^{-1}([a,b])$, $k \leq m$, where $\partial N^{k-1} \subset f^{-1}((-\infty, a])$ if the boundary is nonempty.

- On $M - \bigcup U_i$ we can use the flow of $-\lambda(x) \nabla f|_x$, where $\lambda \geq 0$ and $\lambda^{-1}(0) = \bigcup \bar{U}_i$. This will deform ϕ keeping it fixed on $\bar{U}_i$ and forcing $\max_{f^{-1}([a,b])} f \circ \phi$ to decrease while ensuring that $\max_{\bar{V}_i} f \circ \phi$ is not obtained on ∂V_i.
- Let $S_i = \left\{p \in V_i \mid x^1(p) = \cdots = x^m(p) = 0\right\}$. The restriction $k \leq m$ allows us to use transversality to ensure that there is a homotopy ϕ_t, $t \in [0, \epsilon)$, where $\phi_0 = \phi$; ϕ_t does not intersect S_i for $t > 0$; and $t \mapsto \phi_t$ is constant on $M - V_i$. Moreover, for sufficiently small t $\max_{\bar{V}_i} f \circ \phi_t$ is still not obtained on ∂V_i.
- Finally, when ϕ doesn't intersect the submanifold S_i, the flow for the radial field $\sum_{j=1}^m x^j \partial_j$ on $\bar{V}_i$ decreases the value of $\max_{\bar{V}_i} f \circ \phi$ and moves ϕ outside $\bar{U}_i$.

With these three types of deformations it is possible to continuously deform ϕ until its image lies in $f^{-1}((-\infty, a])$. □

In analogy with $\Omega_{p,q}(M)$ define

$$\Omega_{A,B}(M) = \{c : [0,1] \to M \mid c(0) \in A,\ c(1) \in B\}.$$

If $A, B \subset M$ are compact, then the energy functional $E : \Omega_{A,B}(M) \to [0, \infty)$ is reasonably nice in the sense that it behaves like a proper smooth function on a manifold. If in addition A and B are submanifolds, then the variational fields for variations in $\Omega_{A,B}(M)$ consist of fields along the curve that are tangent to A and B at the endpoints. Therefore, critical points are naturally identified with geodesics that are perpendicular to A and B at the endpoints. We say that the index of such a geodesic $\geq k$ if there is a k-dimensional space of fields along the geodesic such that the second variation of the these fields is negative.

One can now either try to reprove the above theorem in a suitable infinite dimensional context (see [30] or [69]) or use finite dimensional approximations to $\Omega_{A,B}(M)$ (see [75]). Both routes are technical but fairly straightforward.

Theorem 6.5.3. *Let M be a complete Riemannian manifold and $A \subset M$ a compact submanifold. If every geodesic in $\Omega_{A,A}(M)$ that is perpendicular to A at the end points has index $\geq k$, then $A \subset M$ is k-connected.*

Outline of Proof. See also [30] or [69, Theorem 2.5.16] for a proof. Identify $A = E^{-1}(0)$ and use the previous theorem as a guide for how to deform maps. This shows that $A \subset \Omega_{A,A}(M)$ is $(k-1)$-connected. Next we note that

$$\pi_l(\Omega_{A,A}(M), A) = \pi_{l+1}(M, A).$$

This proves the result. □

This theorem can be used to prove a sphere theorem by Berger.

Theorem 6.5.4 (Berger, 1958). *Let M be a closed n-manifold with* $\sec \geq 1$. *If* $\mathrm{inj}_p > \pi/2$ *for some $p \in M$, then M is $(n-1)$-connected and hence a homotopy sphere.*

Proof. We'll use theorem 6.5.3 with $A = \{p\}$. First note that every geodesic loop at p is either the constant curve or has length $> \pi$ since $\mathrm{inj}_p > \pi/2$. We showed in lemma 6.3.1 that geodesics of length $> \pi$ have proper variations whose second derivative is negative. In fact each parallel field along the geodesic could be modified to create such a variation. As there is an $(n-1)$-dimensional space of such parallel fields we conclude that the index of such geodesics is $\geq (n-1)$. This shows that $p \in M$ is $(n-1)$-connected and consequently that M is $(n-1)$-connected.

Finally, to see that M is a homotopy sphere we select a map $F : M \to S^n$ of degree 1. Since M is $(n-1)$-connected this map must be an isomorphism on π_k for $k < n$ as S^n is also $(n-1)$-connected. We claim that

$$\pi_n(M) \simeq H_n(M) \to H_n(S^n) \simeq \pi_n(S^n)$$

is an isomorphism. Hurewicz's result shows that the homotopy and homology groups are isomorphic, while the fact that F has degree 1 implies that $H_n(S^n) \to H_n(M)$ is an isomorphism. A theorem of Whitehead then implies that F is a homotopy equivalence. □

This theorem is even more interesting in view of the injectivity radius estimate in positive curvature that we discussed in section 6.5.1. We can extend this to odd dimensions using theorem 6.5.3.

Theorem 6.5.5 (Klingenberg, 1961). *A compact simply connected Riemannian n-manifold (M, g) with* $1 \leq \sec < 4$ *has* $\mathrm{inj} > \pi/2$.

Proof. It is more convenient to show that simply connected manifolds with $1 < \sec \leq 4$ have $\mathrm{inj} \geq \pi/2$. A simple scaling shows that this implies the statement of the theorem. We can also assume that $n \geq 3$ as we know the theorem to be true in even dimensions. The lower curvature bound implies that there is a $\delta > 0$ such that geodesics of length $\geq \pi - \delta$ have index $\geq n - 1 \geq 2$. In particular, any map $[0, 1] \to \Omega_{p,p}(M)$ of constant speed loops based at p is homotopic to a map where the loops have length $< \pi$. It is easy to force the loops to have constant speed as we can replace them by nearby loops that are piecewise segments and therefore shorter. This can be done uniformly along a fixed homotopy by selecting the break points on S^1 independently of the variational parameter.

The proof proceeds by contradiction so assume that $\mathrm{inj}_p < \pi/2$. Then lemma 6.4.7 shows that there is a geodesic loop at p of length $< \pi$ that realizes the injectivity radius. Next use simple connectivity to find a homotopy of loops based at p to the constant loop and further assume that all the loops in the homotopy have constant speed and length $< \pi$. For each $s \in [0, 1]$ parametrize the corresponding loop $c_s(t) : [0, 1] \to M$ so that $c_s(0) = c_s(1) = p$; $c_0(t) = p$ for all t; and c_1 the closed geodesic of length $< \pi$. As each c_s has length $< \pi$ it must be contained in $B(p, \pi/2)$.

Note that the exponential map $\exp_p : B(0, \pi/2) \to B(p, \pi/2) \subset M$ is nonsingular and a diffeomorphism when restricted to $B(0, \mathrm{inj}_p)$. We shall further use the pull back metric on $B\left(0, \frac{\pi}{2}\right)$ so that $\exp_p$ becomes a local isometry. This tells us that any of the loops $c_s : [0, 1] \to B(p, \pi/2)$ with $c(0) = p$ have a unique lift to a curve $\bar{c}_s : [0, b_s] \to \bar{B}(0, \pi/2)$ with $\bar{c}_s(0) = 0$. Here either $\bar{c}_s(b_s) \in \partial B(0, \pi/2)$ or $b = 1$. Note that when c_s is a piecewise geodesic, then we can easily create such a lift by lifting the velocity vectors at break points.

Let $A \subset [0, 1]$ be the set of s such that c_s lifts to a loop $\bar{c}_s : [0, 1] \to B(0, \pi/2)$ based at 0.

Clearly $0 \in A$, and as $\exp_p$ is a diffeomorphism near 0 loops c_s with s near 0 also lift to loops.

A is closed: Let $s_i \in A$ converge to s. Then $\bar{c}_{s_i}(1)$ is defined and $\bar{c}_{s_i}(1) = 0$. The unique lift $\bar{c}_s$ must be the limit of the curves $\bar{c}_{s_i}$. Thus it is defined on $[0, 1]$ and is a loop. Finally observe that the limit curve $\bar{c}_s$ clearly lies $\bar{B}(0, \pi/2)$ and is forced to lie in the interior as it has length $< \pi$.

A is open: Fix $s_0 \in A$ and let the lift be $\bar{c}_{s_0}$. Select $\epsilon > 0$ so that $\exp_p : B(\bar{c}_{s_0}(t), \epsilon) \to B(c_{s_0}(t), \epsilon)$ is an isometry for all t and $B(\bar{c}_{s_0}(t), \epsilon) \subset B(0, \pi/2)$. For s near s_0, the loops c_s must be contained in $\bigcup_{t \in [0,1]} B(c_{s_0}(t), \epsilon)$. But then they have unique lifts to loops in $\bigcup_{t \in [0,1]} B(\bar{c}_{s_0}(t), \epsilon) \subset B(0, \pi/2)$. Thus $\bar{c}_s(1) \in B(0, \epsilon)$ is a lift of p and consequently $\bar{c}_s(1) = 0$. This shows that a neighborhood of s_0 is contained in A.

All in all we've concluded that $A = [0, 1]$. However, the geodesic c_1 lifts to a line that starts at 0 and consequently is not a loop. The establishes the contradiction. $\square$

This gives us the classical version of the sphere theorem.

Corollary 6.5.6 (Rauch, Berger, and Klingenberg, 1951–61). *Let M be a closed simply connected n-manifold with* $4 > \sec \geq 1$. *Then M is $(n-1)$-connected and hence a homotopy sphere.*

The conclusion can be strengthened to say that M is homeomorphic to a sphere. This follows from the solution to the (generalized) Poincaré conjecture given what we have already proven. In section 12.3 we exhibit an explicitly constructed homeomorphism.

Using an analysis similar to the proof of theorem 6.5.4 one also gets the more modest result.

Corollary 6.5.7. *If M is a closed n-manifold with* $\mathrm{Ric} \geq (n-1)$ *and* $\mathrm{inj}_p > \pi/2$ *for some $p \in M$, then M is simply connected.*

Finally we mention a significant result that allows us to make strong conclusions about connectedness in positive curvature. The result will be enhanced in lemma 8.3.6.

Lemma 6.5.8 (The Connectedness Principle, Wilking, 2003). *Let M^n be a compact n-manifold with positive sectional curvature.*

(a) *If* $N^{n-k} \subset M^n$ *is a closed codimension* k *totally geodesic submanifold, then* $N \subset M$ *is* $(n-2k+1)$*-connected.*
(b) *If* $N_1^{n-k_1}$ *and* $N_2^{n-k_2}$ *are closed totally geodesic submanifolds of* M *with* $k_1 \leq k_2$ *and* $k_1 + k_2 \leq n$*, then* $N_1 \cap N_2$ *is a nonempty totally geodesic submanifold and* $N_1 \cap N_2 \to N_2$ *is* $(n-k_1-k_2)$*-connected.*

Proof. (a) Let $c \in \Omega_{N,N}(M)$ be a geodesic and E a parallel field along c such that E is tangent to N at the endpoints. Then we can construct a variation $\bar{c}(s,t)$ such that $\bar{c}(0,t) = c(t)$ and $s \mapsto \bar{c}(s,t)$ is a geodesic with initial velocity $E|_{c(t)}$. Since N is totally geodesic we see that $\bar{c}(s,0), \bar{c}(s,1) \in N$. Thus the variational curves lie in $\Omega_{N,N}(M)$. The second variation formula for this variation tells us that

$$\begin{aligned}\frac{d^2E(c_s)}{ds^2}|_{s=0} &= \int_0^1 \left|\dot{E}\right|^2 dt - \int_a^b g(R(E,\dot{c})\dot{c},E)\,dt + g\left(\frac{\partial^2\bar{c}}{\partial s^2},\dot{c}\right)\Big|_0^1 \\ &= -\int_0^1 g(R(E,\dot{c})\dot{c},E)\,dt \\ &< 0\end{aligned}$$

since $\dot{E} = 0$, $\frac{\partial^2\bar{c}}{\partial s^2} = 0$, and E is perpendicular to $\dot{c}$. Thus each such parallel field gives us a negative variation. This shows that the index of c is bigger than the set of parallel variational fields.

Let $V \subset T_{c(1)}M$ be the subspace of vectors $v = E(1)$, where E is a parallel field along c with $E(0) \in T_{c(0)}N$. The space of parallel fields used to get negative variations is then identified with $V \cap T_{c(1)}N$. To find the dimension of that space we note that T_pN and hence also V have dimension $n-k$. Moreover, V and $T_{c(1)}N$ lie in the orthogonal complement to $\dot{c}(1)$. Putting this together gives us

$$\begin{aligned}2n-2k &= \dim\left(T_{c(1)}N\right) + \dim(V) \\ &= \dim\left(V \cap T_{c(1)}N\right) + \dim\left(V + T_{c(1)}N\right) \\ &\leq \dim\left(V \cap T_{c(1)}N\right) + n - 1.\end{aligned}$$

(b) It is easy to show that $N_1 \cap N_2$ is also totally geodesic. The key is to guess that for $p \in N_1 \cap N_2$ we have $T_p(N_1 \cap N_2) = T_pN_1 \cap T_pN_2$. To see that $N_1 \cap N_2 \neq \varnothing$ select a geodesic from N_1 to N_2. The dimension conditions imply that there is a $(n-k_1-k_2+1)$-dimensional space of parallel field along this geodesic that are tangent to N_1 and N_2 at the end points. Since $k_1 + k_2 \leq n$ we get a variation with negative second derivative, thus nearby variational curves are shorter. This shows that there can't be a nontrivial geodesic of shortest length joining N_1 and N_2.

Using $E : \Omega_{N_1,N_2}(M) \to [0,\infty)$ we can identify $N_1 \cap N_2 = E^{-1}(0)$. So we have in fact shown that $N_1 \cap N_2 \subset \Omega_{N_1,N_2}(M)$ is $(n-k_1-k_2)$-connected. Using that $N_1 \subset M$ is $(n-2k_1+1)$-connected shows that $\Omega_{N_1,N_2}(M) \subset \Omega_{M,N_2}(M)$ is also $(n-2k_1+1)$-connected. Since $k_1 \leq k_2$ this shows that $N_1 \cap N_2 \subset \Omega_{M,N_2}(M)$ is $(n-k_1-k_2)$-connected. Finally observe that $\Omega_{M,N_2}(M)$ can be retracted to N_2 and is homotopy equivalent to N_2. This proves the claim. □

What is commonly known as Frankel's theorem is included in part (b). The statement is simply that under the conditions in (b) the intersection is nonempty.

6.6 Further Study

Several textbooks treat the material mentioned in this chapter, and they all use variational calculus. We especially recommend [23, 30, 47] and [65]. The latter also discusses in more detail closed geodesics and, more generally, minimal maps and surfaces in Riemannian manifolds.

As we won't discuss manifolds of nonpositive curvature in detail later in the text some references for this subject should be mentioned here. With the knowledge we have right now, it shouldn't be too hard to read the books [10] and [8]. For a more advanced account we recommend the survey by Eberlein-Hammenstad-Schroeder in [51]. At the moment the best, most complete, and up to date book on the subject is probably [38].

For more information about the injectivity radius in positive curvature the reader should consult the article by Abresch and Meyer in [54].

All of the necessary topological background material used in this chapter can be found in [75] and [96].

6.7 Exercises

EXERCISE 6.7.1. Show that in even dimensions the sphere and real projective space are the only closed manifolds with constant positive curvature.

EXERCISE 6.7.2. Consider a rotationally symmetric metric $dr^2 + \rho^2(r)\, d\theta^2$. We wish to understand parallel translation along a latitude, i.e., a curve with $r = a$. To this end construct a cone $dr^2 + (\rho(a) + \dot{\rho}(a)(r-a))^2\, d\theta^2$ that is tangent to this surface at the latitude $r = a$. 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 $r = a$.

(1) Show that in the standard coordinates (r, θ) on these two 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.
(2) Now take a piece of paper and try to figure out what parallel translation along a latitude on a cone looks like. If you unwrap the paper, then it is flat; thus parallel translation is what it is in the plane. Now rewrap the paper and observe that parallel translation along a latitude does not necessarily generate a closed parallel field.
(3) Show that in the above example the parallel field along $r = a$ closes up when $\dot{\rho}(a) = 0$.

EXERCISE 6.7.3 (Fermi-Walker transport). Related to parallel transport there is a more obscure type of transport sometimes used in physics. Let $c : [a, b] \to M$ be a curve into a Riemannian manifold whose speed never vanishes and

$$T = \frac{\dot{c}}{|\dot{c}|}$$

the unit tangent of c. We say that V is a *Fermi-Walker field* along c if

$$\begin{aligned}\dot{V} &= g(V, T)\,\dot{T} - g\left(V, \dot{T}\right) T \\ &= \left(\dot{T} \wedge T\right)(V).\end{aligned}$$

(1) Show that given $V(t_0)$ there is a unique Fermi-Walker field V along c whose value at t_0 is $V(t_0)$.
(2) Show that T is a Fermi-Walker field along c.
(3) Show that if V, W are Fermi-Walker fields along c, then $g(V, W)$ is constant along c.
(4) If c is a geodesic, then Fermi-Walker fields are parallel.

EXERCISE 6.7.4. 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 $k \le 0$ 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 theorem 6.2.2.)

EXERCISE 6.7.5. Let $c(s, t) : [0, 1]^2 \to (M, g)$ be a variation where $R\left(\frac{\partial c}{\partial s}, \frac{\partial c}{\partial t}\right) = 0$. Show that for each $v \in T_{c(0,0)}M$, there is a parallel field $V : [0, 1]^2 \to TM$ along c, i.e., $\frac{\partial V}{\partial s} = \frac{\partial V}{\partial t} = 0$ everywhere.

EXERCISE 6.7.6. Use the formula

$$R\left(\frac{\partial c}{\partial s}, \frac{\partial c}{\partial t}\right)\frac{\partial c}{\partial u} = \frac{\partial^3 c}{\partial s \partial t \partial u} - \frac{\partial^3 c}{\partial t \partial s \partial u}$$

to show that the two skew-symmetry properties and Bianchi's first identity from proposition 3.1.1 hold for the curvature tensor.

EXERCISE 6.7.7. Let c be a geodesic and X a Killing field in a Riemannian manifold. Show that the restriction of X to c is a Jacobi field.

EXERCISE 6.7.8. Let $c : [0, 1] \to M$ be a geodesic. Show that $\exp_{c(0)}$ has a critical point at $t\dot{c}(0)$ if and only if there is a nontrivial Jacobi field J along c such that $J(0) = 0$, $\dot{J}(0) \perp \dot{c}(0)$, and $J(t) = 0$.

EXERCISE 6.7.9. Fix $p \in M$ and $v \in \mathrm{seg}_p^0$. Consider a geodesic $c(t) = \exp_p(tv)$ and geodesic variation $\bar{c}(s,t) = \exp_p(t(v + sw))$ with variational Jacobi field $J(t)$. Show that if $f_0(x) = \frac{1}{2}|xp|^2$, then

$$\nabla f_0|_{c(1)} = \dot{c}(1),$$
$$\mathrm{Hess} f_0(J(1), J(1)) = g\left(\dot{J}(1), J(1)\right).$$

Use this equation to prove lemma 6.2.5 without first estimating Hess r.

EXERCISE 6.7.10. Let c be a geodesic in a Riemannian manifold and J_1, J_2 Jacobi fields along c.

(1) Show that $g\left(\dot{J}_1, J_2\right) - g\left(J_1, \dot{J}_2\right)$ is constant.
(2) Show that $g(J_1(t), \dot{c}(t)) = g(J_1(0), \dot{c}(0)) + g\left(\dot{J}_1(0), \dot{c}(0)\right)t$.

EXERCISE 6.7.11. Let J be a nontrivial Jacobi field along a unit speed geodesic c with $J(0) = 0$, $\dot{J}(0) \perp \dot{c}(0)$. Assume that the Riemannian manifold has sectional curvature $\leq K$.

(1) Define

$$\lambda = \frac{|J|^2}{g\left(J, \dot{J}\right)}$$

and show that $\dot{\lambda} \leq 1 + K\lambda^2$, $\lambda(0) = 0$ for as long as λ is defined.
(2) Show that if $J(b) = 0$ for some $b > 0$, then $g\left(J(t), \dot{J}(t)\right) = 0$ for some $t \in (0, b)$. Give an explicit example where this occurs.

EXERCISE 6.7.12. Let c be a geodesic in a Riemannian manifold and $\mathfrak{J}$ a space of Jacobi fields along c. Further assume that $\mathfrak{J}$ is self-adjoint, i.e., $g\left(\dot{J}_1, J_2\right) = g\left(J_1, \dot{J}_2\right)$ for all $J_1, J_2 \in \mathfrak{J}$. Consider the subspace

$$\mathfrak{J}(t) = \left\{\dot{J}(t) \mid J \in \mathfrak{J},\ J(t) = 0\right\} + \{J(t) \mid J \in \mathfrak{J}\} \subset T_{c(t)}M.$$

(1) Show that the two subspaces in this sum are orthogonal.
(2) Show that the space $\{J \in \mathfrak{J} \mid J(t) = 0\} \subset \mathfrak{J}$ is naturally isomorphic to the first summand in the decomposition.
(3) Show that $\dim \mathfrak{J} = \dim \mathfrak{J}(t)$ for all t. Hint: Consider a basis for $\mathfrak{J}$ where the first part of the basis spans $\{J \in \mathfrak{J} \mid J(t) = 0\}$.

EXERCISE 6.7.13. 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 $|p_i p_j| = |q_i q_j|$ there is an isometry F with $F(p_i) = q_i$. When $k = 1$ we simply say that the space is homogeneous.

(1) Show that a homogenous space has constant scalar curvature.
(2) Show that if $k > 1$ and (M, g) is k-point homogeneous, then M is also $(k-1)$-point homogeneous.
(3) Show that if (M, g) is two-point homogeneous, then (M, g) is an Einstein metric.
(4) Show that if (M, g) is three-point homogeneous, then (M, g) has constant curvature.
(5) Show that $\mathbb{RP}^2$ is not three-point homogeneous by finding two equilateral triangles of side lengths $\frac{\pi}{3}$ that are not congruent by an isometry.

It is possible to show that the simply connected space forms are the only three-point homogeneous spaces. Moreover, all 2-point homogeneous spaces are symmetric with rank 1 (see [106]).

EXERCISE 6.7.14. Starting with a geodesic on a two-dimensional space form, discuss how the equidistant curves change as they move away from the original geodesic.

EXERCISE 6.7.15. Let $r(x) = |xp|$ in a Riemannian manifold with $-K \le \sec \le K$. Write the metric as $g = dr^2 + g_r$ on $B(p, R) - \{p\} = (0, R) \times S^{n-1}$, where $2R < \mathrm{inj}_p$.

(1) Show that
$$\mathrm{sn}_K^2(r)\, ds_{n-1}^2 \le g_r \le \mathrm{sn}_{-K}^2(r)\, ds_{n-1}^2.$$
Hint: Estimate $|J|^2$, where J is a Jacobi field along a geodesic c with $c(0) = 0$, $J(0) = 0$, $\dot{J}(0) \perp \dot{c}(0)$, and $|\dot{J}(0)| = 1$.
(2) Show that there is a universal constant C such that
$$\left|\mathrm{Hess}\,\tfrac{1}{2}r^2 - g\right| \le CK^2 R^2$$
as long as $R < \frac{\pi}{2\sqrt{K}}$.

EXERCISE 6.7.16. Let (M, g) be a complete Riemannian manifold. Show that every element of $\pi_1(M, p)$ contains a shortest loop at p and that this shortest loop is a geodesic loop.

EXERCISE 6.7.17. Let (M, g) be a complete Riemannian manifold with $\mathrm{inj}_p < R$, where $\exp_p : B(0, R) \to B(p, R)$ is nonsingular. Show that the geodesic loop c at p that realizes the injectivity radius has index 0. Hint: When c is trivial as an element in $\pi_1(M, p)$, show that it does not admit a homotopy through loops that are all shorter than c. When c is nontrivial as an element in $\pi_1(M, p)$, show that it is a local minimum for the energy functional.

EXERCISE 6.7.18 (Frankel). Let M be an n-dimensional Riemannian manifold of positive curvature and A, B two closed totally geodesic submanifolds. Show directly that A and B must intersect if $\dim A + \dim B \geq n$. 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 second variation formula to get a shorter curve from A to B.

EXERCISE 6.7.19. Let M be a complete n-dimensional Riemannian manifold and $A \subset M$ a compact submanifold. Establish the following statements without using Wilking's connectedness principle.

(1) Show that curves in $\Omega_{A,A}(M)$ that are not stationary for the energy functional can be deformed to shorter curves in $\Omega_{A,A}(M)$.
(2) Show that the stationary curves for the energy functional on $\Omega_{A,A}(M)$ consist of geodesics that are perpendicular to A at the end points.
(3) If M has positive curvature, $A \subset M$ is totally geodesic, and $2\dim A \geq \dim M$, then all stationary curves can be deformed to shorter curves in $\Omega_{A,A}(M)$.
(4) *(Wilking)* Conclude using (3) that any curve $c : [0, 1] \to M$ that starts and ends in A is homotopic through such curves to a curve in A, i.e., $\pi_1(M, A)$ is trivial.

EXERCISE 6.7.20. Generalize Preissmann's theorem to show that any solvable subgroup of the fundamental group of a compact negatively curved manifold must be cyclic. Hint: Recall that the group is torsion free. Use contradiction and solvability to find a subgroup generated by deck transformations F, G with $F \circ G = G^k \circ F$, $k \neq 0$. Then show that if c is an axis for G, then $F \circ c$ is an axis for G^k and use uniqueness of axes for G^k to reach a contradiction.

EXERCISE 6.7.21. Let (M, g) be a compact manifold of positive curvature and $F : M \to M$ an isometry of finite order without fixed points. Show that if $\dim M$ is even, then F must be orientation reversing, while if $\dim M$ is odd, it must be orientation preserving. Weinstein has proven that this holds even if we don't assume that F has finite order.

EXERCISE 6.7.22. Use an analog of theorem 6.2.3 to show that any closed manifold of constant curvature $= 1$ must either be the standard sphere or have diameter $\leq \frac{\pi}{2}$. Generalize this to show that any closed manifold with $\sec \geq 1$ is either simply connected or has diameter $\leq \frac{\pi}{2}$. In section 12.3 we shall show the stronger statement that a closed manifold with $\sec \geq 1$ and diameter $> \frac{\pi}{2}$ must in fact be homeomorphic to a sphere.

EXERCISE 6.7.23. Consider a complete Riemannian n-manifold (M, g) with $|\sec| \leq K$. Fix n points p_i and a ball $B(p, \epsilon)$ such that the distance functions $r^i(x) = |xp_i|$ are smooth on $B(p, \epsilon)$ with $g\left(\nabla r^i, \nabla r^j\right)|_p = \delta^{ij}$ and $|pp_i| \geq 2\epsilon$.

(1) Let $g^{ij} = g\left(\nabla r^i, \nabla r^j\right) = g\left(dr^i, dr^j\right)$. Show that there exists $C(K,\epsilon) > 0$ such that $\left|dg^{ij}\right| \leq C = C(n,K,\epsilon)$.
(2) Show further that $C(n,K,\epsilon)$ can be chosen so that $C\left(n,\lambda^{-2}K,\lambda\epsilon\right) \to 0$ as $\lambda \to \infty$.
(3) Show that there is a $\delta = \delta(n,C) > 0$ such that g^{ij} is invertible on $B(p,\delta)$ and the inverse g_{ij} satisfies: $\left|\left[g_{ij} - \delta_{ij}\right]\right| \leq \frac{1}{9}$ and $\left|dg_{ij}\right| \leq C'(n,K,\epsilon)$. Hint: Find δ such that $\left|\left[g^{ij} - \delta^{ij}\right]\right| \leq \frac{1}{10}$ on $B(p,\delta)$ and use a geometric series of matrices to calculate the inverse.
(4) Show that $\left(r^1(x) - r^1(p), \ldots, r^n(x) - r^n(p)\right)$ form a coordinate system on $B(p,\delta)$ and that the image contains the ball $B\left(0,\frac{\delta}{4}\right)$. Hint: Inspect the proof of the inverse function theorem.

EXERCISE 6.7.24 (The Index Form). Below we shall use the second variation formula to prove several results established in section 5.7.3. If V, W are vector fields along a geodesic $c : [0,1] \to (M,g)$, then the *index form* is the symmetric bilinear form

$$I_0^1(V,W) = I(V,W) = \int_0^1 \left(g\left(\dot{V},\dot{W}\right) - g\left(R(V,\dot{c})\dot{c},W\right)\right) dt.$$

In case the vector fields come from a proper variation of c this is equal to the second variation of energy. Assume below that $c : [0,1] \to (M,g)$ locally minimizes the energy functional. This implies that $I(V,V) \geq 0$ for all proper variations.

(1) If $I(V,V) = 0$ for a proper variation, then V is a Jacobi field. Hint: Let W be any other variational field that also vanishes at the end points and use that

$$0 \leq I(V + \varepsilon W, V + \varepsilon W) = I(V,V) + 2\varepsilon I(V,W) + \varepsilon^2 I(W,W)$$

for all small ε to show that $I(V,W) = 0$. Then use that this holds for all W to show that V is a Jacobi field.
(2) Let V and J be variational fields along c such that $V(0) = J(0)$ and $V(1) = J(1)$. If J is a Jacobi field show that

$$I(V,J) = I(J,J).$$

(3) *(The Index Lemma)* Assume in addition that there are no Jacobi fields along c that vanish at both end points. If V and J are as in (2) show that $I(V,V) \geq I(J,J)$ with equality holding only if $V = J$ on $[0,1]$. Hint: Prove that if $V \neq J$, then

$$0 < I(V-J, V-J) = I(V,V) - I(J,J).$$

(4) Assume that there is a nontrivial Jacobi field J that vanishes at 0 and 1, show that $c : [0, 1+\varepsilon] \to M$ is not locally minimizing for $\varepsilon > 0$. Hint: For sufficiently small ε there is a Jacobi field $K : [1-\varepsilon, 1+\varepsilon] \to TM$ such that $K(1+\varepsilon) = 0$

and $K(1-\varepsilon) = J(1-\varepsilon)$. Let V be the variational field such that $V|_{[0,1-\varepsilon]} = J$ and $V|_{[1-\varepsilon,1+\varepsilon]} = K$. Finally extend J to be zero on $[1, 1+\varepsilon]$. Now show that

$$\begin{aligned} 0 = I_0^1(J,J) &= I_0^{1+\varepsilon}(J,J) = I_0^{1-\varepsilon}(J,J) + I_{1-\varepsilon}^{1+\varepsilon}(J,J) \\ &> I_0^{1-\varepsilon}(J,J) + I_{1-\varepsilon}^{1+\varepsilon}(K,K) = I(V,V). \end{aligned}$$

EXERCISE 6.7.25 (Index Comparison). Let J be a nontrivial Jacobi field along a unit speed geodesic c with $J(0) = 0$, $\dot{J}(0) \perp \dot{c}(0)$. Assume that the Riemannian manifold has sectional curvature $\geq k$. The index form on $c|_{[0,b]}$ is given by

$$I_0^b(V,V) = \int_0^b \left(\left|\dot{V}\right|^2 - g(R(V,\dot{c})\dot{c},V)\right) dt$$

and we assume that there are no Jacobi fields on $c|_{[0,b]}$ that vanish at the ends points as in part (3) of exercise 6.7.24.

(1) Show that $I_0^b(J,J) = g\left(J(b), \dot{J}(b)\right)$.
(2) Define

$$V(t) = \frac{\mathrm{sn}_k(t)}{\mathrm{sn}_k(b)} E(t)$$

where E is a parallel field with $E(b) = J(b)$. Show that

$$I_0^b(V,V) \leq \frac{\mathrm{sn}_k'(b)}{\mathrm{sn}_k(b)} |J(b)|^2.$$

Hint: Differentiate $\mathrm{sn}_k(t)\,\mathrm{sn}_k'(t)$.
(3) Conclude that $g\left(J(b), \dot{J}(b)\right) \leq \frac{\mathrm{sn}_k'(b)}{\mathrm{sn}_k(b)} |J(b)|^2$ and use this to prove the part of theorem 6.4.3 that relates to lower curvature bounds.

EXERCISE 6.7.26. Consider a subgroup $\mathrm{G} \subset \mathrm{Iso}(M,g)$ of a Riemannian manifold. The topology of $\mathrm{Iso}(M,g)$ is the compact-open topology discussed in exercise 5.9.41.

(1) Show that if M is complete, simply connected, has nonpositive curvature, and G is compact, then G has a fixed point, i.e., there exists $p \in M$ that is fixed by all elements in G. Hint: Imitate the proof of theorem 6.2.3.
(2) Given $p \in M$ and $\epsilon > 0$, we say that G is (p,ϵ)-small, if $\mathrm{G}p \subset \bar{B}(p,\epsilon)$. Show that for sufficiently small $\epsilon(p)$ the closure $\bar{\mathrm{G}} \subset \mathrm{Iso}(M,g)$ of a (p,ϵ)-small group is compact and also (p,ϵ)-small. Note: We do not assume that M is complete so closed balls are not necessarily compact.
(3) Show that if G is (p,ϵ)-small, then it is $(q, 2|pq| + \epsilon)$-small.

(4) Assume that G is (p, ϵ)-small and that ϵ is much smaller than the convexity radius for all points in $\bar{B}(p, 4\epsilon)$. Show that G has a fixed point. Hint: Imitate (1) after noting all of the necessary distance functions are convex on suitable domains.
(5) Given a Riemannian manifold show that for all $p \in M$ there exists $\epsilon > 0$ such that no subgroup $G \subset \mathrm{Iso}(M, g)$ can be (p, ϵ)-small. Hint: As in the proof of theorem 5.6.19 make G act freely on a suitable subset of $M \times \cdots \times M$.
(6) A topological group is said to have no small subgroups if there a neighborhood around the identity that contains no nontrivial subgroups. Show that $\mathrm{Iso}(M, g)$ has no small subgroups.

Bochner-Montgomery showed more generally that a locally compact subgroup of $\mathrm{Diff}(M)$ has no small subgroups. Gleason and Yamabe then later proved that a locally compact topological group without small subgroups is a Lie group. See also [79] for the complete story of this fascinating solution to Hilbert's 5th problem. It is still unknown whether (locally) compact subgroups of the homeomorphism group of a topological manifold are necessarily Lie groups.

EXERCISE 6.7.27. 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. Hint: Construct a suitable horizontal distribution by declaring that for a given curve in M all parallel fields along this curve correspond to the horizontal lifts of this curve.

EXERCISE 6.7.28. 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 $\mathrm{O}(n)$. Hint: Construct a suitable horizontal distribution by declaring that for a given curve in M all orthonormal parallel frames along this curve correspond to the horizontal lifts of this curve.

Chapter 7
Ricci Curvature Comparison

In this chapter we prove some of the fundamental results 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. Later some of the analytic estimates we develop here will be used to estimate Betti numbers for manifolds with lower curvature bounds.

The goal is to develop several techniques to help us understand lower Ricci curvature bounds. In the 50s Calabi discovered that one has weak upper bounds for the Laplacian of distance function given lower Ricci curvature bounds, even at points where this function isn't smooth. However, it wasn't until after 1970, when Cheeger and Gromoll proved their splitting theorem, that this was fully appreciated. 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 [14]. At the time, however, one only considered balls of radius less than the injectivity radius. 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 (see chapter 9) using only topological methods and volume comparison. Anderson 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 have proven to be false through the construction of 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 focus in this chapter will be on the fundamental comparison techniques and how they are used to prove a few rigidity theorems. In subsequent chapters there will be many further results related to lower Ricci curvature bounds that depend on more analytical techniques.

P. Petersen, *Riemannian Geometry*, Graduate Texts in Mathematics 171,
DOI 10.1007/978-3-319-26654-1_7

7.1 Volume Comparison

7.1.1 The Fundamental Equations

Throughout this section, assume that we have a complete Riemannian manifold (M, g) of dimension n and a distance function $r(x)$ that is smooth on an open set $U \subset M$. In subsequent sections we shall further assume that $r(x) = |xp|$ so that it is smooth on the image of the interior of the segment domain (see section 5.7.3). Recall the following fundamental equations for the metric from proposition 3.2.11:

(1) $L_{\partial_r} g = 2\mathrm{Hess} r$,
(2) $(\nabla_{\partial_r} \mathrm{Hess} r)(X, Y) + \mathrm{Hess}^2 r(X, Y) = -R(X, \partial_r, \partial_r, Y)$.

There is a similar set of equations for the volume form.

Proposition 7.1.1. *The volume form* vol *and Laplacian* Δr *of a smooth distance function* r *are related by:*

(tr1) $L_{\partial_r} \mathrm{vol} = \Delta r \,\mathrm{vol}$,

(tr2) $\partial_r \Delta r + \frac{(\Delta r)^2}{n-1} \leq \partial_r \Delta r + |\mathrm{Hess}\, r|^2 = -\mathrm{Ric}(\partial_r, \partial_r)$.

Proof. The first equation was established in section 2.1.3 as one of the definitions of the Laplacian of r.

To establish the second equation we take traces in (2). More precisely, select an orthonormal frame E_i, set $X = Y = E_i$, and sum over i. In addition it is convenient to assume that this frame is parallel: $\nabla_{\partial_r} E_i = 0$. On the right-hand side

$$\sum_{i=1}^{n} R(E_i, \partial_r, \partial_r, E_i) = \mathrm{Ric}(\partial_r, \partial_r).$$

While on the left-hand side

$$\begin{aligned}\sum_{i=1}^{n} (\nabla_{\partial_r} \mathrm{Hess}\, r)(E_i, E_i) &= \sum_{i=1}^{n} \partial_r \mathrm{Hess}\, r(E_i, E_i) \\ &= \partial_r \Delta r\end{aligned}$$

and

$$\begin{aligned}\sum_{i=1}^{n} \mathrm{Hess}^2 r(E_i, E_i) &= \sum_{i=1}^{n} g(\nabla_{E_i}\partial_r, \nabla_{E_i}\partial_r) \\ &= \sum_{i,j=1}^{n} g\left(\nabla_{E_i}\partial_r, g\left(\nabla_{E_i}\partial_r, E_j\right) E_j\right) \\ &= \sum_{i,j=1}^{n} g\left(\nabla_{E_i}\partial_r, E_j\right) g\left(\nabla_{E_i}\partial_r, E_j\right) \\ &= |\mathrm{Hess}\, r|^2.\end{aligned}$$

Finally we need to show that

$$\frac{(\Delta r)^2}{n-1} \leq |\mathrm{Hess}\, r|^2 .$$

To this end also assume that $E_1 = \partial_r$, then

$$\begin{aligned}
|\mathrm{Hess}\, r|^2 &= \sum_{i,j=1}^{n} \left(g\left(\nabla_{E_i}\partial_r, E_j\right)\right)^2 \\
&= \sum_{i,j=2}^{n} \left(g\left(\nabla_{E_i}\partial_r, E_j\right)\right)^2 \\
&\geq \frac{1}{n-1}\left(\sum_{i=2}^{n} g\left(\nabla_{E_i}\partial_r, E_i\right)\right)^2 \\
&= \frac{1}{n-1}(\Delta r)^2 .
\end{aligned}$$

The inequality

$$|A|^2 \geq \frac{1}{k}|\mathrm{tr}\,(A)|^2$$

for a $k \times k$ matrix A is a direct consequence of the Cauchy-Schwarz inequality

$$|(A, I_k)|^2 \leq |A|^2 |I_k|^2 = |A|^2 k,$$

where I_k is the identity $k \times k$ matrix. □

If we use the polar coordinate decomposition $g = dr^2 + g_r$ and vol_{n-1} is the standard volume form on $S^{n-1}(1)$, then $\mathrm{vol} = \lambda(r,\theta)\, dr \wedge \mathrm{vol}_{n-1}$, where θ indicates a coordinate on S^{n-1}. If we apply (tr1) to this version of the volume form we get

$$L_{\partial_r}\mathrm{vol} = L_{\partial_r}\left(\lambda(r,\theta)\, dr \wedge \mathrm{vol}_{n-1}\right) = \partial_r(\lambda)\, dr \wedge \mathrm{vol}_{n-1}$$

as both $L_{\partial_r} dr = 0$ and $L_{\partial_r}\mathrm{vol}_{n-1} = 0$. This allows us to simplify (tr1) to the formula

$$\partial_r \lambda = \lambda \Delta r.$$

In constant curvature k we know that $g_k = dr^2 + \mathrm{sn}_k^2(r)\, ds_{n-1}^2$, thus the volume form is

$$\mathrm{vol}_k = \lambda_k(r)\, dr \wedge \mathrm{vol}_{n-1} = \mathrm{sn}_k^{n-1}(r)\, dr \wedge \mathrm{vol}_{n-1} .$$

This conforms with the fact that

$$\Delta r = (n-1)\frac{\mathrm{sn}_k'(r)}{\mathrm{sn}_k(r)},$$
$$\partial_r\left(\mathrm{sn}_k^{n-1}(r)\right) = (n-1)\frac{\mathrm{sn}_k'(r)}{\mathrm{sn}_k(r)}\mathrm{sn}_k^{n-1}(r).$$

7.1.2 Volume Estimation

With the above information we can prove the estimates that are analogous to our basic comparison estimates for the metric and Hessian of $r(x) = |xp|$ assuming lower sectional curvature bounds (see section 6.4).

Lemma 7.1.2 (Ricci Comparison). *If* (M, g) *has* $\mathrm{Ric} \geq (n-1)\cdot k$ *for some* $k \in \mathbb{R}$, *then*

$$\Delta r \leq (n-1)\frac{\mathrm{sn}_k'(r)}{\mathrm{sn}_k(r)},$$
$$\partial_r\left(\frac{\lambda}{\lambda_k}\right) \leq 0,$$
$$\lambda(r,\theta) \leq \lambda_k(r) = \mathrm{sn}_k^{n-1}(r).$$

Proof. Notice that the right-hand sides of the inequalities correspond exactly to what one would obtain in constant curvature k. Thus the first inequality is a direct consequence of corollary 6.4.2 if we use $\rho = \frac{\Delta r}{n-1}$.

For the second inequality use that $\partial_r\lambda = \lambda\Delta r$ to conclude that

$$\partial_r\lambda \leq (n-1)\frac{\mathrm{sn}_k'(r)}{\mathrm{sn}_k(r)}\lambda$$

and

$$\partial_r\lambda_k = (n-1)\frac{\mathrm{sn}_k'(r)}{\mathrm{sn}_k(r)}\lambda_k.$$

This means that

$$\partial_r\left(\frac{\lambda}{\lambda_k}\right) \leq 0.$$

The last inequality follows from the second after the observation that $\lambda = r^{n-1} + O(r^n)$ at $r = 0$ so that

$$\lim_{r\to 0}\frac{\lambda}{\lambda_k} = 1.$$

□

Our first volume comparison yields the obvious upper volume bound coming from the upper bound on the volume density.

Lemma 7.1.3. *If* (M, g) *has* $\text{Ric} \geq (n-1) \cdot k$, *then* $\text{vol}\, B(p, r) \leq v(n, k, r)$, *where* $v(n, k, r)$ *denotes the volume of a ball of radius* r *in the constant curvature space form* S_k^n.

Proof. In polar coordinates

$$
\begin{aligned}
\text{vol}\, B(p, r) &= \int_{\text{seg}_p \cap B(0,r)} \lambda(r)\, dr \wedge \text{vol}_{n-1} \\
&\leq \int_{\text{seg}_p \cap B(0,r)} \lambda_k(r)\, dr \wedge \text{vol}_{n-1} \\
&\leq \int_{B(0,r)} \text{vol}_k \\
&= v(n, k, r). \qquad \square
\end{aligned}
$$

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 throughout the rest of the text.

Lemma 7.1.4 (Relative Volume Comparison, Bishop, 1964 and Gromov, 1980). *Let* (M, g) *be a complete Riemannian manifold with* $\text{Ric} \geq (n-1) \cdot k$. *The volume ratio*

$$
r \mapsto \frac{\text{vol}\, B(p, r)}{v(n, k, r)}
$$

is a nonincreasing function whose limit is 1 as $r \to 0$.

Proof. We will use exponential polar coordinates. The volume form $\lambda dr \wedge \text{vol}_{n-1}$ for (M, g) is initially defined only on some star-shaped subset of $T_pM = \mathbb{R}^n$ but we can just set $\lambda = 0$ outside this set. The comparison density λ_k is defined on all of $\mathbb{R}^n$ when $k \leq 0$ and on $B\left(0, \pi/\sqrt{k}\right)$ when $k > 0$. We can likewise extend $\lambda_k = 0$ outside $B\left(0, \pi/\sqrt{k}\right)$. Myers' theorem 6.3.3 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 $k > 0$.

The ratio is

$$
\frac{\text{vol}\, B(p, R)}{v(n, k, R)} = \frac{\int_0^R \int_{S^{n-1}} \lambda dr \wedge \text{vol}_{n-1}}{\int_0^R \int_{S^{n-1}} \lambda_k dr \wedge \text{vol}_{n-1}},
$$

and $0 \leq \lambda(r, \theta) \leq \lambda_k(r) = \text{sn}_k^{n-1}(r)$ everywhere.

Differentiation of this quotient with respect to R yields

$$\begin{aligned}
&\frac{d}{dR}\left(\frac{\operatorname{vol} B(p,R)}{v(n,k,R)}\right)\\
&= \frac{\left(\int_{S^{n-1}} \lambda\,(R,\theta)\operatorname{vol}_{n-1}\right)\left(\int_0^R \int_{S^{n-1}} \lambda_k\,(r)\,dr\wedge \operatorname{vol}_{n-1}\right)}{(v(n,k,R))^2}\\
&\quad -\frac{\left(\int_{S^{n-1}} \lambda_k\,(R)\operatorname{vol}_{n-1}\right)\left(\int_0^R \int_{S^{n-1}} \lambda\,(r,\theta)\,dr\wedge \operatorname{vol}_{n-1}\right)}{(v(n,k,R))^2}\\
&= (v(n,k,R))^{-2}\cdot \int_0^R \left[\left(\int_{S^{n-1}} \lambda\,(R,\theta)\operatorname{vol}_{n-1}\right)\cdot\left(\int_{S^{n-1}} \lambda_k\,(r)\operatorname{vol}_{n-1}\right)\right.\\
&\quad \left. -\left(\int_{S^{n-1}} \lambda_k\,(R)\operatorname{vol}_{n-1}\right)\left(\int_{S^{n-1}} \lambda\,(r,\theta)\operatorname{vol}_{n-1}\right)\right]dr.
\end{aligned}$$

So to see that

$$R \mapsto \frac{\operatorname{vol} B(p,R)}{v(n,k,R)}$$

is nonincreasing, it suffices to check that

$$\frac{\int_{S^{n-1}} \lambda\,(r,\theta)\operatorname{vol}_{n-1}}{\int_{S^{n-1}} \lambda_k\,(r)\operatorname{vol}_{n-1}} = \frac{1}{\omega_{n-1}}\int_{S^{n-1}} \frac{\lambda\,(r,\theta)}{\lambda_k\,(r)}\operatorname{vol}_{n-1}$$

is nonincreasing. This follows from lemma 7.1.2 as $\partial_r\left(\frac{\lambda(r,\theta)}{\lambda_k(r)}\right)\le 0$. □

7.1.3 The Maximum Principle

We explain how one can assign second derivatives to functions at points where the function is not smooth. In section 12.1 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 7.1.5. *If $f, h : (M,g) \to \mathbb{R}$ are C^2 functions such that $f(p) = h(p)$ and $f(x) \ge h(x)$ for all x near p, then*

$$\begin{aligned}
\nabla f\,(p) &= \nabla h\,(p)\,,\\
\operatorname{Hess} f|_p &\ge \operatorname{Hess} h|_p,\\
\Delta f(p) &\ge \Delta h(p).
\end{aligned}$$

Proof. If $(M, g) \subset (\mathbb{R}, g_{\mathbb{R}})$, then the theorem is standard from single variable calculus. In general, let $c : (-\varepsilon, \varepsilon) \to M$ be a curve with $c(0) = p$. Then use this observation on $f \circ c$, $h \circ c$ to see that

$$df(\dot{c}(0)) = dh(\dot{c}(0)),$$
$$\operatorname{Hess} f(\dot{c}(0), \dot{c}(0)) \geq \operatorname{Hess} h(\dot{c}(0), \dot{c}(0)).$$

This clearly implies the lemma if we let $v = \dot{c}(0)$ run over all $v \in T_pM$. □

The lemma implies that a C^2 function $f : M \to \mathbb{R}$ has $\operatorname{Hess} f|_p \geq B$, where B is a symmetric bilinear map on T_pM (or $\Delta f(p) \geq a \in \mathbb{R}$), if and only if 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) $\operatorname{Hess} f_\varepsilon|_p \geq B - \varepsilon \cdot g|_p$ (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 $\operatorname{Hess} f$ and Δf. Support functions are also known as barrier functions in PDE theory.

For a continuous function $f : (M, g) \to \mathbb{R}$ we say that: $\operatorname{Hess} f|_p \geq B$ (or $\Delta f(p) \geq a$) if and only if for all $\varepsilon > 0$ there exist smooth support functions f_ε satisfying (1)-(3). One also says that $\operatorname{Hess} f|_p \geq B$ (or $\Delta f(p) \geq a$) hold in the support or barrier sense. In PDE theory there are other important ways of defining weak derivatives. The notion used here is guided by what we can obtain from geometry.

One can easily check that if $(M, g) \subset (\mathbb{R}, g_{\mathbb{R}})$, then f is convex if $\operatorname{Hess} f \geq 0$ everywhere. Thus, $f : (M, g) \to \mathbb{R}$ is convex if $\operatorname{Hess} f \geq 0$ everywhere. Using this, one can prove

Theorem 7.1.6. *If $f : (M, g) \to \mathbb{R}$ is continuous with* $\operatorname{Hess} f \geq 0$ *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 using the idea of support functions. A continuous function $f : (M, g) \to \mathbb{R}$ with $\Delta f \geq 0$ everywhere is said to be *subharmonic*. If $\Delta f \leq 0$, then f is *superharmonic*.

Theorem 7.1.7 (The Strong Maximum Principle). *If $f : (M, g) \to \mathbb{R}$ is continuous and subharmonic, then f is constant in a neighborhood of every local maximum. In particular, if f has a global maximum, then 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 $\operatorname{Hess} f_\varepsilon(p) \leq 0$, which contradicts (3).

Next assume that $\Delta f \geq 0$ and let $p \in M$ be a local maximum for f. For sufficiently small $r < \operatorname{inj}(p)$ the restriction $f : B(p,r) \to \mathbb{R}$ will have a global maximum at p. If f is constant on $B(p,r)$, then we are done. Otherwise assume (by possibly decreasing r) that $f(x_0) \neq f(p)$ for some

$$x_0 \in \partial B(p,r) = \{x \in M \mid |xp| = r\}$$

and define

$$V = \{x \in \partial B(p,r) \mid f(x) = f(p)\}.$$

Our goal is to construct a smooth function $h = e^{\alpha\varphi} - 1$ such that

$$\begin{aligned} h &< 0 \text{ on } V, \\ h(p) &= 0, \\ \Delta h &> 0 \text{ on } \bar{B}(p,r). \end{aligned}$$

This function is found by first selecting an open disc $U \subset \partial B(p,r)$ that contains V and then ϕ such that

$$\begin{aligned} \phi(p) &= 0, \\ \phi &< 0 \text{ on } U, \\ \nabla\phi &\neq 0 \text{ on } \bar{B}(p,r). \end{aligned}$$

Such a ϕ can be found by letting $\phi = x^1$ in a coordinate system $\left(x^1, \ldots, x^n\right)$ centered at p where U lies in the lower half-plane: $x^1 < 0$ (see also figure 7.1). Lastly, choose α so large that

$$\Delta h = \alpha e^{\alpha\phi}\left(\alpha|\nabla\phi|^2 + \Delta\phi\right) > 0 \text{ on } \overline{B}(p,r).$$

Now consider the function $\bar{f} = f + \delta h$ on $\overline{B}(p,r)$. This function has a local maximum in the interior $B(p,r)$, provided δ is very small, since this forces

$$\begin{aligned} \bar{f}(p) &= f(p) \\ &> \max\left\{\bar{f}(x) \mid x \in \partial B(p,r)\right\}. \end{aligned}$$

On the other hand, we can also show that $\bar{f}$ has positive Laplacian, thus obtaining a contradiction as in the first part of the proof. To see that the Laplacian is positive,

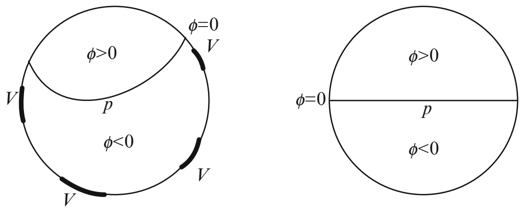


Fig. 7.1 Coordinate function construction

select f_ε as 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 $\bar{f}$ 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 $\operatorname{Hess} f \equiv 0$, i.e., both of the inequalities $\operatorname{Hess} f \geq 0$, $\operatorname{Hess} f \leq 0$ hold everywhere. This easily implies that

$$(f \circ c)(t) = f(c(0)) + \alpha t$$

for each geodesic c as $f \circ c$ is both convex and concave. Thus

$$f \circ \exp_p(x) = f(p) + g(v_p, x)$$

for each $p \in M$ and some $v_p \in T_pM$. In particular, 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 an important fact that any harmonic function is C^∞ if the metric is C^∞. Using the above maximum principle this is a standard result in PDE theory (see also theorem 9.2.7 and section 11.2).

Theorem 7.1.8 (Regularity of harmonic functions). *If $f : (M, g) \to \mathbb{R}$ is continuous and harmonic in the weak sense, then f is smooth.*

Proof. We fix $p \in M$ and a neighborhood Ω around p with smooth boundary. We can in addition assume that Ω is contained in a coordinate neighborhood. It is a standard but nontrivial fact from PDE theory that the following Dirichlet boundary value problem has a solution:

$$\Delta u = 0,$$
$$u|_{\partial\Omega} = f|_{\partial\Omega}.$$

Moreover, such a solution u is smooth on the interior of Ω. Now consider the two functions $u - f$ and $f - u$ on Ω. If they are both nonpositive, then they must vanish and hence $f = u$ is smooth near p. Otherwise one of these functions must be positive somewhere. However, as it vanishes on the boundary and is subharmonic this implies that it has an interior global maximum. The maximum principle then shows that the function is constant, but this is only possible if it vanishes. □

7.1.4 Geometric Laplacian Comparison

The idea of using support functions to estimate the Laplacian is particularly convenient for geometric applications since distance functions always have support functions from above.

Lemma 7.1.9 (Calabi, 1958). *If (M, g) is complete and $\operatorname{Ric}(M, g) \geq (n-1)k$, then any distance function $r(x) = |xp|$ satisfies:*

$$\Delta r(x) \leq (n-1)\frac{\operatorname{sn}_k'(r(x))}{\operatorname{sn}_k(r(x))}.$$

Proof. We know from lemma 7.1.2 that the result is true whenever r is smooth. In general, we can for each $q \in M$ choose a unit speed segment $\sigma : [0, L] \to M$ with $\sigma(0) = p$, $\sigma(L) = q$. Then the triangle inequality implies that $r_\varepsilon(x) = \varepsilon + |\sigma(\varepsilon)x|$ is a support function from above for r at q. If all these support functions are smooth at q, then

$$\begin{aligned}
\Delta r_\varepsilon(q) &\leq (n-1)\,\frac{\operatorname{sn}_k'(r_\epsilon(q))}{\operatorname{sn}_k(r_\epsilon(q))} \\
&= (n-1)\,\frac{\operatorname{sn}_k'(r(q)-\epsilon)}{\operatorname{sn}_k(r(q)-\epsilon)} \\
&\searrow (n-1)\,\frac{\operatorname{sn}_k'(r(q))}{\operatorname{sn}_k(r(q))}
\end{aligned}$$

as $\varepsilon \to 0$ since $\frac{\operatorname{sn}_k'(r)}{\operatorname{sn}_k(r)}$ is decreasing.

Now for the smoothness. Fix $\varepsilon > 0$ and suppose r_ε is not smooth at q. Then we know from lemma 5.7.9 that either

(1) there are two segments from $\sigma(\varepsilon)$ to q,
(2) q is a critical value for $\exp_{\sigma(\varepsilon)} : \operatorname{seg}(\sigma(\varepsilon)) \to M$.

Case (1) would give 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 $\exp_q$ has $\sigma(\varepsilon)$ as a critical value.

Using that q is critical for $\exp_{\sigma(\varepsilon)}$, we find a Jacobi field $J(t) : [\varepsilon, L] \to TM$ along $\sigma|_{[\varepsilon,L]}$ such that $J(\varepsilon) = 0$, $\dot{J}(\varepsilon) \neq 0$ and $J(L) = 0$ (see section 5.7.3). Then also $\dot{J}(L) \neq 0$ as it solves a linear second-order equation. Running backwards from q to $\sigma(\varepsilon)$ then shows that $\exp_q$ is critical at $\sigma(\varepsilon)$. This however contradicts that $\sigma : [0, L] \to M$ is a segment. □

7.1.5 *The Segment, Poincaré, and Sobolev Inequalities*

We shall use the results obtained in section 7.1.2 to prove a some important analytic inequalities that will be used in chapter 9.

Theorem 7.1.10 (The Segment Inequality, Cheeger and Colding, 1996). *Assume that* (M, g) *has* $\mathrm{Ric} \geq (n-1)k$, $k \leq 0$. *Let* $f : M \to [0, \infty)$ *and* $A, B \subset W \subset M$. *Further select segments* $c_{x,y} : [0, 1] \to M$ *between points* $x, y \in M$. *If* $c_{x,y}(t) \in W$ *for all* $x \in A$, $y \in B$, $t \in [0, 1]$, *and* $\operatorname{diam} W \leq D$, *then*

$$\int_{A\times B} \int_0^1 f \circ c_{x,y}(t)\, dt \operatorname{vol}_x \wedge \operatorname{vol}_y \leq C (\operatorname{vol} A + \operatorname{vol} B) \int_W f \operatorname{vol},$$

where $C = C\left(n, kD^2\right)$.

Proof. Define

$$C = \max_{R \leq D} \frac{\operatorname{sn}_k^{n-1}(R)}{\operatorname{sn}_k^{n-1}\left(\frac{1}{2}R\right)}.$$

Note that when $k = 0$ we have $C = 2^{n-1}$ and otherwise one can show that

$$C = \frac{\sinh^{n-1}\left(\sqrt{-k}D\right)}{\sinh^{n-1}\left(\frac{1}{2}\sqrt{-k}D\right)}.$$

Fix $x \in A$, $t \geq \frac{1}{2}$, and use polar coordinates with center x. The map $y \mapsto c_{x,y}(t)$ is a well-defined “scaling” by t inside the segment domain. With that in mind we have:

$$\begin{aligned}
\int_B f \circ c_{x,y}(t) \operatorname{vol}_y &= \int_B f \circ c_{x,y}(t)\, \lambda(y)\, dr \wedge \operatorname{vol}_{n-1} \\
&= \int_B f \circ c_{x,y}(t)\, \lambda\left(c_{x,y}(t)\right) \frac{\lambda(y)}{\lambda\left(c_{x,y}(t)\right)} dr \wedge \operatorname{vol}_{n-1} \\
&\leq C \int_B f \circ c_{x,y}(t)\, \lambda\left(c_{x,y}(t)\right) dr \wedge \operatorname{vol}_{n-1} \\
&\leq C \int_W f \operatorname{vol}.
\end{aligned}$$

This gives us

$$\int_A \int_B \int_{\frac{1}{2}}^1 f \circ c_{x,y}(t)\, dt \operatorname{vol}_y \operatorname{vol}_x \leq \frac{1}{2} C \operatorname{vol} A \int_W f \operatorname{vol}.$$

Similarly

$$\int_B \int_A \int_0^{\frac{1}{2}} f \circ c_{x,y}(t)\, dt \operatorname{vol}_x \operatorname{vol}_y \leq \frac{1}{2} C \operatorname{vol} B \int_W f \operatorname{vol}.$$

Adding these gives the desired result. □

This estimate allows us to establish a weak Poincaré inequality. To formulate the result it'll be convenient to define the L^p norm on a domain B by also averaging the integral:

$$\|u\|_{p,B} = \left(\frac{1}{\operatorname{vol} B} \int_B |u|^p \operatorname{vol} \right)^{\frac{1}{p}}$$

and using the notation $u_B = \frac{1}{\operatorname{vol} B} \int_B u \operatorname{vol}$ for the average value of a function on a bounded domain.

Corollary 7.1.11. *Assume that* (M, g) *has* $\operatorname{Ric} \geq (n-1)\, k$, $k \leq 0$. *Any smooth* $u : M \to [0, \infty)$ *satisfies*

$$\left\| u - u_{B(p,R)} \right\|_{1,B(p,R)} \leq 4C^2 R \, \|du\|_{1,B(p,2R)},$$

where $R \leq D$.

Proof. This proof is due to Cheeger and Colding. We use the segment inequality with $A = B = B(p, R)$, $W = B(p, 2R)$, and $f = |du|$ as well as the observation

$$\begin{aligned}
\int_B |u - u_B| &= \int_B \left| u(x) - \frac{1}{\operatorname{vol} B} \int_B u(y) \operatorname{vol}_y \right| \operatorname{vol}_x \\
&= \int_B \left| \frac{1}{\operatorname{vol} B} \int_B (u(x) - u(y)) \operatorname{vol}_y \right| \operatorname{vol}_x \\
&\leq \frac{1}{\operatorname{vol} B} \int_B \int_B |u(x) - u(y)| \operatorname{vol}_y \operatorname{vol}_x \\
&\leq \frac{1}{\operatorname{vol} B} \int_B \int_B \int_0^1 |xy| \left| |du| \left(c_{x,y}(t) \right) \right| dt \operatorname{vol}_y \operatorname{vol}_x \\
&\leq \frac{2R}{\operatorname{vol} B} \int_B \int_B \int_0^1 \left| |du| \left(c_{x,y}(t) \right) \right| dt \operatorname{vol}_y \operatorname{vol}_x .
\end{aligned}$$

This shows that

$$\left\|u - u_{B(p,R)}\right\|_{1,B(p,R)} \leq 4CR\frac{\operatorname{vol} B(p,2R)}{\operatorname{vol} B(p,R)} \left\||du|\right\|_{1,B(p,2R)} .$$

The result follows by using that the volume ratio is bounded explicitly by the ratio

$$\frac{v(n,k,2R)}{v(n,k,R)} \leq \frac{v(n,k,2D)}{v(n,k,D)}. \qquad \square$$

Remark 7.1.12. Note that the corollary holds for any measurable u with a function G in place of $|du|$ provided

$$|u(x) - u(y)| \leq \int_0^1 G(c(t))\,|\dot{c}|\,dt$$

for all $c \in \Omega_{x,y}$. Such a G is also called an *upper gradient*.

This leads us, surprisingly, to the much stronger Poincaré-Sobolev inequality where the domain is the same on both sides and a stronger norm is used on the left-hand side.

Theorem 7.1.13. *Assume that* (M,g) *has* $\operatorname{Ric} \geq (n-1)k$, $k \leq 0$. *For all smooth* $u : M \to [0,\infty)$ *and* $\nu \in \left[1, \frac{n}{n-1}\right]$

$$\left\|u - u_{B(x,R)}\right\|_{\nu,B(x,R)} \leq C\left(n, kD^2\right) R \left\||du|\right\|_{1,B(x,R)} ,$$

where $R \leq D$.

We offer a proof by Hajłasz and Koskela that can be found in [60]. An even shorter proof is possible when $\nu < \frac{n}{n-1}$. Traditionally, proofs of this theorem required a very deep and difficult theorem from geometric measure theory. Here we only need a few basic concepts from analysis together with the weak Poincaré inequality and relative volume comparison. This proof has the added benefit of easily allowing generalizations to suitable metric spaces. We will for simplicity prove it in case $B(x,R) = M$ and D is an upper bound for the diameter of M. To keep constants at bay we shall also keep writing them as C with the understanding that $C = C\left(n, kD^2\right)$ depends on n and possibly also kD^2. However, the constants might change from line to line in a proof.

The *maximal function* of a function u is defined as

$$M(u)(x) = \sup_{R \in (0,D]} \frac{1}{\operatorname{vol} B(x,R)} \int_{B(x,R)} |u| \operatorname{vol}.$$

We only need the weak version of the maximal function estimate. Note that this estimate does not bound $\|M(u)\|_1$ in terms of $\|u\|_1$, which is in fact impossible, but it can be used to prove the standard bounds $\|M(u)\|_p \leq C\,\|u\|_p$ for all $p > 1$.

Theorem 7.1.14 (Maximal Function Theorem). *There exists a constant* $C = C\left(n, kD^2\right)$ *such that*

$$t \operatorname{vol}\{M(u) > t\} \leq C \int |u| \operatorname{vol}.$$

Proof. Note that for each $x \in \{M(u) > t\}$ there is $R_x \leq D$ such that

$$t \operatorname{vol} B(x, R_x) < \int_{B(x,R_x)} |u| \operatorname{vol}.$$

Now use the basic covering property (see exercise 7.5.5) to cover $\{M(u) > t\}$ by balls $B(x_i, 5R_{x_i})$ with the property that $B(x_i, R_{x_i})$ are pairwise disjoint. Relative volume comparison gives us

$$\frac{\operatorname{vol} B(x, 5R)}{\operatorname{vol} B(x, R)} \leq \frac{v(n, k, 5D)}{v(n, k, D)} = C = C\left(n, kD^2\right).$$

We can then estimate

$$\begin{aligned} t \operatorname{vol}\{M(u) > t\} &\leq \sum t \operatorname{vol} B(x_i, 5R_{x_i}) \\ &\leq C \sum t \operatorname{vol} B(x_i, R_{x_i}) \\ &< C \sum \int_{\operatorname{vol} B\left(x_i, R_{x_i}\right)} |u| \operatorname{vol} \\ &\leq C \int |u| \operatorname{vol}. \end{aligned}$$

□

Theorem 7.1.15. *Assume that* (M, g) *has* $\operatorname{Ric} \geq (n-1)k$*,* $k \leq 0$*, and* $\operatorname{diam} M \leq D$*. Let* $u : M \to [0, \infty)$ *be smooth. There is a weak Poincaré-Sobolev inequality*

$$t^{\frac{n}{n-1}} \operatorname{vol}\left\{\left|u - u_{B(x,R)}\right| > t\right\} \leq C R^{\frac{n}{n-1}} \operatorname{vol} M \, \||du|\|_1^{\frac{n}{n-1}},$$

where $C = C\left(n, kD^2\right)$.

Proof. For simplicity we prove this when $R = D$. Fix $x \in M$ and define $R_i = 2^{-i}D$. If $B_i = B(x, R_i)$, then $M = B_0$. By continuity of u we have $u(x) = \lim u_{B_i}$. This tells us that

$$\begin{aligned} \left|u(x) - u_{B_0}\right| &\leq \sum_{i=0}^{\infty} \left|u_{B_i} - u_{B_{i+1}}\right| \\ &\leq \sum_{i=0}^{\infty} \|u - u_{B_i}\|_{1,B_{i+1}} \end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{i=0}^{\infty} \frac{\operatorname{vol} B_i}{\operatorname{vol} B_{i+1}} \left\|u - u_{B_i}\right\|_{1,B_i} \\
&\leq C \sum_{i=0}^{\infty} \left\|u - u_{B_i}\right\|_{1,B_i} \\
&\leq 2C^3 \sum_{i=0}^{\infty} R_i \left\||du|\right\|_{1,B_{i-1}} .
\end{aligned}
$$

Therefore, it suffices to prove an estimate of the form:

$$
t^{\frac{n}{n-1}} \operatorname{vol}\left\{ \sum_{i=0}^{\infty} R_i \left\||du|\right\|_{1,B_{i-1}} > t \right\} \leq C D^{\frac{n}{n-1}} \operatorname{vol} M \left\||du|\right\|_1^{\frac{n}{n-1}} .
$$

For any $x \in M$ and $r > 0$ split up the sum

$$
\sum_{R_i \leq r} R_i \left\||du|\right\|_{1,B_{i-1}} + \sum_{R_i > r} R_i \left\||du|\right\|_{1,B_{i-1}} .
$$

The first term is controlled by the maximal function

$$
\begin{aligned}
\sum_{R_i \leq r} R_i \left\||du|\right\|_{1,B_{i-1}} &\leq \left(\sum_{R_i \leq r} R_i \right) M\left(|du|\right)(x) \\
&\leq 2rM\left(|du|\right)(x) .
\end{aligned}
$$

The second term is bounded by $\left\||du|\right\|_1$ as follows:

$$
\begin{aligned}
\sum_{R_i > r} R_i \left\||du|\right\|_{1,B_{i-1}} &\leq \left(\sum_{R_i > r} R_i \frac{\operatorname{vol} M}{\operatorname{vol} B_{i-1}} \right) \left\||du|\right\|_1 \\
&\leq C \sum_{R_i > r} R_i \left(\frac{D}{R_{i-1}} \right)^n \left\||du|\right\|_1 \\
&\leq C \sum_{R_i > r} \frac{2^{-i}}{2^{-n(i-1)}} D \left\||du|\right\|_1 \\
&= C 2^{-n} \sum_{R_i > r} 2^{(n-1)i} D \left\||du|\right\|_1 \\
&\leq 2^{1-n} C 2^{(n-1)i_0} D \left\||du|\right\|_1 \\
&= 2^{1-n} C \left(2^{i_0} D^{-1}\right)^{n-1} D^n \left\||du|\right\|_1 \\
&= 2^{1-n} C R_{i_0}^{1-n} D^n \left\||du|\right\|_1 \\
&\leq 2^{1-n} C r^{1-n} D^n \left\||du|\right\|_1 .
\end{aligned}
$$

Thus

$$\sum_{i=0}^{\infty} R_i \,\||du|\|_{1,B_{i-1}} \leq C\left(rM\left(|du|\right)(x) + r^{1-n}D^n \,\||du|\|_1\right)$$

and for $r = D\left(\frac{\||du|\|_1}{M(|du|)(x)}\right)^{\frac{1}{n}}$ yields the estimate:

$$\sum_{i=0}^{\infty} R_i \,\||du|\|_{1,B_{i-1}} \leq CD\left(M\left(|du|\right)(x)\right)^{\frac{n-1}{n}} \,\||du|\|_1^{\frac{1}{n}}.$$

Note that while it is natural to assume $r \leq D$ this estimate is still valid when $r > D$.

The maximal function theorem can now be used to obtain the inequality

$$\begin{aligned}
\mathrm{vol}\left\{\sum_{i=0}^{\infty} R_i \,\||du|\|_{1,B_{i-1}} > t\right\} &= \mathrm{vol}\left\{\left(\sum_{i=0}^{\infty} R_i \,\||du|\|_{1,B_{i-1}}\right)^{\frac{n}{n-1}} > t^{\frac{n}{n-1}}\right\} \\
&\leq \mathrm{vol}\left\{CD^{\frac{n}{n-1}} \,\||du|\|_1^{\frac{1}{n-1}} M\left(|du|\right)(x) > t^{\frac{n}{n-1}}\right\} \\
&\leq t^{-\frac{n}{n-1}} CD^{\frac{n}{n-1}} \,\||du|\|_1^{\frac{1}{n-1}} \int_M |du| \,\mathrm{vol} \\
&\leq t^{-\frac{n}{n-1}} CD^{\frac{n}{n-1}} \,\mathrm{vol}\, M \,\||du|\|_1^{\frac{n}{n-1}}.
\end{aligned}$$

□

The proof of the Poincaré-Sobolev inequality can now be completed as follows.

Proof of theorem 7.1.13. We use the estimate from theorem 7.1.15 to prove the result. First we need two more elementary facts. Note that for any $c \in \mathbb{R}$:

$$\|u - u_M\|_p \leq \|c - u_M\|_p + \|u - c\|_p = |u_M - c| + \|u - c\|_p \leq 2\,\|u - c\|_p$$

and

$$\inf_c \|u - c\|_p \leq \|u - u_M\|_p.$$

So it suffices to estimate $\|u - c\|_p$ for a suitable c.

For a general $u : M \to \mathbb{R}$ find m such that $\mathrm{vol}\{u \geq m\} \geq \frac{\mathrm{vol}\, M}{2}$ and $\mathrm{vol}\{u \leq m\} \geq \frac{\mathrm{vol}\, M}{2}$. Then split u into the two functions $v^+ = \max\{u - m, 0\}$ and $v^- = \max\{m - u, 0\}$. Note that they both satisfy $\mathrm{vol}\left\{v^{\pm} = 0\right\} \geq \frac{\mathrm{vol}\, M}{2}$.

While $v^{\pm}$ is not smooth we can set $\left|dv^{\pm}\right| = 0$ at all points where $v^{\pm}$ vanishes. Thus it suffices to show that

$$\left\|v^{\pm}\right\|_{\frac{n}{n-1}} \leq C\left(n, kD^2\right) D \left\|\left|dv^{\pm}\right|\right\|_1$$

as

$$\|u\|_{\frac{n}{n-1}} \le \left\|v^+\right\|_{\frac{n}{n-1}} + \left\|v^-\right\|_{\frac{n}{n-1}} \text{ and } \left|dv^+\right| + \left|dv^-\right| \le |du| .$$

We first claim that $v = v^{\pm}$ satisfies

$$\operatorname{vol}\{v > t\} \le 2 \operatorname{vol}\left\{|v-c| > \frac{t}{2}\right\}.$$

To see this note that when $\frac{t}{2} \le c$ we have $\left\{c - v > \frac{t}{2}\right\} \subset \{v = 0\}$, while when $\frac{t}{2} \ge c$ we have $\left\{v > c + \frac{t}{2}\right\} \subset \{v > t\}$.

For $0 < a < b$ consider the truncated function

$$v_a^b(x) = \begin{cases} b-a & \text{if } v(x) \ge b, \\ v(x) - a & \text{if } a < v(x) \le b, \\ 0 & \text{if } v(x) \le a, \end{cases}$$

and note that the weak Poincaré inequality holds for v_a^b if we use $|dv| \cdot \chi_{\{a<v\le b\}}$ as an upper gradient. Theorem 7.1.15 can now be used:

$$\begin{aligned} t^{\frac{n}{n-1}} \operatorname{vol}\left\{v_a^b > t\right\} &\le 2t^{\frac{n}{n-1}} \inf_c \operatorname{vol}\left\{\left|v_a^b - c\right| > \frac{t}{2}\right\} \\ &= 2^{\frac{n}{n-1}+1}\left(\frac{t}{2}\right)^{\frac{n}{n-1}} \inf_c \operatorname{vol}\left\{\left|v_a^b - c\right| > \frac{t}{2}\right\} \\ &\le 2^{\frac{n}{n-1}+1}\left(\frac{t}{2}\right)^{\frac{n}{n-1}} \operatorname{vol}\left\{\left|v_a^b - \left(v_a^b\right)_M\right| > \frac{t}{2}\right\} \\ &\le CD^{\frac{n}{n-1}} 2^{\frac{n}{n-1}+1} \operatorname{vol} M \left\||dv| \cdot \chi_{\{a<v\le b\}}\right\|_1^{\frac{n}{n-1}} . \end{aligned}$$

We then get the desired estimate as follows:

$$\begin{aligned} \int v^{\frac{n}{n-1}} \operatorname{vol} &\le \sum_{k=-\infty}^{\infty} 2^{k\frac{n}{n-1}} \operatorname{vol}\left\{2^{k-1} < v \le 2^k\right\} \\ &\le \sum_k 2^{k\frac{n}{n-1}} \operatorname{vol}\left\{v > 2^{k-1}\right\} \\ &\le \sum_k 2^{k\frac{n}{n-1}} \operatorname{vol}\left\{v_{2^{k-2}}^{2^{k-1}} > 2^{k-1} - 2^{k-2}\right\} \\ &= \sum_k 2^{k\frac{n}{n-1}} \operatorname{vol}\left\{v_{2^{k-2}}^{2^{k-1}} > 2^{k-2}\right\} \\ &\le 2^{3\frac{n}{n-1}+1} CD^{\frac{n}{n-1}} \operatorname{vol} M \sum_k \left\||dv| \cdot \chi_{\left\{2^{k-2}<v\le 2^{k-1}\right\}}\right\|_1^{\frac{n}{n-1}} \end{aligned}$$

$$\leq 2^{3\frac{n}{n-1}+1} C D^{\frac{n}{n-1}} \operatorname{vol} M \left\| \sum_k |dv| \cdot \chi_{\left\{2^{k-2}<v\leq 2^{k-1}\right\}} \right\|_1^{\frac{n}{n-1}}$$

$$= 2^{3\frac{n}{n-1}+1} C D^{\frac{n}{n-1}} \operatorname{vol} M \, \||dv|\|_1^{\frac{n}{n-1}} . \qquad \square$$

Remark 7.1.16. See exercise 7.5.17 for the Poincaré inequality for functions with Dirichlet boundary conditions.

Finally we also obtain an entire hierarchy of such inequalities.

Proposition 7.1.17. *Assume that all smooth functions on* (M, g) *satisfy the inequality*

$$\|u - u_M\|_{\frac{s}{s-1}} \leq S \|du\|_1 ,$$

with $s > 1$, *then for* $1 \leq p < s$

$$\|u\|_{\frac{sp}{s-p}} \leq \frac{p(s-1)}{s-p} S \|du\|_p + \|u\|_p .$$

Proof. When $p = 1$, this follows from

$$\begin{aligned} \|u - u_M\|_{\frac{s}{s-1}} &\geq \|u\|_{\frac{s}{s-1}} - \|u_M\|_{\frac{s}{s-1}} \\ &= \|u\|_{\frac{s}{s-1}} - |u_M| \\ &\geq \|u\|_{\frac{s}{s-1}} - \|u\|_1 . \end{aligned}$$

For $p > 1$ first note that

$$\begin{aligned} \|u\|^q_{\frac{qs}{s-1}} &= \|u^q\|_{\frac{s}{s-1}} \\ &\leq S \|du^q\|_1 + \|u^q\|_1 \\ &= Sq \left\|u^{q-1} du\right\|_1 + \left\|u^{q-1} u\right\|_1 \\ &\leq \|u\|^{q-1}_{\frac{p(q-1)}{p-1}} \left(Sq \|du\|_p + \|u\|_p\right) . \end{aligned}$$

Then choose $q = \frac{p(s-1)}{s-p}$ so that $\frac{qs}{s-1} = \frac{p(q-1)}{p-1} = \frac{sp}{s-p}$ to obtain the desired inequality. $\square$

Finally we establish the Rellich compactness theorem. The same strategy can also be used to prove the more general Kondrachov compactness theorem for $L^p(M)$. Define $W^{1,2}(M)$ as the Hilbert space closure of $C^\infty(M)$ with the square norm $\|u\|_2^2 + \|du\|_2^2$. Recall that a sequence $v_i \in (H, (\cdot,\cdot))$ in a Hilbert space is *weakly convergent*, $v_i \rightharpoonup v$ if $(v_i, w) \to (v, w)$ for all $w \in H$. Moreover, any bounded sequence has a weakly convergent subsequence.

Theorem 7.1.18 (Rellich Compactness). *Assume (M^n, g) is a compact Riemannian n-manifold. The inclusion $W^{1,2}(M) \subset L^2(M)$ is compact.*

Outline of Proof. Consider a sequence u_i of smooth functions where $\|u_i\|_2^2 + \|du_i\|_2^2$ is bounded. Then there will be a weakly convergent subsequence $u_i \rightharpoonup u$. In particular, $u_{i,B(x,R)} \to u_{B(x,R)}$ for fixed $x \in M$ and $R > 0$.

By the Lebesgue differentiation theorem we also have that $u_{B(x,R)} \to u(x)$ as $R \to 0$ for almost all $x \in M$. Next note that by theorem 7.1.15

$$\operatorname{vol}\left\{\left|u_i - u_{i,B(x,R)}\right| > \epsilon\right\} \le C\left(\frac{R}{\epsilon}\right)^{\frac{n}{n-1}} \operatorname{vol} M \, \|du_i\|_1^{\frac{n}{n-1}} \le C'\left(\frac{R}{\epsilon}\right)^{\frac{n}{n-1}},$$

where C' is independent of i.

This implies that $\operatorname{vol}\{|u_i(x) - u(x)| > \epsilon\} \to 0$ as $i \to \infty$. We can then extract another subsequence of u_i that converges pointwise to u almost everywhere on M. Since $\|u_i\|_{\frac{2n}{n-2}}$ is bounded Egorov's theorem implies that $u_i \to u$ in L^2. □

7.2 Applications of Ricci Curvature Comparison

7.2.1 Finiteness of Fundamental Groups

Our first application of volume comparison shows how one can control the fundamental group. We start with a result that addresses how fundamental groups can be represented.

Lemma 7.2.1 (Gromov, 1980). *A compact Riemannian manifold M admits generators $\{c_1, \ldots c_m\}$ for the fundamental group $\Gamma = \pi_1(M)$ such that all relations for Γ in are of the form $c_i \cdot c_j \cdot c_k^{-1} = 1$ for suitable i, j, k. Moreover, the generators c_i can be represented by loops of length $\le 3 \operatorname{diam}(M)$.*

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.

Any loop in M based at x is homotopic to a loop in the 1-skeleton of the triangulation, i.e., a loop that is constructed out of juxtaposing edges e_{ij}. Since $e_{ij}e_{jk} = e_{ij}\sigma_j^{-1}\sigma_j e_{jk}$ such loops are the product of loops of the form σ_{ij}. Therefore, Γ is generated by σ_{ij}.

Next observe that if three vertices x_i, x_j, x_k are adjacent to each other, then they span a 2-simplex $\triangle_{ijk}$. Consequently 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 in M. Then σ

also contracts in the 2-skeleton. Thus, a homotopy corresponds to a collection of 2-simplices $\triangle_{ijk}$. In this way we can represent the relation $\sigma = 1$ as a product of elementary relations of the form $\sigma_{ij}\sigma_{jk}\sigma_{ik}^{-1} = 1$.

The generators correspond to loops of length $\leq 2\,\mathrm{diam}\,(M) + \varepsilon$ so the result is proven. □

A simple example might be instructive here.

Example 7.2.2. 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 $k \to \infty$. This is also justified by the next theorem.

For numbers $n \in \mathbb{N}$, $k \in \mathbb{R}$, and $v, D \in (0, \infty)$, let $\mathfrak{M}\,(n, k, v, D)$ denote the class of compact Riemannian n-manifolds with

$$\begin{aligned} \mathrm{Ric} &\geq (n-1)\,k, \\ \mathrm{vol} &\geq v, \\ \mathrm{diam} &\leq D. \end{aligned}$$

We can now prove:

Theorem 7.2.3 (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 $\{c_1, \ldots, c_m\}$ 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. Fix $x \in \tilde{M}$ and consider c_i as deck transformations. The lemma also guarantees that $|xc_i(x)| \leq 3D$. 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 $\mathrm{vol}\,F = \mathrm{vol}\,M$. One could, for example, choose the Dirichlet domain

$$F = \left\{ z \in \tilde{M} \mid |xz| \leq |c(x)z| \text{ for all } c \in \pi_1(M) \right\}.$$

Then the sets $c_i\,(F)$ are disjoint up to sets of measure 0; all have the same volume; and all lie in the ball $B\,(x, 6D)$. Thus,

$$m \leq \frac{\mathrm{vol}\,B\,(x, 6D)}{\mathrm{vol}\,F} \leq \frac{v\,(n, k, 6D)}{v}.$$

In other words, we have bounded the number of generators in terms of n, D, v, k alone. □

A related result shows that groups generated by short loops must in fact be finite.

Lemma 7.2.4 (Anderson, 1990). *For fixed numbers $n \in \mathbb{N}$, $k \in \mathbb{R}$, and $v, D \in (0, \infty)$ there exist $L = L(n, k, v, D)$ 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 subgroup generated by loops $\{c_1, \ldots, c_k\}$ 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 $F \subset \tilde{M}$ as above with $x \in F$. Thus, for any $c_1, c_2 \in \pi_1(M)$, either $c_1 = c_2$ or $c_1(F) \cap c_2(F)$ has measure 0.

Now define $U(r)$ as the set of $c \in \Gamma$ such that c can be written as a product of at most r elements from $\{c_1, \ldots, c_k\}$. Since $|xc_i(x)| \leq L$ for all i it follows that $|xc(x)| \leq r \cdot L$ for all $c \in U(r)$. This means that $c(F) \subset B(x, r \cdot L + D)$. As the sets $c(F)$ are disjoint up to sets of measure zero, we obtain

$$\begin{aligned} |U(r)| &\leq \frac{\operatorname{vol} B(x, r \cdot L + D)}{\operatorname{vol} F} \\ &\leq \frac{v(n, k, r \cdot L + D)}{v}. \end{aligned}$$

Now define

$$\begin{aligned} N &= \frac{v(n, k, 2D)}{v} + 1, \\ L &= \frac{D}{N}. \end{aligned}$$

If Γ has more than N elements we get a contradiction by using $r = N$ as we would have

$$\begin{aligned} \frac{v(n, k, 2D)}{v} + 1 &= N \\ &\leq |U(N)| \\ &\leq \frac{v(n, k, 2D)}{v}. \end{aligned}$$

□

7.2.2 *Maximal Diameter Rigidity*

Next we show how Laplacian comparison can be used. Given Myers' diameter estimate, it is natural to ask what happens when the diameter attains it maximal value. The next result shows that only the sphere has this property.

Theorem 7.2.5 (S. Y. Cheng, 1975). *If (M, g) is a complete Riemannian manifold with $\operatorname{Ric} \geq (n-1)k > 0$ and $\operatorname{diam} = \pi/\sqrt{k}$, then (M, g) is isometric to S^n_k.*

Proof. Fix $p, q \in M$ such that $|pq| = {}^{\pi}/\sqrt{k}$. Define $r(x) = |xp|$, $\tilde{r}(x) = |xq|$. We will show that

(1) $r + \tilde{r} = {}^{\pi}/\sqrt{k},\ x \in M$.
(2) $r, \tilde{r}$ are smooth on $M - \{p, q\}$.
(3) $\operatorname{Hess} r = \frac{\operatorname{sn}_k'}{\operatorname{sn}_k} ds_{n-1}^2$ on $M - \{p, q\}$.
(4) $g = dr^2 + \operatorname{sn}_k^2 ds_{n-1}^2$.

We already know that (3) implies (4) and that (4) implies M must be S_k^n.

Proof of (1): Consider $\tilde{r}(x) = |xq|$ and $r(x) = |xp|$, where $|pq| = \pi/\sqrt{k}$. Then $r + \tilde{r} \geq \pi/\sqrt{k}$, and equality will hold for any $x \in M - \{p, q\}$ that lies on a segment joining p and q. On the other hand lemma 7.1.9 implies

$$
\begin{aligned}
\Delta(r + \tilde{r}) &\leq \Delta r + \Delta \tilde{r} \\
&\leq (n-1)\sqrt{k}\cot(\sqrt{k}r(x)) + (n-1)\sqrt{k}\cot(\sqrt{k}\tilde{r}(x)) \\
&\leq (n-1)\sqrt{k}\cot(\sqrt{k}r(x)) + (n-1)\sqrt{k}\cot\left(\sqrt{k}\left(\frac{\pi}{\sqrt{k}} - r(x)\right)\right) \\
&= (n-1)\sqrt{k}(\cot(\sqrt{k}r(x)) + \cot(\pi - \sqrt{k}r(x))) = 0.
\end{aligned}
$$

Thus $r + \tilde{r}$ is superharmonic on $M - \{p, q\}$ and has a global minimum. Consequently, the minimum principle implies that $r + \tilde{r} = \pi/\sqrt{k}$ on M.

Proof of (2): If $x \in M - \{q, p\}$, then x can be joined to both p and q by segments c_1, c_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 (see proposition 5.4.4). Thus c_1 and c_2 are both subsegments of a larger segment. This implies from our characterization of when distance functions are smooth that both r and $\tilde{r}$ are smooth at $x \in M - \{p, q\}$ (see corollary 5.7.11).

Proof of (3): Since $r(x) + \tilde{r}(x) = {}^{\pi}/\sqrt{k}$, we have $\Delta r = -\Delta \tilde{r}$. On the other hand,

$$
\begin{aligned}
(n-1)\frac{\operatorname{sn}_k'(r(x))}{\operatorname{sn}_k(r(x))} &\geq \Delta r(x) \\
&= -\Delta \tilde{r}(x) \\
&\geq -(n-1)\frac{\operatorname{sn}_k'(\tilde{r}(x))}{\operatorname{sn}_k(\tilde{r}(x))} \\
&= -(n-1)\frac{\operatorname{sn}_k'\left(\frac{\pi}{\sqrt{k}} - r(x)\right)}{\operatorname{sn}_k\left(\frac{\pi}{\sqrt{k}} - r(x)\right)} \\
&= (n-1)\frac{\operatorname{sn}_k'(r(x))}{\operatorname{sn}_k(r(x))}.
\end{aligned}
$$

This implies,

$$\Delta r = (n-1)\frac{\mathrm{sn}_k'}{\mathrm{sn}_k}$$

and

$$\begin{aligned}
-(n-1)k &= \partial_r(\Delta r) + \frac{(\Delta r)^2}{n-1} \\
&\leq \partial_r(\Delta r) + |\operatorname{Hess} r|^2 \\
&\leq -\operatorname{Ric}(\partial_r, \partial_r) \\
&\leq -(n-1)k.
\end{aligned}$$

Hence, all inequalities are equalities, and in particular

$$(\Delta r)^2 = (n-1)|\operatorname{Hess} r|^2.$$

Recall from the proof of tr2 from proposition 7.1.1 that this gives us equality in the Cauchy-Schwarz inequality $k|A|^2 \geq (\operatorname{tr} A)^2$. Thus $A = \frac{\operatorname{tr} A}{k} I_k$. In our case we have restricted $\operatorname{Hess} r$ to the $(n-1)$ dimensional space orthogonal to ∂_r so on this space we obtain:

$$\operatorname{Hess} r = \frac{\Delta r}{n-1} g_r = \frac{\mathrm{sn}_k'}{\mathrm{sn}_k} g_r.$$

□

We now know that a complete manifold with $\operatorname{Ric} \geq (n-1) \cdot k > 0$ has diameter $\leq \pi/\sqrt{k}$, and equality holds only when the space is S_k^n. Therefore, a natural perturbation question is: Do manifolds with $\operatorname{Ric} \geq (n-1) \cdot k > 0$ and $\operatorname{diam} \approx \pi/\sqrt{k}$, have to 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 and $n = 3$ was proven in [95] (but sadly never published). The examples for $n \geq 4$ are divided into two cases: $n = 4$ and $n \geq 5$.

Example 7.2.6 (Anderson, 1990). For $n = 4$ consider metrics on $I \times S^3$ of the form

$$dr^2 + \rho^2\sigma_1^2 + \phi^2(\sigma_2^2 + \sigma_3^2).$$

If we define

$$\rho(r) = \begin{cases} \frac{\sin(ar)}{a} & r \leq r_0, \\ c_1 \sin(r+\delta) & r \geq r_0, \end{cases}$$

$$\phi(r) = \begin{cases} br^2 + c & r \leq r_0, \\ c_2 \sin(r+\delta) & r \geq r_0, \end{cases}$$

and then reflect these function in $r = \pi/2 - \delta$, we get a metric on $\mathbb{CP}^2 \sharp \bar{\mathbb{CP}}^2$. For any small $r_0 > 0$ we can adjust the parameters so that ρ and ϕ become C^1 and generate a metric with $\mathrm{Ric} \geq 3$. 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 7.2.7 (Otsu, 1991). For $n \geq 5$ we consider standard doubly warped products:

$$dr^2 + \rho^2 \cdot ds_2^2 + \phi^2 ds_{n-3}^2$$

on $I \times S^2 \times S^{n-3}$. Similar choices for ρ and ϕ 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 only constructed C^1 functions ρ, ϕ and therefore only C^1 metrics. However, the functions are concave and can easily be smoothed near the break points so as to stay concave. This will not change the values or first derivatives much and only increase the second derivative in absolute value. Thus the lower curvature bound still holds.

7.3 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, it can be used to show that $S^3 \times S^1$ does not admit a Ricci flat metric.

7.3.1 Rays and Lines

We will work only with complete and noncompact manifolds in this section. A *ray* $r(t) : [0, \infty) \to (M, g)$ is a unit speed geodesic such that

$$|r(t)r(s)| = |t - s| \text{ for all } t, s \geq 0.$$

One can think of a ray as a semi-infinite segment or as a segment from $r(0)$ to infinity. A *line* $l(t) : \mathbb{R} \to (M, g)$ is a unit speed geodesic such that

$$|l(t)l(s)| = |t - s| \text{ for all } t, s \in \mathbb{R}.$$

Lemma 7.3.1. *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.*

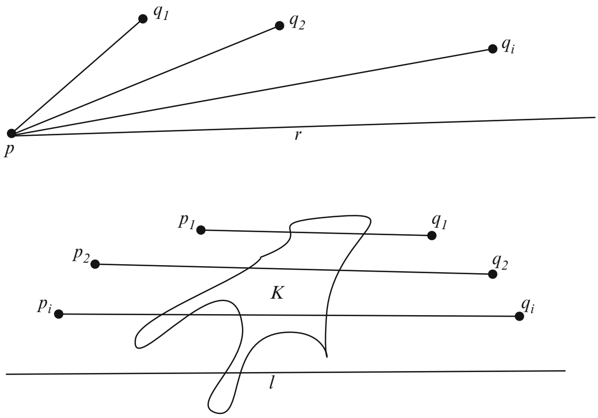


Fig. 7.2 Construction of rays and lines

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 7.2). 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

$$|\sigma(s)\sigma(t)| = \lim |\sigma_i(s)\sigma_i(t)| = |s - t|.$$

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, then there is 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 passes 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 7.2). □

Example 7.3.2. Surfaces of revolution $dr^2 + \rho^2(r)ds_{n-1}^2$, where $\rho : [0, \infty) \to [0, \infty)$ and $\dot{\rho}(t) < 1,\ \ddot{\rho}(t) < 0,\ t > 0$, cannot contain any lines. These manifolds look like paraboloids.

Example 7.3.3. Any complete metric on $S^{n-1} \times \mathbb{R}$ must contain a line since the manifold is disconnected at infinity.

Example 7.3.4. The Schwarzschild metric on $S^{n-2} \times \mathbb{R}^2$ does not contain any lines. This will also follow from our main result in this section as the space is not metrically a product.

Theorem 7.3.5 (The Splitting Theorem, Cheeger and Gromoll, 1971). *If* (M, g) *contains a line and has* $\operatorname{Ric} \geq 0$*, then* (M, g) *is isometric to a product* $(H \times \mathbb{R}, g_0 + dt^2)$.

Outline of Proof. The proof is quite involved and will require several constructions. The main idea is to find a distance function $r : M \to \mathbb{R}$ (i.e. $|\nabla r| \equiv 1$) that is linear (i.e. $\operatorname{Hess} r \equiv 0$). Having found such a function, one can easily see that $M = U_0 \times \mathbb{R}$, where $U_0 = \{r = 0\}$ and $g = dt^2 + g_0$. The maximum principle will play a key role in showing that r, when it has been constructed, is both smooth and linear. Recall that in the proof of the maximal diameter theorem 7.2.5 we used two distance functions $r, \tilde{r}$ placed at maximal distance from each other and then proceeded to show that $r + \tilde{r}$ is constant. This implied that $r, \tilde{r}$ were smooth, except at the two chosen points, and that Δr is exactly what it is in constant curvature. We then used the rigidity part of the Cauchy-Schwarz inequality to compute $\operatorname{Hess} r$. In the construction of our linear distance function we shall use a similar construction. In this situation the two ends of the line play the role of the points at maximal distance. Using this line we will construct two distance functions $b_\pm$ from infinity that are continuous, 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 implies 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 C^∞. At this point in the proof it is shown that they are distance functions, i.e., $|\nabla b_\pm| \equiv 1$. We can then invoke proposition 7.1.1 to conclude that

$$\begin{aligned} 0 &= D_{\nabla b_\pm} \Delta b_\pm + \frac{(\Delta b_\pm)^2}{n-1} \\ &\leq D_{\nabla b_\pm} \Delta b_\pm + |\operatorname{Hess} b_\pm|^2 \\ &= |\operatorname{Hess} b_\pm|^2 \\ &\leq -\operatorname{Ric}(\nabla b_\pm, \nabla b_\pm) \\ &\leq 0. \end{aligned}$$

This shows that $|\operatorname{Hess} b_\pm|^2 = 0$ and $b_\pm$ are the sought after linear distance functions. □

7.3.2 Busemann Functions

For the rest of this section fix a complete noncompact Riemannian manifold (M, g) with nonnegative Ricci curvature. Let $c : [0, \infty) \to (M, g)$ be a unit speed ray, and define

$$b_t(x) = |xc(t)| - t.$$

Proposition 7.3.6. *The functions b_t satisfy:*

(1) *For fixed x, the function $t \mapsto b_t(x)$ is decreasing and bounded in absolute value by $|xc(0)|$.*
(2) $|b_t(x) - b_t(y)| \le |xy|$.
(3) $\Delta b_t(x) \le \frac{n-1}{b_t+t}$ *everywhere.*

Proof. (2) and (3) are obvious since $b_t(x) + t$ is the distance from $c(t)$. For (1), first observe that the triangle inequality implies

$$|b_t(x)| = ||xc(t)| - t| = ||xc(t)| - |c(0)c(t)|| \le |xc(0)|.$$

Second, if $s < t$, then

$$\begin{aligned} b_t(x) - b_s(x) &= |xc(t)| - t - |xc(s)| + s \\ &= |xc(t)| - |xc(s)| - |c(t)c(s)| \\ &\le |c(t)c(s)| - |c(t)c(s)| = 0. \end{aligned}$$ □

This proposition shows that the family of distance decreasing functions $\{b_t\}_{t\ge 0}$ is pointwise bounded and decreasing. Thus, b_t converges pointwise to a distance decreasing function b_c satisfying

$$\begin{aligned} |b_c(x) - b_c(y)| &\le |xy|, \\ |b_c(x)| &\le |xc(0)|, \end{aligned}$$

and

$$b_c(c(r)) = \lim b_t(c(r)) = \lim (|c(r)c(t)| - t) = -r.$$

This function b_c is called the *Busemann function* for c and should be interpreted as renormalized a distance function from "$c(\infty)$."

Example 7.3.7. If $M = (\mathbb{R}^n, g_{\mathbb{R}^n})$, then all Busemann functions are of the form

$$b_c(x) = \dot{c}(0) \cdot (c(0) - x)$$

(see figure 7.3).

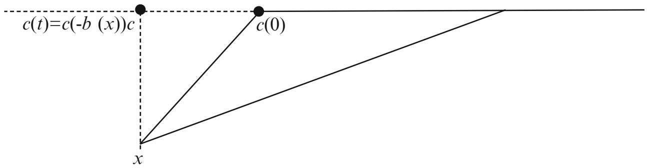


Fig. 7.3 Busemann function in Euclidean space

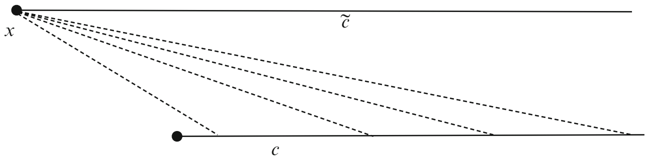


Fig. 7.4 Asymptote construction from a ray

The level sets $b_c^{-1}(t)$ are called *horospheres*. In $\mathbb{R}^n$ these are obviously hyperplanes. In the Poincaré model of hyperbolic space they look like spheres that are tangent to the boundary.

Given our ray c, as before, and $p \in M$, consider a family of unit speed segments $\sigma_t : [0, L_t] \to (M, g)$ from p to $c(t)$. As in the construction of rays this family subconverges to a ray $\tilde{c} : [0, \infty) \to M$, with $\tilde{c}(0) = p$. Such $\tilde{c}$ are called *asymptotes* for c from p (see figure 7.4) and need not be unique.

Proposition 7.3.8. *The Busemann functions are related by:*

(1) $b_c(x) \leq b_c(p) + b_{\tilde{c}}(x)$.
(2) $b_c(\tilde{c}(t)) = b_c(p) + b_{\tilde{c}}(\tilde{c}(t)) = b_c(p) - t$.

Proof. Let $\sigma_i : [0, L_i] \to (M, g)$ be the segments converging to $\tilde{c}$. To check (1), observe that

$$\begin{aligned} |xc(s)| - s &\leq |x\tilde{c}(t)| + |\tilde{c}(t)c(s)| - s \\ &= |x\tilde{c}(t)| - t + |p\tilde{c}(t)| + |\tilde{c}(t)c(s)| - s \\ &\to |x\tilde{c}(t)| - t + |p\tilde{c}(t)| + b_c(\tilde{c}(t)) \text{ as } s \to \infty. \end{aligned}$$

Thus, we see that (1) is true provided that (2) is true. To establish (2), note that

$$|pc(t_i)| = |p\sigma_i(s)| + |\sigma_i(s)c(t_i)|$$

for some sequence $t_i \to \infty$. Then $\sigma_i(s) \to \tilde{c}(s)$ and

$$\begin{aligned} b_c(p) &= \lim \left(|pc(t_i)| - t_i\right) \\ &= \lim \left(|p\tilde{c}(s)| + |\tilde{c}(s)c(t_i)| - t_i\right) \end{aligned}$$

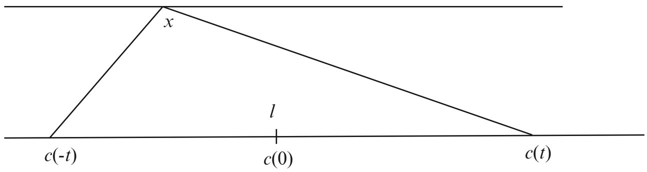


Fig. 7.5 Triangle inequality for two Busemann functions

$$
\begin{aligned}
&= |p\tilde{c}(s)| + \lim\left(|\tilde{c}(s)c(t_i)| - t_i\right) \\
&= s + b_c(\tilde{c}(s)) \\
&= -b_{\tilde{c}}(\tilde{c}(s)) + b_c(\tilde{c}(s)).
\end{aligned}
$$
□

We have shown that b_c has $b_c(p) + b_{\tilde{c}}$ as support function from above at $p \in M$.

Lemma 7.3.9. *If* $\mathrm{Ric}(M, g) \geq 0$, *then* $\Delta b_c \leq 0$ *everywhere.*

Proof. Since $b_c(p) + b_{\tilde{c}}$ is a support function from above at p, we only need to check that $\Delta b_{\tilde{c}} \leq 0$ at p. To see this, observe that the functions $b_t(x) = |x\tilde{c}(t)| - t$ are support functions from above for $b_{\tilde{c}}$ at p. Furthermore, these functions are smooth at p with

$$\Delta b_t(p) \leq \frac{n-1}{t} \to 0 \text{ as } t \to \infty.$$
□

Proof of Theorem 7.3.5. Now suppose (M, g) has $\mathrm{Ric} \geq 0$ and contains a line $c(t) : \mathbb{R} \to M$. Let b^+ be the Busemann function for $c : [0, \infty) \to M$, and b^- the Busemann function for $c : (-\infty, 0] \to M$. Thus,

$$
\begin{aligned}
b^+(x) &= \lim_{t \to +\infty} \left(|xc(t)| - t\right), \\
b^-(x) &= \lim_{t \to +\infty} \left(|xc(-t)| - t\right).
\end{aligned}
$$

Clearly,

$$b^+(x) + b^-(x) = \lim_{t \to +\infty} \left(|xc(t)| + |xc(-t)| - 2t\right),$$

so by the triangle inequality $\left(b^+ + b^-\right)(x) \geq 0$ for all x. Moreover, $\left(b^+ + b^-\right)(c(t)) = 0$ since c is a line (see figure 7.5).

This gives us a function $b^+ + b^-$ with $\Delta(b^+ + b^-) \leq 0$ and a global minimum at $c(t)$. The minimum principle then shows that $b^+ + b^- = 0$ everywhere. In particular, $b^+ = -b^-$ and $\Delta b^+ = \Delta b^- = 0$ everywhere.

To finish the proof of the splitting theorem, we still need to show that $b^\pm$ are distance functions, i.e. $\left|\nabla b^\pm\right| \equiv 1$. To see this, let $p \in M$ and construct asymptotes $\tilde{c}^\pm$ for $c^\pm$ from p. Then consider $b_t^\pm(x) = \left|x\tilde{c}^\pm(t)\right| - t$, and observe:

$$b_t^+(x) \geq b^+(x) - b^+(p) = -b^-(x) + b^-(p) \geq -b_t^-(x)$$

with equality holding for $x = p$. Since both $b_t^{\pm}$ are smooth at p with unit gradient it follows that $\nabla b_t^+(p) = -\nabla b_t^-(p)$. Then $b^{\pm}$ must also be differentiable at p with unit gradient. Therefore, we have shown (without using that $b^{\pm}$ are smooth from $\Delta b^{\pm} = 0$) that $b^{\pm}$ are everywhere differentiable with unit gradient. The result that harmonic functions are smooth can now be invoked and the proof is finished as explained earlier. □

7.3.3 Structure Results in Nonnegative Ricci Curvature

The splitting theorem gives several nice structure results for compact manifolds with nonnegative Ricci curvature.

Corollary 7.3.10. *$S^k \times S^1$ does not admit any Ricci flat metrics when $k = 2, 3$.*

Proof. The universal covering is $S^k \times \mathbb{R}$. As this space is disconnected at infinity any metric with nonnegative Ricci curvature must split. If the original metric is Ricci flat, then after the splitting we obtain a Ricci flat metric on a k-manifold H that is homotopy equivalent to S^k. In particular, H is compact and simply connected. If $k \leq 3$, such a metric must also be flat and so can't be simply connected as it is compact. □

When $k \geq 4$ it is not known whether any space that is homotopy equivalent to S^k admits a Ricci flat metric, but there do exist Ricci flat metrics on compact simply connected manifolds in dimensions ≥ 4.

Theorem 7.3.11 (Structure Theorem for Nonnegative Ricci Curvature, Cheeger and Gromoll, 1971). *Suppose (M, g) is a compact Riemannian manifold with* $\mathrm{Ric} \geq 0$.

(1) *The universal cover $(\tilde{M}, \tilde{g})$ splits isometrically as a product $N \times \mathbb{R}^k$, where N is a compact manifold.*
(2) *The isometry group splits* $\mathrm{Iso}\left(\tilde{M}\right) = \mathrm{Iso}(N) \times \mathrm{Iso}\left(\mathbb{R}^k\right)$.
(3) *There exists a finite normal subgroup* $\mathrm{G} \subset \pi_1(M)$ *whose factor group is* $\pi_1(M) \cap \mathrm{Iso}\left(\mathbb{R}^k\right)$ *and there is a finite index subgroup* $\mathbb{Z}^k \subset \pi_1(M) \cap \mathrm{Iso}\left(\mathbb{R}^k\right)$.

Proof. First we use the splitting theorem to write $\tilde{M} = N \times \mathbb{R}^k$, where N does not contain any lines. Observe that if $c(t) = (c_1(t), c_2(t)) \in N \times \mathbb{R}^k$ is a geodesic, then both c_i are geodesics, and if c is a line, then both c_i are also lines unless they are constant. Thus, all lines in $\tilde{M}$ must be of the form $c(t) = (x, \sigma(t))$, where $x \in N$ and σ is a line in $\mathbb{R}^k$.

(2) Let $F : \tilde{M} \to \tilde{M}$ be an isometry. If $L(t)$ is a line in $\tilde{M}$, then $F \circ L$ is also a line in $\tilde{M}$. Since all lines in $\tilde{M}$ lie in $\mathbb{R}^k$ and every vector tangent to $\mathbb{R}^k$ is the velocity of some line, we see that for each $x \in N$ we can find $F_1(x) \in N$ such that

$$F : \{x\} \times \mathbb{R}^k \to \{F_1(x)\} \times \mathbb{R}^k.$$

This implies that F must be of the form $F = (F_1, F_2)$, where $F_1 : N \to N$ is an isometry. Since DF preserves the tangents spaces to $\mathbb{R}^k$ it must also preserve the tangent spaces to N. Thus $F_2 : \mathbb{R}^k \to \mathbb{R}^k$. This shows that $\mathrm{Iso}\left(\tilde{M}\right) = \mathrm{Iso}(N) \times \mathrm{Iso}\left(\mathbb{R}^k\right)$.

(1) Since the deck transformations π_1 act by isometries we can consider the group $\pi_1 \cap \mathrm{Iso}(N)$ that comes from the projection $N \times \mathbb{R}^k \to N$. As π_1 acts discretely and cocompactly on $\tilde{M}$, it follows that $\pi_1 \cap \mathrm{Iso}(N)$ also acts cocompactly on N. In particular, for any sequence $p_i \in N$, it is possible to select $F_i \in \pi_1 \cap \mathrm{Iso}(N)$ such that all the points $F_i(p_i)$ lie in a fixed compact subset of N.

If N is not compact, then it must contain a ray $c(t) : [0, \infty) \to N$. We can then choose a sequence $t_i \to \infty$ and $F_i \in \pi_1 \cap \mathrm{Iso}(N)$ such that $F_i(c(t_i))$ lie in a compact set. We can then choose a subsequence so that $DF_i(\dot{c}(t_i))$ converges to a unit vector $v \in TN$. This implies that the geodesics $c_i : \mathbb{R} \to N$ defined by $c_i(t) = F_i(c(t + t_i))$ converge to the geodesic $\exp(tv)$. Moreover, for a fixed $a \in \mathbb{R}$ the geodesics c_i are rays on $[a, \infty)$ when $t_i \geq -a$ so it follows that $\exp(tv)$ is also a ray on $[a, \infty)$. But this shows that $\exp(tv) : \mathbb{R} \to N$ is a line which contradicts that N does not contain any lines.

(3) Let G be the kernel that comes from the map $\pi_1(M) \to \pi_1(M) \cap \mathrm{Iso}\left(\mathbb{R}^k\right)$ induced by the projection $N \times \mathbb{R}^k \to \mathbb{R}^k$. This group acts freely and discretely on $N \times \mathbb{R}^k$ without acting in the second factor. Thus it acts freely and discretely on N and must be finite as N is compact.

The translations form a normal subgroup $\mathbb{R}^k \subset \mathrm{Iso}\left(\mathbb{R}^k\right)$ whose factor group is $\mathrm{O}(k)$. The intersection $\pi_1(M) \cap \mathbb{R}^k$ is a finitely generated Abelian group with finite index in $\pi_1 \cap \mathrm{Iso}\left(\mathbb{R}^k\right)$ that acts discretely and cocompactly on $\mathbb{R}^k$. In particular, it is of the form $\mathbb{Z}^m$. When $m < k$ it is not possible for $\mathbb{Z}^m$ to act cocompactly on $\mathbb{R}^k$ since it will generate a proper subspace of the space of translations on $\mathbb{R}^k$. On the other hand if $m > k$, then $\mathbb{Z}^m$ will contain two elements that are linearly independent over $\mathbb{Q}$ but not over $\mathbb{R}$ inside the space of translations. The subgroup in $\mathbb{Z}^m$ generated by these two elements will generate orbits that are contained in line, but it can't act discretely on these lines (see also the end of the proof of theorem 6.2.6.) We conclude that $\pi_1(M) \cap \mathbb{R}^k = \mathbb{Z}^k$. (For more details about discrete actions on $\mathbb{R}^n$ see also [38] and [106].) □

Remark 7.3.12. Wilking in [103] has in fact shown that any group G that admits a finite normal subgroup $\mathrm{H} \subset \mathrm{G}$ so that G/H acts discretely and cocompactly on a Euclidean space must be the fundamental group of a compact manifold with nonnegative sectional curvature.

We next prove some further results about the structure of compact manifolds with nonnegative Ricci curvature.

Corollary 7.3.13. *Suppose (M, g) is a compact Riemannian manifold with* $\mathrm{Ric} \geq 0$. *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 $\tilde{M} = \mathbb{R}^k \times C$, 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. □

Corollary 7.3.14. *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 $\mathbb{R}^k \times C$, where $k \geq 1$. Thus, the universal covering is compact. □

This result is a bit stronger than simply showing that $H^1(M, \mathbb{R}) = 0$ as we shall prove using the Bochner technique (see 9.2.3). The next result is equivalent to Bochner's theorem, but the proof is quite a bit different.

Corollary 7.3.15. *If (M, g) is compact and has* Ric ≥ 0*, then $b_1(M) \leq \dim M = n$, with equality holding if and only if (M, g) is a flat torus.*

Proof. There is a natural surjection

$$h : \pi_1(M) \to H_1(M, \mathbb{Z}) \simeq \mathbb{Z}^{b_1} \times \mathrm{T},$$

that maps loops to cycles, and where T is a finite Abelian group. The structure of the fundamental group shows that $h(\mathrm{G}) \subset \mathrm{T}$ since G is finite. Thus we obtain a surjective homomorphism $\pi_1(M)/\mathrm{G} \to \mathbb{Z}^{b_1}$, where $\pi_1(M)/\mathrm{G} = \pi_1(M) \cap \mathrm{Iso}\left(\mathbb{R}^k\right)$. Moreover, the image of $\pi_1(M) \cap \mathbb{R}^k = \mathbb{Z}^k$ in $\mathbb{Z}^{b_1}$ has finite index. This shows that $b_1 \leq k \leq n$.

When $b_1 = n$ it follows that $\tilde{M} = \mathbb{R}^n$. In particular, G is trivial. Moreover, the restriction of h to $\mathbb{Z}^n$ must be injective as the image otherwise couldn't have finite index in $H_1(M, \mathbb{Z})$. Thus the kernel of h cannot intersect the finite index subgroup $\mathbb{Z}^n \subset \pi_1(M)$ and so must be finite. However, any isometry on $\mathbb{R}^n$ of finite order has a fixed point so it follows that $\ker h$ is trivial. Thus $\pi_1(M) \simeq \mathbb{Z}^n \times \mathrm{T}$ and consequently T is trivial. This shows that $M = \mathbb{R}^n/\mathbb{Z}^n$ is a torus. Note, however, that the action of $\mathbb{Z}^n$ on $\mathbb{R}^n$ might not be the standard action so we don't necessarily end up with the square torus. □

Finally we prove a similar structure result for homogeneous spaces.

Theorem 7.3.16. *Let (M, g) be a Riemannian manifold that is homogeneous. If* Ric ≥ 0*, then*

$$(M, g) = \left(N \times \mathbb{R}^k, g_N + g_{\mathbb{R}^k}\right),$$

where (N, g_N) is a compact homogeneous space.

Proof. First split $(M, g) = \left(N \times \mathbb{R}^k, g_N + g_{\mathbb{R}^k}\right)$ so that N does not contain any lines. Then note that the isometry group splits as in theorem 7.3.11 thus forcing N to become homogeneous.

The claim will then follow from the splitting theorem provided we can show that any noncompact homogeneous space contains a line. To see this choose a unit speed ray $c : [0, \infty) \to M$ and isometries F_s such that $F_s(c(s)) = c(0)$. Now consider the unit speed rays $c_s : [-s, \infty)$ defined by $c_s(t) = F_s(c(t+s))$. Then $c_s(0) = c(0)$ and $\dot{c}_s(0) = \dot{c}(0)$ so c_s is simply the extension of c. As c_s is a ray it follows that the extension of c to $\mathbb{R}$ must be a line. □

7.4 Further Study

The adventurous reader could consult [53] for further discussions. Anderson's article [2] contains some interesting examples of manifolds with nonnegative Ricci curvature. For the examples with almost maximal diameter we refer the reader to [3] and [81]. It is also worthwhile to consult the original paper on the splitting theorem [31] and the elementary proof of it in [41]. The reader should also consult the articles by Colding, Perel'man, and Zhu in [54] to get an idea of how the subject has developed.

7.5 Exercises

EXERCISE 7.5.1. With notation as in section 7.1.1 and using $\mathrm{vol} = \lambda dr \wedge \mathrm{vol}_{n-1}$ show that $\mu = \lambda^{\frac{1}{n-1}}$ satisfies

$$\begin{aligned} \partial_r^2 \mu &\leq -\frac{\mu}{n-1} \operatorname{Ric}(\partial_r, \partial_r), \\ \mu(0, \theta) &= 0, \\ \lim_{r \to 0} \partial_r \mu(r, \theta) &= 1. \end{aligned}$$

This can also be used to show the desired estimates for the volume form.

EXERCISE 7.5.2 (Calabi and Yau). Let (M, g) be a complete noncompact manifold with $\operatorname{Ric} \geq 0$ and fix $p \in M$.

(1) Show that for each $R > 1$ there is an $x \in M$ such that

$$\begin{aligned} \operatorname{vol} B(p, 1) &\leq \operatorname{vol} B(x, R+1) - \operatorname{vol} B(x, R-1) \\ &\leq \frac{(R+1)^n - (R-1)^n}{(R+1)^n} \operatorname{vol} B(p, 2R). \end{aligned}$$

(2) Show that there is a constant $C > 0$ so that $\operatorname{vol} B(p, R) \geq CR$.

EXERCISE 7.5.3. Let $f : I \to \mathbb{R}$ be continuous, where $I \subset \mathbb{R}$ an interval. Show that the following conditions are equivalent.

(1) f is convex.
(2) f has a "linear" support function from below of the form $a(x - x_0) + f(x_0)$ at every $x_0 \in I$.
(3) $f'' \geq 0$ in the support sense at all points $x_0 \in I$.

EXERCISE 7.5.4. Show that on a compact Riemannian manifold it is not possible to find $S(s) < \infty$ such that $\|f - f_M\|_{\frac{s}{s-1}} \leq S\|df\|_1$ when $1 < s < \dim M$.

EXERCISE 7.5.5 (Basic Covering Lemma). Given a separable metric space (X, d) and a bounded positive function $R : X \to (0, D]$, show that there is a countable subset $A \subset X$, such that the balls $B(p, R(p))$ are pairwise disjoint for $p \in A$ and $X = \bigcup_{p \in A} B(p, 5R(p))$. Hint: Select the points in A successively so that $R(p_{k+1}) \geq \frac{1}{2}\sup_{p \in X - \bigcup_{i=1}^{k} B(p_i, 2R(p_i))} R(p)$.

EXERCISE 7.5.6. Assume the distance function $r(x) = |xp|$ is smooth on $B(p, R)$. Show that if

$$\operatorname{Hess} r = \frac{\operatorname{sn}_k'(r)}{\operatorname{sn}_k(r)} g_r$$

in polar coordinates, then all sectional curvatures on $B(p, R)$ are equal to k.

EXERCISE 7.5.7. Construct convex surfaces in $\mathbb{R}^3$ by capping off cylinders $[-R, R] \times S^1$ to show that the Sobolev-Poincaré constants increase as R increases. Hint: Consider test functions that are constant except on $[-1, 1] \times S^1$.

EXERCISE 7.5.8. Show that if (M, g) has $\operatorname{Ric} \geq (n-1)k$ and for some $p \in M$ we have $\operatorname{vol} B(p, R) = v(n, k, R)$, then the metric has constant curvature k on $B(p, R)$.

EXERCISE 7.5.9. Let X be a vector field on a Riemannian manifold and consider $F_t(p) = \exp_p\left(tX|_p\right)$.

(1) For $v \in T_pM$ show that $J(t) = DF_t(v)$ is a Jacobi field along $t \mapsto c(t) = \exp(tX)$ with the initial conditions $J(0) = v$, $\dot{J}(0) = \nabla_v X$.
(2) Select an orthonormal basis e_i for T_pM and let $J_i(t) = DF_t(e_i)$. Show that

$$(\det[DF_t])^2 = \det\left[g\left(J_i(t), J_j(t)\right)\right].$$

(3) Show that as long as $\det(DF_t) \neq 0$ it satisfies

$$\frac{d^2(\det(DF_t))^{\frac{1}{n}}}{dt^2} \leq -\frac{(\det(DF_t))^{\frac{1}{n}}}{n}\operatorname{Ric}(\dot{c}, \dot{c}).$$

Hint: Use that any $n \times n$ matrix satisfies $(\operatorname{tr}(A))^2 \leq n \operatorname{tr}(A^*A)$.

EXERCISE 7.5.10. 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.

EXERCISE 7.5.11. Show that any function on an n-dimensional Riemannian manifold satisfies

$$|\mathrm{Hess}\, u|^2 \geq \frac{1}{n} |\Delta u|^2$$

with equality holding only when $\mathrm{Hess}\, u = \frac{\Delta u}{n} g$. What can you say about M when $\mathrm{Hess}\, u = \frac{\Delta u}{n} g$?

EXERCISE 7.5.12. Show that if $u, v : M \to \mathbb{R}$ are compactly supported functions that are both smooth on open dense sets in M, then the following integrals make sense and are equal

$$\int u\Delta v \,\mathrm{vol} = \int v\Delta u \,\mathrm{vol} = -\int g(du,dv)\,\mathrm{vol} = -\int g(\nabla u, \nabla v)\,\mathrm{vol}\,.$$

EXERCISE 7.5.13. Show that if $\Delta u = \lambda u$ on a closed Riemannian manifold, then $\lambda \leq 0$ and when $\lambda = 0$, then u is constant.

EXERCISE 7.5.14. Show that the modified distance functions $u_k = \cos\left(\sqrt{k} r\right)$ on $S^n_k = S^n\left(\frac{1}{\sqrt{k}}\right)$, satisfy $\Delta u_k = -(nk)\, u_k$ and $\int u_k \,\mathrm{vol} = 0$.

EXERCISE 7.5.15 (Lichnerowicz). Let (M^n, g) be closed with $\mathrm{Ric} \geq (n-1)\,k > 0$. Use the Bochner formula to show that all functions with $\Delta u = -\lambda u$, $\lambda > 0$, satisfy $\lambda \geq nk$.

The spectral theorem for Δ then implies that all functions with $\int u \,\mathrm{vol} = 0$ satisfy the Poincaré inequality

$$\int u^2 \,\mathrm{vol} \leq \frac{1}{nk} \int |du|^2 \,\mathrm{vol}\,.$$

EXERCISE 7.5.16 (Obata). Let (M^n, g) be closed with $\mathrm{Ric} \geq (n-1)\,k > 0$. Use the Bochner formula as in exercise 7.5.15 to show that if there is a function such that $\Delta u = -(nk)\,u$, then $\mathrm{Hess}\, u = -kug$. Conclude that $(M^n, g) = S^n_k$.

EXERCISE 7.5.17 (P. Li and Schoen). The goal of this exercise is to show a Poincaré inequality for functions that vanish on the boundary of a ball. Let (M,g) be a complete Riemannian n-manifold with $\mathrm{Ric} \geq -(n-1)\,k^2$, $k \geq 0$; $p \in M$; $R > 0$ chosen so that $\partial B(p, 2R) \neq \varnothing$; $q \in \partial B(p, 2R)$; and $r(x) = |xq|$.

(1) Show that $\Delta r \leq (n-1)\left(R^{-1}+k\right)$ on $B(p,R)$.
(2) Let $f(x) = a\exp(-ar(x))$, $a > 0$. Show that

$$\Delta f \geq a\exp(-a3R)\left(a-(n-1)\left(R^{-1}+k\right)\right).$$

(3) Let $u \geq 0$ be a smooth function with compact support in $B(p,R)$ and choose $a = n\left(R^{-1}+k\right)$. Use

$$\int_{B(p,R)} u\Delta f \,\mathrm{vol} = -\int_{B(p,R)} g(du, df)\,\mathrm{vol}$$

to show that

$$\int_{B(p,R)} u \,\mathrm{vol} \leq C\int_{B(p,R)} |du|\,\mathrm{vol},$$

where $C = \frac{R}{1+kR}\exp(2n(1+kR))$.
(4) Prove this inequality for all smooth functions u with compact support in $B(p,R)$.
(5) Let $s \geq 1$ and u have compact support in $B(p,R)$. Show that

$$\int_{B(p,R)} |u|^s \,\mathrm{vol} \leq (sC)^s \int_{B(p,R)} |du|\,\mathrm{vol}.$$

EXERCISE 7.5.18 (Cheeger). The relative volume comparison estimate can be generalized as follows: Suppose (M^n, g) has $\mathrm{Ric} \geq (n-1)k$.

(1) Select points $p_1, \ldots, p_k \in M$. Then the function

$$r \mapsto \frac{\mathrm{vol}\left(\bigcup_{i=1}^k B(p_i, r)\right)}{v(n,k,r)}$$

is nonincreasing and converges to k as $r \to 0$.
(2) If $A \subset M$, then

$$r \mapsto \frac{\mathrm{vol}\left(\bigcup_{p\in A} B(p, r)\right)}{v(n,k,r)}$$

is nonincreasing. To prove this, use the above with the finite collection of points taken to be very dense in A.

EXERCISE 7.5.19. The absolute volume comparison can be generalized to hold for cones. Namely, for $p \in M$ and a subset $\Gamma \subset T_pM$ of unit vectors, consider the cones defined in polar coordinates:

$$B^{\Gamma}(p, R) = \{(t, \theta) \in M \mid t \leq R \text{ and } \theta \in \Gamma\}.$$

If $\operatorname{Ric} M \geq (n-1)k$, show that

$$\operatorname{vol} B^{\Gamma}(p, R) \leq \operatorname{vol} \Gamma \cdot \int_0^R (\operatorname{sn}_k(t))^{n-1} dt.$$

EXERCISE 7.5.20. Let G be a compact connected Lie group with a biinvariant metric such that $\operatorname{Ric} \geq 0$. Use the results from this chapter to prove

(1) If G has finite center, then G has finite fundamental group.
(2) A finite covering of G looks like $G' \times T^k$, where G' is compact simply connected, and T^k is a torus.
(3) If G has finite fundamental group, then the center is finite.

EXERCISE 7.5.21. 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 $M = \mathbb{R}^n$.

EXERCISE 7.5.22. Show that if $\operatorname{Ric} \geq n-1$, then $\operatorname{diam} \leq \pi$, by showing that if $|pq| > \pi$, then

$$e_{p,q}(x) = |px| + |xq| - |pq|$$

has negative Laplacian at a local minimum.

Chapter 8
Killing Fields

In this chapter we begin with a section on some general results about Killing fields and their relationship to the isometry group. This is used in the subsequent section to prove Bochner's theorems about the lack of Killing fields on manifolds with negative Ricci curvature. In the last section we present several results about how Killing fields influence the topology of manifolds with positive sectional curvature. This is a somewhat more recent line of inquiry.

8.1 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 translates into the following simple characterization:

Proposition 8.1.1. *A vector field X on a Riemannian manifold (M, g) is a Killing field if and only if $L_X g = 0$.*

Proof. Let F^t be the local flow for X. Recall that

$$(L_X g)(v, w) = \frac{d}{dt} g\left(DF^t(v), DF^t(w)\right)|_{t=0}.$$

P. Petersen, *Riemannian Geometry*, Graduate Texts in Mathematics 171,
DOI 10.1007/978-3-319-26654-1_8

Thus we have

$$\begin{aligned}\frac{d}{dt}g\left(DF^{t}(v),DF^{t}(w)\right)|_{t=t_0} &= \frac{d}{dt}g\left(DF^{t-t_0}DF^{t_0}(v),DF^{t-t_0}DF^{t_0}(w)\right)|_{t=t_0}\\ &= \frac{d}{ds}g\left(DF^{s}DF^{t_0}(v),DF^{s}DF^{t_0}(w)\right)|_{s=0}\\ &= (L_Xg)\left(DF^{t_0}(v),DF^{t_0}(w)\right).\end{aligned}$$

This shows that $L_Xg = 0$ if and only if $t \mapsto g\left(DF^{t}(v),DF^{t}(w)\right)$ is constant. Since F^0 is the identity map this is equivalent to assuming the flow acts by isometries. □

We can use this characterization to show

Proposition 8.1.2. *X is a Killing field if and only if $v \mapsto \nabla_v X$ is a skew symmetric $(1,1)$-tensor.*

Proof. Let $\theta_X(v) = g(X,v)$ be the 1-form dual to X. Recall that

$$d\theta_X(V,W) + (L_Xg)(V,W) = 2g(\nabla_V X, W).$$

Thus $L_Xg \equiv 0$ if and only if $v \mapsto \nabla_v X$ is skew-symmetric. □

Proposition 8.1.3. *If $X \in \mathfrak{iso}$, i.e., X is a Killing field, then*

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

If $X, Y \in \mathfrak{iso}$, then

$$[\nabla X, \nabla Y](V) + \nabla_V[X,Y] = R(X,Y)V.$$

Proof. The fact that X is a Killing field implies that $L_X\nabla = 0$. Using this with the identity

$$(L_X\nabla)_V W = R(X,V)W + \nabla^2_{V,W}X$$

from the proof of the first Bianchi identity in proposition 3.1.1 implies the first claim.

The second identity is a direct calculation that uses the first Bianchi identity

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

□

Proposition 8.1.4. *For a given $p \in M$ a Killing field $X \in \mathfrak{iso}$ is uniquely determined by $X|_p$ and $(\nabla X)|_p$. In particular, we obtain a short exact sequence*

$$0 \to \mathfrak{iso}_p \to \mathfrak{iso} \to \mathfrak{t}_p \to 0,$$

where

$$\begin{aligned} \mathfrak{iso}_p &= \left\{ X \in \mathfrak{iso} \mid X|_p = 0 \right\}, \\ \mathfrak{t}_p &= \left\{ X|_p \in T_pM \mid X \in \mathfrak{iso} \right\}. \end{aligned}$$

Proof. The equation $L_X g \equiv 0$ is linear in X, so the space of Killing fields is a vector space. Therefore, it suffices to show that $X \equiv 0$ on M provided $X|_p = 0$ and $(\nabla X)|_p = 0$. Using an open-closed argument we can reduce our considerations to a neighborhood of p.

Let F^t be the local flow for X near p. The condition $X|_p = 0$ implies that $F^t(p) = p$ for all t. Thus $DF^t : T_pM \to T_pM$. We claim that also $DF^t = I$. The assumptions show that X commutes with any vector field at p since

$$[X, Y]|_p = \nabla_{X(p)} Y - \nabla_{Y(p)} X = 0.$$

If $Y|_p = v$, then the definition of the Lie derivative implies

$$0 = L_X Y|_p = \lim_{t \to 0} \frac{DF^t(v) - v}{t}.$$

Applying this to the vector field $F_*^{t_0}(Y)$ yields

$$\begin{aligned} 0 &= L_X DF^{t_0}(Y)|_p \\ &= \lim_{s \to 0} \frac{DF^s DF^{t_0}(v) - DF^{t_0}(v)}{s} \\ &= \lim_{t - t_0 \to 0} \frac{DF^{t-t_0} DF^{t_0}(v) - DF^{t_0}(v)}{t - t_0} \\ &= \lim_{t \to t_0} \frac{DF^t(v) - DF^{t_0}(v)}{t - t_0}. \end{aligned}$$

In other words $t \mapsto DF^t(v)$ is constant. As $DF^0(v) = v$ it follows that $DF^t = I$.

Since the flow diffeomorphisms act by isometries, proposition 5.6.2 implies that they must be the identity map, and hence $X = 0$ in a neighborhood of p.

Alternatively we could also have used that X when restricted to a geodesic c must be a Jacobi field as the flow of X generates a geodesic variation. Thus $X = 0$ along this geodesic if $X|_{c(0)}$ and $\nabla_{\dot{c}(0)} X$ both vanish.

The second part of the claim is immediate from the fact that $\mathfrak{t}_p$ is defined as the image and $\mathfrak{iso}_p$ as the kernel of the evaluation map $\mathfrak{iso} \to T_pM$. □

These properties lead to two important general results about Killing fields.

Theorem 8.1.5. *The zero set of a Killing field is a disjoint union of totally geodesic submanifolds each of even codimension.*

Proof. The flow generated by a Killing field X on (M, g) acts by isometries so we know from proposition 5.6.5 that the fixed point set of these isometries is a union of totally geodesic submanifolds. We next observe that the fixed point set of all of these isometries is precisely the set of points where the Killing field vanishes. Finally assume that $X|_p = 0$ and let $V = \ker(\nabla X)_p$. Then V is the Zariski tangent space to the zero set and hence also the tangent space as in the proof of proposition 5.6.5. Thus $w \mapsto \nabla_w X$ is an isomorphism on $V^\perp$. As it is also a skew-symmetric map it follows that $V^\perp$ is even-dimensional. □

Theorem 8.1.6. *The set of Killing fields* $\mathfrak{iso}(M, g)$ *is a Lie algebra of dimension* $\leq \binom{n+1}{2}$. *Furthermore, if M is complete, 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 $L_X g = L_Y g = 0$, then $L_{[X,Y]} g = 0$. Thus, $\mathfrak{iso}(M, g)$ forms a Lie algebra. From proposition 8.1.4 it follows that the map $X \mapsto (X|_p, (\nabla X)|_p)$ is linear with trivial kernel. Thus

$$\begin{aligned}\dim(\mathfrak{iso}(M, g)) &\leq \dim T_p M + \dim \mathfrak{so}\left(T_p M\right) \\ &= n + \frac{n(n-1)}{2} = \binom{n+1}{2}.\end{aligned}$$

The last statement depends crucially on knowing that $\mathrm{Iso}(M, g)$ is a Lie group in the first place. We endow $\mathrm{Iso}(M, g)$ with the compact-open topology so that convergence is equivalent to uniform convergence on compact sets (see exercise 5.9.41). We saw in theorem 5.6.19 that this makes $\mathrm{Iso}(M, g)$ into a Lie group. One can also appeal to the profound theorem of Bochner and Montgomery that any group of diffeomorphisms that is also locally compact with respect to the compact-open topology is a Lie group in that topology (see exercise 6.7.26 and [79]).

Since M is complete the Killing fields have flows that are defined for all time (see exercise 8.4.5). These flows consist of isometries and thus yield differentiable one-parameter subgroups of $\mathrm{Iso}(M, g)$. Conversely each differentiable one-parameter subgroup of $\mathrm{Iso}(M, g)$ also gives a Killing field. This correspondence between one-parameter subgroups and Killing fields shows that the Lie algebra $\mathfrak{iso}(M, g)$ is the Lie algebra of $\mathrm{Iso}(M, g)$.

There is an alternate proof of this theorem in [83]. However, it requires another very subtle result about Lie algebras called Ado's theorem: Every finite dimensional Lie algebra is the Lie algebra of a Lie group. Using this and the fact that the flows of Killing fields are defined for all time and consist of isometries shows that there is a connected subgroup $\mathrm{Iso}_0(M, g) \subset \mathrm{Iso}(M, g)$, which is a Lie group with Lie algebra $\mathfrak{iso}(M, g)$. Given $X \in \mathfrak{iso}(M, g)$ with flow F^t and $F \in \mathrm{Iso}(M, g)$ observe that the flow of $F_* X$ is $F \circ F^t \circ F^{-1}$. Thus conjugation by elements $F \in \mathrm{Iso}(M, g)$ defines an automorphism on $\mathrm{Iso}_0(M, g)$ whose differential at the identify is given by F_*.

This shows that $\mathrm{Iso}_0(M,g)$ is a normal subgroup of $\mathrm{Iso}(M,g)$. We can then define the topology on $\mathrm{Iso}(M,g)$ so that $\mathrm{Iso}_0(M,g)$ becomes the connected component of $\mathrm{Iso}(M,g)$ containing the identity. This will make $\mathrm{Iso}(M,g)$ into a Lie group whose component containing the identity is $\mathrm{Iso}_0(M,g)$. It is a general fact from the theory of Lie groups that the differentiable Lie group structure is unique if we know the group structure and the smooth 1-parameter subgroups. This means that the topology just introduced is forced to be the same as the compact-open topology. Note that it is not otherwise immediately clear from this construction that $\mathrm{Iso}(M,g)$ has a countable number of connected components. □

Recall that $\dim(\mathrm{Iso}(S^n_k)) = \binom{n+1}{2}$. Thus, all simply connected 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^n_k. The Killing fields on the quotient S^n_k/Γ can be identified with the Killing fields on S^n_k that are invariant under Γ. The corresponding connected subgroup $\mathrm{G} \subset \mathrm{Iso}(S^n_k)$ will then commute with all elements in Γ. So if $\dim(\mathrm{Iso}(S^n_k/\Gamma))$ is maximal, then $\dim \mathrm{G} = \binom{n+1}{2}$ and $\mathrm{G} = \mathrm{Iso}(S^n_k)_0$. As we know the possibilities for $\mathrm{Iso}(S^n_k)_0$ (see section 1.3.1) it is not hard to check that this forces Γ to consist of homotheties. Thus, Γ can essentially only be $\{\pm I\}$ if it is nontrivial. But $-I$ acts freely only on the sphere. Thus, only one other constant curvature space form has maximal dimension for the isometry group, namely $\mathbb{RP}^n$.

More generally, when $\dim(\mathfrak{iso}(M,g)) = \binom{n+1}{2}$, then (M,g) has constant curvature. To prove this we use that for each $p \in M$ the map $X \mapsto (X|_p, (\nabla X)|_p)$ is surjective. First note that for each $S \in \mathfrak{so}(T_pM)$ there is a Killing field X with $X|_p = 0$ and $(\nabla X)|_p = S$. The flow fixes p, i.e., $F^t_X(p) = p$, and thus $DF^t_X|_p$ defines a local one-parameter group of orthogonal transformations with $\frac{d}{dt}|_{t=0} DF^t_X|_p = S$. This implies that $DF^t_X|_p = \exp(tS)$, where exp is the usual operator or matrix exponential map. Since $\exp : \mathfrak{so}(T_pM) \to \mathrm{SO}(T_pM)$ is a local diffeomorphism near the identity it follows that any two planes in T_pM that are sufficiently close to each other can be mapped to each other by an isometry F^t_X. This shows that these planes have the same sectional curvature. We can now use an open-closed argument to show that all sectional curvatures at p are the same.

Finally, for each $v \in T_pM$ there is a unique Killing field X such that $X|_p = v$ and $(\nabla X)|_p = 0$. If F^t_X is the (local) flow of X, then we obtain an "exponential" map $E_p : O_p \subset T_pM \to M$ by $E_p(v) = F^1_X(p)$, where O_p is a neighborhood of the origin. Note that $E_p(tv) = F^t_X(p)$ so it follows that $\frac{d}{dt}|_{t=0} E_p(tv) = v$. In particular, the differential is the identity map at the origin and E_p is locally a diffeomorphism. This implies that for every point q in a neighborhood of p, there is a local isometry that maps p to q. This means that the curvatures are constant on a neighborhood of p. An open-closed argument shows as before that the curvature is constant on M.

8.2 Killing Fields in Negative Ricci Curvature

We start by proving a general result that will be used throughout the chapter.

Proposition 8.2.1. *Let X be a Killing field on (M,g) and consider the function $f = \frac{1}{2}g(X,X) = \frac{1}{2}|X|^2$. Then*

(1) $\nabla f = -\nabla_X X$.
(2) $\mathrm{Hess} f\,(V,V) = |\nabla_V X|^2 - R\,(V,X,X,V)$.
(3) $\Delta f = |\nabla X|^2 - \mathrm{Ric}\,(X,X)$.

Proof. To see (1) observe that

$$\begin{aligned} g\,(V,\nabla f) &= D_V f \\ &= g(\nabla_V X, X) \\ &= -g(V,\nabla_X X). \end{aligned}$$

(2) is proven directly by repeatedly using that $V \mapsto \nabla_V X$ is skew-symmetric:

$$\begin{aligned} \mathrm{Hess} f\,(V,V) &= g\,(\nabla_V\,(-\nabla_X X)\,,V) \\ &= -\,g\,(R\,(V,X)\,X,V) - g\,(\nabla_X\nabla_V X,V) - g\left(\nabla_{[V,X]}X,V\right) \\ &= -\,R\,(V,X,X,V) - g\,(\nabla_X\nabla_V X,V) \\ &\quad +\,g\,(\nabla_{\nabla_X V}X,V) - g\,(\nabla_{\nabla_V X}X,V) \\ &= -\,R_X\,(V) + g\,(\nabla_V X,\nabla_V X) - g\,(\nabla_X\nabla_V X,V) - g\,(\nabla_V X,\nabla_X V) \\ &= -\,R_X\,(V) + g\,(\nabla_V X,\nabla_V X) - D_X g\,(\nabla_V X,V) \\ &= -\,R_X\,(V) + g\,(\nabla_V X,\nabla_V X)\,. \end{aligned}$$

For (3) we select an orthonormal frame E_i and see that

$$\begin{aligned} \Delta f &= \sum_{i=1}^{n} \mathrm{Hess} f\,(E_i,E_i) \\ &= \sum_{i=1}^{n} g\,(\nabla_{E_i}X,\nabla_{E_i}X) - \sum_{i=1}^{n} R\,(E_i,X,X,E_i) \\ &= \sum_{i=1}^{n} g\,(\nabla_{E_i}X,\nabla_{E_i}X) - \mathrm{Ric}\,(X,X) \\ &= |\nabla X|^2 - \mathrm{Ric}\,(X,X)\,. \end{aligned}$$

□

Formula (3) in this proposition is called a Bochner formula. We shall meet many more of these types of formulas in the next chapter.

Theorem 8.2.2 (Bochner, 1946). *Suppose* (M, g) *is compact and has* $\mathrm{Ric} \leq 0$. *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 the condition $\mathrm{Ric} \leq 0$ gives us

$$\Delta f = |\nabla X|^2 - \mathrm{Ric}\,(X, X) \geq 0.$$

The maximum principle then shows that f is constant and that $|\nabla X| \equiv 0$, i.e., X is parallel. In addition $\mathrm{Ric}\,(X, X) \equiv 0$. When $\mathrm{Ric} < 0$ this implies that $X \equiv 0$. □

Corollary 8.2.3. *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. As any Killing field is parallel, the linear map: $X \mapsto X|_p$ from $\mathfrak{iso}(M, g)$ to T_pM is injective. This gives the result. For the second part use that $\mathrm{Iso}(M, g)$ is compact, since M is compact, and that the identity component is trivial. □

Corollary 8.2.4. *With* (M, g) *as before and* $k = \dim(\mathfrak{iso}(M, g))$, *we have that the universal covering splits isometrically as* $\tilde{M} = \mathbb{R}^k \times N$.

Proof. On $\tilde{M}$ there are k linearly independent parallel vector fields, which we can assume to be orthonormal. Since $\tilde{M}$ is simply connected, each of these vector fields is the gradient field for a distance function. If we consider just one of these distance functions we see that the metric splits as $g = dr^2 + g_r = dr^2 + g_0$ since the Hessian of this distance function vanishes. As we get such a splitting for k distance functions with orthonormal gradients we get the desired splitting of $\tilde{M}$. □

We can now say more about the homogeneous situation discussed in theorem 7.3.16.

Corollary 8.2.5. *A compact homogeneous space with* $\mathrm{Ric} \leq 0$ *is flat. In particular, any Ricci flat homogeneous space is flat.*

Proof. We know that every Killing field is parallel and the assumption that the space is homogeneous tells us that every tangent vector is part of a Killing field (see also exercise 8.4.9). Thus the curvature vanishes.

The second part of the result comes from applying theorem 7.3.16. □

The result about nonexistence of Killing fields can actually be slightly improved to yield

Theorem 8.2.6. *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

$$0 = \Delta f(p) = -\mathrm{Ric}(X, X)(p) > 0. \qquad \square$$

Bochner's theorem has been generalized by X. Rong to a more general statement asserting that a closed Riemannian manifold with negative Ricci curvature can't admit a pure F-structure of positive rank (see [93] for the definition of F structure and proof of this). An F-structure on M is essentially a finite covering of open sets U_i on some finite covering space $\hat{M} \to M$, such that we have a Killing field X_i on each U_i. Furthermore, these Killing fields must commute whenever they are defined at the same point, i.e., $\left[X_i, X_j\right] = 0$ on $U_i \cap U_j$. The idea of the proof is to consider the function

$$f = \det\left[g\left(X_i, X_j\right)\right].$$

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 is a reasonably nice function that has a Bochner formula.

8.3 Killing Fields in Positive Curvature

It is also possible to say quite a bit about Killing fields in positive sectional curvature. This is a much more recent development in Riemannian geometry.

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 compact manifold with positive sectional curvature has positive Euler characteristic. If the curvature operator is positive, then this is certainly true as it follows from theorem 9.4.6 that the Euler characteristic is 2. From corollary 6.3.2 we know that the fundamental group is finite provided the Ricci curvature is positive. In particular $H_1(M, \mathbb{R}) = 0$. This shows that the conjecture holds in dimension 2. In dim $= 4$, Poincaré duality implies that $H_1(M, \mathbb{R}) = H_3(M, \mathbb{R}) = 0$. Hence

$$\chi(M) = 1 + \dim H_2(M, \mathbb{R}) + 1 \geq 2.$$

In higher dimensions we have the following partial justification for the Hopf conjecture.

Theorem 8.3.1 (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, then f will have a positive minimum at some point $p \in M$. In particular, $\text{Hess}f|_p \geq 0$. We also know from proposition 8.2.1 that

$$\text{Hess}f\,(V, V) = |\nabla_V X|^2 - R\,(V, X, X, V)\,.$$

By assumption, $g(R(V, X)X, V) > 0$ if X and V are linearly independent. Using this, we seek V such that $\text{Hess}f\,(V, V) < 0$ near p, thus arriving at a contradiction.

Recall that the linear endomorphism $v \mapsto \nabla_v X$ is skew-symmetric. Furthermore, $(\nabla_X X)|_p = 0$, since $\nabla f|_p = -(\nabla_X X)|_p$, and f has a minimum at p. Thus $(\nabla X)|_p : T_pM \to T_pM$ has at least one zero eigenvalue. However, the rank of a skew-symmetric map is always even, so the kernel must also have even dimension as T_pM is even dimensional. If $v \in T_pM$ is an element in the kernel linearly independent from X, then

$$\begin{aligned}\text{Hess}f\,(v, v) &= |\nabla_V X|^2 - R\,(v, X, X, v)\\ &= -R\,(v, X, X, v) < 0.\end{aligned}$$

□

In odd dimensions this result is not true as the unit vector field that generates the Hopf fibration $S^3(1) \to S^2(1/2)$ is a Killing field.

Having an isometric torus action implies that $\mathfrak{iso}\,(M, g)$ contains a certain number of linearly independent commuting Killing fields. By Berger's result we know that in even dimensions these Killing fields must vanish somewhere. Moreover, the structure of these zero sets is so that each component is a totally geodesic submanifold of even codimension. A type of induction on dimension can be now used to extract information about these manifolds.

To understand how this works some important topological results on the zero set for a Killing field are needed. The Euler characteristic is defined as the alternating sum

$$\chi\,(M) = \sum_{p=0}^{n} (-1)^p \dim H_p\,(M, \mathbb{R})\,.$$

Theorem 8.3.2. *Let X be a Killing field on a compact Riemannian manifold. If $N_i \subset M$ are the components of the zero set for X, then $\chi\,(M) = \sum_i \chi\,(N_i)$.*

Proof. The proof is a modification of the classical result of Poincaré and Hopf where the Euler characteristic is calculated as a sum of indices for the isolated zeros of a vector field. The Meyer-Vietoris sequence can be used show that

$$\chi\,(M) = \chi\,(A) + \chi\,(B) - \chi\,(A \cap B)$$

for nice subsets A, B, $A \cap B \subset M$.

Note that the flow F^t of X fixes points in N_i and in particular, $|F^t(x)N_i| = |xN_i|$. This shows that X is tangent to the level sets of $|xN_i|$. Now choose tubular neighborhoods around each N_i of the form $T_i = \{p \in M \mid |xN_i| \leq \epsilon\}$. Then X is tangent to the smooth boundary ∂T_i. Now both $\chi\left(M - \bigcup \mathrm{int} T_i\right)$ and $\chi\left(\bigcup \partial T_i\right)$ vanish by the Poincaré-Hopf theorem as X is a nonzero vector field on these manifolds. Thus $\chi(M) = \sum \chi(T_i)$. Finally, $N_i \subset T_i$ is a deformation retraction and so they have the same Euler characteristic. This proves the theorem. □

This implies the following corollary.

Corollary 8.3.3. *If M is a compact 6-manifold with positive sectional curvature that admits a Killing field, then $\chi(M) > 0$.*

Proof. We know that the zero set for a Killing field is nonempty and that each component has even codimension. Thus each component is a 0, 2, or 4-dimensional manifold with positive sectional curvature. This shows that M has positive Euler characteristic. □

If we consider 4-manifolds we get a much stronger result (see also [63]). The proof uses techniques that appear later in the text and is only given in outline.

Theorem 8.3.4 (Hsiang and Kleiner, 1989). *If M^4 is a compact orientable positively curved 4-manifold that admits a Killing field, then the Euler characteristic is ≤ 3. In particular, M is topologically equivalent to S^4 or $\mathbb{CP}^2$.*

Proof. We assume for simplicity that $\sec \geq 1$ so that we can use a specific comparison space for Toponogov's theorem (see theorem 12.2.2).

In case the zero set of the Killing field contains a component of dimension 2 the result will follow from lemma 8.3.7 below. Otherwise all zeros are isolated. It is then necessary to obtain a contradiction if there are at least isolated 4 zeros.

Assume p is an isolated zero for a Killing field X on a Riemannian 4-manifold. Then the flow F^t of X induces an isometric action $DF^t|_p = R^t$ on the unit sphere $S^3 \subset T_pM$ that has no fixed points. We can decompose $T_pM = V_1 \oplus V_2$ into the two orthogonal and invariant subspaces for this action. This decomposition allows us to exhibit S^3 as a doubly warped product over the interval $[0, \pi/2]$ as in example 1.4.9. The natural distance function r for this doubly warped decomposition measures the angle to say V_1. It follows that for any $v_j \in S^3, j = 1, 2, 3$ we have

$$|r(v_1)r(v_2)| + |r(v_1)r(v_3)| + |r(v_2)r(v_3)| \leq \pi.$$

The rotations R^t preserve the levels of r and form nontrivial rotations by $\theta_i t$ on each V_i. When θ_1 and θ_2 are irrationally related the orbits are in fact dense in the level sets for r. This shows that for every $\epsilon > 0$ and $v_j \in S^3, j = 1, 2, 3$ there exist t_j such that

$$\angle\left(R^{t_1}v_1, R^{t_2}v_2\right) + \angle\left(R^{t_1}v_1, R^{t_3}v_3\right) + \angle\left(R^{t_2}v_2, R^{t_3}v_3\right) \leq \pi + \epsilon.$$

When θ_1 and θ_2 are rationally related, this will also be true (in fact with $\epsilon = 0$) and can be shown by approximating such a rotation by irrational rotations.

Let $\overrightarrow{\overrightarrow{pq}}$ be the set of all unit vectors tangent to segments from p to q. Define $\inf \angle pxq$ as the infimum of the angles between vectors in $\overrightarrow{\overrightarrow{xp}}$ and $\overrightarrow{\overrightarrow{xq}}$. Assume that p_i, $i = 1, 2, 3, 4$ are isolated zeros for X and note that the flow of X maps segments between any two such zeros to segments between the same two zeros. Thus we have shown that

$$\inf \angle p_2 p_1 p_3 + \inf \angle p_2 p_1 p_4 + \inf \angle p_4 p_1 p_3 \leq \pi.$$

If we add up over all 4 possibilities of points the sum is $\leq 4\pi$.

On the other hand, by Toponogov's theorem any specific angle $\angle pxq$ can be bounded from below by the corresponding angle in $S^2(1)$ for a triangle with the same sides $|xp|, |xq|, |pq|$. This implies that

$$\inf \angle p_1 p_2 p_3 + \inf \angle p_3 p_1 p_2 + \inf \angle p_2 p_3 p_1 > \pi.$$

Adding up over all 4 choices gives a total sum $> 4\pi$. So we have reached a contradiction. □

The conclusion has been improved by Grove-Wilking in [57] and there are similar results for 5-manifolds with torus actions in [42] and [45].

Below we discuss generalizations to higher dimensions. The results, however, do not generalize the Hsiang-Kleiner classification as they require more isometries even in dimension 4.

Two important tools in the proofs below are a generalization of Berger's result about Killing fields in even dimensions and Wilking's connectedness principle (see lemma 6.5.8) as well as an enhancement also due to Wilking.

Theorem 8.3.5 (Grove and Searle, 1994). *Let M be a compact n-manifold with positive sectional curvature. Two commuting Killing fields must be linearly dependent somewhere on M.*

Proof. Let X, Y be commuting Killing fields on M. We have from proposition 8.1.3 that $[\nabla X, \nabla Y] = R(X, Y)$. This gives us the formula

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

If Y vanishes somewhere on M, then we are finished, so assume otherwise and consider the function

$$f = \frac{1}{2}\left(|X|^2 - \frac{g(X, Y)^2}{|Y|^2}\right).$$

If f vanishes somewhere we are again finished. Otherwise, f will have a positive minimum at some point $p \in M$. We scale Y to be a unit vector at p, and adjust X to be $\bar{X} = X - g\left(X|_p, Y|_p\right) Y$ so that $\bar{X}|_p \perp Y|_p$. Neither change will affect f. We now have

$$f = \frac{1}{2}\left(\left|\bar{X}\right|^2 - \frac{g\left(\bar{X}, Y\right)^2}{|Y|^2}\right) \leq \frac{1}{2}\left|\bar{X}\right|^2$$

with equality at p. This means that the function on the right also has a positive minimum at p. In particular, $\nabla_{\bar{X}}\bar{X} = 0$ at p. Since $g\left(\bar{X}, Y\right)$ vanishes at p the Hessian of f at p is simply given by

$$\mathrm{Hess} f\left(v, v\right) = \left|\nabla_v \bar{X}\right|^2 - R\left(v, \bar{X}, \bar{X}, v\right) - \left(D_v g\left(\bar{X}, Y\right)\right)^2$$

for $v \in T_pM$. The last term can be altered to look more like the first

$$\begin{aligned} D_v g\left(\bar{X}, Y\right) &= g\left(\nabla_v \bar{X}, Y\right) + g\left(\bar{X}, \nabla_v Y\right) \\ &= g\left(\nabla_v \bar{X}, Y\right) - g\left(v, \nabla_{\bar{X}} Y\right) \\ &= g\left(\nabla_v \bar{X}, Y\right) - g\left(v, \nabla_Y \bar{X}\right) \\ &= g\left(\nabla_v \bar{X}, 2Y\right). \end{aligned}$$

If there is a $v \perp \bar{X}$ with $\nabla_v \bar{X} = 0$, then the Hessian becomes negative and we have a contradiction. If no such v exists, then $\ker\left(\nabla \bar{X}\right)|_p = \mathrm{span}\left\{\bar{X}|_p\right\}$ since $\nabla_{\bar{X}}\bar{X} = 0$ at p. As $Y \perp \bar{X}$ at p it follows that we can find $v \perp \bar{X}$ such that $\nabla_v \bar{X} = 2Y$. Then we obtain a contradiction again

$$\mathrm{Hess} f\left(v, v\right) = -R\left(v, \bar{X}, \bar{X}, v\right) < 0. \qquad \square$$

Lemma 8.3.6 (Connectedness principle with symmetries, Wilking, 2003). *Let M^n be a compact n-manifold with positive sectional curvature and X a Killing field. If N^{n-k} is a component for the zero set of X, then $N \subset M$ is $(n - 2k + 2)$-connected.*

Proof. Consider a unit speed geodesic c that is perpendicular to N at the endpoints. It is hard to extract more information from parallel fields along c as we did in lemma 6.5.8. Instead we consider fields that are orthogonal to both c and the action and have derivative tangent to the action.

Specifically, consider fields that satisfy the linear ODE

$$\begin{aligned} E\left(0\right) &\in T_{c(0)}N, \\ \dot{E} &= -\frac{g\left(E, \nabla_{\dot{c}} X\right)}{|X|^2} X. \end{aligned}$$

Note that since

$$g(E, \nabla_{\dot{c}} X) = -g(\dot{c}, \nabla_E X)$$

and $E(0) \in T_{c(0)}N$, it follows that $g\left(E(0), \nabla_{\dot{c}(0)} X\right) = 0$. In particular, the differential equation is not singular at $t = 0$.

We claim that these fields satisfy the properties

$$\begin{aligned} g(E, X) &= 0, \\ g\left(E(1), \nabla_{\dot{c}(1)} X\right) &= 0, \\ g(E, \dot{c}) &= 0. \end{aligned}$$

The first condition follows from $X|_{c(0)} = 0$ and

$$\begin{aligned} \frac{d}{dt} g(E, X) &= g\left(\dot{E}, X\right) + g(E, \nabla_{\dot{c}} X) \\ &= g\left(-\frac{g(E, \nabla_{\dot{c}} X)}{|X|^2} X, X\right) + g(E, \nabla_{\dot{c}} X) \\ &= 0. \end{aligned}$$

As $X|_{c(1)} = 0$ this also implies the second property. For the third property first note that

$$\begin{aligned} \frac{d}{dt} g(X, \dot{c}) &= g(\nabla_{\dot{c}} X, \dot{c}) = 0, \\ X|_{c(0)} &= 0 \end{aligned}$$

so $g(X, \dot{c}) = 0$. It then follows that

$$\frac{d}{dt} g(E, \dot{c}) = g\left(\dot{E}, \dot{c}\right) = -\frac{g(E, \nabla_{\dot{c}} X)}{|X|^2} g(X, \dot{c}) = 0.$$

Now note that also $E(1) \perp \operatorname{span}\left\{\dot{c}(1), \nabla_{\dot{c}(1)} X\right\}$. The space $\operatorname{span}\left\{\dot{c}(1), \nabla_{\dot{c}(1)} X\right\}$ is 2-dimensional as $\dot{c}$ is perpendicular to the component N of the zero set for X. This means that the space of such fields E, where in addition $E(1)$ is tangent to N, must have dimension at least $n - 2k + 2$ (see also the proof of part (a) of lemma 6.5.8).

We now need to check that such fields give us negative second variation. This is not immediately obvious as $\left|\dot{E}\right|$ doesn't vanish. However, we can resort to a trick that forces it down in size without losing control of the curvatures $\sec(E, \dot{c})$. In section 4.5.4 we showed that the metric g can be perturbed to a metric g_λ, where X has been squeezed to have size $\to 0$ as $\lambda \to 0$. At the same time, directions orthogonal to X remain unchanged and the curvatures $\sec(E, \dot{c})$ become larger as both E and $\dot{c}$ are perpendicular to X. Finally c remains a geodesic since $\nabla_Y Y$ is unchanged for $Y \perp X$ (see proposition 4.5.1).

The second variation formula for g_λ looks like

$$\begin{aligned}
\frac{d^2E}{ds^2}|_{s=0} &= \int_a^b \left|\dot{E}(t)\right|^2_{g_\lambda} dt - \int_a^b g_\lambda\left(R\left(E,\dot{c}\right)\dot{c},E\right)dt \\
&\leq \int_a^b \left|\frac{g\left(E,\nabla_{\dot{c}}X\right)}{|X|^2_g}X\right|^2_{g_\lambda} dt - \int_a^b \sec_g\left(E,\dot{c}\right)|E|^2_g\,dt \\
&= \int_a^b \frac{\left|g\left(E,\nabla_{\dot{c}}X\right)\right|^2}{|X|^4_g}|X|^2_{g_\lambda}\,dt - \int_a^b \sec_g\left(E,\dot{c}\right)|E|^2_g\,dt \\
&\to -\int_a^b \sec_g\left(E,\dot{c}\right)|E|^2_g\,dt \text{ as } \lambda\to 0.
\end{aligned}$$

This shows that all of the fields E must have negative second variation in the metric g_λ for sufficiently small λ.

This perturbation is independent of $c\in\Omega_{N,N}(M)$ and so we have found a new metric where all such geodesics have index $\geq n-2k+2$. This shows that $N\subset M$ is $(n-2k+2)$-connected. □

To get a feel for how this new connectedness principle can be used we prove.

Lemma 8.3.7 (Grove and Searle, 1994). *Let M be a closed n-manifold with positive sectional curvature. If M admits a Killing field such that the zero set has a component N of codimension 2, then M is diffeomorphic to S^n, $\mathbb{CP}^{\frac{n}{2}}$, or a cyclic quotient of a sphere $S^n/\mathbb{Z}_q$.*

Proof. We only prove a (co-)homology version of this result for simply connected manifolds.

The previous lemma shows that $N\subset M$ is $(n-2)$-connected. Thus, for $k<n-2$, $H_k(N)\to H_k(M)$ and $H^k(M)\to H^k(N)$ are isomorphisms. Using this together with Poincaré duality $H_k(M)\simeq H^{n-k}(M)$ and $H_k(N)\simeq H^{n-2-k}(N)$ shows that for $0<k<n-2$ we have isomorphisms:

$$\begin{gathered}
H_{k+2}(M)\to H^{n-k-2}(M)\to H^{n-k-2}(N)\to H_k(N)\to H_k(M),\\
H^k(M)\to H^k(N)\to H_{n-2-k}(N)\to H_{n-2-k}(M)\to H^{k+2}(M).
\end{gathered}$$

Using that M is simply connected shows that when n is even we have

$$0\simeq H_1(M)\simeq H_3(M)\simeq\cdots\simeq H_{n-1}(M)$$

and when n is odd

$$0\simeq H_1(M)\simeq H_3(M)\simeq\cdots\simeq H_{n-2}(M)\simeq H^2(M)\simeq\cdots\simeq H^{n-1}(M).$$

This shows that M is a homotopy sphere when n is odd.

When n is even we still have to figure out the possibilities for the even dimensional homology groups. This uses that we have

$$H^2(M) \simeq H^4(M) \simeq \cdots \simeq H^{n-2}(M) \to H^{n-2}(N) \simeq \mathbb{Z}.$$

The last map is injective since $N \subset M$ is $(n-2)$-connected. Thus these even dimensional cohomology groups are either all trivial or isomorphic to $\mathbb{Z}$. This gives the claim.

When M has nontrivial fundamental group the proof works for the universal covering, but more work is needed to classify the space itself. □

Proposition 8.3.8. *Let (M, g) be compact and assume that $X, Y \in \mathfrak{iso}(M, g)$ commute.*

(1) *Y is tangent to the level sets of $|X|^2$ and in particular to the zero set of X.*
(2) *If X and Y both vanish on a totally geodesic submanifold $N \subset M$, then some linear combination vanishes on a larger submanifold.*

Proof. (1) Since $L_Y X = 0$ and Y is a Killing field we get

$$\begin{aligned} 0 &= (L_Y g)(X, X) \\ &= D_Y |X|^2 - 2g(L_Y X, X) \\ &= D_Y |X|^2. \end{aligned}$$

Hence flow of Y preserves the level sets for $|X|^2$.

(2) We can assume that N has even codimension. Fix $p \in N$ and observe that proposition 8.1.3 shows that $(\nabla X)|_p$ and $(\nabla Y)|_p$ commute. Thus we obtain a splitting

$$T_pM = T_pN \oplus E_1 \oplus \cdots \oplus E_k,$$

where E_i are 2-dimensional and invariant under both $(\nabla X)|_p$ and $(\nabla Y)|_p$. As the space of skew-symmetric transformations on E_1, say, is one-dimensional some linear combination $\alpha(\nabla X)|_p + \beta(\nabla Y)|_p$ vanishes on E_1. Let $\bar{N} \supset N$ be the connected set on which $\alpha X + \beta Y$ vanishes. Then $T_p\bar{N} = \ker(\nabla(\alpha X + \beta Y))|_p$ contains $T_pN \oplus E_1$ and it follows that $N \subsetneq \bar{N}$. □

We are finally ready to present and prove the higher-dimensional versions alluded to earlier. The *symmetry rank* of a Riemannian manifold is the dimension of a maximal subspace of commuting Killing fields.

Theorem 8.3.9 (Grove and Searle, 1994). *Let M be a compact n-manifold with positive sectional curvature and symmetry rank k. If $k \geq n/2$, then M is diffeomorphic to either a sphere, complex projective space or a cyclic quotient of a sphere $S^n/\mathbb{Z}_q$, where $\mathbb{Z}_q$ is a cyclic group of order q acting by isometries on the unit sphere.*

Proof. We select an Abelian subalgebra $\mathfrak{a} \subset \mathfrak{iso}(M, g)$ of dimension $k \geq {}^{n}/_{2}$. Proposition 8.3.5 shows that with respect to inclusion there is a maximal and nontrivial totally geodesic submanifold $N \subset M$ and $X \in \mathfrak{a}$ that vanishes on N. Theorem 8.3.8 implies that $\mathfrak{a}|_N$ has dimension $k - 1$ (see also exercise 8.4.16). If N does not have codimension 2, then we can continue this construction and construct a totally geodesic 1- or 2-submanifold $S \subset M$ with $\dim(\mathfrak{a}|_S) \geq 2$. A 1-manifold has 1-dimensional isometry group so that does not happen. The 2-dimensional case is eliminated as follows. Select $X, Y \in \mathfrak{a}|_S$. Since X is nontrivial we can find an isolated zero p. As Y preserves the component $\{p\}$ of the zero set for X it also vanishes at p. Then

$$(\nabla X)|_p, (\nabla Y)|_p : T_pM \to T_pM$$

completely determine the Killing fields. As the set of skew-symmetric transformations on T_pM is 1-dimensional they must be linearly dependent. This shows that $\dim(\mathfrak{a}|_S) = 1$.

This means that we can use lemma 8.3.7 to finish the proof. □

With fewer symmetries we also have.

Theorem 8.3.10 (Püttmann and Searle, 2002 and Rong, 2002). *If M^{2n} is a compact $2n$-manifold with positive sectional curvature and symmetry rank $k \geq {}^{2n}/_{4} - 1$, then $\chi(M) > 0$.*

Proof. When $2n = 2, 4$ there are no assumptions about the symmetry rank and we know the theorem holds. When $2n = 6$ it is Berger's result. Next consider the case where M is 8-dimensional. The proof is as in the 6-dimensional situation unless the zero set for the Killing field has a 6-dimensional component. In that case lemma 8.3.7 establishes the claim.

In general we would like to use induction on dimension, but this requires that we work with the stronger statement: If $\mathfrak{a} \subset \mathfrak{iso}(M, g)$ is an Abelian subalgebra of dimension $k \geq {}^{n}/_{2} - 1$, then any component of the zero set for any $X \in \mathfrak{a}$ has positive Euler characteristic. Note that when $2n = 2, 4, 6, 8$ this stronger statement holds.

There are two cases: First assume that every zero set $N \subset M$ for any $X \in \mathfrak{a}$ has codimension ≥ 4. When N is maximal and nontrivial, then $\mathfrak{a}|_N$ has dimension $k - 1$. Since any component of a zero set is contained in a nontrivial maximal element the stronger induction hypothesis can be invoked to prove the induction step.

The other situation is when there is a zero set N for some $X \in \mathfrak{a}$ that has codimension 2. Lemma 8.3.7 then shows that the odd Betti numbers of M vanish. This in turn implies that the odd Betti numbers for any component of a zero set of any Killing field must also vanish (see [17]). □

This theorem unfortunately does not cover all known examples as there is a positively curved 24-manifold $F_4/\mathrm{Spin}(8)$ that has symmetry rank 4 (see [105]).

It is tempting to suppose that one could show that the odd homology groups with real coefficients vanish given the assumptions of the previous theorem. In fact, all known even dimensional manifolds with positive sectional curvature have vanishing odd dimensional homology groups.

Next we mention without proof an extension of the Grove-Searle result by Wilking, (see also [105]).

Theorem 8.3.11 (Wilking, 2003). *Let M be a compact simply connected positively curved n-manifold with symmetry rank k and $n \geq 10$. If $k \geq n/4 + 1$, then M has the topology of a sphere, complex projective space or quaternionic projective space. Moreover, when M isn't simply connected its fundamental group is cyclic.*

The proof is considerably more complicated than the above theorems. When $n = 7$ there are several spaces with positive curvature and symmetry rank 3 (see [105]).

Finally, we mention some recent extensions of the above theorems.

Theorem 8.3.12 (Kennard, 2013 [67]). *If M^{4n} is a compact $4n$-manifold with positive sectional curvature and symmetry rank $k > 2\log_2(4n) - 2$, then $\chi(M) > 0$.*

Theorem 8.3.13 (Kennard, 2012 [68] and Amann and Kennard, 2014 [1]). *Let M^{2n} be a compact positively curved manifold with symmetry rank $k \geq \log_{4/3}(2n)$.*

(1) *The Betti numbers satisfy: $b_2 \leq 1$ and $b_4 \leq 1$.*
(2) *If $b_4 = 0$, then $\chi(M) = 2$.*
(3) (1) *and* (2) *imply that M can't have the homotopy type of a product $N \times N$ where N is compact and simply connected.*

8.4 Exercises

EXERCISE 8.4.1. Show that theorem 8.1.6 does not necessarily extend to incomplete Riemannian manifolds.

EXERCISE 8.4.2. Let N be a component of the zero set for a Killing field X. Show that $\nabla_V(\nabla X) = 0$ for vector fields V tangent to N.

EXERCISE 8.4.3. Show that a coordinate vector field ∂_k is a Killing field if and only if $\partial_k g_{ij} = 0$.

EXERCISE 8.4.4. Let X be a Killing field on (M, g) and $N \subset M$ a submanifold with the induced metric.

(1) Show that if X is tangent to N, then $X|_N$ is a Killing field on N.
(2) Show that if N is totally geodesic (see exercise 5.9.20), then $X^\top$, the projection of X onto N, is a Killing field.

EXERCISE 8.4.5. Let (M, g) be a complete Riemannian manifold and X a Killing field on M.

(1) Let $c : [a, b] \to M$ be a geodesic. Show that there is an $\epsilon > 0$ and a geodesic variation $\bar{c} : (-\epsilon, \epsilon) \times [a, b] \to M$ such that the curves $s \mapsto \bar{c}(s, t)$ are integral curves of X.

(2) Use completeness to extend the geodesic variation $\bar{c} : (-\epsilon, \epsilon) \times \mathbb{R} \to M$. Show that for all $t \in \mathbb{R}$ the curves $s \mapsto \bar{c}(s, t)$ are integral curves of X.
(3) Show that the integral curve of X through any point in M exists on $(-\epsilon, \epsilon)$.
(4) Show that X is complete.

EXERCISE 8.4.6. Let (M^n, g) be a compact Riemannian n-manifold such that $\dim(\mathfrak{iso}) = n$ and $\mathrm{Ric} \leq 0$. Show that M is flat and that $\pi_1 = \mathbb{Z}^n$. Hint: Show that $\tilde{M} = \mathbb{R}^n$ and that the deck transformations must have differential I since they leave a parallel orthonormal frame invariant.

EXERCISE 8.4.7. Let x^i be the standard Cartesian coordinates on $\mathbb{R}^n$ and consider $W = \mathrm{span}_{\mathbb{R}}\left\{1, x^1, \dots, x^n\right\}$ and $V = \mathrm{span}_{\mathbb{R}}\left\{x^1, \dots, x^n\right\}$.

(1) Show that $\mathfrak{iso}(\mathbb{R}^n)$ is naturally isomorphic to $\Lambda^2 W$ if we identify $u \wedge v$ with the vector field $u\nabla v - v\nabla u$.
(2) Show similarly that $\mathfrak{iso}\left(S^{n-1}(R)\right)$ is naturally isomorphic to $\Lambda^2 V$.
(3) Use that $\nabla u = g^{ij}\partial_i u \partial_j$ for a pseudo-Riemannian space to redo (1) for $\mathbb{R}^{p,q}$ and (2) for $H^{n-1}(R) \subset \mathbb{R}^{n,1}$.

EXERCISE 8.4.8. Let x^i be the standard Cartesian coordinates on $\mathbb{R}^{n+1}$ and consider $V = \mathrm{span}_{\mathbb{R}}\left\{1, x^1, \dots, x^{n+1}\right\}$. Finally restrict all functions in V to S^n. Show that for $u, v \in V$ the fields $u\nabla v - v\nabla u$ are conformal. A field X is conformal if $L_X g = \lambda g$ for some function λ.

EXERCISE 8.4.9. Let (M, g) be a Riemannian manifold and consider the subspaces $\mathfrak{k}_p = \left\{X|_p \in T_pM \mid X \in \mathfrak{iso}\right\}$.

(1) Show that $\mathfrak{k}_p$ defines an integrable distribution on an open set $O \subset M$. Hint: Show that $\left\{p \in M \mid \dim \mathfrak{k}_p = \max_{p \in M} \dim \mathfrak{k}_p\right\}$ is open.
(2) Give an example where $O \neq M$.
(3) Show that $DF\left(\mathfrak{k}_p\right) = \mathfrak{k}_{F(p)}$ for all $F \in \mathrm{Iso}$.
(4) Assume (M, g) is complete. Show that the leaves of this distribution are homogeneous and properly embedded. Hint: Show that the connected subgroup $\mathrm{Iso}_0 \subset \mathrm{Iso}$ that contains the identity is closed and preserves the leaves (see also section 5.6.4.)
(5) Show that if (M, g) is homogeneous, then $\mathfrak{k}_p = T_pM$ for all $p \in M$.

EXERCISE 8.4.10. Let $\mathfrak{t} \subset \mathfrak{iso}(M, g)$ be an Abelian subalgebra corresponding to a torus subgroup $T^k \subset \mathrm{Iso}(M, g)$. Define $\mathfrak{p} \subset \mathfrak{t}$ as the set of Killing fields that correspond to circle actions, i.e., actions induced by homomorphisms $S^1 \to T^k$. Show that $\mathfrak{p}$ is a vector space over the rationals with $\dim_{\mathbb{Q}} \mathfrak{p} = \dim_{\mathbb{R}} \mathfrak{t}$.

EXERCISE 8.4.11. 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 frame consisting of Killing fields.

EXERCISE 8.4.12. Let X be a vector field on a Riemannian manifold.

(1) Show that

$$|L_X g|^2 = 2|\nabla X|^2 + 2\operatorname{tr}(\nabla X \circ \nabla X).$$

(2) Establish the following integral formulae on a closed oriented Riemannian manifold:

$$\int_M \left(\operatorname{Ric}(X,X) + \operatorname{tr}(\nabla X \circ \nabla X) - (\operatorname{div} X)^2\right) = 0,$$

$$\int_M \left(\operatorname{Ric}(X,X) + g\left(\operatorname{tr}\nabla^2 X, X\right) + \frac{1}{2}|L_X g|^2 - (\operatorname{div} X)^2\right) = 0.$$

(3) Finally, show that X is a Killing field if and only if

$$\operatorname{div} X = 0,$$
$$\operatorname{tr}\nabla^2 X = -\operatorname{Ric}(X).$$

EXERCISE 8.4.13 (Yano). If X is an affine vector field (see exercise 2.5.13) 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.

EXERCISE 8.4.14. Let X be a vector field on a Riemannian manifold. Show that X is a Killing field if and only if L_X and Δ commute on functions.

EXERCISE 8.4.15. Let (M,g) be a compact n-manifold with positive sectional curvature and $\mathfrak{a} \subset \mathfrak{iso}$ an Abelian subalgebra.

(1) Show that $\dim\mathfrak{a} \leq {}^{n}/{}_{2}$.
(2) Show that spheres and complex projective spaces have maximal symmetry rank.
(3) Show that the flat torus T^n has symmetry rank n.

EXERCISE 8.4.16. Let (M,g) be a compact n-manifold with positive sectional curvature and $\mathfrak{a} \subset \mathfrak{iso}$ an Abelian subalgebra. Let $\mathscr{Z}(\mathfrak{a})$ be the set of nontrivial connected components of the zero sets of Killing fields in $\mathfrak{a}$. Show that if $N \in \mathscr{Z}(\mathfrak{a})$ is maximal under inclusion, then $\dim(\mathfrak{a}|_N) = \dim\mathfrak{a} - 1$ and $\dim N \geq 2(\dim\mathfrak{a} - 1)$.

EXERCISE 8.4.17 (Kennard). Let (M^n, g) be a simply connected compact n-manifold with positive sectional curvature and symmetry rank $\geq {}^{3n}/{}_{8} + 1$.

(1) When $n \leq 8$ conclude that the assumptions are covered by theorem 8.3.9.
(2) Use exercise 8.4.16 to show that if $N \in \mathscr{Z}(\mathfrak{a})$ is maximal under inclusion, then $\dim N \geq {}^{3n}/{}_{4}$.
(3) Show that a maximal $N \in \mathscr{Z}(\mathfrak{a})$ either has codimension 2 or has symmetry rank $\geq \frac{3\dim N}{8} + 1$.

(4) Use induction on n to show that M has the homology groups of a sphere or complex projective space. Hint: When the maximal $N \in \mathscr{X}(\mathfrak{a})$ has codimension ≥ 4 use the connectedness principle to show that N is also simply connected. Then use the connectedness principle to calculate the homology/homotopy groups in dimensions $\leq n/2$. Finally use Poincaré duality to find all the homology groups of M.

EXERCISE 8.4.18. Let $\tilde{M} \to M$ be the universal covering, with deck transformations $\Gamma = \pi_1(M)$ acting as isometries on $\tilde{M}$.

(1) Show that we can identify

$$\mathfrak{iso}(M) = \left\{X \in \mathfrak{iso}\left(\tilde{M}\right) \mid G_* X = X \text{ for all } G \in \Gamma\right\}.$$

(2) Show that the connected Lie subgroup corresponding to $\mathfrak{iso}(M) \subset \mathfrak{iso}\left(\tilde{M}\right)$ is the connected component of the centralizer

$$\mathrm{C}(\Gamma) = \left\{F \in \mathrm{Iso}\left(\tilde{M}\right) \mid FG = GF \text{ for all } G \in \Gamma\right\}$$

that contains the identity.

(3) Show that $\mathrm{Iso}(M)$ can be identified with $\mathrm{N}(\Gamma)/\Gamma$, where $\mathrm{N}(\Gamma)$ is the normalizer

$$\mathrm{N}(\Gamma) = \left\{F \in \mathrm{Iso}\left(\tilde{M}\right) \mid F\Gamma F^{-1} = \Gamma\right\}.$$

Chapter 9
The Bochner Technique

Aside from the variational techniques we've used in prior sections one of the oldest and most important techniques in modern Riemannian geometry is that of the Bochner technique. In this chapter we prove the classical theorem of Bochner about obstructions to the existence of harmonic 1-forms. We also explain in detail how the Bochner technique extends to forms and other tensors by using Lichnerowicz Laplacians. This leads to a classification of compact manifolds with nonnegative curvature operator in chapter 10. To establish the relevant Bochner formula for forms, we have used a somewhat forgotten approach by Poor. It appears to be quite simple and intuitive. It can, as we shall see, also be generalized to work on other tensors including the curvature tensor.

The classical focus of the Bochner technique lies in establishing certain vanishing results for suitable tensors in positive curvature. This immediately leads to rigidity results when the curvature is nonnegative. In the 1970s P. Li discovered that it can be further generalized to estimate the dimension of the kernel of the Laplace operators under more general curvature assumptions. This became further enhanced when Gallot realized that the necessary analytic estimates work with only lower Ricci curvature bounds. This will all be explained here and uses in a crucial way results from sections 7.1.5 and 7.1.3.

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

$$\Delta \frac{1}{2} |\nabla u|^2 = |\mathrm{Hess} u|^2 ,$$

P. Petersen, *Riemannian Geometry*, Graduate Texts in Mathematics 171,
DOI 10.1007/978-3-319-26654-1_9

where $u : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ and $\Delta u = 0$. It was Bochner who 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

$$\Delta \frac{1}{2} |\nabla u|^2 = |\mathrm{Hess} u|^2 + \mathrm{Ric}(\nabla u, \nabla u).$$

With this in mind it is clear that curvature influences the behavior of harmonic functions. The next nontrivial step Bochner took was to realize that one can compute $\Delta \frac{1}{2} |\omega|^2$ 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 cohomology class can be uniquely represented by a harmonic form. Yano further refined the Bochner technique, but it seems to be Lichnerowicz who really put things into motion when he presented his formulas for the Laplacian on forms and spinors around 1960. After this work, Berger, D. Meyer, Gallot, Gromov-Lawson, Witten, and many others have made significant contributions to this tremendously important subject.

Prior to Bochner's work Weitzenböck developed a formula very similar to the Bochner formula. We shall also explain this related formula and how it can be used to establish the Bochner formulas we use. It appears that Weitzenböck never realized that his work could have an impact on geometry and only thought of his work as an application of algebraic invariant theory.

9.1 Hodge Theory

We start by giving a brief account of Hodge theory to explain why it calculates the homology of a manifold.

Recall that on a manifold M we have the *de Rham complex*

$$0 \to \Omega^0(M) \xrightarrow{d^0} \Omega^1(M) \xrightarrow{d^1} \Omega^2(M) \to \cdots \xrightarrow{d^{n-1}} \Omega^n(M) \to 0,$$

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

$$H^k(M) = \frac{\ker(d^k)}{\mathrm{im}(d^{k-1})}$$

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 there is a pairing,

$$\Omega^k(M) \times \Omega^{n-k}(M) \to \mathbb{R},$$
$$(\omega_1, \omega_2) \to \int_M \omega_1 \wedge \omega_2,$$

that induces a nondegenerate pairing on the cohomology groups

$$H^k(M) \times H^{n-k}(M) \to \mathbb{R}.$$

This shows that the two vector spaces $H^k(M)$ and $H^{n-k}(M)$ are dual to each other and in particular have the same dimension.

When M is endowed with a Riemannian metric g we also obtain an adjoint $\delta = \nabla^*$ to the differential (see proposition 2.2.8 and section 2.2.2.2). Specifically,

$$\delta^k : \Omega^{k+1}(M) \to \Omega^k(M)$$

is adjoint to d^k via the formula

$$\int_M g(\delta^k \omega_1, \omega_2)\text{vol} = \int_M g(\omega_1, d^k \omega_2)\text{vol}.$$

It is often convenient to use the notation

$$(T_1, T_2) = \frac{1}{\text{vol}M} \int_M g\,(T_1, T_2)\,\text{vol}.$$

This defines an inner product on any space of tensors of the same type. The map δ is also the adjoint of d with respect to this normalized inner product.

The *Laplacian on forms,* also called the *Hodge Laplacian,* is 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, hence the need for the slightly different symbol $\triangle$ instead of Δ.

Lemma 9.1.1. *$\triangle\omega = 0$ if and only if $d\omega = 0$ and $\delta\omega = 0$. In particular, $\omega = 0$, if $\triangle\omega = 0$ and $\omega = d\theta$.*

Proof. The proof just uses that the maps are adjoints to each other:

$$\begin{aligned} (\triangle\omega, \omega) &= (d\delta\omega, \omega) + (\delta d\omega, \omega) \\ &= (\delta\omega, \delta\omega) + (d\omega, d\omega). \end{aligned}$$

Thus, $\triangle\omega = 0$ implies $(\delta\omega, \delta\omega) = (d\omega, d\omega) = 0$, which shows that $\delta\omega = 0$ and $d\omega = 0$. The opposite direction is obvious.

Note that when $\omega = d\theta$ and $\triangle\omega = 0$, then $\delta d\theta = 0$, which in turn shows that $(\omega, \omega) = (\theta, \delta d\theta) = 0$. □

We can now introduce the *Hodge cohomology*:

$$\mathscr{H}^k(M) = \{\omega \in \Omega^k(M) \mid \triangle\omega = 0\}.$$

Theorem 9.1.2 (Hodge, 1935). *The natural inclusion map $\mathscr{H}^k(M) \to H^k(M)$ is an isomorphism.*

Proof. The fact that $\mathscr{H}^k(M) \to H^k(M)$ is well-defined and injective follows from lemma 9.1.1. To show that it is surjective requires a fair bit of work that is standard in the theory of partial differential equations (see [99] or [92]). Some of the results we prove later will help to establish part of this result in a more general context (see exercises 9.6.4 and 9.6.5). The essential idea is the claim; since $\triangle$ is self-adjoint there is an orthogonal decomposition

$$\Omega^k(M) = \mathrm{im}\triangle \oplus \ker\triangle = \mathrm{im}\triangle \oplus \mathscr{H}^k(M).$$

If $\omega \in \Omega^k$, then we can write $\omega = d\delta\theta + \delta d\theta + \tilde{\omega}$, where $\triangle\tilde{\omega} = 0$. When in addition $d\omega = 0$ it follows that $\triangle d\theta = d\delta d\theta = 0$. Since $d\theta$ is also harmonic it follows that $\delta d\theta = 0$. In particular, $\omega = d\delta\theta + \tilde{\omega}$ and $\tilde{\omega}$ represents the same cohomology class as ω. $\square$

9.2 1-Forms

We shall see how Hodge theory can be used to get information about the first Betti number $b_1(M) = \dim\mathscr{H}^1(M)$. In the next section we generalize this to other forms and tensors.

9.2.1 The Bochner Formula

Let θ be a harmonic 1-form on (M, g) and $f = \frac{1}{2}|\theta|^2$. To get a better feel for this function consider the vector field X field dual to θ, i.e., $\theta(v) = g(X, v)$ for all v. Then

$$f = \frac{1}{2}|\theta|^2 = \frac{1}{2}|X|^2 = \frac{1}{2}\theta(X).$$

Proposition 9.2.1. *If X and θ are related by $\theta(v) = g(v, X)$, then*

(1) $v \mapsto \nabla_v X$ *is symmetric if and only if* $d\theta = 0$ *and*
(2) $\mathrm{div}X = -\delta\theta$.

Proof. Recall that

$$d\theta(V, W) + (L_X g)(V, W) = 2g(\nabla_V X, W).$$

Since $L_X g$ is symmetric and $d\theta$ is skew-symmetric the result immediately follows. The second part was proven in proposition 2.2.7. □

Therefore, when θ is harmonic, then $\mathrm{div}X = 0$ and ∇X is a symmetric $(1,1)$-tensor.

We present the Bochner formula for closed 1-forms formulated through vector fields.

Proposition 9.2.2. *Let X be a vector field so that ∇X is symmetric (i.e. corresponding 1-form is closed). If $f = \frac{1}{2}|X|^2$ and X is the gradient of u near p, then*

(1)

$$\nabla f = \nabla_X X.$$

(2)

$$\begin{aligned}\mathrm{Hess} f(V,V) &= \mathrm{Hess}^2 u(V,V) + (\nabla_X \mathrm{Hess} u)(V,V) + R(V,X,X,V) \\ &= |\nabla_V X|^2 + g\left(\nabla^2_{X,V} X, V\right) + R(V,X,X,V)\end{aligned}$$

(3)

$$\begin{aligned}\Delta f &= |\mathrm{Hess} u|^2 + D_X \Delta u + \mathrm{Ric}(X,X) \\ &= |\nabla X|^2 + D_X \mathrm{div} X + \mathrm{Ric}(X,X)\end{aligned}$$

Proof. For (1) simply observe that

$$g(\nabla f, V) = D_V \frac{1}{2}|X|^2 = g(\nabla_V X, X) = g(\nabla_X X, V).$$

(2) is a direct consequence of theorem 3.2.2 applied to the function u.

For (3) take traces in (2). As in the proof of proposition 8.2.1 this gives us the first and third terms. The second term comes from commuting traces and covariant derivatives. Specifically, either $X|_p = 0$ or E_i can be chosen to parallel along X. In either case

$$\begin{aligned}\sum g\left(\nabla^2_{X,E_i} X, E_i\right) &= \sum (\nabla_X \mathrm{Hess} u)(E_i,E_i) \\ &= D_X \sum \mathrm{Hess} u(E_i,E_i) \\ &= D_X \Delta u.\end{aligned}$$

□

9.2.2 *The Vanishing Theorem*

We can now easily establish the other Bochner theorem for 1-forms.

Theorem 9.2.3 (Bochner, 1948). *If* (M, g) *is compact and has* $\mathrm{Ric} \geq 0$*, then every harmonic 1-form is parallel.*

Proof. Suppose ω is a harmonic 1-form and X the dual vector field. Then proposition 9.2.2 implies

$$\Delta \left(\frac{1}{2}|X|^2\right) = |\nabla X|^2 + \mathrm{Ric}(X, X) \geq 0,$$

since $\mathrm{div}X = \Delta u = 0$. The maximum principle then shows that $\frac{1}{2}|X|^2$ must be constant and $|\nabla X| = 0$. □

Corollary 9.2.4. *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 $\mathrm{Ric}(X, X) \equiv 0$, 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 9.2.5. *If* (M, g) *is compact and satisfies* $\mathrm{Ric} \geq 0$*, then* $b_1(M) \leq n = \dim M$*, with equality holding if and only if* (M, g) *is a flat torus.*

Proof. We know from Hodge theory that $b_1(M) = \dim \mathscr{H}^1(M)$. Now, all harmonic 1-forms are parallel, so the linear map: $\mathscr{H}^1(M) \to T_p^*M$ that evaluates ω at p is injective. In particular, $\dim \mathscr{H}^1(M) \leq n$.

If equality holds, then there are n linearly independent parallel fields E_i, $i = 1, \ldots, n$. This clearly implies that (M, g) is flat. Thus the universal covering is $\mathbb{R}^n$ with $\pi_1(M)$ 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 therefore constant vector fields. This means that we can think of them as the usual Cartesian coordinate vector fields ∂_i. In addition, they are invariant under the action of $\pi_1(M)$, i.e., for each $F \in \pi_1(M)$ we have $DF\left(\partial_i|_p\right) = \partial_i|_{F(p)}$, $i = 1, \ldots, n$. But only translations leave all of the coordinate fields invariant. Thus, $\pi_1(M)$ consists entirely of translations. This means that $\pi_1(M)$ is finitely generated, Abelian, and torsion free. Hence $\Gamma = \mathbb{Z}^k$ for some k. To see that M is a torus, we need $k = n$. If $k < n$, then $\mathbb{Z}^k$ generates a proper subspace of the space of translations and can't act cocompactly on $\mathbb{R}^n$. If $k > n$, then $\mathbb{Z}^k$ can't act discretely on $\mathbb{R}^n$. Thus, it follows that $\Gamma = \mathbb{Z}^n$ and generates $\mathbb{R}^n$. □

9.2.3 The Estimation Theorem

The goal is to generalize theorem 9.2.3 to manifolds with a negative lower bound for the Ricci curvature. The techniques were first developed by P. Li in the late '70s and then improved by Gallot to give the results we present. Gallot's contribution was in part to obtain a suitable bound for Sobolev constants as in theorem 7.1.13.

We start with a very general analysis lemma. Assume we have a compact Riemannian manifold (M,g) and a vector bundle $E \to M$ where the fibers are endowed with a smoothly varying inner product and the dimension of the fibers is m. Sections of this bundle are denoted $\Gamma(E)$ and have several natural norms

$$\|s\|_\infty = \max_{x\in M} |s(x)|,$$

$$\|s\|_p = \left(\frac{1}{\mathrm{vol}M}\int_M |s|^p \,\mathrm{vol}\right)^{\frac{1}{p}}.$$

The normalization is consistent with earlier definitions and guarantees that $\|s\|_p$ increases to $\|s\|_\infty$. Now fix a finite dimensional subspace $V \subset \Gamma(E)$. All of the norms are then equivalent on this space and we can define

$$C(V) = \max_{s\in V-\{0\}} \frac{\|s\|_\infty}{\|s\|_2}.$$

The dimension of V can be estimated by this constant and the dimension of the fibers of E.

Lemma 9.2.6 (P. Li). *With notation as above*

$$\dim V \leq m \cdot C(V).$$

Proof. Note that V has a natural inner product

$$(s_1, s_2) = \frac{1}{\mathrm{vol}M}\int_M \langle s_1, s_2\rangle \,\mathrm{vol}$$

such that $(s,s) = \|s\|_2^2$. Select an orthonormal basis $e_1, \dots, e_l \in V$ with respect to this inner product and observe that the function

$$f(x) = \sum_{i=1}^{l} |e_i(x)|^2$$

does not depend on the choice of orthonormal basis. Moreover,

$$\frac{1}{\mathrm{vol}M}\int_M f\mathrm{vol} = l = \dim V.$$

Let x_0 be the point where f is maximal. Consider the map $V \to E_{x_0}$ that evaluates a section at x_0. We can then assume that the basis is chosen so that the last $l-k$ elements span the kernel. This implies that $k \leq m$ and

$$\dim V \leq f(x_0) \leq k \cdot C(V)$$

since each section had unit L^2-norm. This proves the claim. □

Next we extend the maximum principle to a situation where we can bound $\|u\|_\infty$ in terms of $\|u\|_p$.

Theorem 9.2.7 (Moser iteration). *Let (M, g) be a compact Riemannian manifold such that*

$$\|u\|_{2\nu} \leq S\,\|\nabla u\|_2 + \|u\|_2$$

for all smooth functions, where $\nu > 1$. If $f : M \to [0, \infty)$ is continuous, smooth on $\{f > 0\}$, and $\Delta f \geq -\lambda f$, then

$$\|f\|_\infty \leq \exp\left(\frac{S\sqrt{\lambda\nu}}{\sqrt{\nu} - 1}\right)\|f\|_2 .$$

Proof. Since f is minimized on the set where it vanishes we can assume that all of its derivatives vanish there. In fact, Δf is nonnegative at those points both in the barrier and distributional sense.

First note that Green's formula implies

$$\begin{aligned}\left(f^{2q-1}, \Delta f\right) &= -\left(df^{2q-1}, df\right)\\ &= -\left(2q-1\right)\left(f^{2q-2}df, df\right).\end{aligned}$$

This shows that

$$\begin{aligned}\|df^q\|_2^2 &= q^2\left(f^{2q-2}df, df\right)\\ &= -\frac{q^2}{2q-1}\left(f^{2q-1}, \Delta f\right)\\ &\leq \frac{q^2\lambda}{2q-1}\left(f^{2q-1}, f\right)\\ &= \frac{q^2\lambda}{2q-1}\|f^q\|_2^2 .\end{aligned}$$

We can then use the Sobolev inequality to conclude that

$$\|f^q\|_{2\nu} \leq S\,\|df^q\|_2 + \|f^q\|_2 \leq \left(Sq\left(\frac{\lambda}{2q-1}\right)^{\frac{1}{2}} + 1\right)\|f^q\|_2$$

and

$$\|f\|_{2\nu q} \leq \left(Sq\left(\frac{\lambda}{2q-1}\right)^{\frac{1}{2}} + 1\right)^{\frac{1}{q}}\|f\|_{2q}\,.$$

Letting $q = \nu^k$ gives

$$\|f\|_{2\nu^{k+1}} \leq \left(S\nu^k\left(\frac{\lambda}{2\nu^k-1}\right)^{\frac{1}{2}} + 1\right)^{\nu^{-k}}\|f\|_{2\nu^k}\,.$$

Consequently, by starting at $k = 0$ and letting $k \to \infty$ we obtain

$$\|f\|_{\infty} \leq \left(\prod_{k=0}^{\infty}\left(S\nu^k\left(\frac{\lambda}{2\nu^k-1}\right)^{\frac{1}{2}} + 1\right)^{\nu^{-k}}\right)\|f\|_2\,.$$

The infinite product is estimated by taking logarithms and using $\log(1+x) \leq x$

$$\begin{aligned}\sum_{k=0}^{\infty}\nu^{-k}\log\left(S\nu^k\left(\frac{\lambda}{2\nu^k-1}\right)^{\frac{1}{2}} + 1\right) &\leq S\sqrt{\lambda}\sum_{k=0}^{\infty}\left(\frac{1}{2\nu^k-1}\right)^{\frac{1}{2}}\\ &\leq S\sqrt{\lambda}\sum_{k=0}^{\infty}\left(\frac{1}{\nu^k}\right)^{\frac{1}{2}}\\ &= \frac{S\sqrt{\lambda\nu}}{\sqrt{\nu}-1}.\end{aligned}$$

□

Together these results imply

Theorem 9.2.8 (Gromov, 1980 and Gallot, 1981). *If M is a compact Riemannian manifold of dimension n such that* $\mathrm{Ric} \geq (n-1)k$ *and* $\mathrm{diam}(M) \leq D$*, then there is a function $C\left(n, k\cdot D^2\right)$ such that*

$$b_1(M) \leq C\left(n, k\cdot D^2\right).$$

Moreover, $\lim_{\varepsilon\to 0} C(n,\varepsilon) = n$*. In particular, there is $\varepsilon(n) > 0$ such that when $k\cdot D^2 \geq -\varepsilon(n)$, then $b_1(M) \leq n$.*

Proof. Gromov's proof centered on understanding how covering spaces of M control the Betti number. Gallot's proof has the advantage of also being useful in a wider context as we shall explore below.

The goal is clearly to estimate $\dim \mathscr{H}^1$. Lemma 9.2.6 implies that

$$\dim \mathscr{H}^1 \leq n \cdot C\left(\mathscr{H}^1\right),$$

So we have to estimate the ratios $\frac{\|\omega\|_\infty}{\|\omega\|_2}$. To do so consider $f = |\omega|$. This function is smooth except possibly at points where $\omega = 0$, which also happen to be minimum points for f. Note that

$$2fdf = df^2 = 2g\left(\nabla\omega, \omega\right) \leq 2\left|\nabla\omega\right| f$$

so we obtain Kato's inequality $df \leq |\nabla\omega|$. If X is the dual vector field to ω the Bochner formula implies

$$\begin{aligned} |df|^2 + f\Delta f &= \frac{1}{2}\Delta f^2 \\ &= |\nabla\omega|^2 + \operatorname{Ric}(X, X) \\ &\geq |\nabla\omega|^2 + (n-1)\, kf^2. \end{aligned}$$

It follows by Kato's inequality that $\Delta f \geq (n-1)\, kf$. Theorem 9.2.7 then shows that

$$\|f\|_\infty \leq \exp\left(\frac{S\sqrt{-(n-1)\, k\nu}}{\sqrt{\nu}-1}\right)\|f\|_2 .$$

Since $\frac{\|f\|_\infty}{\|f\|_2} = \frac{\|\omega\|_\infty}{\|\omega\|_2}$ we have proven that

$$\dim \mathscr{H}^1 \leq n \cdot C\left(\mathscr{H}^1\right) \leq n \cdot \exp\left(\frac{S\sqrt{-(n-1)\, k\nu}}{\sqrt{\nu}-1}\right),$$

where $S = D \cdot C\left(n, kD^2\right)$ is estimated in theorem 7.1.13 and proposition 7.1.17. The specific nature of the bound proves the theorem. □

9.3 Lichnerowicz Laplacians

We introduce a natural class of Laplacians and show how the Bochner technique works for these operators. In the next section we then show that there are several natural Laplacians of this type including the Hodge Laplacian.

9.3.1 The Connection Laplacian

We start by collecting the results from the previous section in a more general context.

Fix a tensor bundle $E \to M$ with m-dimensional fibers. This could the bundle whose sections are p-forms, symmetric tensors, curvature tensors etc.

First the vanishing result.

Proposition 9.3.1. *Let (M, g) be a Riemannian manifold and $T \in \Gamma(E)$ a section such that $g(\nabla^*\nabla T, T) \leq 0$. If $|T|$ has a maximum, then T is parallel.*

Proof. Note that

$$\Delta \frac{1}{2} |T|^2 = |\nabla T|^2 - g\left(\nabla^*\nabla T, T\right) \geq 0.$$

In case $|T|$ has a maximum we can apply the maximum principle to the function $|T|^2$ and conclude that it must be constant and that T itself is parallel. □

Next we present the estimating result.

Theorem 9.3.2 (Gallot, 1981). *Assume (M, g) is a compact manifold that satisfies the assumption of theorem 9.2.7. Let $V \subset \Gamma(M)$ be finite dimensional. If*

$$g\left(\nabla^*\nabla T, T\right) \leq \lambda |T|^2$$

for all $T \in V$, then

$$\dim V \leq m \cdot \exp\left(\frac{S\sqrt{\lambda\nu}}{\sqrt{\nu} - 1}\right).$$

Proof. This is proven as in theorem 9.2.8 using $f = |T|$. Instead of the Bochner formula we simply use the equation

$$|df|^2 + f\Delta f = \frac{1}{2}\Delta f^2 = |\nabla T|^2 - g\left(\nabla^*\nabla T, T\right)$$

to conclude via Kato's inequality that $\Delta f \geq -\lambda f$. We can then finish the proof in the same fashion. □

9.3.2 The Weitzenböck Curvature

The *Weitzenböck curvature operator* on a tensor is defined by

$$\mathrm{Ric}(T)(X_1, \ldots, X_k) = \sum \left(R\left(e_j, X_i\right) T\right)\left(X_1, \ldots, e_j, \ldots, X_k\right).$$

We use the Ricci tensor to symbolize this as it is the Ricci tensor when evaluated on vector fields and 1-forms. Specifically:

$$\mathrm{Ric}\,(\omega)\,(X) = \sum \left(R\left(e_j, X\right)\omega\right)\left(e_j\right) = -\omega\left(\sum R\left(e_j, X\right)e_j\right) = \omega\,(\mathrm{Ric}\,(X))\,.$$

Often it is referred to as W, but this can be confused with the Weyl tensor.

The *Lichnerowicz Laplacian* is defined as

$$\Delta_L T = \nabla^* \nabla T + c\mathrm{Ric}\,(T)$$

for a suitable constant $c > 0$. We shall see below that the Hodge Laplacian on forms is of this type with $c = 1$. In addition, interesting information can also be extracted for symmetric $(0, 2)$-tensors as well as the curvature tensor via this operator when we use $c = \frac{1}{2}$.

The Bochner technique works for tensors that lie in the kernel of some Lichnerowicz Laplacian

$$\Delta_L T = \nabla^* \nabla T + c\mathrm{Ric}\,(T) = 0.$$

The idea is to use the maximum principle to show that T is parallel. In order to apply the maximum principle we need $g\,(\nabla^* \nabla T, T) \leq 0$ which by the equation for T is equivalent to showing $g\,(\mathrm{Ric}\,(T)\,, T) \geq 0$.

The two assumptions $\Delta_L T = 0$ and $g\,(\mathrm{Ric}\,(T)\,, T) \geq 0$ we make about T require some discussion.

The first assumption is usually implied by showing that the Lichnerowicz Laplacian has an alternate expression such as we have seen for the Hodge Laplacian. The fact that $\Delta_L T = 0$ might come from certain natural restrictions on the tensor or even as a consequence of having nontrivial topology. In the next section several natural Laplacians are rewritten as Lichnerowicz Laplacians.

The second assumption $g\,(\mathrm{Ric}\,(T)\,, T) \geq 0$, is often difficult to check and in many cases it took decades to sort out what curvature assumptions gave the best results. The goal in this section is to first develop a different formula for $\mathrm{Ric}\,(T)$ and second to change T in a suitable fashion so as to create a significantly simpler formula for $g\,(\mathrm{Ric}\,(T)\,, T)$. This formula will immediately show that $g\,(\mathrm{Ric}\,(T)\,, T)$ is nonnegative when the curvature operator is nonnegative. It will also make it very easy to calculate precisely what happens when T is a $(0, 1)$- or $(0, 2)$-tensor, a task we delay until the next section. It is worthwhile mentioning that the original proofs of some of these facts were quite complicated and only came to light long after the Bochner technique had been introduced.

9.3.3 *Simplification of* Ric (T)

Since $R_{X,Y} : T_pM \to T_pM$ is always skew-symmetric it can be decomposed using an orthonormal basis of skew-symmetric transformations $\Xi_\alpha \in \mathfrak{so}(T_pM)$. A tricky point enters our formulas at this point. It comes from the fact that if v and w are orthonormal, then $v \wedge w \in \Lambda^2 T_pM$ is a unit vector, while the corresponding skew-symmetric operator, a counter clockwise rotation of $\pi/2$ in span$\{v, w\}$, has Euclidean norm $\sqrt{2}$. To avoid confusion and unnecessary factors we assume that $\mathfrak{so}(T_pM)$ is endowed with the metric that comes from $\Lambda^2 T_pM$. With that in mind we have

$$
\begin{aligned}
R_{X,Y} &= \sum g(R_{X,Y}, \Xi_\alpha)\, \Xi_\alpha \\
&= \sum g(\mathfrak{R}(X \wedge Y), \Xi_\alpha)\, \Xi_\alpha \\
&= \sum g(\mathfrak{R}(\Xi_\alpha), X \wedge Y)\, \Xi_\alpha \\
&= -\sum g(R(\Xi_\alpha) X, Y)\, \Xi_\alpha .
\end{aligned}
$$

This allows us to rewrite the Weitzenböck curvature operator.

Lemma 9.3.3. *For any* $(0,k)$*-tensor* T

$$
\begin{aligned}
\mathrm{Ric}(T) &= -\sum R(\Xi_\alpha)(\Xi_\alpha T), \\
\Delta_L T &= \nabla^* \nabla T - c \sum R(\Xi_\alpha)(\Xi_\alpha T).
\end{aligned}
$$

Moreover, Ric *is self-adjoint.*

Proof. This is a straightforward calculation:

$$
\begin{aligned}
\mathrm{Ric}(T)(X_1, \ldots, X_k) &= \sum \left(R\left(e_j, X_i\right) T\right)\left(X_1, \ldots, e_j, \ldots, X_k\right) \\
&= -\sum g\left(R(\Xi_\alpha) e_j, X_i\right)(\Xi_\alpha T)\left(X_1, \ldots, e_j, \ldots, X_k\right) \\
&= -\sum (\Xi_\alpha T)\left(X_1, \ldots, g\left(R(\Xi_\alpha) e_j, X_i\right) e_j, \ldots, X_k\right) \\
&= \sum (\Xi_\alpha T)(X_1, \ldots, R(\Xi_\alpha) X_i, \ldots, X_k) \\
&= -\sum (R(\Xi_\alpha)(\Xi_\alpha T))(X_1, \ldots, X_i, \ldots, X_k).
\end{aligned}
$$

To check that Ric is self-adjoint select an orthonormal basis Ξ_α of eigenvectors for $\mathfrak{R}$, i.e., $\mathfrak{R}(\Xi_\alpha) = \lambda_\alpha \Xi_\alpha$. In this case,

$$\begin{aligned} g(\text{Ric}(T), S) &= -\sum g(R(\Xi_\alpha)(\Xi_\alpha T), S) \\ &= -\sum \lambda_\alpha g(\Xi_\alpha(\Xi_\alpha T), S) \\ &= \sum \lambda_\alpha g(\Xi_\alpha T, \Xi_\alpha S) \end{aligned}$$

which is symmetric in T and S. □

At first sight we have replaced a simple sum over j and i with a possibly more complicated sum. The next result justifies the reformulation.

Corollary 9.3.4. *If $\mathfrak{R} \geq 0$, then $g(\text{Ric}(T), T) \geq 0$. More generally, If $\mathfrak{R} \geq k$, where $k < 0$, then $g(\text{Ric}(T), T) \geq kC|T|^2$, where C depends only on the type of the tensor.*

Proof. As above assume $\mathfrak{R}(\Xi_\alpha) = \lambda_\alpha \Xi_\alpha$ and note that

$$g(\text{Ric}(T), T) = \sum \lambda_\alpha |\Xi_\alpha T|^2 \geq k \sum |\Xi_\alpha T|^2.$$

This shows that the curvature term is nonnegative when $k = 0$. Clearly there is a constant $C > 0$ depending only on the type of the tensor and dimension of the manifold so that

$$C|T|^2 \geq \sum |\Xi_\alpha T|^2.$$

When $k < 0$ this implies:

$$g(\text{Ric}(T), T) \geq kC|T|^2.$$ □

This allows us to obtain vanishing and estimation results for all Lichnerowicz Laplacians on manifolds.

Theorem 9.3.5. *If $\mathfrak{R} \geq k$ and* diam $\leq D$*, then the dimension of*

$$V = \{T \in \Gamma(E) \mid \Delta_L T = \nabla^*\nabla T + c\text{Ric}(T) = 0\}$$

is bounded by

$$m \cdot \exp\left(D \cdot C\left(n, kD^2\right) \frac{\sqrt{-kcCv}}{\sqrt{v}-1}\right),$$

and when $k = 0$ all $T \in V$ are parallel tensors.

9.4 The Bochner Technique in General

The goal in this section is to show that there are several natural Lichnerowicz Laplacians on Riemannian manifolds.

9.4.1 Forms

The first obvious case is that of the Hodge Laplacian on k-forms as we already know that harmonic forms compute the topology of the underlying manifold.

Theorem 9.4.1 (Weitzenböck, 1923). *The Hodge Laplacian is the Lichnerowicz Laplacian with $c = 1$. Specifically,*

$$\triangle\omega = (d\delta + \delta d)(\omega) = \nabla^*\nabla\omega + \mathrm{Ric}(\omega).$$

Proof. We shall follow the proof discovered by W. A. Poor. To perform the calculations we need

$$\begin{aligned}\delta\omega(X_2,\dots,X_k) &= -\sum(\nabla_{E_i}\omega)(E_i,X_2,\dots,X_k),\\ d\omega(X_0,\dots,X_k) &= \sum(-1)^i(\nabla_{X_i}\omega)\left(X_0,\dots,\hat{X}_i,\dots,X_k\right).\end{aligned}$$

We this in mind we get

$$\begin{aligned}d\delta\omega(X_1,\dots,X_k) &= -\sum(-1)^{i+1}\left(\nabla^2_{X_i,E_j}\omega\right)\left(E_j,X_1,\dots,\hat{X}_i,\dots,X_k\right)\\ &= -\sum\left(\nabla^2_{X_i,E_j}\omega\right)\left(X_1,\dots,E_j,\dots,X_k\right),\end{aligned}$$

$$\begin{aligned}\delta d\omega(X_1,\dots,X_k) &= -\sum\left(\nabla^2_{E_j,E_j}\omega\right)(X_1,\dots,X_k)\\ &\quad-\sum(-1)^i\left(\nabla^2_{E_j,X_i}\omega\right)\left(E_j,X_1,\dots,\hat{X}_i,\dots,X_k\right)\\ &= \left(\nabla^*\nabla\omega\right)(X_1,\dots,X_k)\\ &\quad+\sum\left(\nabla^2_{E_j,X_i}\omega\right)\left(X_1,\dots,E_j,\dots,X_k\right).\end{aligned}$$

Thus

$$\begin{aligned}\triangle\omega &= \nabla^*\nabla\omega + \sum\left(R\left(E_j,X_i\right)\omega\right)\left(X_1,\dots,E_j,\dots,X_k\right)\\ &= \nabla^*\nabla\omega + \mathrm{Ric}(\omega).\end{aligned}$$

□

9.4.2 The Curvature Tensor

We show that a suitably defined Laplacian on curvature tensors is in fact a Lichnerowicz Laplacian. This Laplacian is a symmetrized version of $(\nabla_X(\nabla^* R))(Y,Z,W)$ so as to make it have the same symmetries as R. It appears as the right-hand side in the formula below.

Theorem 9.4.2. *The curvature tensor R on a Riemannian manifold satisfies*

$$\begin{aligned}
&\left(\nabla^*\nabla R\right)(X,Y,Z,W) + \tfrac{1}{2}\mathrm{Ric}(R)(X,Y,Z,W)\\
&= \tfrac{1}{2}\left(\nabla_X\nabla^* R\right)(Y,Z,W) - \tfrac{1}{2}\left(\nabla_Y\nabla^* R\right)(X,Z,W)\\
&\quad + \tfrac{1}{2}\left(\nabla_Z\nabla^* R\right)(W,X,Y) - \tfrac{1}{2}\left(\nabla_W\nabla^* R\right)(Z,X,Y).
\end{aligned}$$

Proof. By far the most important ingredient in the proof is that we have the second Bianchi identity at our disposal. We will begin the calculation by considering the (0,4)-curvature tensor R. Fix a point p, let X,Y,Z,W be vector fields with $\nabla X = \nabla Y = \nabla Z = \nabla W = 0$ at p and let E_i be a normal frame at p. Then

$$\begin{aligned}
&\left(\nabla^*\nabla R\right)(X,Y,Z,W)\\
&= -\sum_{i=1}^{n}\left(\nabla^2_{E_i,E_i}R\right)(X,Y,Z,W)\\
&= \sum_{i=1}^{n}\left(\nabla^2_{E_i,X}R\right)(Y,E_i,Z,W) + \left(\nabla^2_{E_i,Y}R\right)(E_i,X,Z,W)\\
&= \sum_{i=1}^{n}\left(\nabla^2_{X,E_i}R\right)(Y,E_i,Z,W) + \left(\nabla^2_{Y,E_i}R\right)(E_i,X,Z,W)\\
&\quad + \sum_{i=1}^{n}\left(R(E_i,X)(R)\right)(Y,E_i,Z,W) + \left(R(E_i,Y)(R)\right)(E_i,X,Z,W)\\
&= \left(\nabla_X\nabla^* R\right)(Y,Z,W) - \left(\nabla_Y\nabla^* R\right)(X,Z,W)\\
&\quad - \sum_{i=1}^{n}\left(R(E_i,X)(R)\right)(E_i,Y,Z,W) + \left(R(E_i,Y)(R)\right)(X,E_i,Z,W).
\end{aligned}$$

Note that the last two terms are half of the expected terms in $-\mathrm{Ric}(R)(X,Y,Z,W)$.

Using that R is symmetric in the pairs X,Y and Z,W we then obtain

$$\begin{aligned}
&\left(\nabla^*\nabla R\right)(X,Y,Z,W)\\
&= \tfrac{1}{2}\left(\nabla^*\nabla R\right)(X,Y,Z,W) + \tfrac{1}{2}\left(\nabla^*\nabla R\right)(Z,W,X,Y)
\end{aligned}$$

$$
\begin{aligned}
&= \tfrac{1}{2}\left(\left(\nabla_X \nabla^* R\right)(Y,Z,W) - \left(\nabla_Y \nabla^* R\right)(X,Z,W)\right) \\
&\quad + \tfrac{1}{2}\left(\left(\nabla_Z \nabla^* R\right)(W,X,Y) - \left(\nabla_W \nabla^* R\right)(Z,X,Y)\right) \\
&\quad - \tfrac{1}{2}\sum_{i=1}^{n} \left(R\left(E_i,X\right)(R)\right)(E_i,Y,Z,W) + \left(R\left(E_i,Y\right)(R)\right)(X,E_i,Z,W) \\
&\quad - \tfrac{1}{2}\sum_{i=1}^{n} \left(R\left(E_i,Z\right)(R)\right)(E_i,W,X,Y) + \left(R\left(E_i,W\right)(R)\right)(Z,E_i,X,Y) \\
&= \tfrac{1}{2}\left(\left(\nabla_X \nabla^* R\right)(Y,Z,W) - \left(\nabla_Y \nabla^* R\right)(X,Z,W)\right) \\
&\quad + \tfrac{1}{2}\left(\left(\nabla_Z \nabla^* R\right)(W,X,Y) - \left(\nabla_W \nabla^* R\right)(Z,X,Y)\right) \\
&\quad - \tfrac{1}{2}\mathrm{Ric}\,(R)\,(X,Y,Z,W)\,. \qquad \square
\end{aligned}
$$

One might expect that, as with the Hodge Laplacian, there should also be terms where one takes the divergence of certain derivatives of R. However, the second Bianchi identity shows that these terms already vanish for R. In particular, R is harmonic if it is divergence free: $\nabla^* R = 0$.

9.4.3 *Symmetric* $(0,2)$*-Tensors*

Let h be a symmetric $(0,2)$-tensor. If we consider the corresponding $(1,1)$-tensor H, then we have defined $\left(d^{\nabla} H\right)(X,Y) = (\nabla_X H)(Y) - (\nabla_Y H)(X)$. Changing the type back allows us to define

$$
d^{\nabla} h\,(X,Y,Z) = (\nabla_X h)(Y,Z) - (\nabla_Y h)(X,Z)\,.
$$

In this form the definition is a bit mysterious but it does occur naturally in differential geometry. Originally it comes from considering the second fundamental II for an immersed hypersurface $M^n \to \mathbb{R}^{n+1}$. In this case the Codazzi-Mainardi equations can be expressed as $d^{\nabla}\mathrm{II} = 0$. Another natural situation is the Ricci tensor where exercise 3.4.8 shows that

$$
\left(d^{\nabla}\mathrm{Ric}\right)(X,Y,Z) = \left(\nabla^* R\right)(Z,X,Y)\,.
$$

This formula also has a counter part relating Schouten and Weyl tensors discussed in exercise 3.4.26.

Using this exterior derivative we obtain a formula that is similar to what we saw for forms and the curvature tensor.

Theorem 9.4.3. *Any symmetric* $(0,2)$*-tensor* h *on a Riemannian manifold satisfies*

$$\left(\nabla_X \nabla^* h\right)(X) + \left(\nabla^* d^\nabla h\right)(X,X) = \left(\nabla^* \nabla h\right)(X,X) + \tfrac{1}{2}\left(\mathrm{Ric}\,(h)\right)(X,X).$$

Proof. Observe that on the left-hand side the terms are

$$\left(\nabla_X \nabla^* h\right)(X) = -\left(\nabla^2_{X,E_i} h\right)(E_i, X)$$

and

$$\begin{aligned}\left(\nabla^* d^\nabla h\right)(X,X) &= -\left(\nabla_{E_i} d^\nabla h\right)(E_i, X, X)\\ &= -\left(\nabla^2_{E_i,E_i} h\right)(X,X) + \left(\nabla^2_{E_i,X} h\right)(E_i, X).\end{aligned}$$

Adding these we obtain

$$\begin{aligned}&\left(\nabla_X \nabla^* h\right)(X) + \left(\nabla^* d^\nabla h\right)(X,X)\\ &= \left(\nabla^* \nabla h\right)(X,X) + \left(\nabla^2_{E_i,X} h\right)(E_i,X) - \left(\nabla^2_{X,E_i} h\right)(E_i,X)\\ &= \left(\nabla^* \nabla h\right)(X,X) + \left(R\,(E_i,X)\,h\right)(E_i,X).\end{aligned}$$

Using that h is symmetric we finally conclude that

$$\left(R\,(E_i,X)\,h\right)(E_i,X) = \tfrac{1}{2}\left(\mathrm{Ric}\,(h)\right)(X,X),$$

thus finishing the proof. □

A symmetric $(0,2)$-tensor is called a *Codazzi tensor* if $d^\nabla h$ vanishes and harmonic if in addition it is divergence free. This characterization can be simplified slightly.

Proposition 9.4.4. *A symmetric* $(0,2)$*-tensor is harmonic if and only if it is a Codazzi tensor with constant trace.*

Proof. In general we have that

$$\begin{aligned}\left(\nabla^* h\right)(X) &= -\left(\nabla_{E_i} h\right)(E_i, X)\\ &= -\left(\nabla_{E_i} h\right)(X, E_i)\\ &= -\left(\nabla_X h\right)(E_i, E_i) + \left(d^\nabla h\right)(X, E_i, E_i)\\ &= -D_X\,(\mathrm{tr}\,h) + \left(d^\nabla h\right)(X, E_i, E_i).\end{aligned}$$

Thus Codazzi tensors are divergence free if and only if their trace is constant. □

This shows that hypersurfaces with constant mean curvature have harmonic second fundamental form. This fact has been exploited by both Lichnerowicz and

Simons. For the Ricci tensor to be harmonic it suffices to assume that it is Codazzi, but this in turn is a strong condition as it is the same as saying that the full curvature tensor is harmonic.

Corollary 9.4.5. *The Ricci tensor is harmonic if and only if the curvature tensor is harmonic.*

Proof. We know that the Ricci tensor is a Codazzi tensor precisely when the curvature tensor has vanishing divergence (see exercise 3.4.8). The contracted Bianchi identity (proposition 3.1.5) together with the proof of the above proposition then tells us

$$\begin{aligned} 2D_X\mathrm{scal} &= -\left(\nabla^*\mathrm{Ric}\right)(X) \\ &= D_X\left(\mathrm{trRic}\right) \\ &= D_X\left(\mathrm{scal}\right). \end{aligned}$$

Thus the scalar curvature must be constant and the Ricci tensor divergence free. □

9.4.4 Topological and Geometric Consequences

Theorem 9.4.6 (D. Meyer, 1971, D. Meyer-Gallot, 1975, and Gallot, 1981). *Let (M,g) be a closed Riemannian n-manifold. If the curvature operator is nonnegative, then all harmonic forms are parallel. When the curvature operator is positive the only parallel l-forms have $l=0,n$. Finally when $\mathfrak{R}\geq k$ and* diam $\leq D$,

$$b_l(M)\leq\binom{n}{l}\exp\left(D\cdot C\left(n,kD^2\right)\sqrt{-kC}\right).$$

Proof. The first statement is immediate given the Weitzenböck formula for forms. For the second part we note that when the curvature operator is positive, then the formula

$$0=g\left(\mathrm{Ric}\left(\omega\right),\omega\right)=\sum\lambda_\alpha\left|\Xi_\alpha\omega\right|^2$$

shows that $\Xi_\alpha\omega=0$ for all α. Hence by linearity $L\omega=0$ for all skew-symmetric L. If we assume $m<n$ and select L so that $L(e_i)=0$ for $i<m$, $L(e_m)=e_{m+1}$, then

$$0=(L\omega)(e_1,\ldots,e_m)=-\omega(e_1,\ldots,e_{m-1},e_{m+1}).$$

Since the basis was arbitrary this shows that $\omega=0$.

The last part follows from our general estimate from theorem 9.3.5. □

We now have a pretty good understanding of manifolds with nonnegative (or positive) curvature operator.

H. Hopf is, among other things, famous for the following 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 metric with positive curvature operator. It is also interesting to observe that $\mathbb{CP}^2$ has positive sectional curvature but doesn't admit a metric with positive curvature operator either. Thus, even among 4-manifolds, there seems to be a big difference between simply connected manifolds that admit $\text{Ric} > 0$, $\sec > 0$, and $\mathfrak{R} > 0$. We shall in chapter 12 describe a simply connected manifold that has $\text{Ric} > 0$ but doesn't even admit a metric with $\sec \geq 0$.

Manifolds with nonnegative curvature operator can in fact be classified (see theorem 10.3.7). 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 9.4.7. We can exhibit a metric with nonnegative sectional curvature on $\mathbb{CP}^2 \sharp \overline{\mathbb{C}}\mathbb{P}^2$ by observing that it is an S^1 quotient of $S^2 \times S^3$. Namely, let S^1 act on the 3-sphere by the Hopf action and on the 2-sphere by rotations. If the total rotation on the 2-sphere is $2\pi k$, then the quotient is $S^2 \times S^2$ if k is even, and $\mathbb{CP}^2 \sharp \overline{\mathbb{C}}\mathbb{P}^2$ if k is odd. In all cases O'Neill's formula tells us that the sectional curvature is nonnegative. 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 $\mathbb{CP}^2$. These examples were first discovered by Cheeger but with a very different construction that also lead to other examples.

The Bochner technique has found many generalizations. It has, for instance, proven very successful in the study of manifolds with nonnegative scalar curvature. Briefly, what happens is that *spin manifolds* admit certain *spinor bundles*. These bundles come with a natural first-order operator called the *Dirac operator*. The square of this operator has a Weitzenböck formula of the form

$$\nabla^* \nabla + \tfrac{1}{4}\text{scal}.$$

This formula was discovered and used by Lichnerowicz (as well as I. Singer, as pointed out in [107]) to show that a sophisticated invariant called the $\hat{A}$-genus vanishes for spin manifolds with positive scalar curvature. Using some generalizations of this formula, Gromov-Lawson showed that any metric on a torus with $\text{scal} \geq 0$ is in fact flat. We just proved this for metrics with $\text{Ric} \geq 0$. Dirac operators and their Weitzenböck formulas 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, the work of Seiberg-Witten, which has had a revolutionary impact on 4-manifolds, 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 non-simply connected manifolds with positive scalar curvature that don't admit even nonnegative Ricci curvature. Specifically, 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.

Finally, we present a more geometric result for the curvature tensor. It was first established in [98], and then with a modified proof in [48]. The proof is quite simple and based on the generalities developed above. In chapter 10 we will also show that this result basically characterizes compact symmetric spaces as they all have nonnegative curvature operator.

Theorem 9.4.8 (Tachibana, 1974). *Let (M, g) be a closed Riemannian manifold. If the curvature operator is nonnegative and $\nabla^* R = 0$, then $\nabla R = 0$. If in addition the curvature operator is positive, then (M, g) has constant curvature.*

Proof. We know from above that

$$\nabla^* \nabla R + \tfrac{1}{2} \mathrm{Ric}\,(R) = 0.$$

So if the curvature operator is nonnegative, then $\nabla R = 0$.

Moreover, when the curvature operator is positive it follows as in the case of forms, that $LR = 0$ for all $L \in \mathfrak{so}\left(T_p M\right)$. This condition implies, as we shall show below, that $R(x, y, y, z) = 0$ and $R(x, y, v, w) = 0$ when the vectors are perpendicular. This in turn shows that any bivector $x \wedge y$ is an eigenvector for $\mathfrak{R}$, but this can only happen if $\mathfrak{R} = kI$ for some constant k.

To show that the mixed curvatures vanish first select L so that $L(y) = 0$ and $L(x) = z$, then

$$0 = LR(x, y, y, x) = -R(L(x), y, y, x) - R(x, y, y, L(x)) = -2R(x, y, y, z).$$

Polarizing in $y = v + w$, then shows that

$$R(x, v, w, z) = -R(x, w, v, z).$$

The Bianchi identity then implies

$$\begin{aligned} R(x, v, w, z) &= R(w, v, x, z) - R(w, x, v, z) \\ &= -2R(w, x, v, z) \\ &= 2R(x, w, v, z) \\ &= -2R(x, v, w, z) \end{aligned}$$

showing that $R(x, v, w, z) = 0$. □

9.4.5 *Simplification of* $g(\mathrm{Ric}(T), T)$

Finally we mention an alternate method that recovers the formula for 1-forms and also gives a formula for general $(0,2)$-tensors.

Having redefined the Weitzenböck curvature of tensors, we take it a step further and also discard the orthonormal basis Ξ_α. To assist in this note that a $(0,k)$-tensor T can be changed to a tensor $\hat{T}$ with values in $\Lambda^2 TM$. Implicitly this works as follows

$$g\left(L, \hat{T}(X_1,\ldots,X_k)\right) = (LT)(X_1,\ldots,X_k) \text{ for all } L \in \mathfrak{so}(TM) = \Lambda^2 TM.$$

Lemma 9.4.9. *For all* $(0,k)$*-tensors* T *and* S

$$g(\mathrm{Ric}(T), S) = g\left(\mathfrak{R}\left(\hat{T}\right), \hat{S}\right).$$

Proof. This is a straight forward calculation

$$\begin{aligned}
g(\mathrm{Ric}(T), S) &= \sum g(\Xi_\alpha T, R(\Xi_\alpha) S) \\
&= \sum (\Xi_\alpha T)(e_{i_1},\ldots,e_{i_k})(R(\Xi_\alpha) S)(e_{i_1},\ldots,e_{i_k}) \\
&= \sum g\left(\Xi_\alpha, \hat{T}(e_{i_1},\ldots,e_{i_k})\right) g\left(R(\Xi_\alpha), \hat{S}(e_{i_1},\ldots,e_{i_k})\right) \\
&= \sum g\left(\mathfrak{R}\left(g\left(\Xi_\alpha, \hat{T}(e_{i_1},\ldots,e_{i_k})\right)\Xi_\alpha\right), \hat{S}(e_{i_1},\ldots,e_{i_k})\right) \\
&= \sum g\left(\mathfrak{R}\left(\hat{T}(e_{i_1},\ldots,e_{i_k})\right), \hat{S}(e_{i_1},\ldots,e_{i_k})\right) \\
&= g\left(\mathfrak{R}\left(\hat{T}\right), \hat{S}\right)
\end{aligned}$$

This shows again that Ric is self-adjoint as $\mathfrak{R}$ is self-adjoint on $\Lambda^2 TM$. □

This new expression for $g(\mathrm{Ric}(T), T)$ is also clearly nonnegative when the curvature operator is nonnegative. In addition, it also occasionally allows us to show that it is nonnegative under less restrictive hypotheses.

Proposition 9.4.10. *If* ω *is a 1-form and* X *the dual vector field, then*

$$\hat{\omega}(Z) = X \wedge Z$$

and

$$g(\mathfrak{R}(\hat{\omega}), \hat{\omega}) = \mathrm{Ric}(X, X).$$

Proof. In this case

$$\begin{aligned}(L\omega)(Z) &= -\omega(L(Z))\\ &= -g(X, L(Z))\\ &= -g(L, Z \wedge X)\end{aligned}$$

so

$$\hat{\omega}(Z) = X \wedge Z.$$

This shows that the curvature term in the Bochner formula becomes

$$\begin{aligned}-\sum g(R(\Xi_\alpha)(\Xi_\alpha\omega), \omega) &= \sum g(\Xi_\alpha\omega, R(\Xi_\alpha)\omega)\\ &= \sum g(\hat{\omega}(E_i), \mathfrak{R}(\hat{\omega}(E_i)))\\ &= \sum g(\mathfrak{R}(E_i \wedge X), E_i \wedge X)\\ &= \sum R(X, E_i, E_i, X)\\ &= \operatorname{Ric}(X, X).\end{aligned}$$

□

More generally, one can show that if ω is a p-form and

$$g\left(\Omega\left(X_1, \ldots, X_{p-1}\right), X_p\right) = \omega\left(X_1, \ldots, X_p\right),$$

then

$$\hat{\omega}\left(X_1, \ldots, X_p\right) = \sum_{i=1}^{p} (-1)^{p-i} X_i \wedge \Omega\left(X_1, \ldots, \hat{X}_i, \ldots, X_p\right).$$

Moreover, note that $\hat{\omega}$ can only vanish when ω vanishes.

Next we focus on understanding $\operatorname{Ric}\left(\hat{h}\right)$ for $(0,2)$-tensors. Given a $(0,2)$-tensor h there is a corresponding $(1,1)$-tensor called H

$$h(z, w) = g(H(z), w).$$

The adjoint of H is denoted H^*.

Proposition 9.4.11. *With that notation*

$$\hat{h}(z, w) = H(z) \wedge w - z \wedge H^*(w)$$

and $\hat{h} = 0$ *if and only if* $h = \lambda g$.

Proof. We start by observing that

$$\begin{aligned}
(Lh)(z,w) &= -h(L(z),w) - h(z,L(w)) \\
&= -g(H(L(z)),w) - g(H(z),L(w)) \\
&= -g\left(L(z),H^*(w)\right) - g(L(w),H(z)) \\
&= -g\left(L,z\wedge H^*(w)\right) - g(L,w\wedge H(z)) \\
&= g\left(L,H(z)\wedge w - z\wedge H^*(w)\right).
\end{aligned}$$

Note that if $h = \lambda g$ then $H = \lambda I = H^*$, thus $\hat{h} = 0$. Next assume that $\hat{h} = 0$. Then for all z, w we have

$$\begin{aligned}
z\wedge H^*(H(w)) &= H(z)\wedge H(w) \\
&= -H(w)\wedge H(z) \\
&= -w\wedge H^*(H(z)) \\
&= H^*(H(z))\wedge w.
\end{aligned}$$

But that can only be true if $H^*H = \lambda^2 I$ and $H = \lambda I$. $\square$

This indicates that we have to control curvatures of the type

$$g\left(\mathfrak{R}\left(H(z)\wedge w - z\wedge H^*(w)\right), \overline{H(z)\wedge w - z\wedge H^*(w)}\right).$$

If H is normal, then it can be diagonalized with respect to an orthonormal basis in the complexified tangent bundle. Assuming that $H(z) = \lambda z$ and $H(w) = \mu w$ where $z, w \in T_pM \otimes \mathbb{C}$ are orthonormal we obtain

$$\left(\mathfrak{R}\left(H(z)\wedge w - z\wedge H^*(w)\right), \overline{H(z)\wedge w - z\wedge H^*(w)}\right) = |\lambda - \bar{\mu}|^2 g\left(\mathfrak{R}(z\wedge w), \overline{z\wedge w}\right).$$

The curvature term $g\left(\mathfrak{R}(z\wedge w), \overline{z\wedge w}\right)$ looks like a complexified sectional curvature and is in fact called the *complex sectional curvature*. It can be recalculated without reference to the complexification. If we consider $z = x + \mathrm{i}y$ and $w = u + \mathrm{i}v$, $x, y, u, v \in TM$, then

$$\begin{aligned}
g(\mathfrak{R}(z\wedge w), \bar{z}\wedge\bar{w}) &= g(\mathfrak{R}(x\wedge u - y\wedge v), x\wedge u - y\wedge v) \\
&\quad + g(\mathfrak{R}(x\wedge v + y\wedge u), x\wedge v + y\wedge u) \\
&= g(\mathfrak{R}(x\wedge u), x\wedge u) + g(\mathfrak{R}(y\wedge v), y\wedge v) \\
&\quad + g(\mathfrak{R}(x\wedge v), x\wedge v) + g(\mathfrak{R}(y\wedge u), y\wedge u) \\
&\quad - 2g(\mathfrak{R}(x\wedge u), y\wedge v) + 2g(\mathfrak{R}(x\wedge v), y\wedge u)
\end{aligned}$$

$$
\begin{aligned}
&= R(x,u,u,x) + R(y,v,v,y) + R(x,v,v,x) + R(y,u,u,y) \\
&\quad +2R(x,u,y,v) - 2R(x,v,y,u) \\
&= R(x,u,u,x) + R(y,v,v,y) + R(x,v,v,x) + R(y,u,u,y) \\
&\quad -2\left(R(v,y,x,u) + R(x,v,y,u)\right) \\
&= R(x,u,u,x) + R(y,v,v,y) + R(x,v,v,x) + R(y,u,u,y) \\
&\quad +2R(y,x,v,u) \\
&= R(x,u,u,x) + R(y,v,v,y) + R(x,v,v,x) + R(y,u,u,y) \\
&\quad +2R(x,y,u,v).
\end{aligned}
$$

The first line in this derivation shows that complex sectional curvatures are nonnegative when $\mathfrak{R} \geq 0$. Thus we see that it is weaker than working with the curvature operator. On the other hand it is stronger than sectional curvature.

There are three special cases depending on the dimension of $\mathrm{span}_{\mathbb{R}}\{x,y,u,v\}$. When $y = v = 0$ we obtain the standard definition of sectional curvature. When x, y, u, v are orthonormal we obtain the so called *isotropic curvature*, and finally if $u = v$ we get a sum of two sectional curvatures

$$2R(x,u,u,x) + 2R(y,u,u,y)$$

also called a *second Ricci curvature* when x, y, u are orthonormal.

The next result is a general version of two separate theorems. Simons and Berger did the case of symmetric tensors and Micallef-Wang the case of 2-forms.

Proposition 9.4.12. *Let h be a $(0,2)$-tensor such that H is normal. If the complex sectional curvatures are nonnegative, then $g\left(\mathfrak{R}\left(\hat{h}\right), \hat{h}\right) \geq 0$.*

Proof. We can use complex orthonormal bases as well as real bases to compute $g\left(\mathfrak{R}\left(\hat{h}\right), \hat{h}\right)$. Using that H is normal we obtain a complex orthonormal basis e_i of eigenvectors $H(e_i) = \lambda_i e_i$ and $H^*(e_i) = \bar{\lambda}_i e_i$. From that we quickly obtain

$$
\begin{aligned}
g\left(\mathfrak{R}\left(\hat{h}\right), \hat{h}\right) &= \sum g\left(\mathfrak{R}\left(\hat{h}(e_i,e_j)\right), \overline{\hat{h}(e_i,e_j)}\right) \\
&= \sum g\left(\mathfrak{R}\left(H(e_i)\wedge e_j - e_i \wedge H^*(e_j)\right), \overline{H(e_i)\wedge e_j - e_i \wedge H^*(e_j)}\right) \\
&= \sum \left|\lambda_i - \bar{\lambda}_j\right|^2 g\left(\mathfrak{R}(e_i\wedge e_j), \overline{e_i \wedge e_j}\right). \qquad \square
\end{aligned}
$$

In the special case where H is self-adjoint the eigenvalues/vectors are real and we need only use the real sectional curvatures. When H is skew-adjoint the eigenvectors are purely imaginary unless they correspond to zero eigenvalues. This shows that we must use the isotropic curvatures and also the second Ricci curvatures when M is odd dimensional. However, in this case none of the terms involve real sectional curvatures.

These characterizations can be combined to show

Proposition 9.4.13. $g\left(\mathfrak{R}\left(\hat{h}\right),\hat{h}\right)\geq 0$ *for all* $(0,2)$*-tensors on* T_pM *if all complex sectional curvatures on* T_pM *are nonnegative.*

Proof. We decompose $h=h_s+h_a$ into symmetric and skew symmetric parts. Then

$$\begin{aligned} g\left(\mathfrak{R}\left(\hat{h}\right),\hat{h}\right) &= g\left(\mathfrak{R}\left(\hat{h}_s\right),\hat{h}_s\right)+g\left(\mathfrak{R}\left(\hat{h}_a\right),\hat{h}_a\right)+g\left(\mathfrak{R}\left(\hat{h}_s\right),\hat{h}_a\right)\\ &\quad +g\left(\mathfrak{R}\left(\hat{h}_a\right),\hat{h}_s\right)\\ &= g\left(\mathfrak{R}\left(\hat{h}_s\right),\hat{h}_s\right)+g\left(\mathfrak{R}\left(\hat{h}_a\right),\hat{h}_a\right)+2g\left(\mathfrak{R}\left(\hat{h}_s\right),\hat{h}_a\right). \end{aligned}$$

However,

$$\begin{aligned} g\left(\mathfrak{R}\left(\hat{h}_s\right),\hat{h}_a\right) &= \sum g\left(\mathfrak{R}\left(\hat{h}_s\left(e_i,e_j\right)\right),\hat{h}_a\left(e_i,e_j\right)\right)\\ &= -\sum g\left(\mathfrak{R}\left(\hat{h}_s\left(e_j,e_i\right)\right),\hat{h}_a\left(e_j,e_i\right)\right)\\ &= -g\left(\mathfrak{R}\left(\hat{h}_s\right),\hat{h}_a\right). \end{aligned}$$

So

$$g\left(\mathfrak{R}\left(\hat{h}\right),\hat{h}\right)=g\left(\mathfrak{R}\left(\hat{h}_s\right),\hat{h}_s\right)+g\left(\mathfrak{R}\left(\hat{h}_a\right),\hat{h}_a\right)$$

and the result follows from the previous proposition. □

9.5 Further Study

For more general and complete accounts of the Bochner technique and spin geometry we recommend the two texts [107] and [71]. The latter book also has a complete proof of the Hodge theorem. Other sources for this particular result are [65], [92], and [101].

For other generalizations to manifolds with integral curvature bounds the reader should consult [46]. In there the reader will find a complete discussion on generalizations of the above mentioned results about Betti numbers.

9.6 Exercises

EXERCISE 9.6.1. Suppose (M^n, g) is compact and has $b_1 = k$. If $\mathrm{Ric} \geq 0$, then the universal covering splits:

$$\left(\tilde{M}, g\right) = (N, h) \times \left(\mathbb{R}^k, g_{\mathbb{R}^n}\right).$$

Give an example where $b_1 < n$ and $\left(\tilde{M}, g\right) = (\mathbb{R}^n, g_{\mathbb{R}^n})$.

EXERCISE 9.6.2. Show directly that if E_i is an orthonormal frame and $V \mapsto \nabla_V X$ is symmetric, then $\sum_i g\left(\nabla_{E_i} X, \nabla_X E_i\right) = 0$ without assuming that E_i are parallel in the direction of X.

EXERCISE 9.6.3. Show that for an oriented Riemannian manifold: $\left(\mathrm{im} d^{k-1}\right)^{\perp} = \ker \delta^{k-1}$ and $\mathrm{im} d^{k-1} \subset \left(\ker \delta^{k-1}\right)^{\perp}$ in $\Omega^k(M)$.

EXERCISE 9.6.4. Let (M^n, g) be a compact Riemannian manifold and $E \to M$ a tensor bundle. Let $W^{k,2}(E)$ denote the Hilbert space completion of $\Gamma(E)$ with square norm $\sum_{i=0}^{k} \left\|\nabla^i T\right\|_2^2$, e.g., if $T \in W^{1,2}(E)$, then T is defined as an element in L^2 and its derivative ∇T as an L^2 tensor that satisfies $(\nabla T, \nabla S) = (T, \nabla^* \nabla S)$ for all $S \in \Gamma(E)$. It follows that $\Gamma(E) \subset W^{k,2}(E)$ is dense. The Sobolev inequality can be used to show that $\bigcap_{k\geq 0} W^{k,2}(E) = \bigcap_{k\geq 0} W^{k,p}(E)$ for all $p < \infty$. The techniques from section 7.1.5 can easily be adapted to show that a tensor $T \in W^{1,p}$ is Hölder continuous when $p > n$ (see also [60]). This in turn shows that $\Gamma(E) = \bigcap_{k\geq 0} W^{k,2}(E)$.

(1) Show that for all $T \in \Gamma(E)$ there is a commutation relationship

$$\nabla^* \nabla \left(\nabla^k T\right) - \nabla^k \left(\nabla^* \nabla T\right) = \sum_{i=0}^{k} C_i^k \left(\nabla^{k-i} R \otimes \nabla^i T\right),$$

where $C_i^k \left(\nabla^{k-i} R \otimes \nabla^i T\right)$ is a suitable contraction.

(2) Assume that $T \in W^{1,2}(E)$ and $T' \in W^{k,2}(E)$ satisfy $(T', S) = (\nabla T, \nabla S)$ for all $S \in \Gamma(E)$, i.e., $\nabla^* \nabla T = T'$ weakly, show that $T \in W^{k+1,2}(E)$. Hint: Define the weak derivatives $\nabla^{l+1} T$ inductively using a relationship of the form:

$$\left(\nabla^{l+1} T, \nabla^{l+1} S\right) = \left(\nabla^l T', \nabla^l S\right) + \sum_{i=0}^{l} \left(C_i^l \left(\nabla^{l-i} R \otimes \nabla^i T\right), \nabla^l S\right).$$

(3) Conclude that if $T' \in \Gamma(E)$, then $T \in \Gamma(E)$ and $\nabla^* \nabla T = T'$.

EXERCISE 9.6.5. Let (M^n, g) be a compact Riemannian manifold with $\mathrm{diam} M \leq D$, $\mathfrak{R} \geq k$, and $E \to M$ a tensor bundle with m-dimensional fibers and a Lichnerowicz Laplacian Δ_L. The goal is to establish *the spectral theorem* for Δ_L and as a consequence obtain the orthogonal decomposition $\Gamma(E) = \ker \Delta_L \oplus \mathrm{im} \Delta_L$.

(1) Consider the Hilbert space completion $W^{1,2}(E)$ of $\Gamma(E)$ as in exercise 9.6.4. Show that the right-hand side in $(\Delta_L T, S) = (\nabla T, \nabla S) + c(\mathrm{Ric}(T), S)$ is symmetric and well-defined for all $T, S \in W^{1,2}(E)$.
(2) Show that $\inf_{T\in W^{1,2}, \|T\|_2=1} (\Delta_L T, T) > cCk$, where $k \leq 0$ and C is the constant in corollary 9.3.4.
(3) Show that a sequence $T_i \subset W^{1,2}(E)$, where $\|T_i\|_2 = 1$ and $(\Delta_L T_i, T_i)$ is bounded, will have an L^2-convergent subsequence that is also weakly convergent in $W^{1,2}(E)$. Hint: Use theorem 7.1.18.
(4) Consider a closed subspace $V \subset W^{1,2}(E)$ that is invariant under Δ_L. Show that the infimum $\lambda = \inf_{T\in V, \|T\|=1} (\Delta_L T, T)$ is achieved by a $T \in V$, then use exercise 9.6.4 to show that $T \in \Gamma(E)$ and $\Delta_L T = \lambda T$. Hint: Prove and use that

$$\|T\|_2^2 + \|\nabla T\|_2^2 \leq \liminf_{i\to\infty} \left(\|T_i\|_2^2 + \|\nabla T_i\|_2^2\right)$$

if $T_i \rightharpoonup T$ (weak convergence) in $W^{1,2}(E)$.
(5) Consider a finite dimensional subspace $V \subset \Gamma(E)$ that is spanned by eigen-tensors T_i with $\Delta_L T_i = \lambda_i T_i$. Show that $\dim V \leq mC\left(n, \max_i \lambda_i, c, D^2 k\right)$.
(6) Show that all eigenspaces for Δ_L are finite dimensional and that the set of eigenvalues is discrete. Conclude that they can be ordered $\lambda_1 < \lambda_2 < \cdots$ with $\lim_{i\to\infty} \lambda_i = \infty$.
(7) Show that the eigenspaces for Δ_L are orthogonal and that their direct sum is dense in $\Gamma(E)$.
(8) Show that $\Gamma(E) = \ker \Delta_L \oplus \mathrm{im}\Delta_L$. Hint: Use exercise 9.6.4.

EXERCISE 9.6.6. 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 $\mathrm{Ric}_g \geq 0$, show that $M = \mathbb{R}^n$. Hint: Find a metric on the n-torus that is isometric to a neighborhood of $K \subset M$ somewhere and otherwise flat. 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 $\mathrm{Ric}_g \geq 0$, and use this to conclude that the manifold is flat.

EXERCISE 9.6.7. Let (M, g) be an Einstein metric. Show that all harmonic 1-forms are eigen-forms for the connection Laplacian $\nabla^*\nabla$.

EXERCISE 9.6.8. Given two vector fields X and Y on (M, g) such that ∇X and ∇Y are symmetric, develop Bochner formulas for $\mathrm{Hess}\frac{1}{2}g(X, Y)$ and $\Delta\frac{1}{2}g(X, Y)$.

EXERCISE 9.6.9. For general tensors s_1 and s_2 of the same type show in analogy with the formula

$$\Delta\frac{1}{2}|s|^2 = |\nabla s|^2 - g\left(\nabla^*\nabla s, s\right)$$

that:

$$\Delta g(s_1, s_2) = 2g(\nabla s_1, \nabla s_2) + g\left(\nabla^*\nabla s_1, s_2\right) + g\left(s_1, \nabla^*\nabla s_2\right).$$

Use this on forms to develop Bochner formulas for inner products of such sections.

More generally consider the 1-form defined by $\omega(v) = g(\nabla_v s_1, s_2)$ that represents half of the differential of $g(s_1, s_2)$. Show that

$$\begin{aligned} -\delta\omega &= g(\nabla s_1, \nabla s_2) - g\left(\nabla^*\nabla s_1, s_2\right) \\ d\omega(X, Y) &= g(R(X, Y)s_1, s_2) - g(\nabla_X s_1, \nabla_Y s_2) + g(\nabla_Y s_1, \nabla_X s_2). \end{aligned}$$

EXERCISE 9.6.10. Let (M, g) be n-dimensional.

(1) Show that

$$L\omega = 0$$

if L is skew-symmetric and ω is an n-form.

(2) When $n = 2$,

$$LR = 0$$

for all skew-symmetric L.

(3) For general L

$$L\mathrm{vol} = \mathrm{tr}(L)\,\mathrm{vol}.$$

EXERCISE 9.6.11 (Simons). Let (M, g) be a compact Riemannian manifold with a $(0, 2)$-tensor field h that is a symmetric Codazzi tensor with constant trace.

(1) Show that if $\sec \geq 0$, then $\nabla h = 0$.
(2) Moreover, if $\sec > 0$, then $h = c \cdot g$ for some constant c.
(3) If the Gauss equations

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

are satisfied and the trace of h vanishes, then

$$\Delta\frac{1}{2}|h|^2 \geq |\nabla h|^2 - |h|^4.$$

EXERCISE 9.6.12. Let $(M^n, g) \looparrowright \mathbb{R}^{n+1}$ be an isometric immersion of a manifold.

(1) Show that the second fundamental form II is a Codazzi tensor.
(2) Show Liebmann's theorem: If (M, g) has constant mean curvature and nonnegative second fundamental from, then (M, g) is a constant curvature sphere.

On the other hand, Wente has exhibited immersed tori with constant mean curvature (see Wente's article in [51]).

EXERCISE 9.6.13 (Berger). Show that a compact manifold with harmonic curvature and nonnegative sectional curvature has parallel Ricci curvature.

EXERCISE 9.6.14. Suppose we have a Killing field K on a closed Riemannian manifold (M, g) . Assume that ω is a harmonic form.

(1) Show that $L_K\omega = 0$. Hint: Show that $L_K\omega$ is also harmonic.
(2) Show that $i_K\omega$ is closed, but not necessarily harmonic.
(3) Show that all harmonic forms are invariant under $\mathrm{Iso}_0(M)$.
(4) Give an example where a harmonic form is not invariant under all of $\mathrm{Iso}(M)$.

EXERCISE 9.6.15. Let (M, g) be a closed Kähler manifold with Kähler form ω, i.e., a parallel nondegenerate 2-form. Show

$$\omega^k = \underbrace{\omega \wedge \cdots \wedge \omega}_{k \text{ times}}$$

is closed but not exact by showing that $\omega^{\frac{\dim M}{2}}$ is proportional to the volume form. Conclude that none of the even homology groups vanish.

EXERCISE 9.6.16. Let $E \to M$ be a tensor bundle.

(1) 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.
(2) Show that there is a natural wedge product

$$\Omega^p(M, \mathrm{Hom}(E, E)) \times \Omega^q(M, E) \to \Omega^{p+q}(M, E) .$$

(3) 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: $d^\nabla s = \nabla s$.
(4) Think of $R(X, Y)s \in \Omega^2(M, \mathrm{Hom}(E, E))$. 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 $d^\nabla R = 0$.

EXERCISE 9.6.17. If we let $E = TM$ in the previous exercise, then

$$\Omega^1 (M, TM) = \mathrm{Hom}(TM, TM)$$

will simply consist of all $(1, 1)$-tensors.

(1) Show that in this case $d^\nabla s = 0$ if and only if s is a Codazzi tensor.
(2) The entire chapter seems to indicate that whenever we have a tensor bundle E 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?
(3) Show that if X is a vector field, then ∇X is a Codazzi tensor if and only if $R(\cdot, \cdot) X = 0$. 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?

EXERCISE 9.6.18 (Thomas). Show that in dimensions $n > 3$ the Gauss equations $(\mathfrak{R} = S \wedge S)$ imply the Codazzi equations $\left(d^\nabla S = 0\right)$ 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

$$\mathrm{Hom}\left(\Lambda^2 V, V\right) \to \mathrm{Hom}\left(\Lambda^3 V, \Lambda^2 V\right),$$
$$T \mapsto T \wedge S$$

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.

Chapter 10
Symmetric Spaces and Holonomy

In this chapter we give an overview of (locally) symmetric spaces and holonomy. Most standard results are proved or at least mentioned. We give a few 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.

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. One person managed to take all the glory for classifying symmetric spaces; Elie Cartan. He started out in his thesis with cleaning up and correcting Killing's classification of simple complex Lie algebras and several years later 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 real simple Lie algebras. This in itself takes so much work that most books on Lie algebras give up after having settled the complex case.

After Cartan's work, a few people worked on getting a better conceptual understanding of some of these new geometries and also on offering a more global classification. Still, not much happened until the 1950s, when people realized a interesting connection between symmetric spaces and holonomy: The de Rham decomposition theorem and Berger's classification of holonomy groups. It then became clear that almost all holonomy groups occurred for symmetric spaces and consequently gave good approximating geometries to most holonomy groups.

P. Petersen, *Riemannian Geometry*, Graduate Texts in Mathematics 171,
DOI 10.1007/978-3-319-26654-1_10

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 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 more recently become better understood with D. Joyce's work.

10.1 Symmetric Spaces

There are many ways of representing symmetric spaces. Below we shall see how they can be described as homogeneous spaces, Lie algebras with involutions, or by their curvature tensor.

10.1.1 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 A_p such that $DA_p : T_pM \to T_pM$ is the antipodal map $-I$. Since isometries preserve geodesics, any geodesic $c(t)$ with $c(0) = p$ has the property that: $A_p \circ c(t) = c(-t)$. This quickly shows that symmetric spaces are homogeneous and hence complete. Specifically, 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.

A homogeneous space $\mathrm{G}/\mathrm{H} = \mathrm{Iso}/\mathrm{Iso}_p$ is symmetric provided that the symmetry A_p exists for just one p. In this case we can use $A_q = g \circ A_p \circ g^{-1}$, where g is an isometry that takes p to q. This means, in particular, that any Lie group G with biinvariant metric is a symmetric space, as $g \to g^{-1}$ is the desired symmetry around the identity element. Tables 10.1, 10.2, 10.3, 10.4 list some of the important families of homogeneous spaces that are symmetric. They always come in dual pairs of compact and noncompact spaces. There are many more families and several exceptional examples as well.

Table 10.1 Compact Groups

group	rank	dim
$\mathrm{SU}(n+1)$	n	$n(n+2)$
$\mathrm{SO}(2n+1)$	n	$n(2n+1)$
$\mathrm{Sp}(n)$	n	$n(2n+1)$
$\mathrm{SO}(2n)$	n	$n(2n-1)$

Table 10.2 Noncompact Analogues of Compact Groups

(complexified group)/group	rank	dim
$\mathrm{SL}(n+1,\mathbb{C})/\mathrm{SU}(n+1)$	n	$n(n+2)$
$\mathrm{SO}(2n+1,\mathbb{C})/\mathrm{SO}(2n+1)$	n	$n(2n+1)$
$\mathrm{Sp}(n,\mathbb{C})/\mathrm{Sp}(n)$	n	$n(2n+1)$
$\mathrm{SO}(2n,\mathbb{C})/\mathrm{SO}(2n)$	n	$n(2n-1)$

Table 10.3 Compact Homogeneous Spaces

Iso	Iso_p	dim	rank	description
$\mathrm{SO}(n+1)$	$\mathrm{SO}(n)$	n	1	Sphere
$\mathrm{O}(n+1)$	$\mathrm{O}(n)\times\{1,-1\}$	n	1	$\mathbb{RP}^n$
$\mathrm{U}(n+1)$	$\mathrm{U}(n)\times\mathrm{U}(1)$	$2n$	1	$\mathbb{CP}^n$
$\mathrm{Sp}(n+1)$	$\mathrm{Sp}(n)\times\mathrm{Sp}(1)$	$4n$	1	$\mathbb{HP}^n$
F_4	$\mathrm{Spin}(9)$	16	1	$\mathbb{OP}^2$
$\mathrm{SO}(p+q)$	$\mathrm{SO}(p)\times\mathrm{SO}(q)$	pq	$\min(p,q)$	Real Grassmannian
$\mathrm{SU}(p+q)$	$\mathrm{S}(\mathrm{U}(p)\times\mathrm{U}(q))$	$2pq$	$\min(p,q)$	Complex Grassmannian
$\mathrm{SU}(n)$	$\mathrm{SO}(n)$	$(n^2-1)/2$	$n-1$	$\mathbb{R}^n$'s in $\mathbb{C}^n$

Table 10.4 Noncompact Homogeneous Spaces

Iso	Iso_p	dim	rank	description
$\mathrm{SO}(n,1)$	$\mathrm{SO}(n)$	n	1	Hyperbolic space
$\mathrm{O}(n,1)$	$\mathrm{O}(n)\times\{1,-1\}$	n	1	Hyperbolic $\mathbb{RP}^n$
$\mathrm{U}(n,1)$	$\mathrm{U}(n)\times\mathrm{U}(1)$	$2n$	1	Hyperbolic $\mathbb{CP}^n$
$\mathrm{Sp}(n,1)$	$\mathrm{Sp}(n)\times\mathrm{Sp}(1)$	$4n$	1	Hyperbolic $\mathbb{HP}^n$
F_4^{-20}	$\mathrm{Spin}(9)$	16	1	Hyperbolic $\mathbb{OP}^2$
$\mathrm{SO}(p,q)$	$\mathrm{SO}(p)\times\mathrm{SO}(q)$	pq	$\min(p,q)$	Hyperbolic Grassmannian
$\mathrm{SU}(p,q)$	$\mathrm{S}(\mathrm{U}(p)\times\mathrm{U}(q))$	$2pq$	$\min(p,q)$	Complex hyperbolic Grassmannian
$\mathrm{SL}(n,\mathbb{R})$	$\mathrm{SO}(n)$	$(n^2-1)/2$	$n-1$	Euclidean structures on $\mathbb{R}^n$

Here $\mathrm{Spin}(n)$ is the universal double covering of $\mathrm{SO}(n)$ for $n>2$. We also have the following special identities in low dimensions:

$$\begin{aligned}\mathrm{SO}(2)&=\mathrm{U}(1)\,,\\ \mathrm{Spin}(3)&=\mathrm{SU}(2)=\mathrm{Sp}(1)\,,\\ \mathrm{Spin}(4)&=\mathrm{Spin}(3)\times\mathrm{Spin}(3)\,.\end{aligned}$$

Note that all of the compact examples have $\sec\geq 0$ by O'Neill's formula (see theorem 4.5.3). 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 sectional curvatures is $\frac{1}{4}$ (see also section 4.5.3). These remarks are further justified below.

In the tables there is a column called *rank*. This is related to the rank of a Lie group as discussed in section 8.3. Here, however, we need a rank concept for more general spaces. The *rank* of a geodesic $c : \mathbb{R} \to M$ is the dimension of parallel fields E along c such that $R(E(t), \dot{c}(t))\dot{c}(t) = 0$ for all t. The rank of a geodesic is, in particular, always ≥ 1. The rank of a Riemannian manifold is defined as the minimum rank over all of the geodesics in M. For symmetric spaces the rank can be computed from knowledge of Abelian subgroups in Lie groups. For a general manifold there might naturally 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 any Cartesian product has rank ≥ 2, and also many symmetric spaces have rank ≥ 2. It is unclear to what extent other manifolds can also have rank ≥ 2. All of the rank 1 symmetric spaces are listed in tables 10.3 and 10.4. The compact ones are also known as CROSSes.

10.1.2 Isometries and Parallel Curvature

Another interesting property for symmetric spaces is that they have parallel curvature tensor. This is because the symmetries A_p leave the curvature tensor and its covariant derivative invariant. In particular, we have

$$DA_p\left(\left(\nabla_X R\right)(Y,Z,W)\right) = \left(\nabla_{DA_pX} R\right)\left(DA_pY, DA_pZ, DA_pW\right),$$

which at p implies

$$\begin{aligned} -\left(\nabla_X R\right)(Y,Z,W) &= \left(\nabla_{-X} R\right)(-Y,-Z,-W) \\ &= \left(\nabla_X R\right)(Y,Z,W). \end{aligned}$$

Thus, $\nabla R = 0$. This almost characterizes symmetric spaces.

Theorem 10.1.1 (Cartan). *If (M,g) is a Riemannian manifold with parallel curvature tensor, then for each $p \in M$ there is an isometry A_p defined in a neighborhood of p with $DA_p = -I$ on T_pM. Moreover, if (M,g) is simply connected and complete, then the symmetry is defined on all of M, and the space is symmetric.*

Proof. The global statement follows from the local one using an analytic continuation argument as in the proof of theorem 5.6.7 and the next theorem 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 $A_p(x) = -x$ in these coordinates. It remains to see why this is an isometry when we have parallel curvature tensor. Equivalently, we must show that in these coordinates the metric has to be the same at x and $-x$. To this end we switch to polar coordinates and use the fundamental equations relating curvature and the metric. 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 that at p

$$R(\cdot, v)\, v = R(\cdot, -v)\,(-v)\,.$$

So the curvatures start out being the same. If ∂_r is the radial field, we also have

$$(\nabla_{\partial_r} R) = 0.$$

Thus, the curvature tensors not only start out being equal, but also satisfy the same simple first-order equation. Consequently, they remain the same as we go equal distance in opposite directions. □

A Riemannian manifold with parallel curvature tensor is called a *locally symmetric space*.

It is worth mentioning that there are left-invariant metrics that are not locally symmetric. The Berger spheres ($\varepsilon \neq 1$) and the Heisenberg group do not have parallel curvature tensor. In fact, as they are 3-dimensional they can't even have parallel Ricci tensor.

With very little extra work we can generalize the above theorem on the existence of local symmetries. Recall that in the discussion about existence of isometries with a given differential prior to theorem 5.6.7 we decided that they could exist only when the spaces had the same constant curvature. However, there is a generalization to symmetric spaces. 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 10.1.2 (Cartan). *Suppose we have a simply connected symmetric space (M, g) and a complete locally symmetric space $(N, \bar{g})$ of the same dimension. Given a linear isometry $L : T_pM \to T_qN$ such that*

$$L\left(R^g(x, y)\, z\right) = R^{\bar{g}}(Lx, Ly)\, Lz$$

for all $x, y, z \in T_pM$, there is a unique Riemannian isometry $F : M \to N$ such that $D_pF = L$.

Proof. The proof of this is, as in the constant curvature case, by analytic continuation. So we need only find these isometries locally. Given that there is an isometry defined locally, we know that it must look like

$$F = \exp_q \circ L \circ \exp_p^{-1}.$$

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 determine the metrics. In addition, the curvatures are parallel and satisfy the same first-order equation. We assume that initially 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. It also tells us what the isometry group must be in case the symmetric space is simply connected. This will be investigated further below.

10.1.3 The Lie Algebra Description

Finally, we offer a more algebraic description of symmetric spaces. There are many ways of writing homogeneous spaces as quotients G/H, e.g.,

$$S^3 = \mathrm{SU}(2) = \mathrm{SO}(4)/\mathrm{SO}(3) = \mathrm{O}(4)/\mathrm{O}(3).$$

But only one of these, $\mathrm{O}(4)/\mathrm{O}(3)$, tells us directly that S^3 is a symmetric space. This is because the isometry A_p modulo conjugation lies in $\mathrm{O}(4)$ as it is orientation reversing. In this section we present two related descriptions based on Killing fields and curvatures.

To begin we must understand how the map A_p acts on $\mathfrak{iso}$. The push forward $(A_p)_*$ preserves Killing fields as A_p is an isometry so there is a natural map $(A_p)_* = \sigma_p : \mathfrak{iso} \to \mathfrak{iso}$.

Throughout the section let (M, g) be a symmetric space.

Proposition 10.1.3. *Let $p \in M$. The map $\sigma_p = (A_p)_*$ defines an involution on $\mathfrak{iso}$. The 1-eigenspace is $\mathfrak{iso}_p$ and the (-1)-eigenspace consists of X such that $(\nabla X)|_p = 0$.*

Proof. Since $A_p^2 = id$ it is clear that also $\sigma_p^2 = id$. This shows that $\mathfrak{iso}$ is a direct sum decomposition of the (± 1)-eigenspaces for σ_p. Moreover, we have:

$$\sigma(X) = (A_p)_* X = DA_p\left(X|_{A_p^{-1}}\right) = DA_p\left(X|_{A_p}\right)$$

and as A_p is an isometry

$$(A_p)_*(\nabla_V X) = \nabla_{(A_p)_* V}(A_p)_* X = \nabla_{(A_p)_* V}\sigma(X).$$

At p we know that $DA_p = -I$ so if $\sigma(X) = X$, then $X|_p = -X|_p$, showing that $X \in \mathfrak{iso}_p$. Conversely, if $X \in \mathfrak{iso}_p$, then also $\sigma(X) \in \mathfrak{iso}_p$ and at p

$$\begin{aligned}\nabla_v \sigma(X) &= -\nabla_{DA_p(v)}\sigma(X)\\ &= -DA_p(\nabla_v X)\\ &= \nabla_v X\end{aligned}$$

showing that $\sigma(X) = X$.

On the other hand, if $\sigma_p(X) = -X$, then $(\nabla X)|_p = 0$ since at p

$$-\nabla_v X = \nabla_{-v}\sigma_p(X) = \nabla_v X.$$

Conversely, if $(\nabla X)|_p = 0$, then $-X = \sigma_p(X)$ as the Killing fields agree at p and both have vanishing derivative at p. □

Recall from proposition 8.1.4 that there is a short exact sequence

$$0 \to \mathfrak{iso}_p \to \mathfrak{iso} \to \mathfrak{t}_p \to 0$$

where

$$\mathfrak{t}_p = \left\{X|_p \in T_pM \mid X \in \mathfrak{iso}\right\}.$$

As M is homogeneous it follows that $\mathfrak{t}_p = T_pM$ and since M is symmetric the (-1)-eigenspace for σ_p is mapped isomorphically onto T_pM. We can then redefine $\mathfrak{t}_p$ as the subspace

$$\mathfrak{t}_p = \left\{X \in \mathfrak{iso} \mid (\nabla X)|_p = 0\right\}.$$

This gives us the natural decomposition $\mathfrak{iso} = \mathfrak{t}_p \oplus \mathfrak{iso}_p$ and by evaluating at p the alternate representation

$$\mathfrak{iso} \simeq T_pM \oplus \mathfrak{s}_p \subset T_pM \oplus \mathfrak{so}\left(T_pM\right),$$

where

$$\mathfrak{s}_p = \left\{(\nabla X)|_p \in \mathfrak{so}\left(T_pM\right) \mid X \in \mathfrak{iso}_p\right\}.$$

This leads to

Proposition 10.1.4. *If we identify $\mathfrak{iso} \simeq T_pM \oplus \mathfrak{s}_p$, then the Lie algebra structure is determined entirely by the curvature tensor and satisfies:*

(1) *If $X, Y \in T_pM$, then $[X, Y] = R(X, Y) \in \mathfrak{s}_p$.*
(2) *If $X \in T_pM$ and $S \in \mathfrak{s}_p$, then $[X, S] = -S(X)$.*
(3) *If $S, T \in \mathfrak{s}_p$, then $[S, T] = -(S \circ T - T \circ S) \in \mathfrak{s}_p$.*

Proof. We rely on proposition 8.1.3: For $X, Y \in \mathfrak{iso}$ we have

$$[\nabla X, \nabla Y](V) + \nabla_V [X, Y] = R(X, Y)V.$$

(1) When $X, Y \in \mathfrak{t}_p$ note that $[X, Y] = \nabla_Y X - \nabla_X Y$ vanishes at p so $[X, Y] \in \mathfrak{iso}_p$. The Lie derivative is then represented by $(\nabla [X, Y])|_p \in \mathfrak{s}_p$. To calculate this note that $[\nabla X, \nabla Y]$ vanishes at p so $(\nabla [X, Y])|_p = R\left(X|_p, Y|_p\right)$.

(2) When $X \in \mathfrak{t}_p$ and $Y \in \mathfrak{iso}_p$ we have $[X, Y] = \nabla_Y X - \nabla_X Y$ which at p reduces to $-\nabla_{X|_p} Y$.
(3) When $X, Y \in \mathfrak{iso}_p$ it follows that $[X, Y] = \nabla_X Y - \nabla_Y X$ also vanishes at p. Moreover, $(\nabla [X, Y]) |_p = - [\nabla X, \nabla Y] |_p$.

□

We just saw how the Lie algebra structure can be calculated from the curvature, but it also shows that the curvature can be calculated from the Lie algebra structure.

On a Lie algebra $\mathfrak{g}$ the *adjoint action* is defined as

$$\begin{aligned} \mathrm{ad}_X : \mathfrak{g} &\to \mathfrak{g} \\ Y &\mapsto \mathrm{ad}_X (Y) = [X, Y] \end{aligned}$$

and the *Killing form* by

$$B (X, Y) = \mathrm{tr} (\mathrm{ad}_X \circ \mathrm{ad}_Y) .$$

It is easy to check that the Killing form is symmetric and that ad_X is skew-symmetric with respect to B:

$$B (\mathrm{ad}_X Y, Z) + B (Y, \mathrm{ad}_X Z) = 0.$$

Remark 10.1.5. In case of the Lie algebra $\mathfrak{iso} \simeq T_p M \oplus \mathfrak{s}_p$ the map $\mathrm{ad}_S : \mathfrak{iso} \to \mathfrak{iso}$, $S \in \mathfrak{s}_p$, is also skew-symmetric with respect to the natural inner product

$$g ((X_1, S_1), (X_2, S_2)) = g (X_1, X_2) + g (S_1, S_2) = g (X_1, X_2) - \mathrm{tr} (S_1 \circ S_2) .$$

Thus

$$B (S, S) = \mathrm{tr} (\mathrm{ad}_S \circ \mathrm{ad}_S) = -\mathrm{tr} \left(\mathrm{ad}_S \circ (\mathrm{ad}_S)^*\right) \leq 0.$$

Moreover, if $B (S, S) = 0$, then $\mathrm{ad}_S = 0$ which in turn implies that $S = 0$ since $\mathrm{ad}_S (X) = [S, X] = S (X)$.

The next result tells us how to calculate the curvature tensor algebraically and is very important for the next two sections.

Theorem 10.1.6. *If $X, Y, Z \in \mathfrak{t}_p$, then $R (X, Y) Z = [Z, [X, Y]]$ at p. In Lie algebraic language on $\mathfrak{iso} \simeq T_p M \oplus \mathfrak{s}_p$:*

$$\begin{aligned} R (X, Y) Z &= -\mathrm{ad}_Z \circ \mathrm{ad}_Y (X) , \\ \mathrm{Ric} (Y, Z) &= -\frac{1}{2} B (Z, Y) . \end{aligned}$$

Moreover, in case Ric $= \lambda g$*, it follows that the curvature operator has the same sign as* λ*.*

Proof. For completeness we offer a proof that does not rely on the previous proposition. Instead it uses proposition 8.1.3: For $X, Y, Z \in \mathfrak{iso}$ we have $\nabla^2_{X,Y} Z = -R(Z,X)Y$. If we additionally assume $X, Y, Z \in \mathfrak{t}_p$, then $\nabla X = \nabla Y = \nabla Z = 0$ at p. Bianchi's first identity then implies

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

For the Ricci tensor formula first note that the operator $\mathrm{ad}_Z \circ \mathrm{ad}_Y$ leaves the decomposition $\mathfrak{iso} \simeq T_pM \oplus \mathfrak{s}_p$ invariant since each of ad_Z and ad_Y interchange the subspaces in this factorization. With this in mind we obtain

$$\begin{aligned} \mathrm{tr}\left(\mathrm{ad}_Z \circ \mathrm{ad}_Y|_{T_pM}\right) &= \mathrm{tr}\left(\mathrm{ad}_Z|_{\mathfrak{s}_p} \circ \mathrm{ad}_Y|_{T_pM}\right) \\ &= \mathrm{tr}\left(\mathrm{ad}_Y|_{T_pM} \circ \mathrm{ad}_Z|_{\mathfrak{s}_p}\right) \\ &= \mathrm{tr}\left(\mathrm{ad}_Y \circ \mathrm{ad}_Z|_{\mathfrak{s}_p}\right). \end{aligned}$$

Using symmetry of B this gives:

$$B(Z,Y) = \mathrm{tr}\left(\mathrm{ad}_Y \circ \mathrm{ad}_Z|_{\mathfrak{s}_p}\right) + \mathrm{tr}\left(\mathrm{ad}_Z \circ \mathrm{ad}_Y|_{T_pM}\right) = 2\mathrm{tr}\left(\mathrm{ad}_Z \circ \mathrm{ad}_Y|_{T_pM}\right).$$

The formula for the curvature tensor then shows that

$$\mathrm{Ric}(Z,Y) = -\mathrm{tr}\left(\mathrm{ad}_Z \circ \mathrm{ad}_Y|_{T_pM}\right) = -\frac{1}{2} B(Z,Y).$$

In case Ric $= \lambda g$, $\lambda \neq 0$ it follows that

$$\begin{aligned} g(R(X,Y)Z,W)|_p &= -g(\mathrm{ad}_Z \circ \mathrm{ad}_Y(X), W)|_p \\ &= -\frac{1}{\lambda}\mathrm{Ric}(\mathrm{ad}_Z \circ \mathrm{ad}_Y(X), W)|_p \\ &= \frac{1}{2\lambda} B(\mathrm{ad}_Z \circ \mathrm{ad}_Y(X), W) \\ &= -\frac{1}{2\lambda} B(\mathrm{ad}_Y(X), \mathrm{ad}_Z(W)). \\ &= -\frac{1}{2\lambda} B([X,Y],[W,Z]) \\ &= \frac{1}{2\lambda}\mathrm{tr}\left(\mathrm{ad}_{[X,Y]} \circ \left(\mathrm{ad}_{[W,Z]}\right)^*\right). \end{aligned}$$

The diagonal terms for the curvature operator then become

$$
\begin{aligned}
g\left(\mathfrak{R}\left(\sum X_i \wedge Y_i\right), \left(\sum X_i \wedge Y_i\right)\right)|_p &= \sum g\left(R\left(X_i, Y_i\right)Y_j, X_j\right)|_p \\
&= -\frac{1}{2\lambda}\sum B\left(\mathrm{ad}_{X_i}\left(Y_i\right), \mathrm{ad}_{X_j}\left(Y_j\right)\right) \\
&= -\frac{1}{2\lambda}B\left(\sum \mathrm{ad}_{X_i}\left(Y_i\right), \sum \mathrm{ad}_{X_i}\left(Y_i\right)\right).
\end{aligned}
$$

Since $\sum \mathrm{ad}_{Y_i}(X_i) \in \mathfrak{s}_p$ it follows from remark 10.1.5 that $B\left(\sum \mathrm{ad}_{X_i}(Y_i), \sum \mathrm{ad}_{X_i}(Y_i)\right) \leq 0$. Thus the eigenvalues of $\mathfrak{R}$ have the same sign as the Einstein constant.

In case $\mathrm{Ric} = 0$ corollary 8.2.5 implies that M is flat in the more general case of homogeneous spaces. □

Note that the formula is similar to the one that was developed for biinvariant metrics in proposition 4.4.2. However, while left-invariant fields on a Lie group are Killing fields as long as the metric is right-invariant, they generally don't have vanishing covariant derivative at the identity.

We can now give a slightly more efficient Lie algebra structure of a symmetric space. Suppose (M, g) is a symmetric space and $p \in M$. We define a bracket operation on $\mathfrak{R}_p = T_pM \oplus \mathfrak{so}\left(T_pM\right)$ by

$$
\begin{aligned}
[X, Y] &= R_{X,Y} \in \mathfrak{so}\left(T_pM\right) \text{ for } X, Y \in T_pM, \\
-[S, X] &= [X, S] = S(X) \in T_pM \text{ for } X \in T_pM \text{ and } S \in \mathfrak{so}\left(T_pM\right), \\
\left[S, S'\right] &= -\left(S \circ S' - S' \circ S\right) \in \mathfrak{so}\left(T_pM\right) \text{ for } S, S' \in \mathfrak{so}\left(T_pM\right).
\end{aligned}
$$

This bracket will in general not satisfy the Jacobi identity on triples that involve precisely two elements from T_pM

$$
\begin{aligned}
&[S, [X, Y]] + [Y, [S, X]] + [X, [Y, S]] \\
&= -S \circ R_{X,Y} + R_{X,Y} \circ S - R_{Y,S(X)} + R_{X,S(Y)} \\
&= -S \circ R_{X,Y} + R_{X,Y} \circ S + R_{S(X),Y} + R_{X,S(Y)}.
\end{aligned}
$$

But the other possibilities for the Jacobi identity do hold. When the triple involves zero or one element from T_pM this is straightforward, while if all three are from T_pM it follows from the Bianchi identity

$$
\begin{aligned}
0 &= R_{X,Y}Z + R_{Z,X}Y + R_{Y,Z}X \\
&= [Z, [X, Y]] + [Y, [Z, X]] + [X, [Y, Z]].
\end{aligned}
$$

Fortunately we have the following modification of theorem 10.1.2.

Corollary 10.1.7. *Let (M, g) be a simply connected symmetric space. If $S \in \mathfrak{so}(T_pM)$, then $S \in \mathfrak{s}_p$ if and only if for all $X, Y \in T_pM$*

$$-S \circ R_{X,Y} + R_{X,Y} \circ S + R_{S(X),Y} + R_{X,S(Y)} = 0.$$

Proof. We start by assuming that $Z \in \mathfrak{iso}_p$ and $S = (\nabla Z)|_p$. Since Z is a Killing field we have $L_Z R = 0$. Since $[V, Z]|_p = S(V)$ for all $V \in T_pM$ this implies that

$$\begin{aligned} 0 &= (L_Z R)_{X,Y} V \\ &= -S(R_{X,Y}V) + R_{X,Y}S(V) + R_{S(X),Y}V + R_{X,S(Y)}V. \end{aligned}$$

The converse is proven using theorem 10.1.2 by constructing the flow of the Killing field corresponding to $S \in \mathfrak{so}(T_pM)$. In T_pM construct the isometric flow for S. Check that the assumption about S implies that the isometries in this flow satisfy theorem 10.1.2 and then conclude that we obtain a global flow on M. This flow will then generate the desired Killing field. □

Let $\mathfrak{r}_p \subset \mathfrak{s}_p$ be the Lie algebra generated by the skew-symmetric endomorphisms $R_{X,Y} \in \mathfrak{so}(T_pM)$. We have basically established the following useful relationship between the curvature tensor and Killing fields on a symmetric space.

Corollary 10.1.8. *Let (M, g) be a simply connected symmetric space. The bracket structure on $\mathfrak{R}_p$ makes $T_pM \oplus \mathfrak{s}_p$ into a Lie algebra with subalgebra $\mathfrak{c}_p = T_pM \oplus \mathfrak{r}_p$. In fact $T_pM \oplus \mathfrak{s}_p$ is characterized as the maximal Lie algebra: $\mathfrak{c}_p \subset T_pM \oplus \mathfrak{s}_p \subset \mathrm{N}_{\mathfrak{R}_p}(\mathfrak{c}_p)$, where $\mathrm{N}_{\mathfrak{R}_p}(\mathfrak{c}_p)$ is the normalizer of $\mathfrak{c}_p$ in $\mathfrak{R}_p$. Moreover, the Lie algebra involution on $T_pM \oplus \mathfrak{s}_p$ (and its restriction on $\mathfrak{c}_p$) has T_pM as the (-1)-eigenspace and $\mathfrak{s}_p$ as the 1-eigenspace.*

Proof. We saw in the above corollary that any subalgebra of $\mathfrak{k} \subset \mathfrak{so}(T_pM)$ such that $T_pM \oplus \mathfrak{k} \subset \mathfrak{R}_p$ becomes a Lie algebra (i.e., also satisfies the Jacobi identity) must be contained in $T_pM \oplus \mathfrak{s}_p$. We also saw that $T_pM \oplus \mathfrak{s}_p \subset \mathrm{N}_{\mathfrak{R}_p}(\mathfrak{c}_p)$. □

We are now ready to attempt to reverse the construction so as to obtain symmetric spaces from suitable Lie algebras. Assume we have a Lie algebra $\mathfrak{g}$ with a Lie algebra involution $\sigma : \mathfrak{g} \to \mathfrak{g}$. First decompose $\mathfrak{g} = \mathfrak{t} \oplus \mathfrak{k}$ where $\mathfrak{t}$ is the (-1)-eigenspace for σ and $\mathfrak{k}$ is the 1-eigenspace for σ. Observe that $\mathfrak{k}$ is a Lie subalgebra as

$$\begin{aligned} \sigma[X, Y] &= [\sigma(X), \sigma(Y)] \\ &= [X, Y]. \end{aligned}$$

Similarly, $[\mathfrak{k}, \mathfrak{t}] \subset \mathfrak{t}$ and $[\mathfrak{t}, \mathfrak{t}] \subset \mathfrak{k}$.

Suppose further that there is a connected compact Lie group K with Lie algebra $\mathfrak{k}$ such that the Lie bracket action of $\mathfrak{k}$ on $\mathfrak{t}$ comes from an action of K on $\mathfrak{t}$. In case K is simply connected this will always be the case. Compactness of K allows us to choose a Euclidean metric on $\mathfrak{t}$ making the action of K isometric. It follows that the decomposition $\mathfrak{g} = \mathfrak{t} \oplus \mathfrak{k}$ is exactly of the type $\mathfrak{iso} = \mathfrak{t}_p \oplus \mathfrak{iso}_p$. Next pick a biinvariant metric on K so that $\mathfrak{g} = \mathfrak{t} \oplus \mathfrak{k}$ is an orthogonal decomposition. Finally, if we can also choose a Lie group $\mathrm{G} \supset \mathrm{K}$ 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 A_p on G/K. When G is simply connected σ will be the differential of a Lie group involution $A : \mathrm{G} \to \mathrm{G}$ that is the identity on K. This defines the desired involution on G/K that fixes the point $p = \mathrm{K}$.

It rarely happens that all of the Lie groups in play are simply connected or even connected. Nevertheless, the constructions can often be verified directly. Without assumptions about connectedness of the groups there is a long exact sequence:

$$\pi_1(\mathrm{K}) \to \pi_1(\mathrm{G}) \to \pi_1(\mathrm{G}/\mathrm{K}) \to \pi_0(\mathrm{K}) \to \pi_0(\mathrm{G}) \to \pi_0(\mathrm{G}/\mathrm{K}) \to 1,$$

where π_0 denotes the set of connected components. As K and G are Lie groups these spaces are in fact groups. From this sequence it follows that G/K is connected and simply connected if $\pi_0(\mathrm{K}) \to \pi_0(\mathrm{G})$ is an isomorphism and $\pi_1(\mathrm{K}) \to \pi_1(\mathrm{G})$ is surjective.

This algebraic approach will in general not immediately give us the isometry group of the symmetric space. 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 $\mathfrak{g}$. For $S^3 = \mathrm{O}(4)/\mathrm{O}(3)$, we see that the algebraic approach can also lead us to the description $S^3 = \mathrm{Spin}(4)/\mathrm{Spin}(3)$. However, as any Lie algebra description can be used to calculate the curvature, corollary 10.1.8 will in principle allow us to determine $\mathfrak{iso}$.

It is important to realize that a Lie algebra $\mathfrak{g}$, in itself, does not give rise to a symmetric space. The involution is 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 *anti-automorphism*. This is particularly interesting if $\mathfrak{g}$ comes from a Lie group G with biinvariant metric. There the involution $A_e(g) = g^{-1}$ is an isometry and makes G a symmetric space. But it's differential on $\mathfrak{g}$ is an anti-automorphism. Instead the algebraic description of G as a symmetric space comes from using $\mathfrak{g} \times \mathfrak{g}$ with $\sigma(X, Y) = (Y, X)$. This will be investigated in the next section.

10.2 Examples of Symmetric Spaces

We explain how some of the above constructions work in the concrete case of the Grassmann manifold and its hyperbolic counterpart. We also look at complex Grassmannians, but there we restrict attention to the complex projective space. Finally, we briefly discuss the symmetric space structure of $\mathrm{SL}(n)/\mathrm{SO}(n)$. After these examples we give a formula for the curvature tensor on compact Lie groups with biinvariant metrics and their noncompact counter parts.

Throughout we use the convention that for $X, Y \in \mathrm{Mat}_{k\times l}$

$$\langle X, Y\rangle = \mathrm{tr}\left(X^*Y\right) = \overline{\mathrm{tr}\left(XY^*\right)}$$

with the conjugation only being relevant when the entries are complex. This inner product is invariant under the natural action of $\mathrm{O}(k) \times \mathrm{O}(l)$ on $\mathrm{Mat}_{k\times l}$ defined by:

$$\begin{aligned}\mathrm{O}(k) \times \mathrm{O}(l) \times \mathrm{Mat}_{k\times l} &\to \mathrm{Mat}_{k\times l},\\ (A, B, X) &\mapsto AXB^{-1} = AXB^*\end{aligned}$$

since

$$\begin{aligned}\langle O_1XO_2, O_1YO_2\rangle &= \mathrm{tr}\left(O_2^*X^*O_1^*O_1YO_2\right)\\ &= \mathrm{tr}\left(O_2^*X^*YO_2\right)\\ &= \mathrm{tr}\left(O_2O_2^*X^*Y\right)\\ &= \langle X, Y\rangle\,.\end{aligned}$$

This allows us to conclude that the metrics we study can be extended to the entire space via an appropriate transitive action whose isotropy is a subgroup of the action by $\mathrm{O}(k) \times \mathrm{O}(l)$ on $\mathrm{Mat}_{k\times l}$.

Theorem 10.1.6 is used to calculate the curvatures in specific examples and the relevant Killing forms are calculated in exercise 10.5.6.

10.2.1 The Compact Grassmannian

First consider the Grassmannian of oriented k-planes in $\mathbb{R}^{k+l}$, denoted by $M = \tilde{\mathrm{G}}_k\left(\mathbb{R}^{k+l}\right)$. Each element in M is a k-dimensional subspace of $\mathbb{R}^{k+l}$ together with an orientation, e.g., $\tilde{\mathrm{G}}_1\left(\mathbb{R}^{n+1}\right) = S^n$. 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. We use that $\mathrm{O}(k+l)$ acts on $\mathbb{R}^{k+l}$. If a k-dimensional subspace has the positively oriented orthonormal basis $e_1, \ldots, e_k$, then the image under $O \in \mathrm{O}(k+l)$ will have the positively oriented orthonormal basis $Oe_1, \ldots Oe_k$. This action is clearly transitive. The isotropy group is $\mathrm{SO}(k) \times \mathrm{O}(l) \subset \mathrm{O}(k+l)$.

The tangent space at $p = \mathbb{R}^k$ is naturally identified with the space of $k\times l$ matrices $\mathrm{Mat}_{k\times l}$, or equivalently, with $\mathbb{R}^k\otimes\mathbb{R}^l$. To see this, just observe that any k-dimensional subspace of $\mathbb{R}^{k+l}$ that is close to $\mathbb{R}^k$ can be represented as a linear graph over $\mathbb{R}^k$ with values in the orthogonal complement $\mathbb{R}^l$. The isotropy action of $\mathrm{SO}(k) \times \mathrm{O}(l)$ on $\mathrm{Mat}_{k\times l}$ is:

$$\begin{aligned}\mathrm{SO}(k)\times\mathrm{O}(l)\times\mathrm{Mat}_{k\times l}&\to\mathrm{Mat}_{k\times l},\\(A,B,X)&\mapsto AXB^{-1}=AXB^{t}.\end{aligned}$$

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 $\mathrm{Mat}_{k\times l}$ but does not cover all of the space. This is an example of an irreducible action on Euclidean space that is not transitive on the unit sphere. The representation, when seen as acting on $\mathbb{R}^k\otimes\mathbb{R}^l$, is denoted by $\mathrm{SO}(k)\otimes\mathrm{O}(l)$.

To see that M is a symmetric space we have to show that the isotropy group contains the required involution. On the tangent space $T_pM=\mathrm{Mat}_{k\times l}$ it is supposed to act as -1. Thus, we have to find $(A,B)\in\mathrm{SO}(k)\times\mathrm{O}(l)$ 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 without using the isometry group of the space. In fact $\mathrm{SO}(k+l)$ is a covering of the isometry group, although that requires some work to prove. But we have found a Lie algebra description with

$$\begin{aligned}\mathfrak{so}(k+l)&\subset\mathfrak{iso},\\\mathfrak{so}(k)\times\mathfrak{so}(l)&\subset\mathfrak{iso}_p,\end{aligned}$$

and an involution that fixes $\mathfrak{so}(k)\times\mathfrak{so}(l)$.

We shall use the block decomposition of matrices in $\mathfrak{so}(k+l)$:

$$X=\begin{pmatrix}X_1 & B\\ -B^t & X_2\end{pmatrix},\ X_1\in\mathfrak{so}(k),\ X_2\in\mathfrak{so}(l),\ B\in\mathrm{Mat}_{k\times l}.$$

If

$$\mathfrak{t}_p=\left\{\begin{pmatrix}0 & B\\ -B^t & 0\end{pmatrix}\mid B\in\mathrm{Mat}_{k\times l}\right\},$$

then we have an orthogonal decomposition:

$$\mathfrak{so}(k+l)=\mathfrak{t}_p\oplus\mathfrak{so}(k)\oplus\mathfrak{so}(l),$$

where $\mathfrak{t}_p=T_pM$. Note that $\mathfrak{t}_p$ consists of skew-symmetric matrices so

$$\langle X,Y\rangle=\mathrm{tr}\left(X^tY\right)=-\mathrm{tr}(XY)=-\frac{1}{k+l-2}B(X,Y).$$

This will be our metric on $\mathfrak{t}_p$ and tells us that $\mathrm{Ric} = \frac{k+l-2}{2} g$ and

$$\langle R(X,Y)Y,X\rangle = -\frac{1}{k+l-2} B([X,Y],[X,Y]) = |[X,Y]|^2 \geq 0.$$

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 many zero sectional curvatures.

The calculations also show that in fact $\mathfrak{r}_p = \mathfrak{so}(k) \oplus \mathfrak{so}(l)$ and corollary 10.1.8 can be used to show that $\mathfrak{iso} = \mathfrak{so}(k+l)$.

10.2.2 The Hyperbolic Grassmannian

Next we consider 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:

$$\begin{aligned} v^t I_{k,l} w &= v^t \begin{pmatrix} I_k & 0 \\ 0 & -I_l \end{pmatrix} w \\ &= \sum_{i=1}^{k} v_i w_i - \sum_{i=k+1}^{k+l} v_i w_i. \end{aligned}$$

The group of linear transformations that preserve this form is denoted by $\mathrm{O}(k,l)$. 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 $\mathrm{O}(k,l)$ is not compact. But it clearly contains the (maximal) compact subgroup $\mathrm{O}(k) \times \mathrm{O}(l)$.

The Lie algebra $\mathfrak{so}(k,l)$ of $\mathrm{O}(k,l)$ 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 for $I_{k,l}$, then

$$Y = \begin{pmatrix} Y_1 & B \\ B^t & Y_2 \end{pmatrix}, \; Y_1 \in \mathfrak{so}(k), \; Y_2 \in \mathfrak{so}(l), \; B \in \mathrm{Mat}_{k\times l}.$$

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 $M = \tilde{G}_k(\mathbb{R}^{k,l})$. The selected point is as before $p = \mathbb{R}^k$. One can easily see that topologically: $\tilde{G}_k(\mathbb{R}^{k,l})$ is an open subset of $\tilde{G}_k(\mathbb{R}^{k+l})$. The metric on this space is another story, however. Clearly, $\mathrm{O}(k,l)$ acts transitively on

M, and those elements that fix p are of the form $\mathrm{SO}(k) \times \mathrm{O}(l)$. 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

$$\mathfrak{so}(k,l) = \mathfrak{t}_p \oplus \mathfrak{so}(k) \oplus \mathfrak{so}(l),$$

where

$$\mathfrak{t}_p = \left\{ \begin{pmatrix} 0 & B \\ B^t & 0 \end{pmatrix} \mid B \in \mathrm{Mat}_{k\times l} \right\}.$$

This time, however, $\mathfrak{t}_p$ consists of symmetric matrices so

$$\langle X, Y \rangle = \mathrm{tr}\left(X^t Y\right) = \mathrm{tr}(XY) = \frac{1}{k+l-2} B(X,Y).$$

This will be our metric on $\mathfrak{t}_p$. Thus $\mathrm{Ric} = -\frac{k+l-2}{2} g$ and

$$\langle R(X,Y)Y, X \rangle = \frac{1}{k+l-2} B([X,Y],[X,Y]) = -\left|[X,Y]\right|^2 \leq 0.$$

This is exactly the negative of the expression we got in the compact case. Hence, the hyperbolic Grassmannians have nonpositive curvature. When $l = 1$, we have reconstructed the hyperbolic space together with its isometry group.

Again it follows that $\mathfrak{r}_p = \mathfrak{so}(k) \oplus \mathfrak{so}(l)$ and $\mathfrak{iso} = \mathfrak{so}(k,l)$.

10.2.3 Complex Projective Space Revisited

We view the complex projective space as a complex Grassmannian. Namely, let $M = \mathbb{CP}^n = G_1\left(\mathbb{C}^{n+1}\right)$, i.e., the complex lines in $\mathbb{C}^{n+1}$. More generally one can consider $G_k\left(\mathbb{C}^{k+l}\right)$ and the hyperbolic counterparts $G_k\left(\mathbb{C}^{k,l}\right)$ of space-like subspaces. We leave this to the reader.

The group $\mathrm{U}(n+1) \subset \mathrm{SO}(2n+2)$ 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 $\mathrm{U}(n+1)$ satisfy $A^{-1} = A^*$. As with the Grassmannian, $\mathrm{U}(n+1)$ 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 $\mathrm{SU}(n+1)$, which still acts transitively, but with a finite kernel consisting of those aI such that $a^{n+1} = 1$.

If $p = \mathbb{C}$ is the first coordinate axis, then the isotropy group is $\mathrm{S}(\mathrm{U}(1) \times \mathrm{U}(n))$, i.e., the matrices in $\mathrm{U}(1) \times \mathrm{U}(n)$ of determinant 1. This group is naturally isomorphic to $\mathrm{U}(n)$ via the map

$$A \mapsto \begin{pmatrix} \det A^{-1} & 0 \\ 0 & A \end{pmatrix}.$$

The involution that makes M symmetric is then given by

$$\begin{pmatrix} (-1)^n & 0 \\ 0 & -I_n \end{pmatrix}.$$

We pass to the Lie algebra level in order to compute the curvature tensor. From above, we have

$$\begin{aligned} \mathfrak{su}(n+1) &= \{A \mid A = -A^*, \operatorname{tr} A = 0\}, \\ \mathfrak{u}(n) &= \{B \mid B = -B^*\}. \end{aligned}$$

The inclusion looks like

$$B \mapsto \begin{pmatrix} -\operatorname{tr} B & 0 \\ 0 & B \end{pmatrix}.$$

Thus if elements of $\mathfrak{su}(n+1)$ are written

$$\begin{pmatrix} -\operatorname{tr} B & -z^* \\ z & B \end{pmatrix},$$

and

$$\mathfrak{t}_p = \left\{ \begin{pmatrix} 0 & -z^* \\ z & 0 \end{pmatrix} \mid z \in \mathbb{C}^n \right\}$$

we obtain

$$\mathfrak{su}(n+1) = \mathfrak{t}_p \oplus \mathfrak{u}(n).$$

For $X, Y \in \mathfrak{t}_p$

$$\langle X, Y \rangle = \operatorname{tr}\left(X^* Y\right) = -\operatorname{tr}(XY) = -\frac{1}{2(n+1)} B(X, Y).$$

So $\mathrm{Ric} = (n+1)\,g$ and

$$\begin{aligned}\langle R(X,Y)Y,X\rangle &= |[X,Y]|^2\\ &= \left|\begin{pmatrix} -(z^*w - w^*z) & 0\\ 0 & wz^* - zw^*\end{pmatrix}\right|^2\\ &= |z^*w - w^*z|^2 - \mathrm{tr}\left(\left(wz^* - zw^*\right)^2\right)\\ &= 4\left|\mathrm{Im}\left(z^*w\right)\right|^2 - 2\mathrm{Re}\left(z^*w\right)^2 + 2|z|^2|w|^2 .\end{aligned}$$

To compute the sectional curvatures we need to pick an orthonormal basis X, Y for a plane. This means that $|z|^2 = |w|^2 = \frac{1}{2}$ and $\mathrm{Re}\,(z^*w) = 0$, which implies $z^*w = \mathrm{iIm}\,(z^*w)$ and

$$\begin{aligned}\sec(X,Y) &= 4\left|\mathrm{Im}\left(z^*w\right)\right|^2 - 2\mathrm{Re}\left(z^*w\right)^2 + 2|z|^2|w|^2\\ &= 6|z^*w|^2 + \frac{1}{2}\\ &\leq 2.\end{aligned}$$

Showing that $\frac{1}{2} \leq \sec \leq 2$, where the minimum value occurs when $z^*w = 0$ and the maximum value when $w = \mathrm{i}z$. Note that this scaling isn't consistent with our discussion in section 4.5.3 but we have still shown that the metric is quarter pinched.

10.2.4 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

$$\mathfrak{sl}(n) = \{X \in \mathrm{Mat}_{n\times n} \mid \mathrm{tr}\,X = 0\}.$$

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 $-I$ on $\mathfrak{k}$ and I on $\mathfrak{so}(n)$ so $\sigma(X) = -X^t$. For $X, Y \in \mathfrak{k}$

$$\langle X, Y\rangle = \mathrm{tr}\left(X^*Y\right) = \mathrm{tr}\,(XY) = \frac{1}{2n}B(X,Y).$$

So $\mathrm{Ric} = -ng$ and $\langle R(X,Y)Z,W\rangle = \langle [X,Y],[Z,W]\rangle$. In particular, the sectional curvatures must be nonpositive.

10.2.5 Lie Groups

Next we check how Lie groups become symmetric spaces.

To this end, start with a compact Lie group G with a biinvariant metric. As usual, the Lie algebra $\mathfrak{g}$ of G is identified with $T_e\mathrm{G}$ as well as the set of left-invariant vector fields on G. Since the left-invariant fields are Killing fields it follows that $\mathrm{ad}_X = 2\nabla X$ is skew-symmetric. In particular, the Killing form is nonpositive and only vanishes on the center of $\mathfrak{g}$. Thus, when $\mathfrak{g}$ has no center, then $g = -B$ defines a biinvariant as ad_X is skew-symmetric with respect to B. This is the situation we are interested in.

The Lie algebra description of G as a symmetric space is given by $(\mathfrak{g} \oplus \mathfrak{g}, \sigma)$ with $\sigma(X, Y) = (Y, X)$. Here the diagonal $\mathfrak{g}^{\Delta} = \{(X, X) \mid X \in \mathfrak{g}\}$ is the 1-eigenspace, while the complement $\mathfrak{g}^{\perp} = \{(X, -X) \mid X \in \mathfrak{g}\}$ is the (-1)-eigenspace. Thus $\mathfrak{k} = \mathfrak{g}^{\Delta} \cong \mathfrak{g}$ and $\mathfrak{t} = \mathfrak{g}^{\perp}$. We already know that $\mathfrak{g}$ corresponds to the compact Lie group G, so we are simply saying that $\mathrm{G} = (\mathrm{G} \times \mathrm{G}) / \mathrm{G}^{\Delta}$. The Ricci tensor is given by $\mathrm{Ric} = -\frac{1}{2}B = \frac{1}{2}g$ and the curvatures are nonnegative. Note that the natural inner product on $\mathfrak{t}$ is scaled by a factor of 2 from the biinvariant metric on $\mathfrak{g}$.

We can also construct a noncompact symmetric space using the same Lie algebra $\mathfrak{g}$ that comes from a compact Lie group without center. Consider: $(\mathfrak{g} \otimes \mathbb{C}, \sigma)$, where $\sigma(X) = \bar{X}$ is complex conjugation. Then $\mathfrak{k} = \mathfrak{g} \subset \mathfrak{g} \otimes \mathbb{C}$ and $\mathfrak{t} = \mathrm{i}\mathfrak{g}$. The inner product on $\mathfrak{k}$ is $-B$ on $\mathfrak{g}$, while on $\mathfrak{t}$ the metric is given by $g(\mathrm{i}X, \mathrm{i}Y) = -B(X, Y) = B(\mathrm{i}X, \mathrm{i}Y)$. This gives us $\mathrm{Ric}(\mathrm{i}X, \mathrm{i}Y) = -\frac{1}{2}B(\mathrm{i}X, \mathrm{i}Y) = -\frac{1}{2}g(\mathrm{i}X, \mathrm{i}Y)$ and nonpositive curvature.

10.3 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.

10.3.1 The Holonomy Group

Let (M, g) be a Riemannian n-manifold. If $c : [a, b] \to M$ is a unit speed curve, then

$$P_{c(a)}^{c(b)} : T_{c(a)}M \to T_{c(b)}M$$

denotes the effect of parallel translating a vector from $T_{c(a)}M$ to $T_{c(b)}M$ along c. 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 the curve is a loop, i.e., $c(a) = c(b) = p$. Then parallel translation yields 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 $\mathrm{O}(T_pM) = \mathrm{O}(n)$. Moreover, it is a Lie group. This takes some work to establish in case the group isn't compact. The *restricted holonomy group* $\mathrm{Hol}_p^0 = \mathrm{Hol}_p^0(M,g)$ is the connected normal subgroup that results from using only contractible loops. This group is compact and consequently a Lie group. Here are some elementary properties that are easy to establish:

(a) $\mathrm{Hol}_p(\mathbb{R}^n) = \{1\}$.
(b) $\mathrm{Hol}_p(S^n(R)) = \mathrm{SO}(n)$.
(c) $\mathrm{Hol}_p(H^n) = \mathrm{SO}(n)$.
(d) $\mathrm{Hol}_p(M,g) \subset \mathrm{SO}(n)$ if and only if M is orientable.
(e) $\mathrm{Hol}_p(\tilde{M},\tilde{g}) = \mathrm{Hol}_p^0(\tilde{M},\tilde{g}) = \mathrm{Hol}_p^0(M,g)$, 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 at $p \in M$ can be extended to a parallel tensor on (M,g) if and only if it is invariant under the holonomy group; e.g., if ω is a 2-form, then we require that $\omega(Pv,Pw) = \omega(v,w)$ for all $P \in \mathrm{Hol}_p(M,g)$ and $v,w \in T_pM$.

We are now ready to study how the Riemannian manifold decomposes according to the holonomy. Guided by (f) we see that being a Cartesian product is 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_p^0 on T_pM. If $E \subset T_pM$ is an invariant subspace, i.e., $\mathrm{Hol}_p^0(E) \subset E$, then the orthogonal complement is also preserved, i.e., $\mathrm{Hol}_p^0(E^\perp) \subset E^\perp$. Thus, T_pM decomposes into irreducible invariant subspaces:

$$T_pM = E_1 \oplus \cdots \oplus E_k.$$

Here, irreducible means that there are no nontrivial invariant subspaces inside E_i. 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 obtain 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 state de Rham's decomposition theorem.

Theorem 10.3.1 (de Rham, 1952). *If we decompose the tangent bundle of a Riemannian manifold (M,g) into irreducible components according to the restricted 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, unfortunately, not a trivial analytic continuation argument and we only offer a general outline. Apparently, one must understand how simple connectivity forces the maximal integral submanifolds to be embedded submanifolds. Let M_i be the maximal integral submanifolds for η_i through a fixed $p \in M$. Consider the abstract Riemannian manifold

$$(M_1 \times \cdots \times M_k, g_1 + \cdots + g_k).$$

Around p, the two manifolds (M, g) and $(M_1 \times \cdots \times M_k, g_1 + \cdots + g_k)$ are isometric to each other. As (M, g) is complete and each M_i is totally geodesic it follows that $(M_1 \times \cdots \times M_k, g_1 + \cdots + g_k)$ is also complete. The goal is to find an isometric embedding

$$(M, g) \to (M_1 \times \cdots \times M_k, g_1 + \cdots + g_k).$$

Completeness will insure us that the map is onto and in fact a Riemannian covering map. We will then have shown that M is isometric to the universal covering of

$$(M_1 \times \cdots \times M_k, g_1 + \cdots + g_k),$$

which is the product manifold

$$\left(\tilde{M}_1 \times \cdots \times \tilde{M}_k, \tilde{g}_1 + \cdots + \tilde{g}_k\right)$$

with the induced pull-back metric. □

Given this decomposition it is 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 a nice characterization of Einstein manifolds.

Theorem 10.3.2. *If (M, g) is an irreducible Riemannian manifold with a parallel $(1, 1)$-tensor T, then both the symmetric $S = \frac{1}{2}(T + T^*)$ and skew-symmetric $A = \frac{1}{2}(T - T^*)$ parts have exactly one (complex) eigenvalue. Moreover, if the skew-symmetric part does not vanish, then it induces a parallel complex structure, i.e., a Kähler structure.*

Proof. The fact that T is parallel implies that the adjoint is also parallel as the metric itself is parallel. More precisely we always have:

$$\begin{aligned} g\left(\left(\nabla_X T\right)(Y), Z\right) &= g\left(\nabla_X T(Y), Z\right) - g\left(T\left(\nabla_X Y\right), Z\right) \\ &= \nabla_X g\left(T(Y), Z\right) - g\left(T(Y), \nabla_X Z\right) - g\left(T\left(\nabla_X Y\right), Z\right) \\ &= \nabla_X g\left(Y, T^*(Z)\right) - g\left(Y, T^*\left(\nabla_X Z\right)\right) - g\left(\nabla_X Y, T^*(Z)\right) \\ &= g\left(Y, \left(\nabla_X T^*\right)(Z)\right). \end{aligned}$$

Thus $\nabla S = 0 = \nabla A$ and both are invariant under parallel translation.

First, decompose $T_pM = E_1 \oplus \cdots \oplus E_k$ into the orthogonal eigenspaces for $S : T_pM \to T_pM$ with respect to distinct eigenvalues $\lambda_1 < \cdots < \lambda_k$. As above, we can parallel translate these eigenspaces to get a global decomposition $TM = \eta_1 \oplus \cdots \oplus \eta_k$ into parallel distributions, with the property that $S|_{\eta_i} = \lambda_i \cdot I$. But then the decomposition theorem tells us that (M, g) is reducible unless $k = 1$.

Second, the skew-symmetric part has purely imaginary eigenvalues, however, we still obtain a decomposition $T_pM = F_1 \oplus \cdots \oplus F_l$ into orthogonal invariant subspaces such that $A|_{F_k}$ has complex eigenvalue $\mathrm{i}\mu_k$, where $\mu_1 < \cdots < \mu_l$. The same argument as above shows that $l = 1$. If $\mu_1 \neq 0$, then $J = \frac{1}{\mu_1}A$ is also parallel and only has i as an eigenvalue. This gives us the desired Kähler structure. □

Corollary 10.3.3. *A simply connected irreducible symmetric space is an Einstein manifold. In particular, it has nonnegative or nonpositive curvature operator according to the sign of the Einstein constant.*

10.3.2 Rough Classification of Symmetric Spaces

We are now in a position to explain the essence of what irreducible symmetric spaces look like. They are all Einstein and come in three basic categories.

Compact Type: If the Einstein constant is positive, then it follows from Myers' diameter bound (theorem 6.3.3) that the space is compact. In this case the curvature operator is nonnegative.

Flat Type: If the space is Ricci flat, then it is flat. Thus, the only Ricci flat irreducible examples are S^1 and $\mathbb{R}^1$.

Noncompact Type: When the Einstein constant is negative, then it follows from Bochner's theorem 8.2.2 on Killing fields that the space is noncompact. In this case 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 tables. Also, there is a further subdivision. Among the compact types there are Lie groups with biinvariant metrics and then all the others. Similarly, in the noncompact regime there are the duals to the biinvariant metrics and then the rest. This gives us the following division explained with Lie algebra pairs that are of the form $(\mathfrak{c}_p, \mathfrak{r}_p) = (\mathfrak{g}, \mathfrak{k})$. In all cases there is an involution with 1-eigenspace given by $\mathfrak{k}$, and $\mathfrak{k}$ is the Lie algebra of a compact group that acts on the (-1)-eigenspace $\mathfrak{t}$. One can further use corollary 10.1.8 to show that for these examples $\mathfrak{r}_p = \mathfrak{iso}_p$.

Type I: Compact irreducible symmetric spaces of the form $(\mathfrak{g}, \mathfrak{k})$, where $\mathfrak{g}$ is simple; the Lie algebra of a compact Lie group; and $\mathfrak{k} \subset \mathfrak{g}$ a maximal subalgebra, e.g., $(\mathfrak{so}(k+l), \mathfrak{so}(k) \times \mathfrak{so}(l))$.

Type II: Compact irreducible symmetric spaces $(\mathfrak{k} \oplus \mathfrak{k}, \Delta\mathfrak{k})$, where $\mathfrak{k}$ is simple and corresponds to a compact Lie group. The space is a compact Lie group with a biinvariant metric.

Type III: Noncompact symmetric spaces $(\mathfrak{g}, \mathfrak{k})$, where $\mathfrak{g}$ is simple; the Lie algebra of a non-compact Lie group; and $\mathfrak{k} \subset \mathfrak{g}$ a maximal subalgebra corresponding to a compact Lie group, e.g., $(\mathfrak{so}(k,l), \mathfrak{so}(k) \times \mathfrak{so}(l))$ or $(\mathfrak{sl}(n), \mathfrak{so}(n))$

Type IV: Noncompact symmetric spaces $(\mathfrak{k} \otimes \mathbb{C}, \mathfrak{k})$, where $\mathfrak{k}$ is simple; corresponds to a compact Lie group; and $\mathfrak{k} \otimes \mathbb{C} = \mathfrak{k} \oplus i\mathfrak{k}$ its complexification, e.g., $(\mathfrak{so}(n, \mathbb{C}), \mathfrak{so}(n))$.

Note that since compact type symmetric spaces have nonnegative curvature operator, it becomes possible to calculate their cohomology algebraically. The Bochner technique tells us that all harmonic forms are parallel. As parallel forms are invariant under the holonomy we are left with a classical invariance problem: 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 [97, vol. 5] and [76] for more on this.

10.3.3 Curvature and Holonomy

We mention, without proof, the general classification of connected irreducible holonomy groups. Berger classified all possible holonomies. Simons gave a direct proof of the fact that spaces with nontransitive holonomy must be locally symmetric, i.e., he did not use Berger's classification of holonomy groups.

Table 10.5 Holonomy Groups

$\dim = n$	Hol_p	Properties
n	$\mathrm{SO}(n)$	Generic case
$n = 2m$	$\mathrm{U}(m)$	Kähler
$n = 2m$	$\mathrm{SU}(m)$	Kähler and Ricci flat
$n = 4m$	$\mathrm{Sp}(1) \cdot \mathrm{Sp}(m)$	Quaternionic-Kähler and Einstein
$n = 4m$	$\mathrm{Sp}(m)$	Hyper-Kähler and Ricci flat
$n = 16$	$\mathrm{Spin}(9)$	Symmetric and Einstein
$n = 8$	$\mathrm{Spin}(7)$	Ricci flat
$n = 7$	G_2	Ricci flat

Theorem 10.3.4 (Berger, 1955 and Simons, 1962). *Let (M, g) be a simply connected irreducible Riemannian n-manifold. The holonomy Hol_p either 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 groups in table 10.5.*

The first important thing to understand is that while we list the groups it is important how they act on the tangent space. The same group, e.g., $\mathrm{SU}(m)$ acts irreducibly in the standard way on $\mathbb{C}^m$, but it also acts irreducibly via conjugation on $\mathfrak{su}(m) = \mathbb{R}^{m^2-1}$. It is only in the former case that the metric is forced to be Ricci flat. The latter situation occurs on the symmetric space $\mathrm{SU}(m)$.

It is curious that all but the two largest irreducible holonomy groups, $\mathrm{SO}(n)$ and $\mathrm{U}(m)$, 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 special properties. One can, using 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 even Ricci flat (see [12] for details). Note that Kähler metrics do not have to be Einstein (see exercise 4.7.24). Quaternionic Kähler manifolds are not necessarily Kähler, as $\mathrm{Sp}(1) \cdot \mathrm{Sp}(m)$ is not contained in $\mathrm{U}(2m)$, in fact $\mathrm{Sp}(1) \cdot \mathrm{Sp}(1) = \mathrm{SO}(4)$. Using a little bit of the theory of Kähler manifolds, it is not hard to see that metrics with holonomy $\mathrm{SU}(n)$ are Ricci flat. Since $\mathrm{Sp}(m) \subset \mathrm{SU}(2m)$, it follows that hyper-Kähler manifolds are Ricci flat. One can also prove that the last two holonomies occur only for Ricci flat manifolds. With the exception of the four types of Ricci flat holonomies all other holonomies occur for symmetric spaces. This follows from the above classification and the fact that the rank one symmetric spaces have holonomy $\mathrm{SO}(n)$, $\mathrm{U}(m)$, $\mathrm{Sp}(1) \cdot \mathrm{Sp}(m)$, or $\mathrm{Spin}(9)$.

This leads to another profound question. Are there compact simply connected Ricci flat spaces with holonomy $\mathrm{SU}(m)$, $\mathrm{Sp}(m)$, G_2, or $\mathrm{Spin}(7)$? The answer is yes. But it is a highly nontrivial yes. Yau got the Fields medal, in part, for establishing the $\mathrm{SU}(m)$ case. Actually, he solved the Calabi conjecture, and the holonomy question was a by-product (see, e.g., [12] for more information on the Calabi conjecture). Note that we have the Eguchi-Hanson metric (exercise 4.7.24 and 4.7.23) which is a complete Ricci flat Kähler metric and therefore has $\mathrm{SU}(2)$ as holonomy group. D. Joyce solved the cases of $\mathrm{Spin}(7)$ and G_2 by methods similar to those employed by Yau. An even more intriguing question is whether there are compact simply connected Ricci flat manifolds with $\mathrm{SO}(n)$ as a holonomy group. Note that the Schwarzschild metric (see section 4.2.5) is complete, Ricci flat, and has $\mathrm{SO}(n)$ as holonomy group. For more in-depth information on these issues we refer the reader to [12].

A general remark about how special ($\neq \mathrm{SO}(n)$) 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 manifolds with holonomy $\mathrm{Spin}(9)$ (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 $\mathrm{Spin}(7)$ and G_2 there are also special 4-forms that do the job of identifying these types of spaces.

From the classification of holonomy groups we immediately get an interesting corollary.

Corollary 10.3.5. *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 Burns-Spatzier (see [7] and [21]).

Theorem 10.3.6. *A compact Riemannian manifold of nonpositive curvature of rank* ≥ 2 *does not have transitive holonomy. In particular, it must be either reducible or locally symmetric.*

It is worthwhile mentioning that in [9] it was shown that the rank of a compact nonpositively 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 nonpositively curved manifolds. The idea is to use the flat parallel fields to show that the holonomy can't be transitive. The Berger-Simons result then shows that the manifold has to be locally symmetric if it is irreducible.

In nonnegative curvature, on the other hand, 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 [48] and in more generality in Chen's article in [51]. 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 10.3.7 (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:*

(a) (M,g) *is either reducible or locally symmetric.*
(b) $\mathrm{Hol}^0(M,g) = \mathrm{SO}(n)$ *and the universal covering is a homology sphere.*
(c) $\mathrm{Hol}^0(M,g) = \mathrm{U}\left(\frac{n}{2}\right)$ *and the universal covering is a homology* $\mathbb{CP}^{\frac{n}{2}}$.

Proof. First we use the structure theory from section 7.3.3 to conclude that the universal covering is isometric to $N \times \mathbb{R}^k$, where $k > 0$ if the fundamental group is infinite. In particular, the manifold is reducible. Therefore, we can assume that we work with a simply connected compact 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 there is a harmonic p-form by the Hodge theorem. The Bochner technique tells us that this form must be parallel, since the curvature operator is nonnegative. The idea of the proof is to check the possibilities for this when we know the holonomy.

We can assume that the manifold is irreducible and has transitive holonomy. The Ricci flat cases are impossible as the nonnegative curvature would then make the manifold flat. Thus, we have only the four possibilities $\mathrm{SO}(n)$,$\mathrm{U}\left(\frac{n}{2}\right)$, $\mathrm{Sp}(1)\,\mathrm{Sp}\left(\frac{n}{4}\right)$, or $\mathrm{Spin}(9)$. In the latter two cases one can show from holonomy considerations that the manifold must be Einstein. Tachibana's result (see theorem 9.4.8) then implies that the metric is locally symmetric. From the classification of symmetric spaces it is further possible to show that the space is isometric to either $\mathbb{HP}^{\frac{n}{4}}$ or $\mathbb{OP}^2$.

Now assume that the holonomy is $\mathrm{SO}(n)$ and that we have a parallel p-form ω. When $0 < p < n$ and $v_1,\ldots,v_p \in T_pM$ it is possible to find an element of $P \in \mathrm{SO}(n)$ such that $P(v_i) = v_i$, $i = 2,\cdots,p$ and $P(v_1) = -v_1$. Therefore, when the holonomy is $\mathrm{SO}(n)$ and ω is invariant under parallel translation, then

$$\begin{aligned}\omega\left(v_1,\ldots,v_p\right) &= \omega\left(Pv_1,\ldots,Pv_p\right)\\ &= \omega\left(-v_1,v_2,\ldots,v_p\right)\\ &= -\omega\left(v_1,\ldots,v_p\right).\end{aligned}$$

This shows that $\omega = 0$.

This leaves us with the case where the holonomy is $\mathrm{U}\left(\frac{n}{2}\right)$, i.e., the metric is Kähler. In this situation we show that the cohomology ring must be the same as that of $\mathbb{CP}^{\frac{n}{2}}$, i.e., there is a homology class $\omega \in H^2(M,\mathbb{R})$ such that any homology class is proportional to some power $\omega^k = \omega \wedge \ldots \wedge \omega$. This can be seen as follows. Since the holonomy is $\mathrm{U}\left(\frac{n}{2}\right)$ there must be an almost complex structure on the tangent spaces that is invariant under parallel translation. After type change this gives us a parallel 2-form ω. Any other parallel 2-from must be a multiple of this form by theorem 10.3.2, so $\dim H^2 = 1$. The odd cohomology groups vanish, like in the case where the holonomy is $\mathrm{SO}(n)$, since the antipodal map $P = -I \in \mathrm{U}\left(\frac{n}{2}\right)$ when n is even. More generally, consider a p-form θ with $1 < p < n$ that is invariant under $\mathrm{U}\left(\frac{n}{2}\right)$. Select an orthonormal basis $e_1, \ldots, e_n$ for T_pM. We claim that θ has the same values on any p vectors $e_{i_1}, \ldots e_{i_p}$, where $i_1 < \cdots < i_p$. There is an element $P \in \mathrm{U}\left(\frac{n}{2}\right)$ such that $P(e_i) = e_i$ for $i = 1, \ldots, p{-}1$ and $P\left(e_p\right) = e_{i_p}$ so it follows that $\theta\left(e_1, \ldots, e_p\right) = \theta\left(e_1, \ldots, e_{p-1}, e_{i_p}\right)$. This can be repeated for $p-1$ etc. This shows that all p-forms must be multiples of each other. Now the powers $\omega^k = \omega \wedge \cdots \wedge \omega$ are all nontrivial parallel forms, so they must generate H^{2k}. This shows that M has the cohomology ring of $\mathbb{CP}^{\frac{n}{2}}$. □

There are two questions left over in this classification. Namely, for the sphere and complex projective space we get only homology 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 $\mathbb{CP}^2$, 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.

There are more profound results that tell us more about the topological structure in cases (b) and (c). For case (b) one can use Ricci flow techniques to show that the space is diffeomorphic to a space of constant curvature. This is a combination of results by Hamilton (see [61]) and Böhm-Wilking (see [16]). In case (c) the universal cover is biholomorphic to $\mathbb{CP}^{\frac{n}{2}}$. This was proven by Mok (see [78]) and can now also be proven using the Ricci flow. In fact the entire result can be generalized using the Ricci flow to hold under weaker assumptions (see [19]).

Theorem 10.3.8 (Brendle and Schoen, 2008). *If (M, g) is a compact Riemannian n-manifold with nonnegative complex sectional curvature, then one of the following cases must occur:*

(a) *(M, g) is either reducible or locally symmetric.*
(b) *M is diffeomeorphic to a space of constant positive curvature.*
(c) *The universal covering of M is biholomorphic to $\mathbb{CP}^{\frac{n}{2}}$.*

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. Nevertheless, there is an example of Aravinda and Farrell showing that there are nonpositively curved manifolds that do not admit metrics with nonpositive curvature operator (see [5]).

10.4 Further Study

We have not covered all important topics about symmetric spaces. For more in-depth information we recommend the texts by Besse, Helgason, and Jost (see [12, Chapters 7,10], [13, Chapter 3], [62], and [65, Chapter 6]). Another very good text which covers the theory of Lie groups and symmetric spaces is [64]. O'Neill's book [80, Chapter 8] also has a nice elementary account of symmetric spaces. Finally, Klingenberg's book [69] has an excellent geometric account of symmetric spaces.

10.5 Exercises

EXERCISE 10.5.1. Let M be a symmetric space and $X \in \mathfrak{k}_p$, i.e., X is a nontrivial Killing field with $(\nabla X)|_p = 0$.

(1) Show that the flow for X is given by $F_s = A_{\exp_p\left(\frac{s}{2}X|_p\right)} \circ A_p$.
(2) Show that $c(t) = \exp\left(tX|_p\right)$ is an axis for F_s, e.g., $F_s(c(t)) = c(t+s)$.
(3) Show that if $c(a) = c(b)$, then $\dot{c}(a) = \dot{c}(b)$.
(4) Conclude that geodesic loops are always closed geodesics.

EXERCISE 10.5.2. Let M be a symmetric space. Show that the action of $A_p : M \to M$ on $\pi_1(M,p)$ is given by $g \mapsto g^{-1}$ and conclude that $\pi_1(M,p)$ is Abelian. Hint: Every element of $\pi_1(M,p)$ is represented by a geodesic loop which by the previous exercise is a closed geodesic.

EXERCISE 10.5.3. Let M be a symmetric space and c a geodesic in M.

(1) Let $E(t)$ be a parallel field along c. Show that $R(E,\dot{c})\dot{c}$ is also parallel.
(2) If $E(0)$ is an eigenvector for $R(\cdot,\dot{c}(0))\dot{c}(0)$ with eigenvalue κ, then $J(t) = \mathrm{sn}_\kappa(t)E(t)$ and $J(t) = \mathrm{sn}'_\kappa(t)E(t)$ are both Jacobi fields along c.
(3) Show that if $J(t)$ is a Jacobi field along c with $J(0) = 0$ and $J(t_0) = 0$, then $\dot{J}\left(\frac{t_0}{2}\right) = 0$.
(4) With J as in (3) construct a geodesic variation $c(s,t)$ such that $\frac{\partial c}{\partial s}(0,t) = J(t)$, $c(s,0) = c(0)$, and $c(s,t_0) = c(t_0)$.

EXERCISE 10.5.4. Let M be a symmetric space.

(1) Show directly that if M is compact, then $\sec \geq 0$. Hint: Argue by contradiction and produce a Jacobi field that is unbounded along a geodesic.
(2) Show that if c is a closed geodesic, then $R(\cdot,\dot{c})\dot{c}$ has no negative eigenvalues.
(3) Show that if M has $\mathrm{Ric} > 0$, then $\sec \geq 0$.

EXERCISE 10.5.5. Assume that M has nonpositive or nonnegative sectional curvature. Let c be a geodesic and E a parallel field along c. Show that the following conditions are equivalent.

(1) $g(R(E,\dot{c})\dot{c},E)=0$ everywhere.
(2) $R(E,\dot{c})\dot{c}=0$ everywhere.
(3) E is a Jacobi field.

EXERCISE 10.5.6. (1) Let $\mathfrak{g}$ be a real Lie algebra with Killing form B. Show that the Killing form of the complixification $\mathfrak{g}\otimes\mathbb{C}$ is simply the complexification of B.
(2) Show that the Killing forms of $\mathfrak{gl}(n,\mathbb{C})$ and $\mathfrak{gl}(n)$ are given by $B(X,Y)=2n\mathrm{tr}(XY)-2\mathrm{tr}X\mathrm{tr}Y$. Hint: As a basis use the matrices $E_{ij}=\left[\delta_{is}\delta_{jt}\right]_{1\le s,t\le n}$.
(3) Show that on $\mathfrak{sl}(n,\mathbb{C})=\mathfrak{sl}(n)\otimes\mathbb{C}$ the Killing form is $B(X,Y)=2n\mathrm{tr}(XY)$. Hint: Use (2) and the fact that $I\in\mathfrak{gl}(n,\mathbb{C})$ commutes with all elements in $\mathfrak{sl}(n,\mathbb{C})$.
(4) Show that $\mathfrak{sl}(k+l,\mathbb{C})=\mathfrak{su}(k,l)\otimes\mathbb{C}$, and conclude that $\mathfrak{sl}(k+l)$ and $\mathfrak{su}(k,l)$ have Killing form $B(X,Y)=2(k+l)\mathrm{tr}(XY)$.
(5) Show that on $\mathfrak{so}(n,\mathbb{C})=\mathfrak{so}(n)\otimes\mathbb{C}$ the Killing form is given by $B(X,Y)=(n-2)\mathrm{tr}(XY)$. Hint: Use the basis $E_{ij}-E_{ji}$, $i<j$.
(6) Show that $\mathfrak{so}(k+l,\mathbb{C})=\mathfrak{so}(k,l)\otimes\mathbb{C}$, and conclude that $\mathfrak{so}(k,l)$ has Killing form $B(X,Y)=(k+l-2)\mathrm{tr}(XY)$.

EXERCISE 10.5.7. Show that $\mathrm{GL}^{+}(p+q,\mathbb{R})/\mathrm{SO}(p,q)$ defines a symmetric space and that it can be identified with the nondegenerate bilinear forms on $\mathbb{R}^{p+q}$ that have index q.

EXERCISE 10.5.8. Show that $\mathrm{U}(p,q)/\mathrm{SO}(p,q)$ defines a symmetric space and that $\mathfrak{u}(p,q)\otimes\mathbb{C}=\mathfrak{gl}(p+q,\mathbb{C})$.

EXERCISE 10.5.9. Show that the holonomy of $\mathbb{CP}^n$ is $\mathrm{U}(n)$.

EXERCISE 10.5.10. Show that a covering space of a symmetric space is also a symmetric space. Show by example that the converse is not necessarily true.

EXERCISE 10.5.11. Show that a manifold is flat if and only if the holonomy is discrete, i.e., $\mathfrak{hol}_p=\{0\}$.

EXERCISE 10.5.12. Show that a compact Riemannian manifold with irreducible restricted holonomy and $\mathrm{Ric}\ge 0$ has finite fundamental group.

EXERCISE 10.5.13. Which known spaces can be described by $\mathrm{SL}(2,\mathbb{R})/\mathrm{SO}(2)$ and $\mathrm{SL}(2,\mathbb{C})/\mathrm{SU}(2)$?

EXERCISE 10.5.14. Show that the holonomy of a Riemannian manifold is contained in $\mathrm{U}(m)$ if and only if it has a Kähler structure.

EXERCISE 10.5.15. Show that if a homogeneous space has $\mathfrak{iso}_p=\mathfrak{so}\left(T_pM\right)$ at some point, then it has constant curvature.

EXERCISE 10.5.16. Show that the subalgebras $\mathfrak{so}(k)\times\mathfrak{so}(n-k)$ and $\mathfrak{u}\left(\frac{n}{2}\right)$ are maximal in $\mathfrak{so}(n)$.

EXERCISE 10.5.17. Show that $\mathfrak{su}(m)\subset\mathfrak{so}\left(m^2-1\right)$. Hint: Let $\mathfrak{su}(m)$ act on itself.

EXERCISE 10.5.18. Show that for any Riemannian manifold $\mathfrak{r}_p \subset \mathfrak{hol}_p$. Give an example where equality does not hold.

EXERCISE 10.5.19. Show that for a symmetric space $\mathfrak{r}_p = \mathfrak{hol}_p$. Use this to show that unless the curvature is constant $\mathfrak{r}_p = \mathfrak{hol}_p = \mathfrak{iso}_p$ provided $\mathfrak{r}_p \subset \mathfrak{so}\left(T_pM\right)$ is maximal.

EXERCISE 10.5.20. Show that $\mathrm{SO}\,(n,\mathbb{C})\,/\mathrm{SO}\,(n)$ and $\mathrm{SL}\,(n,\mathbb{C})\,/\mathrm{SU}\,(n)$ are symmetric spaces with nonpositive curvature operator.

EXERCISE 10.5.21. The *quaternionic projective space* is defined as being the quaternionic lines in $\mathbb{H}^{n+1}$. This was discussed when $n = 1$ in exercise 1.6.22. Define the *symplectic group* $\mathrm{Sp}\,(n) \subset \mathrm{SU}\,(2n) \subset \mathrm{SO}\,(4n)$ as the orthogonal matrices that commute with the three complex structures generated by $\mathrm{i}, \mathrm{j}, \mathrm{k}$ on $\mathbb{R}^{4n}$. An alternative way of looking at this group is by considering $n \times n$ matrices A with quaternionic entries such that

$$A^{-1} = A^*.$$

Show that if we think of $\mathbb{H}^{n+1}$ as a right (or left) $\mathbb{H}$ module, then the space of quaternionic lines can be written as

$$\mathbb{HP}^n = \mathrm{Sp}\,(n+1)\,/\,(\mathrm{Sp}\,(1) \times \mathrm{Sp}\,(n))\,.$$

EXERCISE 10.5.22. Construct the hyperbolic analogues of the complex projective spaces. Show that they have negative curvature and are quarter pinched.

EXERCISE 10.5.23. Give a Lie algebra description of a locally symmetric space (not necessarily complete). Explain why this description corresponds to a global symmetric space. Conclude that a simply connected locally symmetric space admits a monodromy map into a unique global symmetric space. Show that if the locally symmetric space is complete, then the monodromy map is bijective.

EXERCISE 10.5.24. Show that if an irreducible symmetric space has strictly positive or negative curvature operator, then it has constant curvature.

EXERCISE 10.5.25. Let M be a symmetric space. Show that if $X \in \mathfrak{t}_p$ and $Y \in \mathfrak{iso}_p$, then $[X, Y] \in \mathfrak{t}_p$.

EXERCISE 10.5.26. Let M be a symmetric space and $X, Y, Z \in \mathfrak{t}_p$. Show that

$$R\,(X, Y)\,Z = [L_X, L_Y]\,Z,$$
$$\mathrm{Ric}\,(X, Y) = -\,\mathrm{tr}\,([L_X, L_Y])\,.$$

EXERCISE 10.5.27. Consider a Riemannian manifold M and a p-form ω on T_pM. Show that ω has an extension to a parallel form on M if and only if ω is invariant under $\mathrm{Hol}_p\,(M)$.

Chapter 11
Convergence

In this chapter we offer 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 metric spaces to converge to a metric space. The first section centers on the weakest convergence concept: Gromov-Hausdorff convergence. The next section covers some of the elliptic regularity theory needed for the later developments that use stronger types of convergence. In the third section we develop the idea of norms of Riemannian manifolds as an intermediate step towards understanding convergence theory as an analogue to the easier Hölder theory for functions. Finally, in the fourth section we establish the geometric version of the convergence theorem of Riemannian geometry by Cheeger and Gromov as well as its generalizations by Anderson and others. These convergence theorems contain Cheeger's finiteness theorem stating that certain very general 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 used here go back to Shikata's proof of the differentiable sphere theorem. Cheeger's thesis also contains the idea that abstract manifolds can converge to each other. In fact, he proved his finiteness theorem by showing that certain classes of manifolds are precompact in various topologies. Gromov further developed the theory of convergence to the form presented here that starts with the weaker Gromov-Hausdorff convergence of metric spaces. His 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.

P. Petersen, *Riemannian Geometry*, Graduate Texts in Mathematics 171,
DOI 10.1007/978-3-319-26654-1_11

11.1 Gromov-Hausdorff Convergence

11.1.1 Hausdorff Versus Gromov Convergence

At the beginning of the twentieth century, Hausdorff introduced what is now called the *Hausdorff distance* between subsets of a metric space. If $(X, |\cdot\cdot|)$ is the metric space and $A, B \subset X$, then

$$\begin{aligned} d\left(A, B\right) &= \inf\left\{|ab| \mid a \in A,\ b \in B\right\}, \\ B\left(A, \varepsilon\right) &= \left\{x \in X \mid |xA| < \varepsilon\right\}, \\ d_H\left(A, B\right) &= \inf\left\{\varepsilon \mid A \subset B\left(B, \varepsilon\right),\ B \subset B\left(A, \varepsilon\right)\right\}. \end{aligned}$$

Thus, $d\left(A, B\right)$ is small if some points in these sets are close, while the *Hausdorff distance* $d_H\left(A, B\right)$ is small if and only if 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 compact subsets of X and that this collection is compact when X is compact.

We shall concern ourselves only with compact or *proper* metric spaces. The latter by definition have 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 the Hausdorff distance concept to a distance between abstract metric spaces. If X and Y are metric spaces, then an *admissible* metric on the disjoint union $X \cup Y$ is a metric that extends the given metrics on X and Y.

With this the *Gromov-Hausdorff distance* is defined as

$$d_{G-H}\left(X, Y\right) = \inf\left\{d_H\left(X, Y\right) \mid \text{admissible metrics on } X \cup Y\right\}.$$

Thus, we try to place a metric on $X \cup Y$ that extends the metrics on X and Y, such that X and Y are as close as possible in the Hausdorff distance. In other words, we are trying to define distances between points in X and Y without violating the triangle inequality.

Example 11.1.1. If Y is the one-point space, then

$$\begin{aligned} d_{G-H}\left(X, Y\right) &\leq \operatorname{rad} X \\ &= \inf_{y \in X} \sup_{x \in X} |xy| \\ &= \text{radius of smallest ball covering } X. \end{aligned}$$

Example 11.1.2. Using $|xy| = D/2$ for all $x \in X,\ y \in Y$, where $\operatorname{diam} X, \operatorname{diam} Y \leq D$ shows that

$$d_{G-H}\left(X, Y\right) \leq D/2.$$

Let $(\mathscr{M}, d_{G-H})$ denote the collection of compact metric spaces. We wish to consider 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.

Proposition 11.1.3. *If X and Y are compact metric spaces with $d_{G-H}(X, Y) = 0$, then X and Y are isometric.*

Proof. Choose a sequence of metrics $|\cdot\cdot|_i$ on $X \cup Y$ such that the Hausdorff distance between X and Y in this metric is $< i^{-1}$. Then we can find (possibly discontinuous) maps

$$
\begin{aligned}
I_i &: X \to Y, \text{ where } |xI_i(x)|_i \leq i^{-1}, \\
J_i &: Y \to X, \text{ where } |yJ_i(y)|_i \leq i^{-1}.
\end{aligned}
$$

Using the triangle inequality and that $|\cdot\cdot|_i$ restricted to either X or Y is the given metric $|\cdot\cdot|$ on these spaces yields

$$
\begin{aligned}
|I_i(x_1)\, I_i(x_2)| &\leq 2i^{-1} + |x_1x_2|, \\
|J_i(y_1)\, J_i(y_2)| &\leq 2i^{-1} + |y_1y_2|, \\
|x\, J_i \circ I_i(x)| &\leq 2i^{-1}, \\
|y\, I_i \circ J_i(y)| &\leq 2i^{-1}.
\end{aligned}
$$

We construct $I : X \to Y$ and $J : Y \to X$ as limits of these maps in the same way the Arzela-Ascoli lemma is proved. For each x the sequence $(I_i(x))$ in Y has an accumulation point since Y is compact. Let $A \subset X$ be select a countable dense set. Using a diagonal argument select a subsequence I_{i_j} such that $I_{i_j}(a) \to I(a)$ for all $a \in A$. The first inequality shows that I is distance decreasing on A. In particular, it is uniformly continuous and thus has a unique extension to a map $I : X \to Y$, which is also distance decreasing. In a similar fashion we also get a distance decreasing map $J : Y \to X$.

The last two inequalities imply that I and J are inverses to each other. Thus, both I and J are isometries. □

The symmetry and the triangle inequality are easily established for d_{G-H}. Thus, $(\mathscr{M}, d_{G-H})$ becomes a pseudo-metric space, i.e., the equivalence classes form a metric space. We prove below that this metric space is complete and separable. First we show how spaces can be approximated by finite metric spaces.

Example 11.1.4. Let X be compact and $A \subset X$ a finite subset 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 clear that if we use the metric on A induced by X, then $d_{G-H}(X, A) \leq \varepsilon$. The importance of this remark is that for any $\varepsilon > 0$ there exist finite ϵ-dense subsets of X since X is compact. To be consistent with our definition of the abstract distance we should put a metric on $X \cup A$. We can do this by selecting very small $\delta > 0$ and defining $|xa|_{X\cup A} = \delta + |xa|_X$ for $x \in X$ and $a \in A$. Thus $d_{G-H}(X,A) \leq \epsilon + \delta$. Finally, let $\delta \to 0$ to get the estimate.

Example 11.1.5. Suppose we have ε-dense subsets

$$A = \{x_1, \ldots, x_k\} \subset X,\ B = \{y_1, \ldots, y_k\} \subset Y,$$

with the further property that

$$\left|\left|x_i x_j\right| - \left|y_i y_j\right|\right| \leq \varepsilon,\ 1 \leq i, j \leq k.$$

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 on $A \cup B$ that makes A and B ε-Hausdorff close. Define

$$\begin{aligned} |x_i y_i| &= \varepsilon, \\ \left|x_i y_j\right| &= \min_k \left\{|x_i x_k| + \varepsilon + \left|y_j y_k\right|\right\}. \end{aligned}$$

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

$$\begin{aligned} \left|x_i y_j\right| &\leq |x_i z| + \left|y_j z\right|, \\ \left|x_i x_j\right| &\leq |y_k x_i| + \left|y_k x_j\right|, \\ \left|y_i y_j\right| &\leq |x_k y_i| + \left|x_k y_j\right|. \end{aligned}$$

It suffices to check the first two cases as the third is similar to the second. For the first we can assume that $z = x_k$ and find l such that

$$\left|y_j x_k\right| = \varepsilon + \left|y_j y_l\right| + |x_l x_k|.$$

Hence,

$$\begin{aligned} |x_i x_k| + \left|y_j x_k\right| &= |x_i x_k| + \varepsilon + \left|y_j y_l\right| + |x_l x_k| \\ &\geq |x_i x_l| + \varepsilon + \left|y_j y_l\right| \\ &\geq \left|x_i y_j\right|. \end{aligned}$$

For the second case select l, m with

$$\begin{aligned} |y_k x_i| &= |y_k y_l| + \varepsilon + |x_l x_i|, \\ \left|y_k x_j\right| &= |y_k y_m| + \varepsilon + \left|x_m x_j\right|. \end{aligned}$$

The assumption about the metrics on A and B then lead to

$$\begin{aligned}|y_kx_i| + |y_kx_j| &= |y_ky_l| + \varepsilon + |x_lx_i| + |y_ky_m| + \varepsilon + |x_mx_j| \\ &\geq |x_kx_l| + |x_lx_i| + |x_kx_m| + |x_mx_j| \\ &\geq |x_ix_j| .\end{aligned}$$

Example 11.1.6. Suppose $M_k = S^3/\mathbb{Z}_k$ with the usual metric induced from $S^3(1)$. Then we have a Riemannian submersion $M_k \to S^2(1/2)$ whose fibers have diameter $2\pi/k \to 0$ as $k \to \infty$. Using the previous example it follows that $M_k \to S^2(1/2)$ in the Gromov-Hausdorff topology.

Example 11.1.7. One can similarly see that the Berger metrics $(S^3, g_\varepsilon) \to S^2(1/2)$ 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 11.1.8. *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 establish 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 $|\cdot\cdot|_{i,i+1}$ on $X_i \cup X_{i+1}$ making these spaces 2^{-i}-Hausdorff close. Now define a metric $|\cdot\cdot|_{i,i+j}$ on $X_i \cup X_{i+j}$ by

$$|x_ix_{i+j}|_{i,i+j} = \min_{\{x_{i+k} \in X_{i+k}\}} \left\{ \sum_{k=0}^{j-1} |x_{i+k}x_{i+k+1}| \right\} .$$

This defines a metric $|\cdot\cdot|$ on $Y = \cup_i X_i$ with the property that $d_H(X_i, X_{i+j}) \leq 2^{-i+1}$. The metric space is not complete, but the "boundary" of the completion is exactly our desired limit space. To define it, first consider

$$\hat{X} = \left\{ \{x_i\} \mid x_i \in X_i \text{ and } |x_ix_j| \to 0 \text{ as } i,j \to \infty \right\} .$$

This space has a pseudo-metric defined by

$$|\{x_i\}\{y_i\}| = \lim_{i\to\infty} |x_iy_i| .$$

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

$$\{x_i\} \sim \{y_i\} \text{ iff } |\{x_i\}\{y_i\}| = 0.$$

Now we can extend the metric on Y to one on $X \cup Y$ by declaring

$$|x_k \{x_i\}| = \lim_{i\to\infty} |x_k x_i| .$$

Using that $d_H\left(X_j, X_{j+1}\right) \leq 2^{-j}$, we can for any $x_i \in X_i$ find a sequence $\left\{x_{i+j}\right\} \in \hat{X}$ such that $x_{i+0} = x_i$ and $\left|x_{i+j}x_{i+j+1}\right| \leq 2^{-j}$. Then we must have $\left|x_i \left\{x_{i+j}\right\}\right| \leq 2^{-i+1}$. 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 $|y_k \{y_i\}| \leq 2^{-k+1}$ for all k. Thus, X is 2^{-i+1}-close to X_i. □

From the proof of this theorem we obtain 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 $X \cup (\cup_i X_i)$ in which X_i Hausdorff converges to X. With a choice of such a metric it 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 a choice of ambient metric on $X \cup (\cup_i X_i)$.

There is an equivalent way of picturing convergence. For a compact metric space X define $C(X)$ as the continuous functions on X and $L^\infty(X)$ as 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 construct a map $X \to L^\infty(X)$, by sending x to the continuous function $z \mapsto |xz|$. This is usually called the *Kuratowski embedding* when we consider it as a map into $C(X)$. The triangle inequality implies that this is a distance preserving map. Thus, any compact metric space is isometric to a subset of some Banach space $L^\infty(X)$. The important observation 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). Moreover, if $X \subset Y$, then $L^\infty(X) \subset L^\infty(Y)$, by extending a function on X to vanish on $Y - X$. Moreover, any compact metric space is Borel equivalent to a subset of $[0, 1]$. In particular, any compact metric space is isometric to a 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.

11.1.2 Pointed Convergence

So far, we haven't dealt with noncompact spaces. 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

$$d_{G-H}\left((X,x),(Y,y)\right) = \inf\left\{d_H\left(X,Y\right) + |xy|\right\}.$$

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 $\mathscr{M}_* = \{(X, x, |\cdot\cdot|)\}$ in the following way: We say that

$$(X_i, x_i, |\cdot\cdot|_i) \to (X, x, |\cdot\cdot|)$$

in the *pointed Gromov-Hausdorff topology* if for all R there is a sequence $R_i \to R$ such that the closed metric balls

$$\left(\bar{B}\left(x_i, R_i\right), x_i, |\cdot\cdot|_i\right) \to \left(\bar{B}\left(x, R\right), x, |\cdot\cdot|\right)$$

converge with respect to the pointed Gromov-Hausdorff metric.

11.1.3 Convergence of Maps

We also need to address *convergence of maps*. Suppose we have

$$f_k : X_k \to Y_k,$$
$$X_k \to X,$$
$$Y_k \to Y.$$

Then we say that f_k converges to $f : X \to Y$ if for every sequence $x_k \in X_k$ converging to $x \in X$ it follows 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 also a very strong type 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 maps 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

$$F : \left(\bigcup_i X_i \right) \to Y \cup \left(\bigcup_i Y_i \right).$$

The sequence converges if and only if this map has an extension

$$X \cup \left(\bigcup_i X_i \right) \to Y \cup \left(\bigcup_i Y_i \right),$$

in which case the limit map is the restriction to X. Thus, when X_i are compact it follows that a sequence is convergent if and only if the map

$$F : \left(\bigcup_i X_i \right) \to Y \cup \left(\bigcup_i Y_i \right)$$

is uniformly continuous.

A sequence of functions as above is called *equicontinuous,* if for every $\varepsilon > 0$ and $x_k \in X_k$ there is an $\delta > 0$ such that $f_k \left(B \left(x_k, \delta \right) \right) \subset B \left(f_k \left(x_k \right), \varepsilon \right)$ for all k. A sequence is equicontinuous when, for example, all the functions are Lipschitz continuous with the same Lipschitz constant. As for standard equicontinuous sequences, we have the Arzela-Ascoli lemma:

Lemma 11.1.9. *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 pointed we also assume that f_k preserves the base point.*

Proof. The standard proof carries over without much change. Namely, first choose dense subsets $A_i = \left\{ a_1^i, a_2^i, \ldots \right\} \subset X_i$ such that $a_j^i \to a_j \in X$ as $i \to \infty$. Then also, $A = \left\{ a_j \right\} \subset X$ 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. □

11.1.4 Compactness of Classes of Metric Spaces

We now turn our attention to conditions that ensure convergence of spaces. More precisely we want some good criteria for when a collection of (pointed) spaces is precompact (i.e., closure is compact).

For a compact metric space X, define the *capacity* and *covering functions* as follows

$$\begin{aligned}\operatorname{Cap}(\varepsilon) = \operatorname{Cap}_X(\varepsilon) &= \text{ maximum number of disjoint } \tfrac{\epsilon}{2}\text{-balls in } X,\\ \operatorname{Cov}(\varepsilon) = \operatorname{Cov}_X(\varepsilon) &= \text{ minimum number of } \varepsilon\text{-balls it takes to cover } X.\end{aligned}$$

First, note that $\operatorname{Cov}(\varepsilon) \leq \operatorname{Cap}(\varepsilon)$. To see this, select a maximum number of disjoint balls $B(x_i, \epsilon/2)$ and consider the collection $B(x_i, \varepsilon)$. In case the latter balls do not cover X there exists $x \in X - \cup B(x_i, \varepsilon)$. This would imply that $B(x, \epsilon/2)$ is disjoint from all of the balls $B(x_i, \epsilon/2)$. Thus showing that the original $\epsilon/2$-balls did not form a maximal disjoint family.

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:

$$\begin{aligned}\operatorname{Cov}_X(\varepsilon + 2\delta) &\leq \operatorname{Cov}_Y(\varepsilon),\\ \operatorname{Cap}_X(\varepsilon) &\geq \operatorname{Cap}_Y(\varepsilon + 2\delta).\end{aligned}$$

With this information we can characterize precompact classes of compact metric spaces.

Proposition 11.1.10 (Gromov, 1980). *For a class $\mathscr{C} \subset (\mathscr{M}, d_{G-H})$ all of whose diameters are bounded by $D < \infty$, the following statements are equivalent:*

(1) *$\mathscr{C}$ is precompact, i.e., every sequence in $\mathscr{C}$ has a subsequence that is convergent in $(\mathscr{M}, d_{G-H})$.*
(2) *There is a function $N_1(\varepsilon) : (0, \alpha) \to (0, \infty)$ such that $\operatorname{Cap}_X(\varepsilon) \leq N_1(\varepsilon)$ for all $X \in \mathscr{C}$.*
(3) *There is a function $N_2(\varepsilon) : (0, \alpha) \to (0, \infty)$ such that $\operatorname{Cov}_X(\varepsilon) \leq N_2(\varepsilon)$ for all $X \in \mathscr{C}$.*

Proof. (1) $\Rightarrow$ (2): If $\mathscr{C}$ is precompact, then for every $\varepsilon > 0$ we can find $X_1, \ldots, X_k \in \mathscr{C}$ such that for any $X \in \mathscr{C}$ we have that $d_{G-H}(X, X_i) < \epsilon/4$ for some i. Then

$$\operatorname{Cap}_X(\varepsilon) \leq \operatorname{Cap}_{X_i}\left(\tfrac{\epsilon}{2}\right) \leq \max_i \operatorname{Cap}_{X_i}\left(\tfrac{\epsilon}{2}\right).$$

This gives a bound for $\operatorname{Cap}_X(\varepsilon)$ for each $\varepsilon > 0$.

(2) $\Rightarrow$ (3) Use $N_2 = N_1$.

(3) $\Rightarrow$ (1): It suffices to show that $\mathscr{C}$ is totally bounded, i.e., for each $\varepsilon > 0$ we can find finitely many metric spaces $X_1, \ldots, X_k \in \mathscr{M}$ such that any metric space in $\mathscr{C}$ is within ε of some X_i in the Gromov-Hausdorff metric. Since $\operatorname{Cov}_X(\epsilon/2) \leq N(\epsilon/2)$, we know that any $X \in \mathscr{C}$ is within $\frac{\varepsilon}{2}$ of a finite subset with at most $N\left(\frac{\varepsilon}{2}\right)$ elements in it. Using the induced metric we think of these finite subsets as finite metric spaces. The metric on such a finite metric space consists of a matrix $\left(d_{ij}\right)$, $1 \leq i, j \leq N(\epsilon/2)$, where each entry satisfies $d_{ij} \in [0, D]$. From among all such

finite metric spaces, it is possible to select a finite number of them such that any of the matrices $\left(d_{ij}\right)$ is within $\epsilon/2$ of one matrix from the finite selection of matrices. This means that the spaces are within $\epsilon/2$ of each other. We have then found the desired finite collection of metric spaces. □

As a corollary we also obtain a precompactness theorem in the pointed category.

Corollary 11.1.11. *A collection $\mathscr{C} \subset \mathscr{M}_*$ is precompact if and only if for each $R > 0$ the collection*

$$\left\{\bar{B}(x,R) \mid \bar{B}(x,R) \subset (X,x) \in \mathscr{C}\right\} \subset (\mathscr{M}, d_{G-H})$$

is precompact.

In order to achieve compactness we need a condition that is relatively easy to check.

We say that a metric space X satisfies the *metric doubling condition* with constant C, if each metric ball $B(p,R)$ can be covered by at most C balls of radius $R/2$.

Proposition 11.1.12. *If all metric spaces in a class $\mathscr{C} \subset (\mathscr{M}, d_{G-H})$ satisfy the metric doubling condition with constant $C < \infty$ and all have diameters bounded by $D < \infty$, then the class is precompact in the Gromov-Hausdorff metric.*

Proof. Every metric space $X \in \mathscr{C}$ can be covered by at most C^N balls of radius $2^{-N}D$. Consequently, X can be covered by at most C^N balls of radius $\varepsilon \in \left[2^{-N}D, 2^{-N+1}D\right]$. This gives us the desired estimate on $\mathrm{Cov}_X(\varepsilon)$. □

Using the relative volume comparison theorem we can show

Corollary 11.1.13. *For any integer $n \geq 2$, $k \in \mathbb{R}$, and $D > 0$ 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). Fix $R > 0$. We have to show that there can't be too many disjoint balls inside $\bar{B}(x,R) \subset M$. To see this, suppose $B(x_1,\varepsilon), \ldots, B(x_N,\varepsilon) \subset \bar{B}(x,R)$ are disjoint. If $B(x_i,\varepsilon)$ is the ball with the smallest volume, we have

$$N \leq \frac{\mathrm{vol}B(x,R)}{\mathrm{vol}B(x_i,\varepsilon)} \leq \frac{\mathrm{vol}B(x_i,2R)}{\mathrm{vol}B(x_i,\varepsilon)} \leq \frac{v(n,k,2R)}{v(n,k,\varepsilon)}.$$

This gives the desired bound. □

It seems intuitively clear that an n-dimensional space should have $\mathrm{Cov}(\varepsilon) \sim \varepsilon^{-n}$ as $\varepsilon \to 0$. The Minkowski dimension of a metric space is defined as

$$\dim X = \limsup_{\varepsilon \to 0} \frac{\log \mathrm{Cov}(\varepsilon)}{-\log \varepsilon}.$$

This definition will in fact give the right answer for Riemannian manifolds. Some fractal spaces might, however, have non-integral 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.

Lemma 11.1.14. *Let $\mathscr{C}(N(\varepsilon))$ be the collection of metric spaces with* $\mathrm{Cov}(\varepsilon) \le N(\varepsilon)$. *If N is continuous, then $\mathscr{C}(N(\varepsilon))$ 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

$$\mathrm{Cov}_X(\varepsilon) \le \mathrm{Cov}_{X_i}(\varepsilon - 2d_{G-H}(X, X_i)) \le N(\varepsilon - 2d_{G-H}(X, X_i))$$

and

$$N(\varepsilon - 2d_{G-H}(X, X_i)) \to N(\varepsilon) \text{ as } i \to \infty.$$

□

11.2 Hölder Spaces and Schauder Estimates

First, we define the Hölder norms and Hölder spaces, and 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 book by Courant and Hilbert [35], especially chapter IV, and the thorough text [50], especially chapters 1–6. A more modern text that also explains how PDEs are used in geometry, including some of the facts we need is [99], especially vol. III.

11.2.1 Hölder Spaces

Fix a bounded domain $\Omega \subset \mathbb{R}^n$. The bounded continuous functions from Ω to $\mathbb{R}^k$ are denoted by $C^0(\Omega, \mathbb{R}^k)$, and we use the sup-norm

$$\|u\|_{C^0} = \sup_{x \in \Omega} |u(x)|$$

on this space. This makes $C^0\left(\Omega,\mathbb{R}^k\right)$ 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 $C^m\left(\Omega,\mathbb{R}^k\right)$ as the functions with m continuous partial derivatives. Using multi-index notation, we define

$$\partial^I u = \partial_1^{i_1}\cdots\partial_n^{i_n}u = \frac{\partial^{|I|}u}{\partial\left(x^1\right)^{i_1}\cdots\partial\left(x^n\right)^{i_n}},$$

where $I = (i_1,\ldots,i_n)$ and $|I| = i_1+\cdots+i_n$. Then the C^m-norm is

$$\|u\|_{C^m} = \|u\|_{C^0} + \sum_{1\leq|I|\leq m}\left\|\partial^I u\right\|_{C^0}.$$

This norm does result in a Banach space, but the inclusions

$$C^m\left(\Omega,\mathbb{R}^k\right)\subset C^{m-1}\left(\Omega,\mathbb{R}^k\right)$$

are not closed subspaces. For instance, $f(x) = |x|$ is in the closure of

$$C^1\left([-1,1],\mathbb{R}\right)\subset C^0\left([-1,1],\mathbb{R}\right).$$

To accommodate this problem, we define for each $\alpha\in(0,1]$ the C^α-pseudo-norm 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.

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-pseudo-norm. 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, then we write $\|u\|_{C^{m,\alpha},\Omega}$. With this notation we can show

Lemma 11.2.1. *$C^{m,\alpha}\left(\Omega,\mathbb{R}^k\right)$ 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. Since it is also a Cauchy sequence in $C^0\left(\Omega, \mathbb{R}^k\right)$ we have that $u_i \to u \in C^0$ in the C^0-norm. For fixed $x \neq y$ 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 $u \in C^{\alpha}$.

Finally select $\varepsilon > 0$ and N so that for $i, j \geq N$ and $x \neq y$

$$\frac{\left|\left(u_i(x) - u_j(x)\right) - \left(u_i(y) - u_j(y)\right)\right|}{|x-y|^{\alpha}} \leq \varepsilon.$$

If we let $j \to \infty$, this shows that

$$\frac{|(u_i(x) - u(x)) - (u_i(y) - u(y))|}{|x-y|^{\alpha}} \leq \varepsilon.$$

Hence $u_i \to u$ in the C^{α}-topology.

Now for the last statement. A bounded sequence in $C^{\alpha}\left(\Omega, \mathbb{R}^k\right)$ is equicontinuous so the Arzela-Ascoli lemma shows that the inclusion $C^{\alpha}\left(\Omega, \mathbb{R}^k\right) \subset C^0\left(\Omega, \mathbb{R}^k\right)$ 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

$$\|u\|_{\beta} \leq \left(\|u\|_{\alpha}\right)^{\beta/\alpha} \cdot \left(2 \cdot \|u\|_{C^0}\right)^{1-\beta/\alpha}.$$

Therefore, a sequence that converges in C^0 and is bounded in C^{α}, also converges in C^{β}, as long as $\beta < \alpha \leq 1$. □

11.2.2 Elliptic Estimates

We now turn our attention to *elliptic operators* of the form

$$Lu = a^{ij}\partial_i\partial_j u + b^i\partial_i u = f,$$

where $a^{ij} = a^{ji}$ and a^{ij}, b^i are functions. The operator is called *elliptic* when the matrix $\left(a^{ij}\right)$ is positive definite. Throughout we assume that all eigenvalues for $\left(a^{ij}\right)$ lie in some interval $\left[\lambda, \lambda^{-1}\right]$, $\lambda > 0$, and that the coefficients satisfy $\left\|a^{ij}\right\|_\alpha \leq \lambda^{-1}$ and $\left\|b^i\right\|_\alpha \leq \lambda^{-1}$. We state without proof the a priori estimates, usually called the *Schauder* or *elliptic estimates*, that we need.

Theorem 11.2.2. *Let $\Omega \subset \mathbb{R}^n$ be an open domain of diameter $\leq D$ and $K \subset \Omega$ a subdomain such that $d(K, \partial\Omega) \geq \delta$. If $\alpha \in (0,1)$, 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 locally almost constant. A partition of unity type 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, as the Laplacian on manifolds in local coordinates is of that form

$$Lu = \Delta_g u = \frac{1}{\sqrt{\det g_{ij}}} \partial_i \left(\sqrt{\det g_{ij}} \cdot g^{ij} \cdot \partial_j u \right).$$

Thus

$$\int vLu\mathrm{vol} = \int v\partial_i \left(\sqrt{\det g_{ij}} \cdot g^{ij} \cdot \partial_j u \right).$$

The above theorem has an almost immediate corollary.

Corollary 11.2.3. *If in addition we assume that* $\left\|a^{ij}\right\|_{C^{m,\alpha}}, \left\|b^i\right\|_{C^{m,\alpha}} \leq \lambda^{-1}$, *then there is a constant* $C = C(n, m, \alpha, \lambda, \delta, D)$ *such that*

$$\|u\|_{C^{m+2,\alpha},K} \leq C \left(\|Lu\|_{C^{m,\alpha},\Omega} + \|u\|_{C^{\alpha},\Omega} \right).$$

And on a domain with smooth boundary,

$$\|u\|_{C^{m+2,\alpha},\Omega} \leq C \left(\|Lu\|_{C^{m,\alpha},\Omega} + \|\varphi\|_{C^{m+2,\alpha},\partial\Omega} \right).$$

The Schauder estimates can be used to show that the Dirichlet problem always has a unique solution.

Theorem 11.2.4. *Suppose* $\Omega \subset \mathbb{R}^n$ *is a bounded domain with smooth boundary. Then the Dirichlet problem*

$$Lu = f, \ u|_{\partial\Omega} = \varphi$$

always has a unique solution $u \in C^{2,\alpha}(\Omega)$ *iff* $f \in C^{\alpha}(\Omega)$ *and* $\varphi \in C^{2,\alpha}(\partial\Omega)$.

Observe that uniqueness is an immediate consequence of the maximum principle. The existence part requires more work.

11.2.3 Harmonic Coordinates

The above theorems make it possible to introduce *harmonic coordinates* on Riemannian manifolds.

Lemma 11.2.5. *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* $x = \left(x^1, \ldots, x^n\right) : U \to \mathbb{R}^n$, *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 $y = (y^1, \dots, y^n)$ 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 written as $g_{ij} = g(\partial_i, \partial_j) = g\left(\frac{\partial}{\partial y^i}, \frac{\partial}{\partial y^j}\right)$ in the standard Cartesian coordinates $(y^1, \dots, y^n)$. We must then find a coordinate transformation $y \mapsto x$ such that

$$\Delta x^k = \frac{1}{\sqrt{\det g_{ij}}} \partial_i \left(\sqrt{\det g_{ij}} \cdot g^{ij} \cdot \partial_j x^k \right) = 0.$$

To find these coordinates, fix a small ball $B(0, \varepsilon)$ and solve the Dirichlet problem

$$\Delta x^k = 0,\ x^k = y^k \text{ on } \partial B(0, \varepsilon).$$

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

$$\begin{aligned} \|x - y\|_{C^{2,\alpha},B(0,\varepsilon)} &\leq C \left(\|\Delta (x - y)\|_{C^\alpha,B(0,\varepsilon)} + \left\|(x - y)_{|\partial B(0,\varepsilon)}\right\|_{C^{2,\alpha},\partial B(0,\varepsilon)} \right) \\ &= C \|\Delta y\|_{C^\alpha,B(0,\varepsilon)}. \end{aligned}$$

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 coordinate system, we would then also be able to conclude that x formed a coordinate system.

Now 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, $\|\Delta y\|_{C^\alpha,B(0,\varepsilon)} \to 0$ 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 elegant expressions in such coordinates.

Lemma 11.2.6. *Let (M, g) be an n-dimensional Riemannian manifold with a harmonic coordinate system $x : U \to \mathbb{R}^n$. Then*

(1)

$$\Delta u = \frac{1}{\sqrt{\det g_{st}}} \partial_i \left(\sqrt{\det g_{st}} \cdot g^{ij} \cdot \partial_j u \right) = g^{ij} \partial_i \partial_j u.$$

(2)

$$\frac{1}{2} \Delta g_{ij} + Q(g, \partial g) = -\mathrm{Ric}_{ij} = -\mathrm{Ric}(\partial_i, \partial_j).$$

Here Q is a universal rational expression where the numerator is polynomial in the matrix g and quadratic in ∂g, while the denominator depends only on $\sqrt{\det g_{ij}}$.

Proof. (1) By definition:

$$\begin{aligned}
0 &= \Delta x^k \\
&= \tfrac{1}{\sqrt{\det g_{st}}}\partial_i\left(\sqrt{\det g_{st}}\cdot g^{ij}\cdot \partial_j x^k\right) \\
&= g^{ij}\partial_i\partial_j x^k + \tfrac{1}{\sqrt{\det g_{st}}}\partial_i\left(\sqrt{\det g_{st}}\cdot g^{ij}\right)\cdot \partial_j x^k \\
&= g^{ij}\partial_i\delta_j^k + \tfrac{1}{\sqrt{\det g_{st}}}\partial_i\left(\sqrt{\det g_{st}}\cdot g^{ij}\right)\cdot \delta_j^k \\
&= 0 + \tfrac{1}{\sqrt{\det g_{st}}}\partial_i\left(\sqrt{\det g_{st}}\cdot g^{ik}\right) \\
&= \tfrac{1}{\sqrt{\det g_{st}}}\partial_i\left(\sqrt{\det g_{st}}\cdot g^{ik}\right).
\end{aligned}$$

Thus, it follows that

$$\begin{aligned}
\Delta u &= \tfrac{1}{\sqrt{\det g_{st}}}\partial_i\left(\sqrt{\det g_{st}}\cdot g^{ij}\cdot \partial_j u\right) \\
&= g^{ij}\partial_i\partial_j u + \frac{1}{\sqrt{\det g_{st}}}\partial_i\left(\sqrt{\det g_{st}}\cdot g^{ij}\right)\cdot \partial_j u \\
&= g^{ij}\partial_i\partial_j u.
\end{aligned}$$

(2) Recall that if u is harmonic, then the Bochner formula for ∇u is

$$\Delta\left(\tfrac{1}{2}|\nabla u|^2\right) = |\mathrm{Hess}u|^2 + \mathrm{Ric}\left(\nabla u, \nabla u\right).$$

Here the term $|\mathrm{Hess}u|^2$ can be computed explicitly and depends only on the metric and its first derivatives. In particular,

$$\tfrac{1}{2}\Delta g\left(\nabla x^k, \nabla x^k\right) - \left|\mathrm{Hess}x^k\right|^2 = \mathrm{Ric}\left(\nabla x^k, \nabla x^k\right).$$

Polarizing this quadratic expression gives us an identity of the form

$$\tfrac{1}{2}\Delta g\left(\nabla x^i, \nabla x^j\right) - g\left(\mathrm{Hess}x^i, \mathrm{Hess}x^j\right) = \mathrm{Ric}\left(\nabla x^i, \nabla x^j\right).$$

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

$$\tfrac{1}{2}\Delta g^{ij} - g\left(\mathrm{Hess}x^i, \mathrm{Hess}x^j\right) = \mathrm{Ric}\left(\nabla x^i, \nabla x^j\right),$$

which in matrix form looks like

$$\tfrac{1}{2}\left[\Delta g^{ij}\right] - \left[g\left(\mathrm{Hess}x^i, \mathrm{Hess}x^j\right)\right] = \left[g^{ik}\right] \cdot \left[\mathrm{Ric}\left(\partial_k, \partial_l\right)\right] \cdot \left[g^{lj}\right].$$

This is, of course, not the promised formula. Instead, it is a similar formula for the inverse of $[g_{ij}]$. Now use the matrix equation $[g_{ik}] \cdot [g^{kj}] = [\delta_i^j]$ to conclude that

$$\begin{aligned}
0 &= \Delta\left([g_{ik}] \cdot \left[g^{kj}\right]\right) \\
&= [\Delta g_{ik}] \cdot \left[g^{kj}\right] + 2\left[\sum_k g\left(\nabla g_{ik}, \nabla g^{kj}\right)\right] + [g_{ik}] \cdot \left[\Delta g^{kj}\right] \\
&= [\Delta g_{ik}] \cdot \left[g^{kj}\right] + 2\left[\nabla g_{ik}\right] \cdot \left[\nabla g^{kj}\right] + [g_{ik}] \cdot \left[\Delta g^{kj}\right].
\end{aligned}$$

Inserting this in the above equation yields

$$\begin{aligned}
\left[\Delta g_{ij}\right] &= -2\left[\nabla g_{ik}\right] \cdot \left[\nabla g^{kl}\right] \cdot \left[g_{lj}\right] - [g_{ik}] \cdot \left[\Delta g^{kl}\right] \cdot \left[g_{lj}\right] \\
&= -2\left[\nabla g_{ik}\right] \cdot \left[\nabla g^{kl}\right] \cdot \left[g_{lj}\right] \\
&\quad -2\left[g_{ik}\right] \cdot \left[g\left(\mathrm{Hess}x^k, \mathrm{Hess}x^l\right)\right] \cdot \left[g_{lj}\right] \\
&\quad -2\left[g_{ik}\right] \cdot \left[g^{ks}\right] \cdot \left[\mathrm{Ric}\left(\partial_s, \partial_t\right)\right] \cdot \left[g^{tl}\right] \cdot \left[g_{lj}\right] \\
&= -2\left[\nabla g_{ik}\right] \cdot \left[\nabla g^{kl}\right] \cdot \left[g_{lj}\right] - 2\left[g_{ik}\right] \cdot \left[g\left(\mathrm{Hess}x^k, \mathrm{Hess}x^l\right)\right] \cdot \left[g_{lj}\right] \\
&\quad -2\left[\mathrm{Ric}\left(\partial_i, \partial_j\right)\right].
\end{aligned}$$

Each entry in these matrices then satisfies

$$\begin{aligned}
\tfrac{1}{2}\Delta g_{ij} + Q_{ij}\left(g, \partial g\right) &= -\mathrm{Ric}_{ij}, \\
Q_{ij} &= -2\sum_{k,l} g\left(\nabla g_{ik}, \nabla g^{kl}\right) g_{lj} \\
&\quad -2\sum_{k,l} g_{ik} g\left(\mathrm{Hess}x^k, \mathrm{Hess}x^l\right) g_{lj}.
\end{aligned}$$

□

It is interesting to apply this formula to the case of an Einstein metric, where $\mathrm{Ric}_{ij} = (n-1)\, k g_{ij}$. In this case, it reads

$$\tfrac{1}{2}\Delta g_{ij} = -\left(n-1\right) k g_{ij} - Q\left(g, \partial g\right).$$

The right-hand side makes sense as long as g_{ij} is C^1. The equation can then be understood in the weak sense: Multiply by some test function, integrate, and use integration by parts to obtain a formula that uses only first derivatives of g_{ij} on the left-hand side. If g_{ij} is $C^{1,\alpha}$, then the left-hand side lies in some C^β; but then our elliptic estimates show that g_{ij} must be in $C^{2,\beta}$. This can be bootstrapped until we have that the metric is C^∞. In fact, one can even show that it is analytic. Therefore, we can conclude that any metric which 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. A detailed account can be found in [99, vol. III].

11.3 Norms and Convergence of Manifolds

We next explain how the $C^{m,\alpha}$ norm and convergence concepts for functions generalize to Riemannian manifolds. These ideas can be used to prove various compactness and finiteness theorems for classes of Riemannian manifolds.

11.3.1 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 therefore should also have norm zero. In order to create a useful theory, it is often best to have only one space with vanishing norm. Thus we must agree that subsets of Euclidean space cannot have norm zero. To accommodate this problem, we define a family of norms of a Riemannian manifold, i.e., we use a function $N : (0,\infty) \to (0,\infty)$ rather than just a number. 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.

Let (M,g,p) be a pointed Riemannian n-manifold. We say that the $C^{m,\alpha}$*-norm on the scale of* r at p:

$$\|(M,g,p)\|_{C^{m,\alpha},r} \leq Q,$$

provided there exists a $C^{m+1,\alpha}$ chart $\varphi : (B(0,r),0) \subset \mathbb{R}^n \to (U,p) \subset M$ such that

(n1) $|D\varphi| \leq e^Q$ on $B(0,r)$ and $\left|D\varphi^{-1}\right| \leq e^Q$ on U. Equivalently, for all $v \in \mathbb{R}^n$ the metric coefficients satisfy

$$e^{-2Q}\delta_{kl}v^kv^l \leq g_{kl}v^kv^l \leq e^{2Q}\delta_{kl}v^kv^l.$$

(n2) For all multi-indices I with $0 \leq |I| \leq m$

$$r^{|I|+\alpha} \left\| \partial^I g_{kl} \right\|_\alpha \leq Q.$$

Globally we define

$$\|(M,g)\|_{C^{m,\alpha},r} = \sup_{p \in M} \|(M,g,p)\|_{C^{m,\alpha},r}.$$

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 the manifolds differ from the standard objects, namely, Euclidean balls. The first condition tells us that in the chosen coordinates the metric coefficients are bounded from below and above (in particular, we have uniform ellipticity for the Laplacian). The second condition gives us bounds on the derivatives of 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 concepts might not be well-defined. Therefore, we shall have to be a little careful in some situations.

The norm at a point is always finite, but when M is not compact the global norm might not be finite on any scale.

Example 11.3.1. If (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, $\|(\mathbb{R}^n, g_{\mathbb{R}^n})\|_{C^{m,\alpha},r} = 0$ for all r. We will show that these properties characterize flat manifolds and Euclidean space.

11.3.2 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 on A converges in $C^{m,\alpha}$, if they converge in the charts for some fixed finite covering of coordinate patches that are uniformly bi-Lipschitz. 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 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*, $(M_i, g_i, p_i) \to (M, g, p)$, if for every $R > 0$ we can find a domain $\Omega \supset B(p,R) \subset M$ and embeddings $F_i : \Omega \to M_i$ for large i such that $F_i(p) = p_i$, $F_i(\Omega) \supset B(p_i, R)$, and $F_i^* g_i \to g$ on Ω in the $C^{m,\alpha}$ topology. It is easy to see that this type of convergence implies pointed Gromov-Hausdorff convergence. When all manifolds in question are closed with a uniform bound on the diameter, then the maps F_i are diffeomorphisms. For closed manifolds we can also speak about unpointed convergence. In this case, convergence can evidently only occur 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 F_t coming from the flow corresponding to the vector field that is 0 at the two 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. Therefore, away from the poles the metrics $F_t^* g$ will converge to 0 in some fixed coordinates. So they cannot converge in the classical sense. If, however, we pull these metrics back by the diffeomorphisms F_{-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 as the spaces $\left(S^2, F_t^* g\right)$ are all isometric as abstract metric spaces.

11.3.3 Properties of the Norm

Let us now consider some of the elementary properties of norms and their relation to convergence.

Proposition 11.3.2. *Given* (M, g, p), $m \geq 0$, $\alpha \in (0, 1]$ *we have:*

(1) $\|(M, g, p)\|_{C^{m,\alpha},r} = \left\|\left(M, \lambda^2 g, p\right)\right\|_{C^{m,\alpha},\lambda r}$ *for all* $\lambda > 0$.
(2) *The function* $r \mapsto \|(M, g, p)\|_{C^{m,\alpha},r}$ *is increasing, continuous, and converges to* 0 *as* $r \to 0$.
(3) *Suppose* $(M_i, g_i, p_i) \to (M, g, p)$ *in* $C^{m,\alpha}$. *Then*

$$\|(M_i, g_i, p_i)\|_{C^{m,\alpha},r} \to \|(M, g, p)\|_{C^{m,\alpha},r} \text{ for all } r > 0.$$

Moreover, when all the manifolds have uniformly bounded diameter

$$\|(M_i, g_i)\|_{C^{m,\alpha},r} \to \|(M, g)\|_{C^{m,\alpha},r} \text{ for all } r > 0.$$

(4) *If* $\|(M, g, p)\|_{C^{m,\alpha},r} < Q$, *then for all* $x_1, x_2 \in B(0, r)$ *we have*

$$e^{-Q} \min\{|x_1 - x_2|, 2r - |x_1| - |x_2|\} \leq |\varphi(x_1)\, \varphi(x_2)| \leq e^{Q} |x_1 - x_2|.$$

(5) *The norm* $\|(M,g,p)\|_{C^{m,\alpha},r}$ is realized by a $C^{m+1,\alpha}$-chart.
(6) *If* M *is compact, then* $\|(M,g)\|_{C^{m,\alpha},r} = \|(M,g,p)\|_{C^{m,\alpha},r}$ for some $p \in M$.

Proof. (1) If we change the metric g to $\lambda^2 g$, then we can change the chart $\varphi : B(0,r) \to M$ to $\varphi^\lambda(x) = \varphi(\lambda^{-1}x) : B(0,\lambda r) \to M$. Since we scale the metric at the same time, the conditions n1 and n2 will still hold with the same Q.

(2) By restricting $\varphi : B(0,r) \to M$ to a smaller ball we immediately get that $r \mapsto \|(M,g,p)\|_{C^{m,\alpha},r}$ is increasing. Next, consider again the chart $\varphi^\lambda(x) = \varphi(\lambda^{-1}x) : B(0,\lambda r) \to M$, without changing the metric g. If we assume that $\|(M,g,p)\|_{C^{m,\alpha},r} < Q$, then

$$\|(M,g,p)\|_{C^{m,\alpha},\lambda r} \le \max\left\{Q \pm |\log \lambda|, Q \cdot \lambda^2\right\}.$$

Denoting $N(r) = \|(M,g,p)\|_{C^{m,\alpha},r}$, we obtain

$$N(\lambda r) \le \max\left\{N(r) \pm |\log \lambda|, N(r) \cdot \lambda^2\right\}.$$

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,

$$\begin{aligned} N(r) &= N\left(\frac{r}{r_i} r_i\right) \\ &\le \max\left\{N(r_i) \pm \left|\log \frac{r}{r_i}\right|, N(r_i) \cdot \left(\frac{r}{r_i}\right)^2\right\}. \end{aligned}$$

So

$$\begin{aligned} N(r) &\le \liminf \max\left\{N(r_i) \pm \left|\log \frac{r}{r_i}\right|, N(r_i) \cdot \left(\frac{r}{r_i}\right)^2\right\} \\ &= \liminf N(r_i). \end{aligned}$$

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 $g_{kl} = \delta_{kl}$ at p. In particular $\left|D\varphi|_p\right| = \left|D\varphi^{-1}|_p\right| = 1$. Using these coordinates on sufficiently small balls will yield the desired charts.

(3) Fix $r > 0$ and $Q > \|(M,g,p)\|_{C^{m,\alpha},r}$. Pick a domain $\Omega \supset B(p, e^Q r)$ 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 and $F_i(p) = p_i$.

Choose a chart $\varphi : B(0,r) \to M$ with properties n1 and n2. Then define charts in M_i by $\varphi_i = F_i \circ \varphi : B(0,r) \to M_i$ and note that since $F_i^* g_i \to g$ in $C^{m,\alpha}$, these charts satisfy properties n1 and n2 for constants $Q_i \to Q$. This shows that

$$\limsup \|(M_i, g_i, p_i)\|_{C^{m,\alpha},r} \le \|(M, g, p)\|_{C^{m,\alpha},r}.$$

On the other hand, if $Q > \|(M_i, g_i, p_i)\|_{C^{m,\alpha},r}$ for a sufficiently large i, then select a chart $\varphi_i : B(0,r) \to M_i$ and consider $\varphi = F_i^{-1} \circ \varphi_i$ on M. As before, we have

$$\|(M, g, p)\|_{C^{m,\alpha},r} \le Q_i,$$

where Q_i is close to Q. This implies

$$\liminf \|(M_i, g_i, p_i)\|_{C^{m,\alpha},r} \ge \|(M, g, p)\|_{C^{m,\alpha},r}$$

and proves the result.

When all the spaces have uniformly bounded diameter we choose diffeomorphisms $F_i : M \to M_i$ for large i such that $F_i^* g_i \to g$. For every choice of $p \in M$ select $p_i = F_i(p) \in M_i$ and use what we just proved to conclude that

$$\liminf \|(M_i, g_i)\|_{C^{m,\alpha},r} \ge \sup_p \|(M, g, p)\|_{C^{m,\alpha},r}.$$

Similarly, when $p_i \in M_i$ and $p = F_i^{-1}(p_i)$, it follows that

$$\limsup \|(M_i, g_i, p_i)\|_{C^{m,\alpha},r} \le \sup_p \|(M, g)\|_{C^{m,\alpha},r}.$$

(4) The condition $|D\varphi| \le e^Q$, together with convexity of $B(0,r)$, immediately implies the second inequality. For the other, first observe that if any segment from $\varphi(x_1)$ to $\varphi(x_2)$ lies in U, then $\left|D\varphi^{-1}\right| \le e^Q$ implies, that

$$|x_1 - x_2| \le e^Q \, |\varphi(x_1)\,\varphi(x_2)|.$$

So we may assume that $\varphi(x_1)$ and $\varphi(x_2)$ are joined by a segment $c : [0,1] \to M$ that leaves U. Split c into $c : [0,t_1) \to U$ and $c : (t_2, 1] \to U$ with $c(t_i) \in \partial U$. Then we clearly have

$$\begin{aligned} |\varphi(x_1)\,\varphi(x_2)| &= L(c) \ge L(c|_{[0,t_1)}) + L(c|_{(t_2,1]}) \\ &\ge e^{-Q}(L(\varphi^{-1} \circ c|_{[0,t_1)}) + L(\varphi^{-1} \circ c|_{(t_2,1]})) \\ &\ge e^{-Q}(2r - |x_1| - |x_2|). \end{aligned}$$

The last inequality follows from the fact that $\varphi^{-1} \circ c(0) = x_1$ and $\varphi^{-1} \circ c(1) = x_2$, and that $\varphi^{-1} \circ c(t)$ approaches $\partial B(0, r)$ as t approaches t_1 and t_2.

(5) Given a sequence of charts $\varphi_i : B(0, r) \to M$ that satisfy n1 and n2 with $Q_i \to Q$ we can use the Arzela-Ascoli lemma to find a subsequence that converges to a $C^{m+1,\alpha}$ map $\varphi : B(0, r) \to M$. Property (4) shows that φ is injective and becomes a homeomorphism onto its image. This makes φ a chart. We can, after passing to another subsequence, also assume that the metric coefficients converge. This implies that φ satisfies n1 and n2 for Q.

(6) Property (3) implies that $p \mapsto \|(M, g, p)\|_{C^{m,\alpha},r}$ is continuous. Compactness then shows that the supremum is a maximum.

$\square$

Corollary 11.3.3. *If $\|(M, g, p)\|_{C^{m,\alpha},r} \leq Q$, then $B\left(p, e^{-Q}r\right) \subset U$.*

Proof. Let $q \in \partial U$ be the closest point to p so that $B(p, |qp|) \subset U$. If $c : [0, |pq|] \to M$ is a segment from p to q, then $c(s) \in B(p, |qp|)$ for all $s < |qp|$ and we can write $c(s) = \varphi(\bar{c}(s))$, where $\bar{c} : [0, |qp|) \in B(0, r)$ has the property that $\lim_{t \to |qp|} |\bar{c}(t)| = r$. Property (4) from proposition 11.3.2 then shows that

$$\begin{aligned} |qp| &\geq \lim_{s \to |qp|} |\varphi(\bar{c}(s))\, \varphi(0)| \\ &\geq \lim_{s \to |qp|} e^{-Q} \min\{|\bar{c}(s)|, 2r - |\bar{c}(s)|\} \\ &\geq \lim_{s \to |qp|} e^{-Q} |\bar{c}(s)| \\ &= e^{-Q} r. \end{aligned}$$

$\square$

Corollary 11.3.4. *If $\|(M, g, p)\|_{C^{m,\alpha},r} = 0$ for some r, then p is contained in a neighborhood that is flat.*

Proof. It follows from proposition 11.3.2 that there is a $C^{m+1,\alpha}$ chart $\varphi : B(0, r) \to U \supset B\left(p, e^{-Q}r\right)$ with $Q = 0$. This implies that it is a C^1 Riemannian isometry and then by theorem 5.6.15 a Riemannian isometry. $\square$

11.3.4 The Harmonic Norm

We define a more restrictive norm, called the *harmonic norm* and denoted

$$\|(M, g, p)\|^{har}_{C^{m,\alpha},r}.$$

The only change in our previous definition is that $\varphi^{-1} : U \to \mathbb{R}^n$ is also assumed to be harmonic with respect to the Riemannian metric g on M, i.e., for each j

$$\frac{1}{\sqrt{\det[g_{st}]}}\partial_i\left(\sqrt{\det[g_{st}]}\cdot g^{ij}\right)=0.$$

Proposition 11.3.5 (Anderson, 1990). *Proposition 11.3.2 also holds for the harmonic norm when $m \geq 1$.*

Proof. The proof is mostly identical so we only mention the necessary changes.

For the statement in (2) that the norm goes to zero as the scale decreases, just solve the Dirichlet problem as we did when establishing the existence of harmonic coordinates in lemma 11.2.5. There 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 (see exercise 2.5.20).

The proof of (3) is necessarily somewhat different, as we must use and produce harmonic coordinates. Let the set-up be as before. First we show the easy part:

$$\liminf \|(M_i, g_i, p_i)\|^{har}_{C^{m,\alpha},r} \geq \|(M, g, p)\|^{har}_{C^{m,\alpha},r}.$$

To this end, select $Q > \liminf \|(M_i, g_i, p_i)\|^{har}_{C^{m,\alpha},r}$. For large i we can then select charts $\varphi_i : B(0, r) \to M_i$ with the requisite properties. After passing to a subsequence, we can make these charts converge to a chart

$$\varphi = \lim F_i^{-1} \circ \varphi_i : B(0, r) \to M.$$

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 $\|(M, g, p)\|^{har}_{C^{m,\alpha},r} \leq Q$.

For the reverse inequality

$$\limsup \|(M_i, g_i, p_i)\|^{har}_{C^{m,\alpha},r} \leq \|(M, g, p)\|^{har}_{C^{m,\alpha},r},$$

select $Q > \|(M, g, p)\|^{har}_{C^{m,\alpha},r}$. Then, from the continuity of the norm we can find $\varepsilon > 0$ such that also $\|(M, g, p)\|^{har}_{C^{m,\alpha},r+\varepsilon} < Q$. For this scale, select

$$\varphi : B(0, r + \varepsilon) \to U \subset M$$

satisfying the usual conditions. Now define

$$U_i = F_i(\varphi(B(0, r + \varepsilon/2))) \subset M_i.$$

This is clearly a closed disc with smooth boundary

$$\partial U_i = F_i(\varphi(\partial B(0, r + \varepsilon/2))).$$

On each U_i solve the Dirichlet problem

$$\begin{aligned}\psi_i : U_i &\to \mathbb{R}^n,\\ \Delta_{g_i}\psi_i &= 0,\\ \psi_i &= \varphi^{-1}\circ F_i^{-1} \text{ on } \partial U_i.\end{aligned}$$

The inverse of ψ_i, if it exists, will then be a coordinate map $B(0,r) \to U_i$. On the set $B(0, r+\varepsilon/2)$ we can compare $\psi_i \circ F_i \circ \varphi$ 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. Now pull these metrics back to $B\left(0, r+\frac{\varepsilon}{2}\right)$ and refer to them as $g\,(=\varphi^* g)$ and $g_i\left(=\varphi^* F_i^* g_i\right)$. In this way the harmonicity conditions read $\Delta_g I = 0$ and $\Delta_{g_i}\psi_i \circ F_i \circ \varphi = 0$. 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

$$\begin{aligned}\|I - \psi_i \circ F_i \circ \varphi\|_{C^{m+1,\alpha}} &\leq C\left\|\Delta_{g_i}(I - \psi_i \circ F_i \circ \varphi)\right\|_{C^{m-1,\alpha}}\\ &= C\left\|\Delta_{g_i} I\right\|_{C^{m-1,\alpha}}.\end{aligned}$$

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,

$$\|I - \psi_i \circ F_i \circ \varphi\|_{C^{m+1,\alpha}} \to 0.$$

It follows that ψ_i must become coordinates for large i. Also, these coordinates will show that $\|(M_i, g_i, p_i)\|^{har}_{C^{m,\alpha},r} < Q$ for large i. □

11.3.5 Compact Classes of Riemannian Manifolds

We can now state and prove the result that is our manifold equivalent of the Arzela-Ascoli lemma. This theorem is essentially due to J. Cheeger.

Theorem 11.3.6 (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 $\mathscr{M}^{m,\alpha}(n,Q,r)$ of complete, pointed Riemannian n-manifolds (M,g,p) with $\|(M,g)\|_{C^{m,\alpha},r} \leq Q$. The class $\mathscr{M}^{m,\alpha}(n,Q,r)$ is compact in the pointed $C^{m,\beta}$ topology for all $\beta < \alpha$.*

Proof. First we show that $\mathscr{M} = \mathscr{M}^{m,\alpha}(n,Q,r)$ is precompact in the pointed Gromov-Hausdorff topology. Next we prove that $\mathscr{M}$ is closed in the Gromov-Hausdorff topology. The last and longest part is devoted to getting improved convergence from Gromov-Hausdorff convergence.

Setup: Whenever we select $M \in \mathscr{M}$, we can by proposition 11.3.2 assume that it comes equipped with charts around all points satisfying n1 and n2.

(A) $\mathscr{M}$ is precompact in the pointed Gromov-Hausdorff topology.

Define $\delta = e^{-Q}r$ and note that there exists an $N(n,Q)$ such that $B(0,r)$ can be covered by at most N balls of radius $e^{-Q} \cdot \delta/4$. Since $\varphi : B(0,r) \to U$ is a Lipschitz map with Lipschitz constant $\leq e^{Q}$, this implies that $U \supset B(p,\delta)$ can be covered by N balls of radius $\delta/4$.

Next we claim that every ball $B(x, \ell \cdot \delta/2) \subset M$ can be covered by $\leq N^{\ell}$ balls of radius $\delta/4$. For $\ell = 1$ we just proved this. If $B(x, \ell \cdot \delta/2)$ is covered by $B(x_1, \delta/4)$,…, $B(x_{N^\ell}, \delta/4)$, then $B(x, \ell \cdot \delta/2 + \delta/2) \subset \bigcup B(x_i, \delta)$. Now each $B(x_i,\delta)$ can be covered by $\leq N$ balls of radius $\delta/4$, and hence $B(x, (\ell+1)\delta/2)$ can be covered by $\leq N \cdot N^{\ell} = N^{\ell+1}$ balls of radius $\delta/4$.

The precompactness claim is equivalent to showing that we can find a function $C(\varepsilon) = C(\varepsilon, R, K, r, n)$ such that each $B(p,R)$ can contain at most $C(\varepsilon)$ disjoint ε-balls. To check this, let $B(x_1,\varepsilon)$,…,$B(x_s,\varepsilon)$ be a collection of disjoint balls in $B(p,R)$. Suppose that $\ell \cdot \delta/2 < R \leq (\ell+1)\delta/2$. Then

$$\begin{aligned}\mathrm{vol}B(p,R) &\leq N^{\ell+1} \cdot (\text{maximal volume of } \delta/4\text{-ball})\\ &\leq N^{\ell+1} \cdot (\text{maximal volume of chart})\\ &\leq N^{\ell+1} \cdot e^{nK} \cdot \mathrm{vol}B(0,r)\\ &\leq V(R) = V(R,n,K,r).\end{aligned}$$

As long as $\varepsilon < \delta$ each $B(x_i,\varepsilon)$ lies in some chart $\varphi : B(0,r) \to U \subset M$ whose pre-image in $B(0,r)$ contains an $e^{-K} \cdot \varepsilon$-ball. Thus

$$\mathrm{vol}B(p_i,\varepsilon) \geq e^{-nK}\mathrm{vol}B(0,\varepsilon).$$

All in all, we get

$$\begin{aligned} V(R) &\geq \operatorname{vol}B(p,R) \\ &\geq \sum \operatorname{vol}B(p_i,\varepsilon) \\ &\geq s \cdot e^{-nK} \cdot \operatorname{vol}B(0,\varepsilon). \end{aligned}$$

Thus,

$$s \leq C(\varepsilon) = V(R) \cdot e^{nK} \cdot (\operatorname{vol}B(0,\varepsilon))^{-1}.$$

Now select a sequence (M_i, g_i, p_i) in $\mathcal{M}$. From the previous considerations we can assume that $(M_i, g_i, p_i) \to (X, |\cdot\cdot|, p)$ 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 do not reindex the family, but merely assume that the sequence was chosen to have the desired properties from the beginning.

(B) $(X, |\cdot\cdot|, p)$ is a Riemannian manifold of class $C^{m,\alpha}$ with $\|(X,g)\|_{C^{m,\alpha},r} \leq Q$

For each $q \in X$ we need to find a chart $\varphi : B(0,r) \to U \subset X$ with $q = \varphi(0)$. To construct this chart consider $q_i \to q$ and charts $\varphi_i : B(0,r) \to U_i \subset M_i$ with $q_i = \varphi_i(0)$. These charts are uniformly Lipschitz and so must subconverge to a map $\varphi : B(0,r) \to U \subset X$. This map will satisfy property (4) in proposition 11.3.2 and thus be a homeomorphism onto its image. This makes X a topological manifold.

We next construct a compatible Riemannian metric on X that satisfies n1 and n2. For each $q \in X$ consider the metrics $\varphi_i^* g_i = g_{i\cdot\cdot}$ written out in components on $B(0,r)$ with respect to the chart φ_i. Since all of the $g_{i\cdot\cdot}$ satisfy n1 and n2, we can again use Arzela-Ascoli to insure that the components $g_{i\cdot\cdot} \to g_{\cdot\cdot}$ in the $C^{m,\beta}$ topology on $B(0,r)$ to functions $g_{\cdot\cdot}$ that also satisfy n1 and n2. These local Riemannian metrics are possibly only Hölder continuous. Nevertheless, they define a distance as we defined it in section 5.3. Moreover this distance is locally the same as the metric on X. To see this, note that we work entirely on $B(0,r)$ and both the Riemannian structures and the metric structures converge to the limit structures.

Finally, we need to show that the transition function $\varphi^{-1} \circ \psi$ for two such charts $\varphi, \psi : B(0,r) \to X$ with overlapping images are at least C^1 so as to obtain a differentiable structure on X. As it stands $\varphi^{-1} \circ \psi$ is locally Lipschitz with respect to the Euclidean metrics. However, it is distance preserving with respect to the pull back metrics from X. Calabi-Hartman in [22] generalized theorem 5.6.15 to this context. Specifically, they claim that a distance preserving map between C^α Riemannian metrics is $C^{1,\alpha}$. The proof, however, only seems to prove that the map is $C^{1,\frac{\alpha}{2}}$, which is more than enough for our purposes.

(C) $(M_i, g_i, p_i) \to (X, |\cdot\cdot|, p) = (X, g, p)$ in the pointed $C^{m,\beta}$ topology.

We assume that X is equipped with a countable atlas of charts $\varphi_s : B(0,r) \to U_s$, $s = 1, 2.3, \ldots$ that are limits of charts $\varphi_{is} : B(0,r) \to U_{is} \subset M_i$ that also form an atlas for each M_i. We can further assume that transitions converge: $\varphi_{is}^{-1} \circ \varphi_{it} \to \varphi_s^{-1} \circ \varphi_t$ and that the metrics converge: $g_{is\cdot\cdot} \to g_{s\cdot\cdot}$. We say that two maps F_1, F_2 between subsets in M_i and X are $C^{m+1,\beta}$ close if all the coordinate compositions $\varphi_s^{-1} \circ F_1 \circ \varphi_{is}$ and $\varphi_s \circ F_2 \circ \varphi_{is}$ 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

$$\begin{aligned} f_{is} &= \varphi_{is} \circ \varphi_s^{-1} : U_s \to U_{is}, \\ f_{it} &= \varphi_{it} \circ \varphi_t^{-1} : U_t \to U_{it} \end{aligned}$$

"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 assumed. In order to finish the proof, we construct maps

$$F_{i\ell} : \Omega_\ell = \bigcup_{s=1}^{\ell} U_s \to \Omega_{i\ell} = \bigcup_{s=1}^{\ell} U_{is}$$

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 $\ell = 1$ 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 $U_{\ell+1} \cap \Omega_\ell = \varnothing$, 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 $U_{\ell+1}$ by

$$F_{i\ell+1}(x) = \varphi_{i\ell+1} \circ (\mu_1(x) \cdot \varphi_{i\ell+1}^{-1} \circ f_{i\ell+1}(x) + \mu_2(x) \cdot \varphi_{i\ell+1}^{-1} \circ F_{i\ell}(x)),$$

where μ_1, μ_2 are a partition of unity for $U_{\ell+1}$, Ω_ℓ. This map is clearly well-defined on $U_{\ell+1}$, since $\mu_2(x) = 0$ on $U_{\ell+1} - \Omega_\ell$. Now consider this map in coordinates

$$\begin{aligned} \varphi_{i\ell+1}^{-1} \circ F_{i\ell+1} \circ \varphi_{\ell+1}(y) &= (\mu_1 \circ \varphi_{\ell+1}(y)) \cdot \varphi_{\ell+1}^{-1} \circ f_{i\ell+1} \circ \varphi_{\ell+1}(y) \\ &\quad + (\mu_2 \circ \varphi_{\ell+1}(y)) \cdot \varphi_{i\ell+1}^{-1} \circ F_{i\ell} \circ \varphi_{\ell+1}(y) \\ &= \tilde{\mu}_1(y) F_1(y) + \tilde{\mu}_2(y) F_2(y). \end{aligned}$$

Then

$$\begin{aligned} \|\tilde{\mu}_1 F_1 + \tilde{\mu}_2 F_2 - F_1\|_{C^{m+1,\beta}} &= \|\tilde{\mu}_1 (F_1 - F_1) + \tilde{\mu}_2 (F_2 - F_1)\|_{C^{m+1,\beta}} \\ &\leq C(n,m) \|\tilde{\mu}_2\|_{C^{m+1,\beta}} \cdot \|F_2 - F_1\|_{C^{m+1,\beta}}. \end{aligned}$$

This inequality is valid on all of $B(0,r)$, despite the fact that F_2 is not defined on all of $B(0,r)$, since $\tilde{\mu}_1 \cdot F_1 + \tilde{\mu}_2 \cdot F_2 = F_1$ on the region where F_2 is undefined. By assumption

$$\|F_2 - F_1\|_{C^{m+1,\beta}} \to 0 \text{ as } i \to \infty,$$

so $F_{i\ell+1}$ is $C^{m+1,\beta}$-close to f_{is}, $s = 1, \ldots, \ell+1$ as $i \to \infty$.

Finally we see that the closeness of $F_{i\ell}$ to the coordinate charts shows that it is an embedding on all compact subsets of the domain. □

Corollary 11.3.7. *Any subclasses of* $\mathcal{M}^{m,\alpha}(n,Q,r)$, *where the elements in addition satisfy* diam $\leq D$, *respectively* vol $\leq V$, *is compact in the* $C^{m,\beta}$ *topology. In particular, it contains only finitely many diffeomorphism types.*

Proof. We use notation as in the fundamental theorem. If $\operatorname{diam}(M,g,p) \leq D$, then clearly $M \subset B\left(p, k \cdot {}^{\delta}\!/\!{}_2\right)$ for $k > D \cdot {}^2\!/\!{}_\delta$. Hence, each element in $\mathcal{M}^{m,\alpha}(n,Q,r)$ can be covered by $\leq N^k$ charts. Thus, $C^{m,\beta}$-convergence is actually in the unpointed topology, as desired.

If instead, $\operatorname{vol} M \leq V$, then we can use part (A) in the proof to see that we can never have more than $k = V \cdot e^{2nK} \cdot (\operatorname{vol} B(0,\varepsilon))^{-1}$ disjoint ε-balls. In particular, $\operatorname{diam} \leq 2\varepsilon \cdot k$, and we can use the above argument.

Finally, compactness in any $C^{m,\beta}$ topology implies that the class cannot contain infinitely many diffeomorphism types. □

Clearly there is also a harmonic analogue to the fundamental theorem.

Corollary 11.3.8. *Given* $Q > 0$, $n \geq 2$, $m \geq 0$, $\alpha \in (0,1]$, *and* $r > 0$ *the class of complete, pointed Riemannian n-manifolds* (M,g,p) *with* $\|(M,g)\|^{har}_{C^{m,\alpha},r} \leq Q$ *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 $\|(M,g)\|^{har}_{C^{m,\alpha},r} \leq Q$. But one can easily see that harmonic charts converge to harmonic charts as in proposition 11.3.5.

11.3.6 Alternative Norms

Finally, we mention that the norm concept and its properties do not change if n1 and n2 are altered as follows:

(n1') $\quad |D\varphi|, \left|D\varphi^{-1}\right| \leq f_1(n,Q)$,

(n2') $\quad r^{|j|+\alpha} \|\partial^j g_{\cdot\cdot}\|_\alpha \leq f_2(n,Q), \ 0 \leq |j| \leq m$,

where f_1 and f_2 are continuous, $f_1(n,0,r) = 1$, and $f_2(n,0) = 0$. 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 $m = \alpha = 0$. Then n2 no longer makes sense since $\alpha = 0$, however, we still have a C^0-norm concept. 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 only 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 [39]) 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 closely mirrors 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 easily use the fact that the charts converge in $C^\alpha (\alpha < 1)$. But it is possible to do something interesting along these lines. There is an even weaker norm concept called the *Reifenberg norm* that is related to the Gromov-Hausdorff distance. For a metric space $(X, |\cdot\cdot|)$ we define the n-dimensional norm on the scale of r as

$$\|(X, |\cdot\cdot|)\|_r^n = \frac{1}{r} \sup_{p \in X} d_{G-H}\left(B(p,r), B(0,r)\right),$$

where $B(0,R) \subset \mathbb{R}^n$. The the r^{-1} factor insures that we don't have small distance between $B(p,r)$ and $B(0,r)$ just because r is small. Note also that if $(X_i, |\cdot\cdot|_i) \to (X, |\cdot\cdot|)$ in the Gromov-Hausdorff topology then

$$\|(X_i, |\cdot\cdot|_i)\|_r^n \to \|(X, |\cdot\cdot|)\|_r^n$$

for fixed n, r.

For an n-dimensional Riemannian manifold one sees immediately that

$$\lim_{r\to 0} \|(M,g)\|_r^n \to 0 = 0.$$

Cheeger and Colding have proven a converse to this (see [29]). There is an $\varepsilon(n) > 0$ such that if $\|(X, |\cdot\cdot|)\|_r^n \leq \varepsilon(n)$ for all small r, then X is in a weak sense an n-dimensional Riemannian manifold. Among other things, they show that for small r the α-Hölder distance between $B(p,r)$ and $B(0,r)$ is small. Here the α-Hölder distance $d_\alpha(X,Y)$ between metric spaces is defined as the infimum of

$$\log \max \left\{ \sup_{x_1 \neq x_2} \frac{|F(x_1) F(x_2)|}{|x_1 x_2|^{\alpha}}, \sup_{y_1 \neq y_2} \frac{\left|F^{-1}(y_1) F^{-1}(y_2)\right|}{|y_1 y_2|^{\alpha}} \right\},$$

where $F : X \to Y$ runs over all homeomorphisms. They also show that if $(M_i, g_i) \to (X, |\cdot\cdot|)$ in the Gromov-Hausdorff distance and $\|(M_i, g_i)\|_r^n \leq \varepsilon(n)$ for all i and small r, then $(M_i, g_i) \to (X, |\cdot\cdot|)$ in the Hölder distance. In particular, all of the M_is have to be homeomorphic (and in fact diffeomorphic) to X for large i.

This is enhanced by an earlier result of Colding (see [34]) stating that for a Riemannian manifold (M, g) with $\mathrm{Ric} \geq (n-1)k$ we have that $\|(M, g)\|_r^n$ is small if and only if and only if

$$\mathrm{vol} B(p, r) \geq (1 - \delta) \mathrm{vol} B(0, r)$$

for some small δ. Relative volume comparison tells us that the volume condition holds for all small r if it holds for just one r. Thus the smallness condition for the norm holds for all small r provided we have the volume condition for just some r.

11.4 Geometric Applications

To obtain better estimates on the norms it is convenient to use more analysis. The idea of using harmonic coordinates for similar purposes goes back to [37]. In [66] 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 Greene's survey in [51]. Here we shall develop these ideas from a different point of view due to Anderson.

11.4.1 Ricci Curvature

The most important feature about harmonic coordinates is that the metric is apparently 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 harmonic C^1 norm and Ricci curvature.

Lemma 11.4.1 (Anderson, 1990). *Suppose that a Riemannian manifold* (M, g) *has bounded Ricci curvature* $|\mathrm{Ric}| \leq \Lambda$. *For any* $r_1 < r_2$, $K \geq \|(M, g, p)\|_{C^1, r_2}^{har}$, *and* $\alpha \in (0, 1)$ *we can find* $C(n, \alpha, K, r_1, r_2, \Lambda)$ *such that*

$$\|(M,g,p)\|^{har}_{C^{1,\alpha},r_1} \leq C\,(n,\alpha,K,r_1,r_2,\Lambda)\,.$$

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*

$$\|(M,g,p)\|^{har}_{C^{m+1,\alpha},r_1} \leq C\,(n,\alpha,K,r_1,r_2,k,m)\,.$$

Proof. We just need to bound the metric components g_{ij} in some fixed harmonic coordinates. In such coordinates $\Delta = g^{ij}\partial_i\partial_j$. Given that $\|(M,g,p)\|^{har}_{C^1,r_2} \leq K$, we can conclude that we have the necessary conditions on the coefficients of $\Delta = g^{ij}\partial_i\partial_j$ to use the elliptic estimate

$$\left\|g_{ij}\right\|_{C^{1,\alpha},B(0,r_1)} \leq C\,(n,\alpha,K,r_1,r_2)\left(\left\|\Delta g_{ij}\right\|_{C^0,B(0,r_2)} + \left\|g_{ij}\right\|_{C^\alpha,B(0,r_2)}\right).$$

Since

$$\Delta g_{ij} = -2\mathrm{Ric}_{ij} - 2Q\,(g,\partial g)$$

it follows that

$$\left\|\Delta g_{ij}\right\|_{C^0,B(0,r_2)} \leq 2\Lambda \left\|g_{ij}\right\|_{C^0,B(0,r_2)} + \hat{C}\left\|g_{ij}\right\|_{C^1,B(0,r_2)}\,.$$

Using this we obtain

$$\begin{aligned}\left\|g_{ij}\right\|_{C^{1,\alpha},B(0,r_1)} &\leq C\,(n,\alpha,K,r_1,r_2)\left(\left\|\Delta g_{ij}\right\|_{C^0,B(0,r_2)} + \left\|g_{ij}\right\|_{C^\alpha,B(0,r_2)}\right)\\ &\leq C\,(n,\alpha,K,r_1,r_2)\left(2\Lambda + \hat{C} + 1\right)\left\|g_{ij}\right\|_{C^1,B(0,r_2)}\,.\end{aligned}$$

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,

$$\begin{aligned}\left\|g_{ij}\right\|_{C^{2,\alpha},B(0,r_1)} &\leq C\,(n,\alpha,K,r_1,r_2)\left(\left\|\Delta g_{ij}\right\|_{C^\alpha,B(0,r_2)} + \left\|g_{ij}\right\|_{C^\alpha,B(0,r_2)}\right)\\ &\leq C\,(n,\alpha,K,r_1,r_2,k)\cdot C\cdot\left\|g_{ij}\right\|_{C^{1,\alpha},B(0,r_2)}\,.\end{aligned}$$

This gives $C^{2,\alpha}$ bounds on the metric. Then, of course, Δg_{ij} is bounded in $C^{1,\alpha}$, 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 for any m. □

Combining this with the fundamental theorem gives a very interesting compactness result.

Corollary 11.4.2. *For given $n \geq 2$, $Q, r, \Lambda \in (0, \infty)$ consider the class of Riemannian n-manifolds with*

$$\begin{aligned} \|(M,g)\|^{har}_{C^1,r} &\leq Q, \\ |\mathrm{Ric}| &\leq \Lambda. \end{aligned}$$

This class is precompact in the pointed $C^{1,\alpha}$ 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 $m \geq 1$ and $\alpha \in (0,1)$.

Next we show how the injectivity radius can be used to control the harmonic norm.

Theorem 11.4.3 (Anderson, 1990). *Given $n \geq 2$ and $\alpha \in (0,1)$, Λ, $R > 0$, one can for each $Q > 0$ find $r(n, \alpha, \Lambda, R) > 0$ such that any compact Riemannian n-manifold (M, g) with*

$$\begin{aligned} |\mathrm{Ric}| &\leq \Lambda, \\ \mathrm{inj} &\geq R \end{aligned}$$

satisfies $\|(M,g)\|^{har}_{C^{1,\alpha},r} \leq Q$.

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

$$\begin{aligned} |\mathrm{Ric}| &\leq \Lambda, \\ \mathrm{inj} &\geq R, \\ \|(M_i, g_i)\|^{har}_{C^{1,\alpha},i^{-1}} &> Q. \end{aligned}$$

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 $r_i \in \left(0, i^{-1}\right)$ such that $\|(M_i, g_i)\|^{har}_{C^{1,\alpha},r_i} = Q$. Now rescale these manifolds: $\bar{g}_i = r_i^{-2} g_i$. Then we have that $(M_i, \bar{g}_i)$ satisfies

$$\begin{aligned} |\mathrm{Ric}| &\leq r_i \Lambda, \\ \mathrm{inj} &\geq r_i^{-1} R, \\ \|(M_i, \bar{g}_i)\|^{har}_{C^{1,\alpha},1} &= Q. \end{aligned}$$

We can then select $p_i \in M_i$ such that

$$\|(M_i, \bar{g}_i, p_i)\|^{har}_{C^{1,\alpha},1} \in \left[\frac{Q}{2}, Q\right].$$

The first important step is to use the bounded Ricci curvature of $(M_i, \bar{g}_i)$ to conclude that 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 $(M_i, \bar{g}_i, p_i)$ converges in the pointed $C^{1,\alpha}$ topology, to a Riemannian manifold (M, g, p) of class $C^{1,\gamma}$. Since the $C^{1,\alpha}$ norm is continuous in the $C^{1,\alpha}$ topology we can conclude that

$$\|(M, g, p)\|^{har}_{C^{1,\alpha},1} \in \left[\frac{Q}{2}, Q\right].$$

The second thing we can prove is that $(M, g) = (\mathbb{R}^n, g_{\mathbb{R}^n})$. 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 $(M_i, \bar{g}_i)$ are covered by harmonic coordinates that converge to harmonic coordinates in the limit space. In these harmonic coordinates the metric components satisfy

$$\frac{1}{2}\Delta \bar{g}_{kl} + Q(\bar{g}, \partial \bar{g}) = -\mathrm{Ric}_{kl}.$$

But we know that

$$|-\mathrm{Ric}| \leq r_i^{-2}\Lambda \bar{g}_i$$

and that the $\bar{g}_{kl}$ converge in the $C^{1,\alpha}$ topology to the metric coefficients g_{kl} for the limit metric. Consequently, the limit manifold is covered by harmonic coordinates and in these coordinates the metric satisfies:

$$\frac{1}{2}\Delta g_{kl} + Q(g, \partial g) = 0.$$

Thus the limit metric is a weak solution to the Einstein equation $\mathrm{Ric} = 0$ and therefore must be a smooth Ricci flat Riemannian manifold. Finally, we use that: $\mathrm{inj}(M_i, \bar{g}_i) \to \infty$. In the limit space any geodesic is a limit of geodesics from the sequence $(M_i, \bar{g}_i)$, 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 as $\mathrm{inj}(M_i, \bar{g}_i) \to \infty$ any finite length geodesic must be a segment. This, however, implies that $\mathrm{inj}(M, g) = \infty$. The splitting theorem 7.3.5 then shows that the limit space is Euclidean space. □

From this theorem we immediately get

Corollary 11.4.4 (Anderson, 1990). *Let $n \geq 2$ and $\Lambda, D, R > 0$ be given. The class of closed Riemannian n-manifolds satisfying*

$$|\mathrm{Ric}| \leq \Lambda,$$
$$\mathrm{diam} \leq D,$$
$$\mathrm{inj} \geq R$$

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 other similar characterizations of Euclidean space. One of the most interesting ones uses volume pinching.

11.4.2 Volume Pinching

The idea is to use the relative volume comparison (see lemma 7.1.4) rather than the splitting theorem. It is relatively easy to prove that Euclidean space is the only space with

$$\operatorname{Ric} \geq 0,$$
$$\lim_{r\to\infty} \frac{\operatorname{vol} B(p,r)}{\omega_n r^n} = 1,$$

where $\omega_n r^n$ is the volume of a Euclidean ball of radius r (see also exercises 7.5.8 and 7.5.10). This result has a very interesting gap phenomenon associated to it under the stronger hypothesis that the space is Ricci flat.

Lemma 11.4.5 (Anderson, 1990). *For each* $n \geq 2$ *there is an* $\varepsilon(n) > 0$ *such that any complete Ricci flat manifold* (M,g) *that satisfies*

$$\operatorname{vol} B(p,r) \geq (1-\varepsilon)\,\omega_n r^n$$

for some $p \in M$ *is isometric to Euclidean space.*

Proof. First observe that on any complete Riemannian manifold with $\operatorname{Ric} \geq 0$, relative volume comparison can be used to show that

$$\operatorname{vol} B(p,r) \geq (1-\varepsilon)\,\omega_n r^n$$

as long as

$$\lim_{r\to\infty} \frac{\operatorname{vol} B(p,r)}{\omega_n r^n} \geq (1-\varepsilon).$$

Therefore, 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 $\operatorname{Ric} \geq 0$ isn't changed by scaling.

If our assertion is assumed to be false, then for each integer i there is a Ricci flat manifold (M_i, g_i) with

$$\lim_{r\to\infty} \frac{\mathrm{vol}B\left(p_i, r\right)}{\omega_n r^n} \geq \left(1 - i^{-1}\right),$$
$$\|(M_i, g_i)\|^{har}_{C^{1,\alpha},r} \neq 0 \text{ for all } r > 0.$$

By scaling these metrics suitably, it is then possible to arrange it so that we have a sequence of Ricci flat manifolds $(M_i, \bar{g}_i, q_i)$ with

$$\lim_{r\to\infty} \frac{\mathrm{vol}B\left(q_i, r\right)}{\omega_n r^n} \geq \left(1 - i^{-1}\right),$$
$$\|(M_i, \bar{g}_i)\|^{har}_{C^{1,\alpha},1} \leq 1,$$
$$\|(M_i, \bar{g}_i, q_i)\|^{har}_{C^{1,\alpha},1} \in [0.5, 1].$$

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, g, q). 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

$$\lim_{r\to\infty} \frac{\mathrm{vol}B\left(q, r\right)}{\omega_n r^n} = 1.$$

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 11.4.6. *Let $n \geq 2$, $-\infty < \lambda \leq \Lambda < \infty$, and D, $R \in (0,\infty)$ be given. There is a $\delta = \delta\left(n, \lambda \cdot R^2\right)$ such that the class of closed Riemannian n-manifolds satisfying*

$$(n-1)\,\Lambda \geq \mathrm{Ric} \geq (n-1)\,\lambda,$$
$$\mathrm{diam} \leq D,$$
$$\mathrm{vol}B\left(p, R\right) \geq (1-\delta)\, v\left(n, \lambda, R\right)$$

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. Observe that if we have a sequence $(M_i, \bar{g}_i, p_i)$ where $\bar{g}_i = k_i^2 g_i$, $k_i \to \infty$, and the (M_i, g_i) lie in the above class, then the volume condition reads

$$\begin{aligned}
\mathrm{vol}B_{\bar{g}_i}\left(p_i, R\cdot k_i\right) &= k_i^n \mathrm{vol}B_{g_i}\left(p_i, R\right)\\
&\geq k_i^n\,(1-\delta)\, v\left(n, \lambda, R\right)\\
&= (1-\delta)\, v\left(n, \lambda\cdot k_i^{-2}, R\cdot k_i\right).
\end{aligned}$$

From relative volume comparison we can then conclude that for $r \leq R \cdot k_i$ and very large i,

$$\mathrm{vol}B_{\bar{g}_i}(p_i, r) \geq (1-\delta)\, v\left(n, \lambda \cdot k_i^{-2}, r\right) \sim (1-\delta)\,\omega_n r^n.$$

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. □

11.4.3 Sectional Curvature

Given the results for Ricci curvature we immediately obtain.

Theorem 11.4.7 (The Convergence Theorem of Riemannian Geometry). *Given R, $K > 0$, there exist $Q, r > 0$ such that any (M, g) with*

$$\mathrm{inj} \geq R,$$
$$|\sec| \leq K$$

has $\|(M,g)\|^{har}_{C^{1,\alpha},r} \leq Q$. In particular, this class is compact in the pointed $C^{1,\alpha}$ topology for all $\alpha < 1$.

Using the diameter bound in positive curvature and Klingenberg's estimate for the injectivity radius from theorem 6.5.1 we get

Corollary 11.4.8 (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.*

A similar result was also proven by A. Weinstein at the same time. The hypotheses are the same, but Weinstein showed that the class contained finitely many homotopy 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 we work with pointed spaces and also to some extent incomplete manifolds as it isn't clear from the beginning that the complete manifolds in question have global lower bounds for the injectivity radius.

Lemma 11.4.9 (Cheeger, 1967). *Given $n \geq 2$, $v, K > 0$, and a compact n-manifold (M, g) with*

$$|\sec| \leq K,$$
$$\mathrm{vol}B(p, 1) \geq v,$$

for all $p \in M$, then $\mathrm{inj}M \geq R$, *where R depends only on n, K, and v.*

Proof. As for Ricci curvature we can use a contradiction type argument. So assume we have (M_i, g_i) with $\mathrm{inj}M_i \to 0$ and satisfying the assumptions of the lemma. Find $p_i \in M_i$ with $\mathrm{inj}_{p_i} = \mathrm{inj}(M_i, g_i)$ and consider the pointed sequence $(M_i, p_i, \bar{g}_i)$, where $\bar{g}_i = (\mathrm{inj}M_i)^{-2} g_i$ is rescaled so that

$$\mathrm{inj}(M_i, \bar{g}_i) = 1,$$
$$|\sec(M_i, \bar{g}_i)| \leq (\mathrm{inj}(M_i, g_i))^2 \cdot K = K_i \to 0.$$

Now some subsequence of $(M_i, \bar{g}_i, p_i)$ will converge in the pointed $C^{1,\alpha}, \alpha < 1$, topology to a manifold (M, g, p). Moreover, this manifold is flat since $\|(M, g)\|_{C^{1,\alpha},1} = 0$.

The first observation about (M, g, p) is that $\mathrm{inj}(p) \leq 1$. This follows because the conjugate radius for $(M_i, \bar{g}_i)$ is $\geq \pi/\sqrt{K_i} \to \infty$, so Klingenberg's estimate for the injectivity radius (lemma 6.4.7) implies that there must be a geodesic loop of length 2 at $p_i \in M_i$. Since $(M_i, \bar{g}_i, p_i) \to (M, g, p)$ in the pointed $C^{1,\alpha}$ topology, the geodesic loops must converge to a geodesic loop of length 2 in M based at p. Hence, $\mathrm{inj}(M) \leq 1$.

The other contradictory observation is that $(M, g) = (\mathbb{R}^n, g_{\mathbb{R}^n})$. Using the assumption $\mathrm{vol}B(p_i, 1) \geq v$ the relative volume comparison (see lemma 7.1.4) shows that there is a $v'(n, K, v)$ such that $\mathrm{vol}B(p_i, r) \geq v' \cdot r^n$, for $r \leq 1$. The rescaled manifold $(M_i, \bar{g}_i)$ then satisfies $\mathrm{vol}B(p_i, r) \geq v' \cdot r^n$, for $r \leq (\mathrm{inj}(M_i, g_i))^{-1}$. Using again that $(M_i, \bar{g}_i, p_i) \to (M, g, p)$ in the pointed C^α topology, we get $\mathrm{vol}B(p, r) \geq v' \cdot r^n$ for all r. Since (M, g) is flat, this shows that it must be Euclidean space.

To justify the last statement 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). So if M is not simply connected, then there is an intermediate covering $\mathbb{R}^n \to \hat{M} \to M$, where $\pi_1\left(\hat{M}\right) = \mathbb{Z}$. This means that $\hat{M} = \mathbb{R}^{n-1} \times S^1(R)$ for some $R > 0$. Hence, for any $p \in \hat{M}$ we must have

$$\lim_{r\to\infty} \frac{\mathrm{vol}B(p, r)}{r^{n-1}} < \infty.$$

The same must then also hold for M itself, contradicting our volume growth assumption. $\square$

This lemma was proved with a more direct method by Cheeger. We have included this 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 11.4.10 (Cheeger, 1967). *Let $n \geq 2$, $K, D, v > 0$ be given. The class of closed Riemannian n-manifolds with*

$$|\sec| \leq K,$$

$$\operatorname{diam} \leq D,$$

$$\operatorname{vol} \geq v$$

is precompact in the $C^{1,\alpha}$ topology for any $\alpha \in (0,1)$ and in particular, contains only finitely many diffeomorphism types.

11.4.4 Lower Curvature Bounds

It is also possible to obtain similar compactness results for manifolds that only have lower curvature bounds as long as we also assume that the injectivity radius is bounded from below.

We give a proof in the case of lower sectional curvature bounds and mention the analogous result for lower Ricci curvature bounds.

Theorem 11.4.11. *Given R, $k > 0$, there exist Q, r depending on R, k such that any manifold (M, g) with*

$$\sec \geq -k^2,$$

$$\operatorname{inj} \geq R$$

satisfies $\|(M, g)\|_{C^1,r} \leq Q$.

Proof. It suffices to get a Hessian estimate for distance functions $r(x) = |xp|$. Lemma 6.4.3 shows that

$$\operatorname{Hess} r(x) \leq k \cdot \coth(k \cdot r(x)) g_r$$

for all $x \in B(p, R) - \{p\}$. Conversely, if $r(x_0) < R$, then $r(x)$ is supported from below by $f(x) = R - |xy_0|$, where $y_0 = c(R)$ and c is the unique unit speed geodesic that minimizes the distance from p to x_0. Thus

$$\operatorname{Hess} r \geq \operatorname{Hess} f \geq -k \cdot \coth(|x_0 y_0| \cdot k) g_r = -k \cdot \coth(k(R - r(x_0))) g_r$$

at x_0. Hence $|\mathrm{Hess} r| \le Q(k, R)$ on metric balls $B(x, r)$ where $|xp| \ge R/4$ and $r \le R/4$.

For fixed $p \in M$ choose an orthonormal basis $e_1, \dots, e_n$ for T_pM and geodesics $c_i(t)$ with $c_i(0) = p$, $\dot{c}_i(0) = e_i$. We use the distance functions

$$r^i(x) = \left|x\, c_i\left(-\frac{R}{2}\right)\right| : B\left(p, \frac{R}{4}\right) \to \mathbb{R}$$

to create a potential coordinate system

$$\psi(x) = \left(r^1(x), \dots, r^n(x)\right) - \left(r^1(p), \dots, r^n(p)\right).$$

By construction $D\psi|_p(e_i)$ is the standard basis for $T_0\mathbb{R}^n$. In particular, ψ defines a coordinate chart on some neighborhood of p with $g_{ij}|_p = \delta_{ij}$. While we can't define g_{ij} on $B(p, R/4)$, the potential inverse $g^{ij} = g\left(\nabla r^i, \nabla r^j\right)$ is defined on the entire region. The Hessian estimates combined with the fact that $\left|\nabla r^k\right| = 1$ imply that $\left|dg^{ij}\right| \le Q(n, k, R)$ on $B(p, R/4)$. In particular, $\left|\left[\delta^{ij} - g^{ij}|_x\right]\right| < 1/10$ for $x \in B(p, \delta(n, k, R))$. This implies that g^{ij} has a well-defined inverse g_{ij} on $B(p, \delta)$ with the properties that $\left|\left[g_{ij}|_p - g_{ij}|_x\right]\right| \le 1/9$ and $\left|dg_{ij}\right| \le C(n, K, R)$ on $B(p, \delta)$.

By inspecting the proof of the inverse function theorem we conclude that ψ is injective on $B(p, \delta)$ and that $B(0, \delta/4) \subset \psi(B(p, \delta))$ (see also exercise 6.7.23). Moreover, we have also established n1 and n2. □

Example 11.4.12. This theorem is actually optimal. Consider rotationally symmetric metrics $dr^2 + \phi_\varepsilon^2(r)d\theta^2$, where ϕ_ε is concave and satisfies

$$\phi_\varepsilon(r) = \begin{cases} r & \text{for } 0 \le r \le 1 - \varepsilon, \\ \frac{3}{4}r & \text{for } 1 + \varepsilon \le r. \end{cases}$$

These metrics have $\sec \ge 0$ and $\mathrm{inj} \ge 1$. 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 $\mathrm{inj} \ge R$, $\sec \ge -k^2$ can therefore not in general be assumed to be smoother than the above example.

Example 11.4.13. With a more careful construction, we can also find ψ_ε with

$$\psi_\varepsilon(r) = \begin{cases} \sin r & \text{for } 0 \le r \le \frac{\pi}{2} - \varepsilon, \\ 1 & \text{for } \frac{\pi}{2} \le r. \end{cases}$$

Then the metric $dr^2 + \psi_\varepsilon^2(r)d\theta^2$ satisfies $|\sec| \le 4$ and $\mathrm{inj} \ge \frac{1}{4}$. As $\varepsilon \to 0$, we get a limit metric that is $C^{1,1}$. We have, however, only shown that such limit spaces are $C^{1,\alpha}$ for all $\alpha < 1$.

Unlike the situation for bounded curvature we cannot get injectivity radius bounds when the curvature is only bounded from below. The above examples are easily adapted to give the following examples.

EXERCISE 11.4.14. Given $a \in (0,1)$ and $\epsilon > 0$, there is a smooth concave function $\rho_\epsilon(r)$ with the property that

$$\rho_\varepsilon(r) = \begin{cases} r & \text{for } 0 \le r \le \varepsilon, \\ ar & \text{for } 2\varepsilon \le r. \end{cases}$$

The corresponding surfaces $dr^2 + \rho_\varepsilon^2(r)d\theta^2$ have $\sec \ge 0$ and $\operatorname{inj} \le 5\epsilon$, while the volume of any R ball is always $\ge a\pi R^2$.

Finally we mention the Ricci curvature result.

Theorem 11.4.15 (Anderson-Cheeger, 1992). *Given $R, k > 0$ and $\alpha \in (0,1)$ there exist Q, r depending on n, R, k such that any manifold (M^n, g) with*

$$\operatorname{Ric} \ge -(n-1)k^2,$$
$$\operatorname{inj} \ge R$$

satisfies $\|(M,g)\|^{har}_{C^\alpha,r} \le Q$.

The proof of this result is again by contradiction and uses most of the ideas we have already covered. However, since the harmonic norm does not work well without control on the derivatives of the metric it is necessary to use the Sobolev spaces $W^{1,p} \subset C^{1-n/p}$ to define a new harmonic norm with L^p control on the derivatives. For the contradiction part of the argument we need to use distance functions as above, but we only obtain bounds on their Laplacians. By inspecting how these bounds are obtained we can show that they $\to 0$ as $\operatorname{inj} \to \infty$ and $k \to 0$. This will assist in showing that the limit space is Euclidean space. For more details see the original paper [4].

11.4.5 Curvature Pinching

Let us turn our attention to some applications of these compactness theorems. One natural subject to explore is that of *pinching* results. Recall from corollary 5.6.14 that complete constant curvature manifolds have uniquely defined universal coverings. 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 theorems we have already presented.

Theorem 11.4.16. *Given* $n \geq 2,\ R,\ D > 0$*, and* $\lambda \in \mathbb{R}$*, there is an* $\varepsilon(n, \lambda, D, R) > 0$ *such that any closed Riemannian n-manifold* (M, g) *with*

$$\begin{aligned} \operatorname{diam} &\leq D, \\ \operatorname{inj} &\geq R, \\ |\operatorname{Ric} - \lambda g| &\leq \varepsilon \end{aligned}$$

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 is false, there would be 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, $\left|\operatorname{Ric}_{g_i} - \lambda g_i\right| \to 0$. Using harmonic coordinates we 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 volume it follows that the injectivity radius condition could have been replaced with an almost maximal volume condition. Now let us see what happens with sectional curvature.

Theorem 11.4.17. *Given* $n \geq 2,\ v,\ D > 0$*, and* $\lambda \in \mathbb{R}$*, there is an* $\varepsilon(n, \lambda, D, v) > 0$ *such that any closed Riemannian n-manifold* (M, g) *with*

$$\begin{aligned} \operatorname{diam} &\leq D, \\ \operatorname{vol} &\geq v, \\ |\sec - \lambda| &\leq \varepsilon \end{aligned}$$

is $C^{1,\alpha}$ *close to a metric of constant curvature* λ*.*

Proof. In this case first observe that Cheeger's lemma 11.4.9 gives us a lower bound for the injectivity radius. The previous theorem then shows that such metrics must be close to Einstein metrics. We have to check that if $(M_i, g_i) \to (M, g)$, where $\left|\sec_{g_i} - \lambda\right| \to 0$ and $\operatorname{Ric}_g = (n-1)\lambda g$, 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 to write $g_i = dr^2 + g_{r,i}$ and $g = dr^2 + g_r$. Since the metrics converge in $C^{1,\alpha}$, we certainly have that $g_{r,i}$ converge to g_r. Using the curvature pinching, we conclude from theorem 6.4.3

$$\frac{\mathrm{sn}'_{\lambda+\varepsilon_i}(r_i)}{\mathrm{sn}_{\lambda+\varepsilon_i}(r_i)} g_{r,i} \leq \mathrm{Hess} r_i \leq \frac{\mathrm{sn}'_{\lambda-\varepsilon_i}(r_i)}{\mathrm{sn}_{\lambda-\varepsilon_i}(r_i)} g_{r,i}$$

with $\varepsilon_i \to 0$. Using that the metrics converge in $C^{1,\alpha}$ it follows that the limit metric satisfies

$$\mathrm{Hess} r = \frac{\mathrm{sn}'_{\lambda}(r)}{\mathrm{sn}_{\lambda}(r)} g_r.$$

Corollary 4.3.4 then implies that the limit metric has constant curvature λ. □

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 [91], [88], and [36]).

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 theorem 6.5.1, 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 11.4.18. *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 [58]. 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. Using theorem 6.5.5 we can only conclude that.

Corollary 11.4.19. *Given $n \geq 2$, and $\lambda > 0$, there is an $\varepsilon = \varepsilon(n,\lambda) > 0$ such that any closed simply connected Riemannian n-manifold (M,g) with*

$$|\sec - \lambda| \leq \varepsilon$$

is $C^{1,\alpha}$ close to a metric of constant curvature λ.

Also recall the quarter pinching results in positive curvature that we proved in section 12.3. There the conclusions were much weaker and purely topological. These results have more recently been significantly improved using Ricci flow techniques. First in [16] to the situation where the curvature operator is positive and next in [20] to the case where the complex sectional curvatures are positive.

In negative curvature some special things also happen. Namely, Heintze has shown that any complete manifold with $-1 \leq \sec < 0$ has a lower volume bound when the dimension ≥ 4 (see also [52] for a more general statement). The lower volume bound is therefore an extraneous condition when doing pinching in negative curvature. However, unlike the situation in positive curvature the upper diameter bound is crucial. See, e.g., [55] and [43] 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 the diameter bound. The volume condition is also necessary, as the Heisenberg group from the exercise 4.7.22 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

$$|\sec| \leq \varepsilon(n),$$
$$\operatorname{diam} \leq 1$$

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. Interestingly, there is also a Ricci flow type proof of this result in [94]. For more on collapsing in general, the reader can start by reading [44].

11.5 Further Study

Cheeger first proved his finiteness theorem and put down the ideas of C^k convergence for manifolds in [25]. They later appeared in journal form [26], 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 [27]. For more generalities on convergence and their uses we recommend the surveys by Anderson, Fukaya, Petersen, and Yamaguchi in [51]. Also for more on norms and convergence theorems the survey by Petersen in [54] might prove useful. The text [53] should also be mentioned again. It was probably the original french version of 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 [86] 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.

11.6 Exercises

EXERCISE 11.6.1. 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.

EXERCISE 11.6.2. Assume that we have a map (not necessarily continuous) $F : X \to Y$ between metric spaces such that for some $\epsilon > 0$:

$$||x_1x_2| - |F(x_1)F(x_2)|| \leq \epsilon,\ x_1x_2 \in X$$

and

$$F(X) \subset Y \text{ is } \epsilon\text{-dense.}$$

Show that $d_{G-H}(X, Y) < 2\epsilon$.

EXERCISE 11.6.3. C. Croke has shown that there is a universal constant $c(n)$ such that any n-manifold with inj $\geq R$ satisfies vol$B(p, r) \geq c(n) \cdot r^n$ for $r \leq \frac{R}{2}$. Use this to show that the class of n-dimensional manifolds satisfying inj $\geq R$ and vol $\leq V$ is precompact in the Gromov-Hausdorff topology.

EXERCISE 11.6.4. Let (M, g) be a complete Riemannian n-manifold with Ric $\geq (n-1)k$. Show that there exists a constant $C(n, k)$ with the property that for each $\epsilon \in (0, 1)$ there exists a cover of metric balls $B(x_i, \epsilon)$ with the property that no more than $C(n, k)$ of the balls $B(x_i, 5\epsilon)$ can have nonempty intersection.

EXERCISE 11.6.5. Show that there are Bochner formulas for Hess $\left(\frac{1}{2}g(X, Y)\right)$ and $\Delta\frac{1}{2}g(X, Y)$, where X and Y are vector fields with symmetric ∇X and ∇Y. This can be used to prove the formulas relating Ricci curvature to the metric in harmonic coordinates.

EXERCISE 11.6.6. Show that in contrast to 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 div$X$.

EXERCISE 11.6.7. 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 inj $(p) \geq \min\{R, R \cdot d(p, \partial)\}$, $R < 1$, is precompact in the $C^{1,\alpha}$ topology.

EXERCISE 11.6.8. Define a *weighted norm* concept. That is, fix a positive function $\rho(R)$, and assume that in a pointed manifold (M, g, p) the points on the distance spheres $S(p, R)$ have norm $\leq \rho(R)$. Prove the corresponding fundamental theorem.

EXERCISE 11.6.9. Assume $\mathscr{M}$ is a class of compact Riemannian n-manifolds that is compact in the $C^{m,\alpha}$ topology. Show that there is a function $f(r)$, where $f(r) \to 0$ as $r \to 0$, depending on $\mathscr{M}$ such that $\|(M, g)\|_{C^{m,\alpha},r} \leq f(r)$ for all $M \in \mathscr{M}$.

EXERCISE 11.6.10. 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 $\text{inj} \geq R$, 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.

EXERCISE 11.6.11. Consider the singular Riemannian metric $dt^2 + (at)^2\, d\theta^2$, $a > 1$, 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.

EXERCISE 11.6.12. Show that the class of spaces with $\text{inj} \geq R$ and $\left|\nabla^k \text{Ric}\right| \leq \Lambda$ for $k = 0, \ldots, m$ is compact in the $C^{m+1,\alpha}$ topology.

EXERCISE 11.6.13 (S-h. Zhu). Consider the class of complete or compact n-dimensional Riemannian manifolds with

$$\text{conj.rad} \geq R,$$
$$|\text{Ric}| \leq \Lambda,$$
$$\text{vol} B\,(p, 1) \geq v.$$

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.

EXERCISE 11.6.14. Using the Eguchi-Hanson metrics from exercise 4.7.23 show that one cannot in general expect a compactness result for the class

$$|\text{Ric}| \leq \Lambda,$$
$$\text{vol} B\,(p, 1) \geq v.$$

Thus, one must assume either that v is large as we did before or that there a lower bound for the conjugate radius.

EXERCISE 11.6.15. 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.

(1) 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.
(2) Show that the class of manifolds with bounded weak norm is precompact in the Gromov-Hausdorff topology.
(3) Show that (M, g) is flat if and only if the weak norm is zero on all scales.

Chapter 12
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 following section. The first applications of these new ideas are to sphere theorems. We then prove the soul theorem of Cheeger and Gromoll. After that, 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 proof of the quarter pinched sphere theorem. Toponogov himself proved these comparison 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 theorems 7.2.5 and 7.3.5, these results in fact hold in the Ricci curvature setting. The next use of Toponogov's theorem 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 more basic comparison techniques. Still, it is convenient to have 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 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.

P. Petersen, *Riemannian Geometry*, Graduate Texts in Mathematics 171,
DOI 10.1007/978-3-319-26654-1_12

The idea of triangle comparison for surfaces goes back to Alexandrov who in turn influenced Toponogov, however it is interesting to note that in fact Pizzetti had already established the local triangle comparison on surfaces at the beginning of the 20th century (see [84]).

12.1 Critical Point Theory

In the generalized critical point theory developed here, 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 12.1.1. *Let* (M, g) *be a Riemannian manifold and* $f : M \to \mathbb{R}$ *a proper smooth function. If* f *has no critical values in the closed interval* $[a, b]$, *then the pre-images* $f^{-1}([-\infty, b])$ *and* $f^{-1}([-\infty, a])$ *are diffeomorphic. Furthermore, there is a deformation retraction of* $f^{-1}([-\infty, b])$ *onto* $f^{-1}([-\infty, a])$, *in particular, the inclusion*

$$f^{-1}([-\infty, a]) \hookrightarrow f^{-1}([-\infty, b])$$

is a homotopy equivalence.

Proof. The idea for creating such a retraction is to follow the negative gradient field of f. Since there are no critical points for f the gradient $-\nabla f$ is nonzero everywhere on $f^{-1}([a, b])$. Next 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 therefore must be complete (integral curves are defined for all time). Let F^t denote the flow for this vector field. (See figure 12.1)

For fixed $q \in M$ consider the function $t \mapsto f(F^t(q))$. The derivative of this function is $g(X, \nabla f)$, so as long as the integral curve $t \mapsto F^t(q)$ remains in $f^{-1}([a, b])$, the function $t \mapsto f(F^t(q))$ is linear with derivative -1. In particular, the diffeomorphism $F^{b-a} : M \to M$ must carry $f^{-1}([-\infty, b])$ diffeomorphically into $f^{-1}([-\infty, a])$.

The desired retraction is given by:

$$r_t : f^{-1}([-\infty, b]) \to f^{-1}([-\infty, b]),$$

$$r_t(p) = \begin{cases} p & \text{if } f(p) \le a, \\ F^{t(f(p)-a)}(p) & \text{if } a \le f(p) \le b. \end{cases}$$

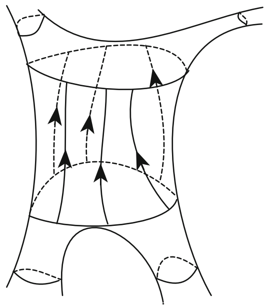


Fig. 12.1 Gradient Flow Deformation

Then $r_0 = id$, and r_1 maps $f^{-1}([-\infty, b])$ diffeomorphically into $f^{-1}([-\infty, a])$. $\square$

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 try to generalize this lemma to functions that are not even C^1. To minimize technicalities we work exclusively with distance functions. There is, however, a more general theory for Lipschitz functions that could be used in this context (see [33]). Suppose (M, g) is complete and $K \subset M$ a compact subset. Then the distance function

$$r(x) = |xK| = \min\{|xp| \mid p \in K\}$$

is proper. Wherever this function is smooth, we know that it has unit gradient and therefore is 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, we list all the possible values it could have (see also exercise 5.9.28 for more details on differentiability of distance functions). Define $\overline{pq}$ to be a choice of a unit speed segment from p to q, its initial velocity is $\overrightarrow{pq}$, and $\overrightarrow{\overrightarrow{pq}}$ the set of all such unit vectors $\overrightarrow{pq}$. We can also replace q by a set K with the understanding that we only consider the segments from p to K of length $|pK|$. In the case where r is smooth at x, we clearly have that $-\nabla r = \overrightarrow{xK}$. At other points, $\overrightarrow{\overrightarrow{xK}}$ might contain more vectors. We say that r is *regular, or noncritical,* at x if the set $\overrightarrow{\overrightarrow{xK}}$ is contained in an open hemisphere of the unit sphere in T_xM. The center of any such hemisphere is then a possible averaged direction for the negative gradient of r at x. Stated differently, we have that r is regular at x if and only if there is a unit vector $v \in T_xM$ such that $\angle\left(v, \overrightarrow{xK}\right) < \pi/2$ for all $\overrightarrow{xK} \in \overrightarrow{\overrightarrow{xK}}$. Clearly v is the center of such a hemisphere. We can quantify being regular by saying that r is α-regular at x if there exists $v \in T_xM$ such that $\angle\left(v, \overrightarrow{xK}\right) < \alpha$ for all $\overrightarrow{xK} \in \overrightarrow{\overrightarrow{xK}}$. Thus, r is regular at x if and only if it is $\pi/2$-regular. Let $R_\alpha(x, K) \subset T_xM$ be the set of all such unit directions v at α-regular points x.

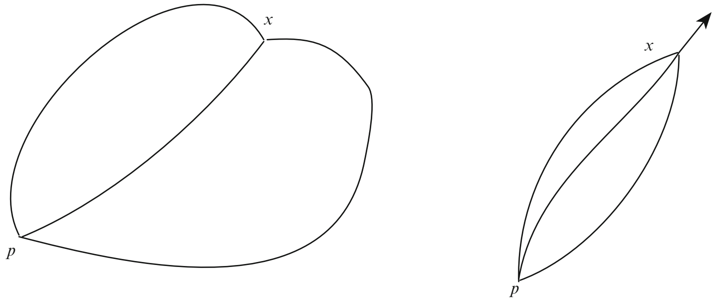


Fig. 12.2 Critical and Regular Points

Evidently, a point x is critical for r if the segments from K to x spread out at x, while it is regular if they more or less point in the same direction (see figure 12.2). It was Berger who first realized and showed that a local maximum must be critical in the above sense. Berger's result is a consequence of the next proposition.

Proposition 12.1.2. *Suppose* (M, g) *and* $r(x) = |xK|$ *are as above. Then:*

(1) $\overrightarrow{\overrightarrow{xK}}$ *is closed and hence compact for all* x.
(2) *The set of* α*-regular points is open in* M.
(3) *The set* $R_\alpha(x, K)$ *is convex for all* $\alpha \le \pi/2$.
(4) *If* U *is an open set of* α*-regular points for* r, *then there is a unit vector field* X *on* U *such that* $X|_x \in R_\alpha(x, K)$ *for all* $x \in U$. *Furthermore, if* c *is an integral curve for* X *and* $s < t$, *then*

$$|c(s)\ K| - |c(t)\ K| > \cos(\alpha)(t - s).$$

Proof. (1) Let $\overline{xq_i}$ be a sequence of unit speed segments from x to K with $\overrightarrow{xq_i}$ converging to some unit vector $v \in T_xM$. Clearly, $\exp_x(tv)$ is the limit of the segments $\overline{xq_i}$ and therefore is a segment itself. Furthermore, since K is closed $\exp_x(|xK|\, v) \in K$.

(2) Suppose $x_i \to x$, and x_i are not α-regular. We shall show that x is not α-regular. This means that for any unit $v \in T_xM$ there is some $\overrightarrow{xK}$ such that $\angle\left(v, \overrightarrow{xK}\right) \ge \alpha$. So fix a unit $v \in T_xM$ and choose a sequence $v_i \in T_{x_i}M$ converging to v. By assumption $\angle\left(v_i, \overrightarrow{x_iK}\right) \ge \alpha$ for some $\overrightarrow{x_iK}$. Now select a subsequence so that the unit vectors $\overrightarrow{x_iK}$ converge to a $w \in T_xM$. Thus $\angle(v, w) \ge \alpha$. Finally note that the segments $\overline{x_i K} = \exp_{x_i}\left(t\,\overrightarrow{x_iK}\right)$, $t \in [0, |x_i K|]$ must converge to the geodesic $\exp_x(tw)$, $t \in [0, |xK|]$ which is then forced to be a segment from x to K.

(3) First observe that for each $w \in T_xM$, the open cone

$$C_\alpha(w) = \{v \in T_xM \mid \angle(v, w) < \alpha\}$$

is convex when $\alpha \leq \pi/2$. Then observe that $R_\alpha(x, K)$ is the intersection of the cones $C_\alpha\left(\overrightarrow{xK}\right)$, $\overrightarrow{xK} \in \overrightarrow{\overrightarrow{xK}}$ and is therefore also convex.

(4) For each $p \in U$ select $v_p \in R_\alpha(p, K)$ and extend v_p to a unit vector field V_p. It follows from the proof of (2) that $V_p(x) \in R_\alpha(x, K)$ 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. Next select a locally finite collection $\{U_i\}$ of U_ps and a corresponding partition of unity λ_i. Then property (3) tells us that the vector field $V = \sum \lambda_i V_i$ is nonzero. Define $X = V/|V|$.

Keep in mind that the flow of X should decrease distances to K so it is easier to consider $-r$ instead of r. Property (4) is clearly true at points where r is smooth, because in that case the derivative of $-(r \circ c)(s) = -|c(s)\ K|$ is

$$g(X, -\nabla r) = \cos\angle(X, -\nabla r) = \cos\angle\left(X, \overrightarrow{xK}\right) > \cos(\alpha).$$

Now observe that since $-r \circ c$ is Lipschitz continuous it is also absolutely continuous. In particular, $-r \circ c$ is almost everywhere differentiable and the integral of its derivative. It might, however, happen that $-r \circ c$ is differentiable at a point x where ∇r is not defined. To see what happens at such points we select a variation $\bar{c}(s,t)$ such that $t \mapsto \bar{c}(0,t)$ is a segment from K to x; $\bar{c}(s,0) = \bar{c}(0,0) \in K$; $\left|\frac{\partial \bar{c}}{\partial t}(s,t)\right|$ is constant in t and hence equal to the length of the t-curves; and $\bar{c}(s,1) = c(s)$ is the integral curve for X through $x = c(0)$. Thus

$$\begin{aligned}
\frac{1}{2}\left|(r \circ c)(s)\right|^2 &\leq \frac{1}{2}\left(\int_0^1 \left|\frac{\partial \bar{c}}{\partial t}\right| dt\right)^2 \\
&\leq \frac{1}{2}\int_0^1 \left|\frac{\partial \bar{c}}{\partial t}\right|^2 dt \\
&= E(\bar{c}_s)
\end{aligned}$$

with equality holding when $s = 0$. In particular, the right-hand side is a support function for the left-hand side. Assuming that $r \circ c$ is differentiable at $s = 0$ we obtain

$$\begin{aligned}
r(x)\frac{d(r \circ c)}{ds}|_{s=0} &= \frac{dE}{ds}|_{s=0} \\
&= g\left(\frac{\partial \bar{c}}{\partial t}(0,1), \frac{\partial \bar{c}}{\partial s}(0,1)\right) \\
&= g\left(\frac{\partial \bar{c}}{\partial t}(0,1), X\right) \\
&= \left|\frac{\partial \bar{c}}{\partial t}(0,1)\right| \cos\left(\angle\left(X, \frac{\partial \bar{c}}{\partial t}\right)\right)
\end{aligned}$$

$$
\begin{aligned}
&= -r(x)\cos\left(\angle\left(X, -\frac{\partial\bar{c}}{\partial t}\right)\right) \\
&= -r(x)\cos\left(\angle\left(X, \overrightarrow{xK}\right)\right).
\end{aligned}
$$

Thus

$$
\begin{aligned}
-\frac{d\,|c(s)\,K|}{ds}|_{s=0} &= \cos\left(\angle\left(X, \overrightarrow{xK}\right)\right) \\
&> \cos\alpha.
\end{aligned}
$$

This proves the desired property.

□

We can now generalize the above retraction lemma.

Lemma 12.1.3 (Grove-Shiohama). *Let (M, g) and $r(x) = |xK|$ be as above. If all points in $r^{-1}([a,b])$ are α-regular for $\alpha < \pi/2$, then $r^{-1}([-\infty, a])$ is homeomorphic to $r^{-1}([-\infty, b])$, and $r^{-1}([-\infty, b])$ deformation retracts onto $r^{-1}([-\infty, a])$.*

Proof. The construction is similar to the first lemma. We can construct a compactly supported "retraction" vector field X such that the flow F^t for X satisfies

$$
r(p) - r\left(F^t(p)\right) > t\cdot\cos(\alpha),\ t \geq 0 \text{ if } p, F^t(p) \in r^{-1}([a,b]).
$$

For each $p \in r^{-1}(b)$ there is a first time $t_p \leq \frac{b-a}{\cos\alpha}$ for which $F^{t_p}(p) \in r^{-1}(a)$. The function $p \mapsto t_p$ is continuous and thus we get the desired retraction

$$
\begin{aligned}
r_t : r^{-1}([-\infty, b]) &\to r^{-1}([-\infty, b]), \\
r_t(p) &= \begin{cases} p & \text{if } r(p) \leq a \\ F^{t\cdot t_p}(p) & \text{if } a \leq r(p) \leq b \end{cases}.
\end{aligned}
$$

□

Remark 12.1.4. The original construction of Grove and Shiohama actually shows something stronger, the distance function can be approximated by smooth functions without critical points on the same region the distance function had no critical points (see [56].) This has also turned out to be important in certain contexts.

The next corollary is our first simple consequence of this lemma.

Corollary 12.1.5. *Suppose K is a compact submanifold of a complete Riemannian manifold (M, g) and that the distance function $|x\,K|$ 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 $|xK|$ increases along the integral curves for $-X$. Moreover, near K the distance function is smooth, and therefore X can be assumed to be equal to $-\overrightarrow{xK}$ near K. Consider the normal exponential map $\exp^{\perp} : T^{\perp}K \to M$. It follows from corollary 5.5.3 that this gives a diffeomorphism from a neighborhood of the zero section in $T^{\perp}K$ onto a neighborhood of K. Also, the curves $t \mapsto \exp(tv)$ for small t coincide with integral curves for $-X$. In particular, for each $v \in T^{\perp}K$ there is a unique integral curve for $-X$ denoted $c_v(t) : (0, \infty) \to M$ such that $\lim_{t \to 0} \dot{c}_v(t) = v$. Now define our diffeomorphism $F : T^{\perp}K \to M$ by

$$F\left(0_p\right) = p \text{ for the origin in } T_p^{\perp}K,$$
$$F(tv) = c_v(t) \text{ where } |v| = 1.$$

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. The integral curves for $-X$ must leave all of the sublevels of the proper function $|xK|$. Consequently they are defined for all $t > 0$. This shows that F is onto. Finally, as it is a diffeomorphism onto a neighborhood of K by the normal exponential map and the flow of a vector field always acts by local diffeomorphisms we see that it has nonsingular differential everywhere. □

12.2 Distance Comparison

In this section we introduce the geometric results that will enable us to check 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 comparison theorem. The proof we present is probably the simplest available and is based upon an idea by H. Karcher (see [32]).

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 $\overline{px}$ and $\overline{xy}$ that form an interior angle α at x, i.e., $\angle\left(\overrightarrow{xp}, \overrightarrow{xy}\right) = \alpha$ if the specified directions are tangent to the given segments. See also figure 12.3.

Triangle: A *triangle* consists of three segments $\overline{xy}$, $\overline{yz}$, $\overline{zx}$ that meet pairwise at the three vertices x, y, z.

In both definitions one could use geodesics instead of segments. It is then possible to have degenerate hinges or triangles where some vertices coincide

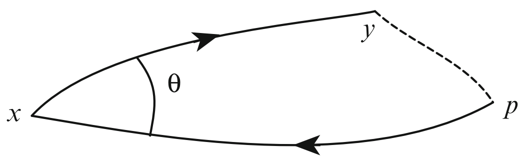


Fig. 12.3 A Hinge

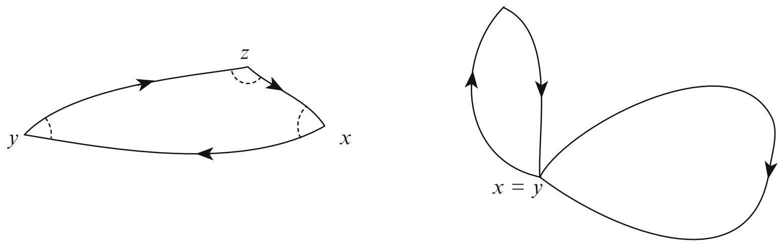


Fig. 12.4 Triangles

without the joining geodesics being trivial. This will be useful in a few situations. In figure 12.4 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 triangle, we can construct *comparison hinges* or *triangles* in the constant curvature spaces S^n_k.

Lemma 12.2.1. *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^n_k *where the corresponding segments have the same length and the angle is the same or all corresponding segments have the same length.*

Proof. Note that when $k > 0$, then corollary 6.3.2 implies $\operatorname{diam} M \leq \pi/\sqrt{k} = \operatorname{diam} S^n_k$. Thus, all segments have length $\leq \pi/\sqrt{k}$.

The hinge case: We have segments $\overline{px}$ and $\overline{xy}$ that form an interior angle $\alpha = \angle\left(\overrightarrow{xp}, \overrightarrow{xy}\right)$ at x. In the space form first choose a segment $\overline{p_k x_k}$ of length $|px|$. At x_k we can then choose a direction $\overrightarrow{x_k y_k}$ so that $\angle\left(\overrightarrow{x_k p_k}, \overrightarrow{x_k y_k}\right) = \alpha$. Then along the unique geodesic going in this direction select y_k so that $|x_k y_k| = |xy|$. This is the desired comparison hinge.

The triangle case: First, pick x_k and y_k such that $|xy| = |x_k y_k|$. Then, consider the two distance spheres $\partial B(x_k, |xz|)$ and $\partial B(y_k, |yz|)$. Since all possible triangle inequalities between x, y, z hold, these distance spheres are nonempty and intersect. Let z_k be any point in the intersection.

To be honest here, we must use Cheng's diameter theorem 7.2.5 in case any of the distances is $\pi/\sqrt{k}$. In this case there is nothing to prove as $(M, g) = S^n_k$. □

The Toponogov comparison theorem can be stated as follows.

Theorem 12.2.2 (Toponogov, 1959). *Let* (M, g) *be a complete Riemannian manifold with* $\sec \geq k$.

Hinge Version: *Given any hinge with vertices* $p, x, y \in M$ *forming an angle* α *at* x, *it follows, that for any comparison hinge in* S^n_k *with vertices* p_k, x_k, y_k *we have:* $|py| \leq |p_k y_k|$ *(see also figure 12.5).*

Triangle Version: *Given any triangle in* M, *it follows that the interior angles are no smaller than the corresponding interior angles for a comparison triangle in* S^n_k.

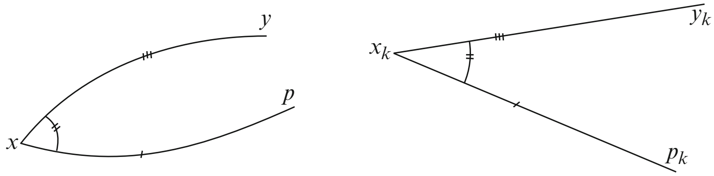
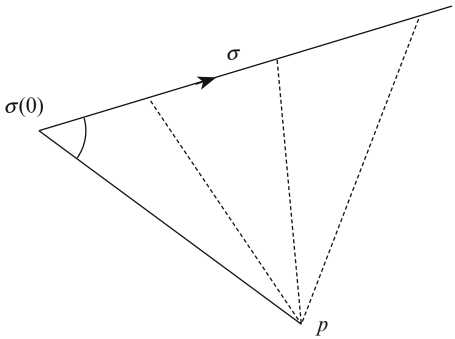


Fig. 12.5 Hinge Comparison

Fig. 12.6 Distance from a point to a line



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 shows that if we have $p, x, y \in S_k^n$ and increase the distance $|py|$ while keeping $|px|$ and $|xy|$ fixed, then the angle at x increases as well. For simplicity, we consider the cases where $k = 1, 0, -1$.

Proposition 12.2.3 (Law of Cosines). *Let a triangle be given in S_k^n with side lengths a, b, c. If α denotes the angle opposite to a, then*

$$k = 0 : a^2 = b^2 + c^2 - 2bc\cos\alpha.$$
$$k = -1 : \cosh a = \cosh b \cosh c - \sinh b \sinh c \cos\alpha.$$
$$k = 1 : \cos a = \cos b \cos c + \sin b \sin c \cos\alpha.$$

Proof. The general setup is the same in all cases. Suppose that a point $p \in S_k^n$ and a unit speed segment $\sigma : [0, c] \to S_k^n$ are given. The goal is to understand the function $|\sigma(t)\, p| = r(\sigma(t))$ (see also figure 12.6). As in corollary 4.3.4 and theorem 5.7.5 we will use the modified distance function f_k. The Hessian is calculated in example 4.3.2 to be $\mathrm{Hess} f_k = (1 - kf_k)\, g$. If we restrict this distance function to $\sigma(t)$, then we obtain a function $\rho(t) = f_k \circ \sigma(t)$ with derivatives

$$\dot{\rho}(t) = g(\dot{\sigma}, \nabla f_k),$$
$$\ddot{\rho}(t) = (1 - k\rho(t)).$$

We now split up into the three cases.

Case $k = 0$: We have more explicitly

$$\rho(t) = \frac{1}{2}(r \circ \sigma(t))^2$$

and

$$\dot{\rho}(t) = g\left(\dot{\sigma}, \nabla \frac{1}{2} r^2\right),$$
$$\ddot{\rho}(t) = 1.$$

So if we define $b = |p\sigma(0)|$ and α as the interior angle between σ and the line joining p with $\sigma(0)$, then

$$\cos(\pi - \alpha) = -\cos\alpha = g(\dot{\sigma}(0), \nabla r).$$

After integration of $\ddot{\rho} = 1$, we get

$$\begin{aligned}\rho(t) &= \rho(0) + \dot{\rho}(0) \cdot t + \frac{1}{2}t^2 \\ &= \frac{1}{2}b^2 - b \cdot \cos\alpha \cdot t + \frac{1}{2}t^2.\end{aligned}$$

Now set $t = c$ and define $a = |p\sigma(c)|$, then

$$\frac{1}{2}a^2 = \frac{1}{2}b^2 - b \cdot c \cdot \cos\alpha + \frac{1}{2}c^2,$$

from which the law of cosines follows.

Case $k = -1$: This time

$$\rho(t) = \cosh(r \circ \sigma(t)) - 1$$

with

$$\dot{\rho}(t) = \sinh(r \circ \sigma(t))\, g(\nabla r, \dot{\sigma}),$$
$$\ddot{\rho}(t) = \rho(t) + 1 = \cosh(r \circ \sigma(t)).$$

As before, we have $b = |p\sigma(0)|$, and the interior angle satisfies

$$\cos(\pi - \alpha) = -\cos\alpha = g(\dot{\sigma}(0), \nabla r).$$

Thus, we must solve the initial value problem

$$\begin{aligned}\ddot{\rho} - \rho &= 1, \\ \rho(0) &= \cosh(b) - 1, \\ \dot{\rho}(0) &= -\sinh(b)\cos\alpha.\end{aligned}$$

The general solution is

$$\begin{aligned}\rho(t) &= C_1 \cosh t + C_2 \sinh t - 1 \\ &= (\rho(0) + 1) \cosh t + \dot{\rho}(0) \sinh t - 1.\end{aligned}$$

So if we let $t = c$ and $a = |pc(c)|$ 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 $k = -1$. Now

$$\rho = 1 - \cos(r \circ \sigma(t))$$

and

$$\begin{aligned}\ddot{\rho} + \rho &= 1, \\ \rho(0) &= 1 - \cos(b), \\ \dot{\rho}(0) &= -\sin b \cos \alpha.\end{aligned}$$

Then,

$$\begin{aligned}\rho(t) &= C_1 \cos t + C_2 \sin t + 1 \\ &= (\rho(0) - 1) \cos t + \dot{\rho}(0) \sin t + 1,\end{aligned}$$

and consequently

$$1 - \cos a = -\cos b \cos c - \sin b \sin c \cos \alpha + 1,$$

which implies the law of cosines. □

The proof of the law of cosines suggests that when working in space forms it is easier to work with a modified distance function, the main advantage being that the Hessian is much simpler.

Lemma 12.2.4 (Hessian Comparison). *Let (M, g) be a complete Riemannian manifold, $p \in M$, and $r(x) = |xp|$. If $\sec M \geq k$, then the Hessian of r satisfies*

$$\mathrm{Hess} f_k \leq (1 - k f_k)\, g$$

in the support sense everywhere.

Proof. We start by noting that this estimate was proven in theorem 6.4.3 when the distance function is smooth. The proof can then be finished in the same way as lemma 7.1.9. □

We are ready to prove the hinge version of Toponogov's theorem. The proof is divided into the three cases: $k = 0, -1, 1$ with the same set-up. Let $p \in M$ and a geodesic $c : [0, L] \to M$ be given. Correspondingly, select $\bar{p} \in S_k^n$ and a segment $\bar{c} : [0, L] \to S_k^n$. With the appropriate initial conditions, we claim that

$$|p\, c\,(t)| \leq |\bar{p}\, \bar{c}\,(t)|\,.$$

If we assume that $|xp|$ is smooth at $c\,(0)$. Then the initial conditions are

$$|p\, c\,(0)| \leq |\bar{p}\, \bar{c}\,(0)|\,,$$
$$g\,(\nabla r, \dot{c}\,(0)) \leq g_k\left(\nabla \bar{r}, \frac{d}{dt}\bar{c}\,(0)\right).$$

In case r is not smooth at $c\,(0)$, we can just slide c down along a segment joining p with $c\,(0)$ and use a continuity argument. This also shows that we can assume the stronger initial condition

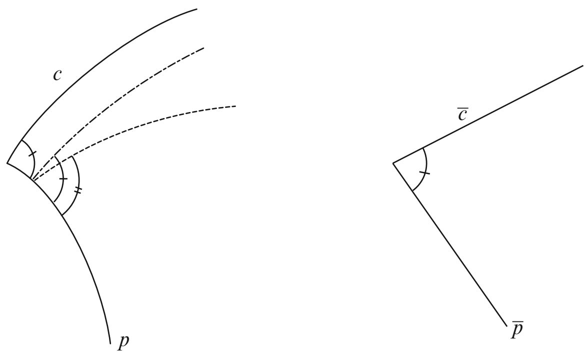
$$|p\, c\,(0)| < |\bar{p}\, \bar{c}\,(0)|\,.$$

In figure 12.7 we have shown how c can be changed by moving it down along a segment joining p and $c\,(0)$. We have also shown how the angles can be slightly decreased. This will be important in the last part of the proof. Note that we could instead have used exercise 5.9.28 to obtain these initial values as the restriction of r to c always has one sided derivatives.

Proof. Case $k = 0$: We consider the modified functions

$$\rho\,(t) = \frac{1}{2}\,(r \circ c\,(t))^2\,,$$
$$\bar{\rho}\,(t) = \frac{1}{2}\,(\bar{r} \circ \bar{c}\,(t))^2\,.$$

Fig. 12.7 Hinge adjustment and comparison hinge



For small t these functions are smooth and satisfy

$$\begin{aligned}\rho(0) &< \bar{\rho}(0), \\ \dot{\rho}(0) &\leq \dot{\bar{\rho}}(0).\end{aligned}$$

Moreover, for the second derivatives we have

$$\begin{aligned}\ddot{\rho} &\leq 1 \text{ in the support sense,} \\ \ddot{\bar{\rho}} &= 1,\end{aligned}$$

whence the difference $\psi(t) = \bar{\rho}(t) - \rho(t)$ satisfies

$$\begin{aligned}\psi(0) &> 0, \\ \dot{\psi}(0) &\geq 0, \\ \ddot{\psi}(t) &\geq 0 \text{ in the support sense.}\end{aligned}$$

This shows that ψ is a convex function that is positive and increasing for small t. Thus, it is increasing and positive for all t. This proves the hinge version.

Case $k = -1$: Consider

$$\begin{aligned}\rho(t) &= \cosh r \circ c(t) - 1, \\ \bar{\rho}(t) &= \cosh \bar{r} \circ \bar{c}(t) - 1.\end{aligned}$$

Then

$$\begin{aligned}\rho(0) &< \bar{\rho}(0), \\ \dot{\rho}(0) &\leq \dot{\bar{\rho}}(0), \\ \ddot{\rho} &\leq \rho + 1 \text{ in the support sense,} \\ \ddot{\bar{\rho}} &= \bar{\rho} + 1.\end{aligned}$$

The difference $\psi = \bar{\rho} - \rho$ satisfies

$$\begin{aligned}\psi(0) &> 0, \\ \dot{\psi}(0) &\geq 0, \\ \ddot{\psi}(t) &\geq \psi(t) \text{ in the support sense.}\end{aligned}$$

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 for small t. It follows that ψ cannot have a positive maximum as that violates convexity. Thus ψ keeps increasing.

Case $k = 1$: This case is considerably harder. We begin as before by defining

$$\rho(t) = 1 - \cos(r \circ c(t)),$$
$$\bar{\rho}(t) = 1 - \cos(\bar{r} \circ \bar{c}(t))$$

and observe that the difference $\psi = \bar{\rho} - \rho$ satisfies

$$\psi(0) > 0,$$
$$\dot{\psi}(0) \geq 0,$$
$$\ddot{\psi}(t) \geq -\psi(t) \text{ in the support sense.}$$

That, however, looks less promising. Even though the function starts out being positive, the last condition only gives a negative lower bound for the second derivative. This is where a standard trick from Sturm-Liouville theory will save us. For that to work well it is best to assume $\dot{\psi}(0) > 0$. Thus, another little continuity argument is necessary as we need to perturb c 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. We compare ψ to a new function $\zeta(t)$ defined by

$$\ddot{\zeta} = -(1+\varepsilon)\zeta,$$
$$\zeta(0) = \psi(0) > 0,$$
$$\dot{\zeta}(0) = \dot{\psi}(0) > 0.$$

For small t we have

$$\begin{aligned}\frac{d^2}{dt^2}(\psi(t) - \zeta(t)) &\geq -\psi(t) + (1+\varepsilon)\zeta(t)\\ &= \zeta(t) - \psi(t) + \varepsilon\zeta(t)\\ &> 0.\end{aligned}$$

This implies that $\psi(t) - \zeta(t) \geq 0$ for small t. To extend this to the interval where $\zeta(t)$ is positive, i.e., for

$$t < \frac{\pi - \arctan\left(\frac{\psi(0)\cdot\sqrt{1+\varepsilon}}{\dot{\psi}(0)}\right)}{\sqrt{1+\varepsilon}},$$

consider the quotient ψ/ζ. This ratio satisfies

$$\frac{\psi}{\zeta}(0) = 1,$$

$$\frac{\psi}{\zeta}(t) \geq 1 \text{ for small } t.$$

Should the ratio dip below 1 before reaching the end of the interval it would have a positive local maximum at some t_0. At this point we can use support functions ψ_δ for ψ from below, and conclude that also ψ_δ/ζ has a local maximum at t_0. Thus, we have

$$\begin{aligned}
0 &\geq \frac{d^2}{dt^2}\left(\frac{\psi_\delta}{\zeta}\right)(t_0) \\
&= \frac{\ddot{\psi}_\delta(t_0)}{\zeta(t_0)} - 2\frac{\dot{\zeta}(t_0)}{\zeta(t_0)} \cdot \frac{d}{dt}\left(\frac{\psi_\delta}{\zeta}\right)(t_0) - \frac{\psi_\delta(t_0)}{\zeta^2(t_0)}\ddot{\zeta}(t_0) \\
&\geq \frac{-\psi_\delta(t_0) - \delta}{\zeta(t_0)} + \frac{\psi_\delta(t_0)}{\zeta(t_0)}(1+\varepsilon) \\
&= \frac{\varepsilon \cdot \psi_\delta(t_0) - \delta}{\zeta(t_0)}.
\end{aligned}$$

But this becomes positive as $\delta \to 0$, since we assumed $\psi_\delta(t_0) > 0$. Thus we have a contradiction. Next, we can let $\varepsilon \to 0$ and finally, let $\psi(0) \to 0$ to get the desired estimate for all $t \leq \pi$ using continuity. □

Remark 12.2.5. 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. When $k > 0$ this means that the geodesic must have length $\leq \pi/\sqrt{k}$. This was precisely the important condition in the last part of the proof.

12.3 Sphere Theorems

Our first applications of the Toponogov theorem are to the case of positively curved manifolds. Using scaling, we can assume throughout this section that we work with a closed Riemannian n-manifold (M, g) with $\sec \geq 1$. For such spaces we have established:

(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 M is simply connected and $\max \sec < 4$, then $\operatorname{inj}(M) \geq \pi/\sqrt{\max \sec}$.
(5) If M is simply connected and $\max \sec < 4$, then M is homotopy equivalent to a sphere.

We can now prove the celebrated Rauch-Berger-Klingenberg sphere theorem, also known as the quarter pinched sphere theorem. Note that the conclusion is stronger than in corollary 6.5.6. The part of the proof presented below is also due the Berger.

Theorem 12.3.1. *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. Select two points $p, q \in M$ such that $|pq| = \operatorname{diam} M$ and note that $\operatorname{diam} M \geq \operatorname{inj} M > \pi/2$. We 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. This certainly makes M look like a sphere as it is the union of two discs. Below we construct an explicit homeomorphism to the sphere in a more general setting.

Fix $x \in M$ and consider the triangle with vertices p, x, q. If, for instance, $|xq| > \pi/2$, then we claim that $|px| < \pi/2$. 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, given a segment $\overline{xq}$ there is a segment $\overline{pq}$ such that the interior angle at q satisfies $\alpha = \angle\left(\overrightarrow{qx}, \overrightarrow{qp}\right) \leq \pi/2$. The hinge version of Toponogov's theorem implies

$$\begin{aligned} \cos|px| &\geq \cos|xq|\cos|pq| + \sin|xq|\sin|pq|\cos\alpha \\ &\geq \cos|xq|\cos|pq|. \end{aligned}$$

Now, both $|xq|, |pq| > \pi/2$, so the left-hand side is positive. This implies that $|px| < \pi/2$ as desired (see also figure 12.8 for the picture on the comparison space). □

Michaleff and Moore in [73] proved a version of this theorem for closed simply connected manifolds that only have positive isotropic curvature (see exercise 3.4.17 and also section 9.4.5). Since quarter pinching implies positive complex sectional curvature and in particular positive isotropic curvature this result is stronger. In fact more recently Brendle and Schoen in [20] have shown that manifolds with positive complex sectional curvature admit metrics with constant curvature. This result uses the Ricci flow.

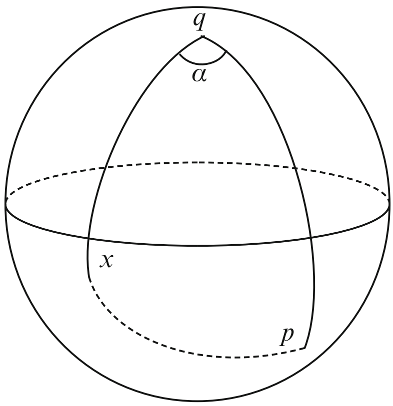


Fig. 12.8 Spherical hinge with long sides

Note that the above theorem does not say anything about the non-simply connected situation. Thus we cannot conclude that such spaces are homeomorphic to spaces of constant curvature. Only that the universal covering is a sphere. The proof in [20], however, does not depend on the fundamental group and thus shows that strictly quarter pinched manifolds admit constant curvature metrics.

The above proof suggests 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 for simply connected manifolds by using Toponogov's theorem to show that there is a point where all geodesic loops have length $> \pi$ and then appealing to the proof of theorem 6.5.4. 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 introduced critical point theory.

Theorem 12.3.2 (Berger, 1962 and 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. We first give Berger's index estimation proof that follows his index proof of the quarter pinched sphere theorem. The goal is to find $p \in M$ such that all geodesic loops at p have length $> \pi$ and then finish by using the proof of theorem 6.5.4. Select $p, q \in M$ such that $|pq| = \operatorname{diam} M > \pi/2$. We claim that p has the desired property. Supposing otherwise we get a geodesic loop $c : [0, 1] \to M$ based at p of length $\leq \pi$. As p is at maximal distance from q we can find a segment $\overline{qp}$, such that the hinge spanned by $\overline{pq}$ and c has interior angle $\leq \pi/2$. While c is not a segment it is sufficiently short that the hinge version of Toponogov's theorem still holds for the degenerate hinge with sides $\overline{qp}$, c and angle $\alpha \leq \pi/2$ at p (see also Fig. 12.9, where we included two geodesics from p to q). Thus

$$\begin{aligned}
0 &> \cos|pq| \\
&\geq \cos|pq| \cos L(c) + \sin|pq| \sin L(c) \cos\alpha \\
&\geq \cos|pq| \cos L(c).
\end{aligned}$$

This is clearly not possible unless $L(c) = 0$.

Next we give the Grove-Shiohama proof. Fix $p, q \in M$ with $|pq| = \operatorname{diam} M > \pi/2$. The claim is that the distance function from p only has q as a critical point (Fig. 12.10). To see this, let $x \in M - \{p, q\}$ and α be the interior angle between any two segments $\overline{xp}$ and $\overline{xq}$. If we suppose that $\alpha \leq \pi/2$, then the hinge version of Toponogov's theorem implies

$$\begin{aligned}
0 &> \cos|pq| \\
&\geq \cos|px| \cos|xq| + \sin|px| \sin|xq| \cos\alpha \\
&\geq \cos|px| \cos|xq|.
\end{aligned}$$

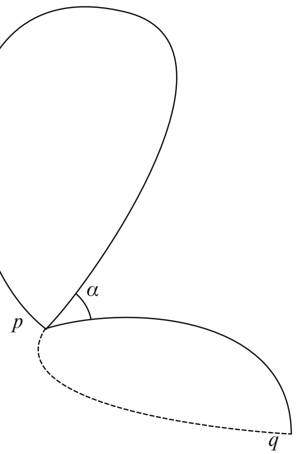


Fig. 12.9 Degenerate hinge where one side is a loop

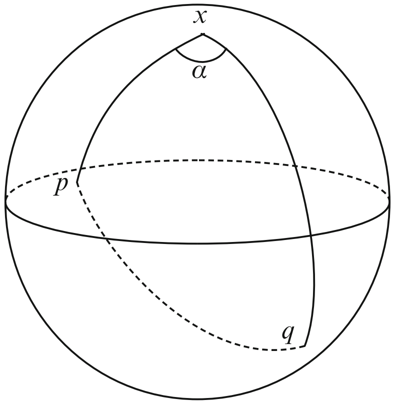


Fig. 12.10 Spherical hinge

But then $\cos|px|$ and $\cos|xq|$ have opposite signs. If, for example, $\cos|px| > 0$ then it follows that $\cos|pq| > \cos|xq|$, which implies $|xq| > |pq| = \operatorname{diam} M$. Thus we have arrived at a contradiction (see also figure 12.10 for the picture on the comparison space).

We construct a vector field X that is the gradient field for $x \mapsto |xp|$ near p and the negative of the gradient field for $x \mapsto |xq|$ 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 $c_x(t)$ for X through x. Suppose that x varies over a small distance sphere $\partial B(p, \varepsilon)$ that is diffeomorphic to S^{n-1}. After time t_x this integral curve will hit the distance sphere $\partial B(q, \varepsilon)$ which can also be assumed to be diffeomorphic to S^{n-1}. The function $x \mapsto t_x$ is continuous and in fact smooth as both distance spheres are smooth submanifolds. Thus we have a diffeomorphism defined by

$$\partial B(p, \varepsilon) \times [0, 1] \to M - (B(p, \varepsilon) \cup B(q, \varepsilon)),$$
$$(x, t) \mapsto c_x(t \cdot t_x).$$

Gluing this map together with the two discs $B(p,\varepsilon)$ and $B(q,\varepsilon)$ then yields a continuous bijection $M \rightarrow S^n$. Note that the construction does not guarantee smoothness of this map on $\partial B(p,\varepsilon)$ and $\partial B(q,\varepsilon)$. □

Aside from the fact that the conclusions in the above theorems could possibly be strengthened to diffeomorphism, we have optimal results. Complex projective space has curvatures in $[1,4]$ and diameter $\pi/2$ and the real projective space has constant curvature 1 and diameter $\pi/2$. If one relaxes the conditions slightly, it is, however, still possible to say something.

Theorem 12.3.3 (Brendle-Schoen 2008 and Petersen-Tao 2009). *Let* (M,g) *be a simply connected of dimension n. There is* $\varepsilon(n) > 0$ *such that if* $1 \leq \sec \leq 4+\varepsilon$*, then M is diffeomorphic to a sphere or one of the projective spaces* $\mathbb{CP}^{n/2}$, $\mathbb{HP}^{n/4}$, $\mathbb{OP}^2$.

The spaces $\mathbb{CP}^{n/2}$, $\mathbb{HP}^{n/4}$, or $\mathbb{OP}^2$ are known as the compact rank 1 symmetric spaces (CROSS). The quaternionic projective space is a quaternionic generalization of complex projective space $\mathbb{HP}^m = S^{4m+3}/S^3$, but the octonion plane is a bit more exotic: $F_4/\mathrm{Spin}(9) = \mathbb{OP}^2$ (see also chapter 10 for more on symmetric spaces spaces). The theorem as stated was proven in [87] and uses convergence theory and the Ricci flow. It relies on a new rigidity result by Brendle and Schoen (see [19]) that generalizes an older result by Berger and several subtle injectivity radius estimates (see also section 6.5.1 for a discussion on this).

For the diameter situation we have:

Theorem 12.3.4 (Grove-Gromoll, 1987 and Wilking, 2001). *If* (M,g) *is closed and satisfies* $\sec \geq 1$, $\operatorname{diam} \geq \pi/2$*, 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* $\mathbb{CP}^m$, $\mathbb{HP}^m$*, or* $\mathbb{CP}^m/\mathbb{Z}_2$ *for m odd.*
(4) *M is isometric to* $\mathbb{OP}^2$.

Grove and Gromoll settled all but part (4), where they only showed that M had to have the cohomology ring of $\mathbb{OP}^2$. It was Wilking who finally settled this last case (see [104]).

12.4 The Soul Theorem

The idea behind the soul theorem is a similar result by Cohn-Vossen for convex surfaces that are complete and noncompact. Such surfaces must contain a core or soul that is either a point or a planar convex circle. In the case of a point the surface is diffeomorphic to a plane. In the case of a circle the surface is isometric to the generalized cylinder over the circle.

Theorem 12.4.1 (Gromoll-Meyer, 1969 and Cheeger-Gromoll, 1972). *If (M, g) is a complete noncompact Riemannian manifold with* $\sec \geq 0$, *then M contains a* soul $S \subset M$. *The soul S is a closed totally convex submanifold and 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 if $\sec > 0$, then M is diffeomorphic to $\mathbb{R}^n$. Soon after, Cheeger-Gromoll established the full theorem. The Gromoll-Meyer theorem is in itself remarkable.

We 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 consider the following less ambitious result.

Lemma 12.4.2 (Gromov's critical point estimate, 1981). *If (M, g) is a complete open manifold of nonnegative sectional curvature, then for every $p \in M$ the distance function $|xp|$ has no critical points outside some ball $B(p, R)$. In particular, M must have the topology of a compact manifold with boundary.*

Proof. Assume we a critical point x for $|xp|$ and that y is chosen so that $\angle\left(\overrightarrow{px}, \overrightarrow{py}\right) \leq \pi/3$. The hinge version of Toponogov's theorem implies that

$$\begin{aligned} |xy|^2 &\leq |py|^2 + |xp|^2 - 2\,|py|\,|xp| \cos\tfrac{\pi}{3} \\ &= |py|^2 + |xp|^2 - |py|\,|xp|\,. \end{aligned}$$

Next use that x is critical for p to select segments $\overline{px}$ and $\overline{xy}$ that form an angle $\leq \pi/2$ at x. Then use the hinge version again to conclude

$$\begin{aligned} |py|^2 &\leq |xp|^2 + |xy|^2 \\ &\leq |xp|^2 + |py|^2 + |xp|^2 - |py|\,|xp| \\ &= 2\,|xp|^2 + |py|^2 - |py|\,|xp|\,. \end{aligned}$$

This forces $|py| \leq 2\,|xp|$.

Now observe that there is a fixed bound on the number of unit vectors at p that mutually form an angle $> \pi/3$. Specifically, use that balls of radius $\pi/6$ around these unit vectors are disjoint inside the unit sphere and note that

$$\frac{v\,(n-1, 1, \pi)}{v\left(n-1, 1, \frac{\pi}{6}\right)} \leq \frac{v\,(n-1, 0, \pi)}{v\left(n-1, 0, \frac{\pi}{6}\right)} \leq 6^{n-1}.$$

Finally, we conclude that there can be at most 6^{n-1} critical points x_i for $|xp|$ such that $|x_{i+1}p| > 2\,|x_i p|$ for all i. This shows, in particular, that the distance function $|xp|$ has no critical points outside some large ball $B(p, R)$. □

The proof of the soul theorem depends on understanding what it means for a submanifold and more generally a subset to be *totally convex*. The notion is similar to being totally geodesic. A subset $A \subset M$ of a Riemannian manifold is said to be *totally convex* if any geodesic in M joining two points in A also lies in A. There are in fact several different kinds of convexity, but as they are not important for any other developments here we confine ourselves to total convexity. The first observation is that this definition agrees with the usual definition of 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. This is not completely trivial, but using the energy functional as in section 6.5.2 we note that if $A \subset M$ is totally convex, then $A \subset M$ is k-connected for any k. It is however, not possible for a closed n-manifold to have n-connected nontrivial subsets as this would violate Poincaré duality. On complete manifolds on the other hand it is sometimes possible to find totally convex sets.

Example 12.4.3. Let (M, g) be the flat cylinder $\mathbb{R} \times S^1$. All of the circles $\{p\} \times S^1$ are geodesics and 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 12.11).

Example 12.4.4. Let (M, g) be a smooth rotationally symmetric metric on $\mathbb{R}^2$ of the form $dr^2 + \rho^2(r)\, d\theta^2$, where $\ddot{\rho} < 0$. 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 12.11).

The way to find totally convex sets is via convexity of functions.

Lemma 12.4.5. *If $f : (M, g) \to \mathbb{R}$ is concave, in the sense that the Hessian is weakly nonpositive everywhere, then every superlevel set $A = \{x \in M \mid f(x) \geq a\}$ is totally convex.*

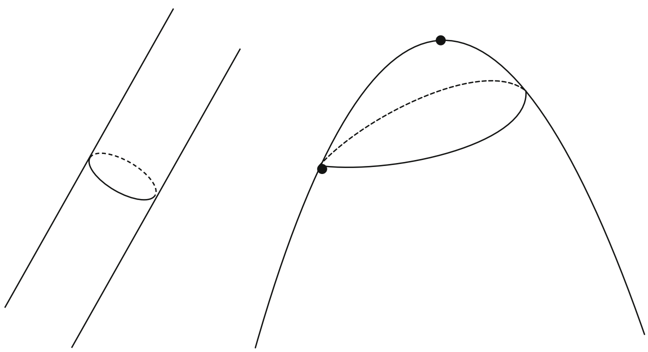


Fig. 12.11 Souls for cylinder and parabola

Proof. Given a geodesic c in M, we have that the function $f \circ c$ has nonpositive weak second derivative. Thus, $f \circ c$ 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. □

We are left with the problem of the existence of proper concave functions on complete manifolds with nonnegative sectional curvature. This requires the notions of rays and Busemann functions from sections 7.3.1 and 7.3.2.

Lemma 12.4.6. *Let* (M, g) *be complete, noncompact, have* $\sec \geq 0$*, and* $p \in M$*. If we take all rays* $R_p = \{c : [0, \infty) \to M \mid c(0) = p\}$ *and construct*

$$f = \inf_{c \in R_p} b_c,$$

where b_c *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 from section 7.3.2 that in nonnegative Ricci curvature Busemann functions are superharmonic. The proof of concavity is almost identical. Instead of the Laplacian estimate for distance functions, we must use a similar Hessian estimate. If $r(x) = |xp|$, then we know that $\operatorname{Hess} r$ vanishes on radial directions $\partial_r = \nabla r$ and satisfies $\operatorname{Hess} r \leq r^{-1} g$ on vectors perpendicular to the radial direction. In particular, $\operatorname{Hess} r \leq r^{-1} g$ at all smooth points. We can then extend this estimate to the points where r isn't smooth as we did for modified distance functions. We can now proceed as in the Ricci curvature case to show that Busemann functions have nonpositive Hessians in the weak sense.

The infimum of a collection of concave functions is clearly also concave. So we must show that the superlevel sets for f are compact. Suppose, on the contrary, that some superlevel set $A = \{x \in M \mid f(x) \geq a\}$ is noncompact. If $a > 0$, then $\{x \in M \mid f(x) \geq 0\}$ is also noncompact. So we can assume that $a \leq 0$. As all of the Busemann functions b_c are zero at p also $f(p) = 0$. In particular, $p \in A$. Using noncompactness select a sequence $p_k \in A$ that goes to infinity. Then consider segments $\overline{pp_k}$, and as in the construction of rays, choose a subsequence so that $\overrightarrow{pp_k}$ converges. This forces the segments to converge to a ray emanating from p. As A is totally convex, all of these segments lie in A. Since A is closed the ray must also lie in A and therefore be one of the rays $c \in R_p$. This leads to a contradiction as

$$a \leq f(c(t)) \leq b_c(c(t)) = -t \to -\infty.$$

□

We need to establish a few fundamental properties of totally convex sets.

Lemma 12.4.7. *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*

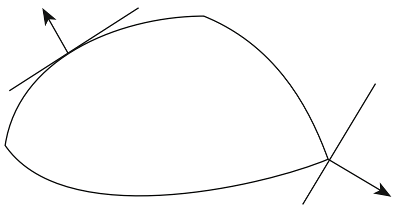


Fig. 12.12 Supporting planes and normals for a convex set

the boundary has the property that for each $x \in \partial A$ there is an inward pointing vector $w \in T_xM$ such that any segment $\overline{xy}$ with $y \in \operatorname{int}A$ has the property that $\angle\left(w, \overrightarrow{xy}\right) < \pi/2$.

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 the topological sense. Most convex sets will in fact not have any topological interior at all. The property about the boundary is called the *supporting hyperplane property*. Namely, the interior of the convex set is supposed to lie on one side of a hyperplane at any of the boundary points. The vector w is the normal to this hyperplane and can be taken to be tangent to a 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 $\operatorname{int}A$ cannot have any critical points on ∂A (see also figure 12.12).

Proof. The convexity radius estimate from theorem 6.4.8 will be used in many places. Specifically we shall use that there is a positive function $\varepsilon(p) : M \to (0, \infty)$ such that $r_p(x) = |xp|$ is smooth and strictly convex on $B(p, \varepsilon(p)) - \{p\}$.

First, let us identify points in the interior and on the boundary. To make the identifications simpler assume that A is closed.

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 submanifold. 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 convex submanifold whose closure is A. This means we can define $\operatorname{int}A = N$ and $\partial A = A - N$.

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. By shrinking N_p if necessary, we can also assume that it is embedded. Thus there exists $\delta \in (0, \varepsilon(p))$ so that $B(p, \delta) \cap N_p = N_p$. The claim is 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 $\delta < \operatorname{inj}_q$. Then we can join each point in $B(p, \delta) \cap N_p$ 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 contained in A (see figure 12.13), thus contradicting maximality of k. This shows that N is an embedded 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 $|pq| < \operatorname{inj}_q$ there is a k-dimensional submanifold $N_p \subset N$ such that $q \in \bar{N}_p$. Specifically, choose a $(k-1)$-dimensional submanifold

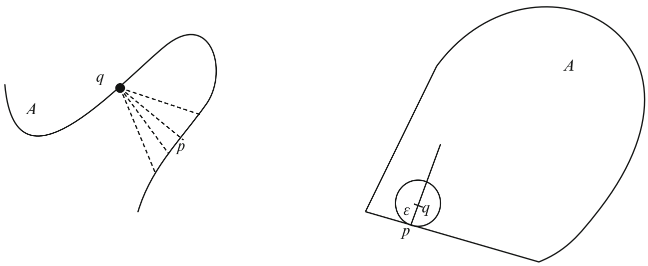


Fig. 12.13 Interior and boundary points of convex sets

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 any segment $\overline{xy}$ with $y \in N$ must, except possibly for x, lie in N. In particular, N is dense in A.

Having identified the interior and boundary, we have to establish the supporting hyperplane property. First, note that since N is totally geodesic its tangent spaces T_qN are preserved by parallel translation along curves in N. For $p \in \partial A$ we then obtain a well-defined k-dimensional tangent space $T_pA \subset T_pM$ coming from parallel translating the tangent spaces to N along curves in N that end at p. Next define the tangent cone at $p \in \partial A$

$$C_pA = \left\{v \in T_pM \mid \exp_p(tv) \in N \text{ for some } t > 0\right\}.$$

Note that if $v \in C_pA$, then in fact $\exp_p(tv) \in N$ for all small $t > 0$. This shows that C_pA is a cone. Clearly $C_pA \subset T_pA$ and is easily seen to be open in T_pA.

In order to prove the supporting half plane property we start by showing that C_pA does not contain antipodal vectors $\pm v$. If it did, then there would be short segments through p whose endpoints line in N. This in turn shows that $p \in N$.

For $p \in \partial A$ and $\varepsilon > 0$ assume that there are $q \in A_\varepsilon = \{x \in A \mid |x\partial A| \geq \varepsilon\}$ with $|qp| = \varepsilon$. 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 $|qp| = \varepsilon$ is dense in ∂A. We start by proving the supporting plane property for such p. We can also assume ε is so small that $r_q(x) = |xq|$ is smooth and convex on a neighborhood containing p. The claim is that $\angle\left(-\nabla r_q, v\right) < \pi/2$ for all $v \in C_pA$. To see this, observe that we have a convex set

$$A' = A \cap \bar{B}(q, \varepsilon),$$

with interior

$$N' = A \cap B(q, \varepsilon) \subset N$$

and $p \in \partial A'$ (see figure 12.13). Thus $C_pA' \subset C_pA$ and $T_pA = T_pA'$. The tangent cone of $\bar{B}(q,\varepsilon)$ is given by

$$C_p\bar{B}(q,\varepsilon) = \left\{v \in T_pM \mid \angle\left(v, -\nabla r_q\right) < \frac{\pi}{2}\right\}$$

as r is smooth at p, thus

$$C_pA' = \left\{v \in T_pA \mid \angle\left(v, -\nabla r_q\right) < \frac{\pi}{2}\right\}.$$

If $C_pA' \subsetneqq C_pA$, then openness of C_pA in T_pA implies C_pA contains antipodal vectors.

At other points $p \in \partial A$ select $p_i \to p$ where

$$C_{p_i}A = \left\{v \in T_{p_i}A \mid \angle(v, -w_i) < \frac{\pi}{2}\right\}.$$

These open half spaces will have an accumulation half space

$$\left\{v \in T_pA \mid \angle(v, -w) < \frac{\pi}{2}\right\}.$$

By continuity $C_pA \subset \left\{v \in T_pA \mid \angle(v,-w) \leq \frac{\pi}{2}\right\}$. As $C_pA \subset T_pA$ is also open it must be contained in an open half space. □

The last lemma we need is

Lemma 12.4.8. *Let* (M,g) *have* $\sec \geq 0$. *If* $A \subset M$ *is totally convex, then the distance function* $r : A \to \mathbb{R}$ *defined by* $r(x) = |x\partial A|$ *is concave on* A. *When* $\sec > 0$, *then any maximum for* r *is unique.*

Proof. We shall show that the Hessian is nonpositive in the support sense. Fix $q \in \text{int}A$, and find $p \in \partial A$ so that $|pq| = |q\partial A|$. Then select a segment $\overline{pq}$ in A. Using exponential coordinates at p we create a hypersurface H which is the image of the hyperplane perpendicular to $\overrightarrow{pq}$. This hypersurface is perpendicular to $\overrightarrow{pq}$, the second fundamental form for H at p is zero, and $H \cap \text{int}A = \varnothing$. (See figure 12.14.) We have that $f(x) = |xH|$ is a support function from above for $r(x) = |x\partial A|$ at all points on $\overline{pq}$.

Select a point $p' \neq p, q$ on the segment $\overline{pq}$, i.e., $|pp'| + |p'q| = |pq|$. One can show as in section 5.7.3 that f is smooth at p' except possibly when $p' = q$. We start by showing that the support function f is concave at $p' \neq q$. Note that $\overline{pq}$ is an integral curve for ∇f. Evaluating the fundamental equation (see 3.2.5) on a parallel field, along $\overline{pq}$, that starts out being tangent to H, i.e., perpendicular to $\overrightarrow{pq}$ therefore yields:

$$\begin{aligned}\frac{d}{dt}\text{Hess}f(E,E) &= -R(E,\nabla f,\nabla f,E) - \text{Hess}^2 f(E,E)\\ &\leq 0.\end{aligned}$$

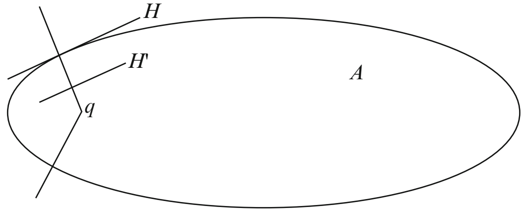


Fig. 12.14 Distance function to the boundary of a convex set

Since $\mathrm{Hess} f(E, E) = 0$ at p we see that $\mathrm{Hess} f(E, E) \leq 0$ along $\overline{pq}$ (and < 0 if $\sec > 0$). This shows that we have a smooth support function for $|x\partial A|$ on an open and dense subset in A.

If f is not smooth at q, we can find a hypersurface H' as above that is perpendicular to $\overrightarrow{p'q}$ at p' and has vanishing second fundamental form at p'. For p' close to q we have that $|xH'|$ is smooth at q and therefore also has nonpositive (negative) Hessian at q. In this case we claim that $|pp'| + |xH'|$ is a support function for $|x\partial A|$. Clearly, the functions are equal at q and we only need to worry about x where $|x\partial A| > |pp'|$. In this case we can select $z \in H'$ with $|x\partial A| = |z\partial A| + |xH'|$. Thus we are reduced to showing that $|z\partial A| \leq |pp'|$ for each $z \in H'$.

As f is smooth at p' it follows that $|x\partial A|$ is concave in a neighborhood of p'. Now select a segment $\overline{p'z}$. By the construction of H' we can assume that $\overline{p'z}$ is contained in H' and therefore perpendicular to $\overrightarrow{p'q}$. Concavity of $x \mapsto |x\partial A|$ along the segment then shows that $|z\partial A| \leq |p'\partial A|$ as it lies under the tangent through p'. This establishes our claim.

Finally, choose a concave $\phi : [0, \infty) \to [0, \infty)$ with $\phi(0) = 0$ and $\phi' > 0$. Then $\phi \circ r$ will clearly also be concave. Moreover, if we select ϕ to be strictly concave and $\sec > 0$, then $\phi \circ r$ will be strictly concave. In case it has a maximum it follows that it is unique as in the construction of a center of mass in section 6.2.2. □

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

$$C_1 = \{x \in M \mid f(x) = \max f\}$$

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 $A = \{x \in M \mid f(x) \geq a\}$ are convex with $\partial A = f^{-1}(a)$, so $f(x) = |x\partial A|$ on A. If C_1 is a submanifold, then we are also done. In this case $|xC_1|$ 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 $|x\partial C_1|$ is concave on C_1. 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 $|xC_2|$ has no critical points, as any point lies on the boundary for a superlevel set for either f or $|x\partial C_1|$. We can iterate this process to obtain 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 S

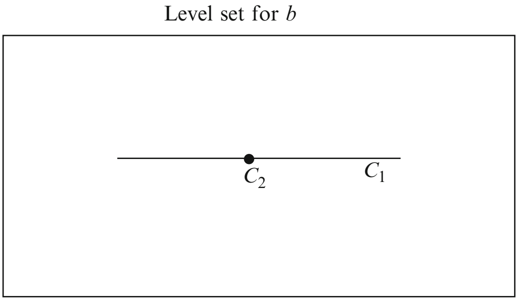


Fig. 12.15 Iteration for soul construction

that we call the soul (see figure 12.15). This is because $\dim C_i > \dim C_{i+1}$. To see this suppose $\dim C_i = \dim C_{i+1}$. Then $\mathrm{int} C_{i+1}$ will be an open subset of $\mathrm{int} C_i$. So if $p \in \mathrm{int} C_{i+1}$, then we can find δ such that

$$B(p,\delta) \cap \mathrm{int} C_{i+1} = B(p,\delta) \cap \mathrm{int} C_i.$$

Now choose a segment c from p to ∂C_i. Clearly $|x\partial C_i|$ is strictly increasing along c. On the other hand, c runs through $B(p,\delta) \cap \mathrm{int} C_i$, thus showing that $|x\partial C_i|$ must be constant on the part of c 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 [54]. We briefly mention two important results:

Theorem 12.4.9. *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 additional follows that S must be a point if all sectional curvatures based at just one point in M 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 [82] 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 theorem 4.5.3 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 [49] 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 [51] describing these and other examples. The construction uses doubly warped product metrics on $I \times S^2 \times S^1$ as described in section 1.4.5.

12.5 Finiteness of Betti Numbers

We prove two results in this section.

Theorem 12.5.1 (Gromov, 1978 and 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 $\mathbb{F}$ of coefficients the Betti numbers are bounded:*

$$\sum_{i=0}^{n} b_i(M, \mathbb{F}) = \sum_{i=0}^{n} \dim H_i(M, \mathbb{F}) \leq C(n).$$

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

$$\underbrace{\left(S^2 \times S^2\right) \sharp \left(S^2 \times S^2\right) \sharp \cdots \sharp \left(S^2 \times S^2\right)}_{k \text{ times}}.$$

For large k, the Betti number bound shows that these connected sums cannot have a metric with nonnegative sectional curvature. Thus, there exist 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 in sections 9.4.4 and 10.3.3 to see how much more is known about these manifolds.

In the context of nonnegative sectional curvature there are three difficult open problems. They were discussed and settled in chapters 9 and 10 for manifolds with nonnegative curvature operator.

(H. Hopf) Does $S^2 \times S^2$ admit a metric with positive sectional curvature?
(H. Hopf) For M^{2n}, does $\sec \geq 0$ (> 0) imply $\chi(M) \geq 0$ (> 0)?
(Gromov) Does $\sec \geq 0$ imply $\sum_{i=0}^{n} b_i(M, \mathbb{F}) \leq 2^n$?

Recall that these questions were also discussed in section 8.3 under additional assumptions about the isometry group.

First we establish part (1) of Gromov's theorem. The proof resembles that of the critical point estimate lemma 12.4.2 from the previous section.

Proof of (1). We construct what is called a *short set* of generators for $\pi_1(M)$. Consider $\pi_1(M)$ as acting by deck transformations on the universal covering $\tilde{M}$ and fix $p \in \tilde{M}$. Inductively select a generating set $\{g_1, g_2, \ldots\}$ such that

(a) $|pg_1(p)| \leq |pg(p)|$ for all $g \in \pi_1(M) - \{e\}$.
(b) $|pg_k(p)| \leq |pg(p)|$ for all $g \in \pi_1(M) - \langle g_1, \ldots, g_{k-1}\rangle$.

We claim that $\angle\left(\overrightarrow{pg_k(p)}, \overrightarrow{pg_l(p)}\right) \geq \pi/3$ for $k < l$. Otherwise, the hinge version of Toponogov's theorem would imply

$$\begin{aligned} |g_l(p)\, g_k(p)|^2 &< |pg_k(p)|^2 + |pg_l(p)|^2 \\ &\quad - |pg_k(p)|\, |pg_l(p)| \\ &\leq |pg_l(p)|^2 . \end{aligned}$$

But then

$$\left|p\,(g_l^{-1} g_k)(p)\right| < |pg_l(p)|,$$

which contradicts our choice of g_l. Therefore, we have produced a generating set with a bounded number of elements. □

The proof of the Betti number estimate is established through several lemmas. First, we need to make three definitions for metric balls. Throughout, fix a Riemannian n-manifold M with $\sec \geq 0$ and a field $\mathbb{F}$ of coefficients for our homology theory

$$H_*(\cdot, \mathbb{F}) = H_*(\cdot) = H_0(\cdot) \oplus \cdots \oplus H_n(\cdot).$$

For $A \subset B \subset M$ define

$$\operatorname{rank}_k(A \subset B) = \operatorname{rank}(H_k(A) \to H_k(B)),$$

$$\operatorname{rank}_*(A \subset B) = \operatorname{rank}(H_*(A) \to H_*(B)).$$

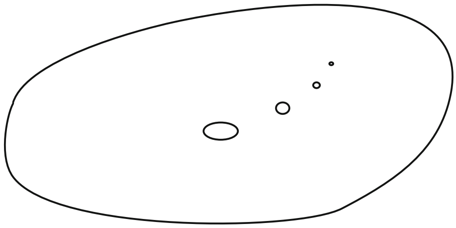


Fig. 12.16 Compact set with infinite topology

Note that when $A \subset B \subset C \subset D$, then

$$\operatorname{rank}_* (A \subset D) \leq \operatorname{rank}_* (B \subset C) .$$

It follows that when A, B are open, bounded, and $\bar{A} \subset B$, then the rank is finite, even when the homology of either set is not finite dimensional. Figure 12.16 pictures a planar domain where infinitely many discs of smaller and smaller size have been extracted. 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.

Content: The *content* of a metric ball $B(p, r) \subset M$ is

$$\operatorname{cont}B(p, r) = \operatorname{rank}_* \left(B\left(p, \tfrac{r}{5}\right) \subset B(p, r)\right) .$$

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 $|p_i x_i| = 10 r_i$.
(b) $r_i \geq 3 r_{i-1}$ for $i = 2, \ldots, k$.
(c) $A \subset \bigcap_{i=1}^k B(p_i, r_i)$.

Compressibility: A ball $B(p, R)$ is said to be *compressible* if it contains a ball $B(x, r) \subset B(p, R)$ such that

(a) $r \leq R/2$.
(b) $\operatorname{cont}B(x, r) \geq \operatorname{cont}B(p, R)$.

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.

We connect these three concepts through a few 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

$$\operatorname{cont}B(p, r) = \sum_i b_i(M) .$$

Also, the corank of such a ball will be zero for large r by lemma 12.4.2. The idea is to compress such a ball until it becomes incompressible and then estimate its content in terms of balls that have corank 1. In this way, we will be able to successively 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, 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 12.5.2. *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 $B(p_1, r_1), \dots, B(p_k, r_k)$ with corresponding critical points $x_1, \dots, x_k$, where $k > 100^n$. Now choose $z \in A$ and select segments $\overline{zx_i}$. One can check that in the unit sphere:

$$\frac{v(n-1,1,\pi)}{v\left(n-1,1,\frac{1}{12}\right)} \le (12\pi)^{n-1} \le 40^n.$$

Since $k > 40^n$ there will be two segments $\overline{zx_i}$ and $\overline{zx_j}$ that form an angle $< 1/6$ at z. Figure 12.17 gives the pictures of the geometry involved.

Note that $|zp_i| \le r_i$ and $|zp_j| \le r_j$. The triangle inequality implies

$$|zx_i| \le 10r_i + |zp_i| \le 11r_i,$$
$$|zx_j| \ge 10r_j - r_j \ge 9r_j.$$

Also, $r_j \ge 3r_i$, so $|zx_j| > |zx_i|$. The hinge version of Toponogov's theorem then implies

$$\begin{aligned} |x_ix_j|^2 &\le |zx_j|^2 + |zx_i|^2 - 2\,|zx_i|\,|zx_j| \cos\tfrac{1}{6} \\ &= |zx_j|^2 + |zx_i|^2 - \frac{31}{16}\,|zx_i|\,|zx_j| \\ &\le \left(|zx_j| - \frac{3}{4}\,|zx_i|\right)^2. \end{aligned}$$

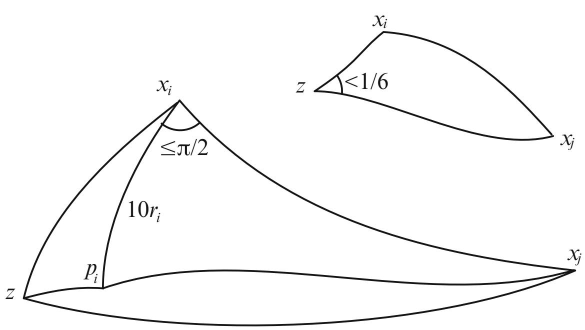


Fig. 12.17 Hinges and triangles

In other words: $\left|x_ix_j\right| \leq \left|zx_j\right| - \frac{3}{4}\left|zx_i\right|$. Now use the triangle inequality to conclude

$$\begin{aligned}
\left|p_ix_j\right| &\geq \left|zx_j\right| - \left|zp_i\right| \\
&\geq 10r_j - \left|zp_j\right| - \left|zp_i\right| \\
&\geq 8r_j \\
&\geq 24r_i \\
&\geq 20r_i = 2\left|p_ix_i\right|.
\end{aligned}$$

Yet another application of the triangle inequality will then imply $\left|x_ix_j\right| \geq \left|p_ix_i\right|$. Since x_i is critical for p_i, we use the hinge version of Toponogov's theorem to conclude

$$\left|p_ix_j\right|^2 \leq \left|p_ix_i\right|^2 + \left|x_ix_j\right|^2 \leq \left(\left|x_ix_j\right| + \frac{1}{2}\left|p_ix_i\right|\right)^2.$$

Thus,

$$\left|p_ix_j\right| \leq \left|x_ix_j\right| + \frac{1}{2}\left|p_ix_i\right| \leq \left|x_ix_j\right| + 5r_i.$$

The triangle inequality implies

$$\left|zx_j\right| \leq \left|p_ix_j\right| + \left|zp_i\right| \leq \left|p_ix_j\right| + r_i \leq \left|x_ix_j\right| + 6r_i.$$

However, we also have

$$\left|zx_i\right| \geq 10r_i - \left|zp_i\right| \geq 9r_i,$$

which together with

$$\left|x_ix_j\right| \leq \left|zx_j\right| - \frac{3}{4}\left|zx_i\right|$$

implies

$$\left|x_ix_j\right| \leq \left|zx_j\right| - \frac{27}{4}r_i.$$

Thus, we have a contradiction:

$$\left|x_ix_j\right| + \frac{27}{4}r_i \leq \left|zx_j\right| \leq \left|x_ix_j\right| + 6r_i.$$

□

Having established a bound on the corank, we next check how the topology changes when we pass from balls of lower corank to balls of higher corank. This requires two lemmas. The first is purely topological and its proof can be skipped.

Lemma 12.5.3. *Assume that we have bounded open sets* B_i^j, $i = 1, \ldots, m$ *and* $j = 0, \ldots, n+1$ *with* $\bar{B}_i^j \subset B_i^{j+1} \subset M^n$, *then*

$$\operatorname{rank}_k\left(\bigcup_i B_i^0 \subset \bigcup_i B_i^{n+1}\right) \leq \operatorname{rank}_k\left(\bigcup_i B_i^0 \subset \bigcup_i B_i^{k+1}\right)$$
$$\leq \sum_{l=0}^{k} \sum_{i_0<\cdots<i_{k-l}} \operatorname{rank}_l\left(\bigcap_{s=0}^{k-l} B_{i_s}^0 \subset \bigcap_{s=0}^{k-l} B_{i_s}^{l+1}\right).$$

Proof. To see why we need multiple intermediate coverings consider the commutative diagram

$$\begin{array}{ccccc} \ker L & \to & V & \to & \operatorname{im} L \\ \downarrow 0 & & \downarrow f & & \downarrow \\ \operatorname{im} L & \to & \operatorname{im} L & \to & 0 \end{array}$$

where the rank of f is clearly not bounded by the ranks of the other two maps between the exact sequences.

We use the generalized Mayer-Vietoris double complex of singular chains with coefficients $\mathbb{F}$ (see also [18, Sections 8 and 15]):

$$C_{p,q}^j = \bigoplus_{i_0<\cdots<i_q} C_p\left(\cap_{s=0}^q B_{i_s}^j\right).$$

This comes with the boundary maps

$$\partial : C_{p,q}^j \to C_{p-1,q}^j,$$
$$\delta : C_{p,q}^j \to C_{p,q-1}^j,$$

where δ comes from the inclusions $C_p\left(\cap_{s=0}^q B_{i_s}^j\right) \to C_p\left(\cap_{s\neq l} B_{i_s}^j\right)$ with sign $(-1)^l$ for $l = 0, \ldots, q$. The choice of sign is consistent with the usual Mayer-Vietoris sequence. Moreover, $\partial\delta = \delta\partial$. Define $\eth = (-1)^p \delta$ (eth) to make them anticommute. We then obtain a new chain complex with vector spaces $\oplus_{l=0}^k C_{l,k-l}^j$ and boundary maps $D = \partial + \eth$ defined as:

$$\oplus_{l=0}^k C_{l,k-l}^j \to \oplus_{l=0}^{k-1} C_{l,k-1-l}^j,$$
$$\left(c_{0,k}^j, c_{1,k-1}^j, \ldots, c_{k,0}^j\right) \mapsto \left(\eth c_{0,k}^j + \partial c_{1,k-1}^j, \ldots, \eth c_{k-1,1}^j + \partial c_{k,0}^j\right).$$

The $\eth$-complex can be augmented by adding the natural inclusions $\eth : C^j_{p,0} \to C_p\left(\cup_i B^j_i\right)$. The images of this map generates the ∂-homology, so let $C^j_p = \operatorname{im}\eth$ to obtain a diagram with exact columns:

$$\begin{array}{ccccccccc}
\vdots & & \vdots & & \vdots & & \vdots & & \iddots \\
 & & \eth\downarrow & & \eth\downarrow & & \eth\downarrow & & \\
0 & \stackrel{\partial}{\leftarrow} & C^j_{0,1} & \stackrel{\partial}{\leftarrow} & C^j_{1,1} & \stackrel{\partial}{\leftarrow} & C^j_{2,1} & \stackrel{\partial}{\leftarrow} & \cdots \\
 & & \eth\downarrow & & \eth\downarrow & & \eth\downarrow & & \\
0 & \stackrel{\partial}{\leftarrow} & C^j_{0,0} & \stackrel{\partial}{\leftarrow} & C^j_{1,0} & \stackrel{\partial}{\leftarrow} & C^j_{2,0} & \stackrel{\partial}{\leftarrow} & \cdots \\
 & & \eth\downarrow & & \eth\downarrow & & \eth\downarrow & & \\
0 & \stackrel{\partial}{\leftarrow} & C^j_0 & \stackrel{\partial}{\leftarrow} & C^j_1 & \stackrel{\partial}{\leftarrow} & C^j_2 & \stackrel{\partial}{\leftarrow} & \cdots \\
 & & \eth\downarrow & & \eth\downarrow & & \eth\downarrow & & \\
 & & 0 & & 0 & & 0 & & \cdots
\end{array}$$

Any D-cycle $\left(c^j_{l,k-l}\right)$ defines a ∂-cycle $\eth c^j_{k,0} \in C^j_k$ since

$$\partial\eth c^j_{k,0} = -\eth\partial c^j_{k,0} = \eth\eth c^j_{k-1,1} = 0.$$

Conversely, any ∂-cycle $c \in C^j_k$ comes from a D-cycle: We need $c^j_{l,k-l} \in C^j_{l,k-l}$ such that $\partial c^j_{l,k-l} = -\eth c^j_{l-1,k-l+1}$. Start by finding $c^j_{k,0}$ such that $\eth c^j_{k,0} = c$. Since $\eth\partial c^j_{k,0} = -\partial\eth c^j_{k,0} = 0$ we can by exactness of $\eth$ find $c^j_{k-1,1}$ with $-\eth c^j_{k-1,1} = \partial c^j_{k,0}$ etc.

This correspondence also preserves being a boundary and thus gives an isomorphism between D-homology and regular ∂-homology. This is crucial for the proof as it shows how to represent homology classes by chains in the intersections. That said, as we don't generate cycles in the intersections, they don't immediately create homology classes.

We use a modified version of Cheeger's cohomology proof in [28]. To prove the lemma consider a finite dimensional subspace $Z_k \subset \oplus^k_{l=0} C^0_{l,k-l}$ of D-cycles that contains no nontrivial D-boundaries. Let $Z^0_k = \eth Z_k \subset C^0_k$ be the isomorphic subspace of ∂-cycles. Specifically: $z = \eth z_{k,0} \in Z^0_k$, where $(z_{l,k-l}) \in Z_k$. We claim that there is a filtration $Z^0_k \supset Z^1_k \supset \cdots \supset Z^{k+1}_k$ with the properties that

$$\dim\left(Z^l_k/Z^{l+1}_k\right) \le \sum_{i_0<\cdots<i_{k-l}} \operatorname{rank}_l\left(\bigcap_{s=0}^{k-l} B^l_{i_s} \subset \bigcap_{s=0}^{k-l} B^{l+1}_{i_s}\right)$$

and Z^{k+1}_k consists of ∂-cycles that are mapped to ∂-boundaries in C^{k+1}_k. This will prove the lemma if Z^0_k is chosen to map isomorphically to its image in $H_k\left(\cup_i B^{k+1}_i\right)$.

For the construction choose inverses $\partial^{-1} : \partial C^j_{p,q} \to C^j_{p,q}$ with $\partial\partial^{-1}c = c$, and name the inclusion maps $f^j : C_*\left(\cup_i B_i^{j-1}\right) \to C_*\left(\cup_i B_i^j\right)$. Note also that the restriction maps $z \mapsto z_{l,k-l}$ are linear.

The construction is inductive and relies on finding suitable linear maps $L^l : Z^l_k \to C^l_{l,k-l}$ whose images consists of ∂-cycles and then define

$$Z^{l+1}_k = \left\{z \in Z^l_k \mid f^{l+1}L^l(z) \in \partial C^{l+1}_{l+1,k-l}\right\}.$$

This will show that

$$\dim\left(Z^l_k / Z^{l+1}_k\right) \leq \sum_{i_0 < \cdots < i_{k-l}} \operatorname{rank}_l \left(\bigcap_{s=0}^{k-l} B^l_{i_s} \subset \bigcap_{s=0}^{k-l} B^{l+1}_{i_s}\right).$$

Starting with $l = 0$ we set $L^0(z) = z_{0,k}$,

$$Z^1_k = \left\{z \in Z^0_k \mid f^1(z_{0,k}) \in \partial C^1_{1,k}\right\},$$

and note that we trivially have $\partial z_{0,k} = 0$.

Now assume we have $L^l : Z^l_k \to C^l_{l,k-l}$ with the properties that

$$\begin{aligned}\eth L^l(z) &= f^l \cdots f^1 (\eth z_{l,k-l}),\\ \partial L^l(z) &= f^l \cdots f^1 (\partial z_{l,k-l}) + f^l\left(\eth L^{l-1}(z)\right) = 0.\end{aligned}$$

This is clearly valid for $l = 0$. We can then define Z^{l+1}_k as above and with that

$$L^{l+1}(z) = f^{l+1} \cdots f^1 (z_{l+1,k-l-1}) + \eth\partial^{-1} f^{l+1} L^l(z).$$

It follows from our induction hypotheses on L^l and the fact that $(z_{l,k-l})$ is a D-cycle that

$$\begin{aligned}\eth L^{l+1}(z) &= f^{l+1} \cdots f^1 (\eth z_{l+1,k-l-1}),\\ \partial L^{l+1}(z) &= f^{l+1} \cdots f^1 (\partial z_{l+1,k-l-1}) + f^{l+1}\left(\eth L^l(z)\right) = 0.\end{aligned}$$

Finally, when $z \in Z^{k+1}_k$ we have that $f^{k+1}L^k(z) \in \partial C^{k+1}_{k+1,0}$ and

$$\eth f^{k+1} L^k(z) = f^{k+1} f^k \cdots f^1 (\eth z_{k,0}) = f^{k+1} \cdots f^1 (z)$$

showing that also $f^{k+1} \cdots f^1(z)$ is a ∂-boundary. □

Let $\mathscr{B}(k)$ denote the set of balls in M of corank $\geq k$, and $\mathscr{C}(k)$ the largest content of any ball in $\mathscr{B}(k)$.

Lemma 12.5.4. *There is a constant $C(n)$ depending only on dimension such that*

$$\mathscr{C}(k) \leq C(n)\,\mathscr{C}(k+1).$$

Proof. Clearly $\mathscr{C}(k)$ is always realized by some incompressible ball $B(p,R)$. Now consider a ball $B(x,r)$ where $x \in B(p, R/4)$ and $r \leq R/100$. We claim that this ball lies in $\mathscr{B}(k+1)$. To see this, first consider the balls

$$B\left(x, \tfrac{R}{10}\right) \subset B\left(p, \tfrac{R}{5}\right) \subset B\left(x, \tfrac{R}{2}\right) \subset B(p,R).$$

If there are no critical points for x in $B(x, R/2) - B(x, R/10)$, then

$$\begin{aligned} &\operatorname{rank}_*\left(B\left(x, \tfrac{R}{10}\right) \subset B\left(x, \tfrac{R}{2}\right)\right) \\ &= \operatorname{rank}_*\left(B\left(p, \tfrac{R}{5}\right) \subset B\left(x, \tfrac{R}{2}\right)\right) \\ &\geq \operatorname{rank}_*\left(B\left(p, \tfrac{R}{5}\right) \subset B(p,R)\right). \end{aligned}$$

This implies that $\operatorname{cont}B(p,R) \leq \operatorname{cont}B(x, R/2)$ and thus contradicts incompressibility of $B(p,R)$. We can now show that $B(x,r) \in \mathscr{B}(k+1)$. Using that $B(p,R) \in \mathscr{B}(k)$, 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 $B(x, R/2) - B(x, R/10)$ and consider the ball $B(x, |xy|/10)$. Then the balls $B(p_1,r_1), \ldots, B(p_l,r_l)$, $B(x, |xy|/10)$ show that $B(x,r)$ has corank $\geq l+1 > k$.

Now cover $B(p, R/5)$ by balls $B\left(p_i, 10^{-n-3}R\right)$, $i = 1, \ldots, m$. If in addition the balls $B\left(p_i, 10^{-n-3}R/2\right)$ are pairwise disjoint, then

$$m \leq \frac{v(n,0,R)}{v(n,0,10^{-n-3}R/2)} = 2^n \cdot 10^{n(n+3)}.$$

Since $B(p_i, R/2) \subset B(p,R)$ it follows that

$$\operatorname{cont}B(p,R) \leq \operatorname{rank}_*\left(\bigcup_{i=1}^m B\left(p_i, 10^{-n-3}R\right) \subset \bigcup_{i=1}^m B\left(p_i, \tfrac{R}{2}\right)\right).$$

To estimate

$$\operatorname{rank}_*\left(\bigcup_{i=1}^m B\left(p_i, 10^{-n-3}R\right) \subset \bigcup_{i=1}^m B\left(p_i, \tfrac{R}{2}\right)\right),$$

use the doubly indexed family $B_i^j = B\left(p_i, 10^{j-n-3}R\right)$, $j = 0, \ldots, n+1$. Note that for fixed j the family covers $B(p, R/5)$. It follows from lemma 12.5.3 that

$$\begin{aligned}
&\operatorname{rank}_*\left(\bigcup_{i=1}^{m} B\left(p_i, 10^{-n-3}R\right) \subset \bigcup_{i=1}^{m} B\left(p_i, \tfrac{R}{2}\right)\right) \\
&\le \operatorname{rank}_*\left(\bigcup_{i=1}^{m} B\left(p_i, 10^{-n-3}R\right) \subset \bigcup_{i=1}^{m} B\left(p_i, 10^{-2}R\right)\right) \\
&\le \sum_{j=0}^{n}\sum_{k=0}^{m-1}\sum_{i_0<\cdots<i_k} \operatorname{rank}_*\left(\bigcap_{s=0}^{k} B\left(p_{i_s}, 10^{j-n-3}\right) \subset \bigcap_{s=0}^{k} B\left(p_{i_s}, 10^{j-n-2}R\right)\right).
\end{aligned}$$

When $\bigcap_{t=0}^{s} B\left(p_{i_t}, 10^{j-n-3}R\right) \neq \varnothing$ the triangle inequality shows that

$$\begin{aligned}
\bigcap_{t=0}^{s} B\left(p_{i_t}, 10^{j-n-3}R\right) &\subset B\left(p_{i_0}, 10^{j-n-3}R\right) \\
&\subset B\left(p_{i_0}, 10^{j-n-2}\tfrac{R}{2}\right) \\
&\subset \bigcap_{t=0}^{s} B\left(p_{i_t}, 10^{j-n-2}R\right).
\end{aligned}$$

Consequently, as long as $j \le n$ we have

$$\operatorname{rank}_*\left(\bigcap_{t=0}^{s} B\left(p_{i_t}, 10^{j-n-3}R\right) \subset \bigcap_{t=0}^{s} B\left(p_{i_t}, 10^{j-n-2}R\right)\right) \le \operatorname{cont} B\left(p_{i_0}, 10^{j-n-2}\tfrac{R}{2}\right)$$

where $B\left(p_{i_0}, 10^{j-n-2}\frac{R}{2}\right) \in \mathscr{B}(k+1)$.

The number of intersections $\bigcap_{t=0}^{s} B\left(p_{i_t}, 10^{j-n-3}R\right)$ with $j \le n$ and $s \le 2^n \cdot 10^{n(n+3)}$ is bounded by

$$C(n) = \sum_{j=0}^{n} 2^{2^n \cdot 10^{n(n+3)}}.$$

So we obtain an estimate of the form $\mathscr{C}(k) \le C(n)\,\mathscr{C}(k+1)$. □

Proof of (2). The above lemma implies that

$$\operatorname{cont} M = \mathscr{C}(0) \le \mathscr{C}(k) \cdot (C(n))^k,$$

where $k \le 100^n$ is the largest possible corank in M. It then remains to check that $\mathscr{C}(k) = 1$. However, it follows from the above that if $\mathscr{B}(k)$ contains an incompressible ball, then $\mathscr{B}(k+1) \neq \varnothing$. Thus, all balls in $\mathscr{B}(k)$ are compressible, but then they must have minimal content 1. □

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 $C\left(n, k^2 D\right)$ such that any closed Riemannian n-manifold (M, g) with $\sec \geq k$ and $\operatorname{diam} \leq D$ has the properties that

(1) $\pi_1(M)$ can be generated by $\leq C\left(n, kD^2\right)$ elements,
(2) $\sum_{i=0}^{n} b_i(M, \mathbb{F}) \leq C\left(n, kD^2\right)$.

It is also possible to reach a stronger conclusion (see [102]). In outline this is done as follows. First one should use simplicial instead of singular homology. If one inspects the proof of lemma 12.5.3 with this in mind, then one can, from a sufficiently fine simplicial subdivision of M relative to the doubly indexed cover, create a CW complex X that uses at most $C\left(n, kD^2\right)$ cells as well as maps $M \to X \to M$ whose composition is the identity. In other words M is dominated by a CW complex with a bounded number of cells. This will also give a bound for the Betti numbers.

12.6 Homotopy Finiteness

This section is devoted to 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 12.6.1 (Grove and Petersen, 1988). *Given an integer $n > 1$ and numbers $v, D, k > 0$, the class of Riemannian n-manifolds with*

$$\begin{aligned}\operatorname{diam} &\leq D,\\ \operatorname{vol} &\geq v,\\ \sec &\geq -k^2\end{aligned}$$

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 12.6.2. *For M as in the theorem, there exists $\alpha = \alpha(n, D, v, k) \in \left(0, \frac{\pi}{2}\right)$ and $\delta = \delta(n, D, v, k) > 0$ such that if $p, q \in M$ satisfy $|pq| \leq \delta$, then either p is α-regular for q or q is α-regular for p.*

Proof. The proof is by contradiction and based on a suggestion by Cheeger. For simplicity assume that $k = -1$. Suppose there are points $p, q \in M$ that are not

α-regular with respect to each other and with $|pq| \leq \delta$. Then the two sets $\overrightarrow{\overrightarrow{pq}} \subset T_pM$ and $\overrightarrow{\overrightarrow{qp}} \subset T_qM$ of unit vectors tangent segments joining p and q are by assumption $(\pi - \alpha)$-dense in the unit spheres. It is a simple exercise to show that if $A \subset S^{n-1}$, then the function

$$t \mapsto \frac{\mathrm{vol} B\,(A, t)}{v\,(n-1, 1, t)}$$

is nonincreasing (see also exercise 7.5.18 for a more general result). In particular, for any $(\pi - \alpha)$-dense set $A \subset S^{n-1}$

$$\begin{aligned} \mathrm{vol}\left(S^{n-1} - B\,(A, \alpha)\right) &= \mathrm{vol} S^{n-1} - \mathrm{vol} B\,(A, \alpha) \\ &\leq \mathrm{vol} S^{n-1} - \mathrm{vol} S^{n-1} \cdot \frac{v\,(n-1, 1, \alpha)}{v\,(n-1, 1, \pi - \alpha)} \\ &= \mathrm{vol} S^{n-1} \cdot \frac{v\,(n-1, 1, \pi - \alpha) - v\,(n-1, 1, \alpha)}{v\,(n-1, 1, \pi - \alpha)}. \end{aligned}$$

Now choose $\alpha < \frac{\pi}{2}$ 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 in M (see exercise 7.5.19) satisfy

$$\begin{aligned} \mathrm{vol} B^{S^{n-1} - B\left(\overrightarrow{\overrightarrow{pq}}, \alpha\right)}\,(p, D) &\leq \frac{v}{6}, \\ \mathrm{vol} B^{S^{n-1} - B\left(\overrightarrow{\overrightarrow{qp}}, \alpha\right)}\,(q, D) &\leq \frac{v}{6}. \end{aligned}$$

We 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 12.18 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

$$v\,(n, -1, r) = \frac{v}{6}.$$

We claim that if δ is sufficiently small, then

$$M = B\,(p, r) \cup B\,(q, r) \cup B^{S^{n-1} - B\left(\overrightarrow{\overrightarrow{pq}}, \alpha\right)}\,(p, D) \cup B^{S^{n-1} - B\left(\overrightarrow{\overrightarrow{qp}}, \alpha\right)}\,(q, D)\,.$$

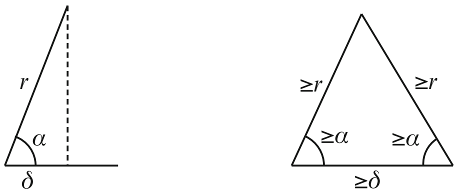


Fig. 12.18 Comparison hinge and triangle

This will, of course, lead to a contradiction, as we would then have

$$\begin{aligned} v &\leq \operatorname{vol} M \\ &\leq \operatorname{vol}\left(B(p,r)\cup B(q,r)\cup B^{S^{n-1}-B\left(\overrightarrow{pq},\alpha\right)}(p,D)\cup B^{S^{n-1}-B\left(\overrightarrow{qp},\alpha\right)}(q,D)\right) \\ &\leq 4\cdot\frac{v}{6}<v. \end{aligned}$$

To see that these sets cover M, observe that if

$$x\notin B^{S^{n-1}-B\left(\overrightarrow{pq},\alpha\right)}(p,D),$$

then there is a hinge $\overline{xp}$ and $\overline{pq}$ with angle $\leq\alpha$ (see figure 12.18).

Thus, we have from Toponogov's theorem that

$$\cosh|xq|\leq\cosh|pq|\cosh|xp|-\sinh|pq|\sinh|xp|\cos(\alpha).$$

If also

$$x\notin B^{S^{n-1}-B\left(\overrightarrow{qp},\alpha\right)}(q,D),$$

we have in addition,

$$\cosh|xp|\leq\cosh|pq|\cosh|xq|-\sinh|pq|\sinh|xq|\cos(\alpha).$$

If $|xp|>r$ and $|xq|>r$, we get

$$\begin{aligned} \cosh|xq| &\leq \cosh|pq|\cosh|xp|-\sinh|pq|\sinh|xp|\cos(\alpha) \\ &\leq \cosh|xp| \\ &\quad+(\cosh|pq|-1)\cosh D-\sinh|pq|\sinh r\cos(\alpha) \end{aligned}$$

and

$$\begin{aligned} \cosh|xp| &\leq \cosh|xq| \\ &\quad+(\cosh|pq|-1)\cosh D-\sinh|pq|\sinh r\cos(\alpha). \end{aligned}$$

However, as $|pq| \to 0$, we see that the quantity

$$\begin{aligned} f(|pq|) &= (\cosh|pq|-1)\cosh D - \sinh|pq|\sinh r\cos(\alpha) \\ &= (-\sinh r\cos\alpha)\,|pq| + O\left(|pq|^2\right) \end{aligned}$$

becomes negative. Thus, we can find $\delta(D,r,\alpha) > 0$ such that for $|pq| \le \delta$ we have

$$(\cosh|pq|-1)\cosh D - \sinh|pq|\sinh r\cos(\alpha) < 0.$$

We have then arrived at another contradiction, as this would imply

$$\cosh|xq| < \cosh|xp|$$

and

$$\cosh|xp| < \cosh|xq|$$

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 $|pq| \le \delta$, 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 simple analysis will, however, enable us to draw some very useful conclusions from it.

Consider the product $M \times M$ with the product metric. Geodesics in this space are of the form (c_1, c_2), where both c_1, c_2 are geodesics in M. In $M \times M$ we have the diagonal $\Delta = \{(x,x) \mid x \in M\}$. Note that

$$T_{(p,p)}\Delta = \left\{(v,v) \mid v \in T_pM\right\},$$

and the normal bundle

$$T^{\perp}_{(p,p)}\Delta = \left\{(v,-v) \mid v \in T_pM\right\}.$$

Therefore, if $(c_1,c_2) : [a,b] \to M \times M$ is a segment from (p,q) to Δ, then $\dot{c}_1(b) = -\dot{c}_2(b)$. Thus these two segments can be joined at the common point $c_1(b) = c_2(b)$ 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, $\sqrt{2}\cdot|(p,q)\,\Delta| = |pq|$.

The above lemma implies

Corollary 12.6.3. *Any point within distance $\delta/\sqrt{2}$ of Δ is α-regular for Δ.*

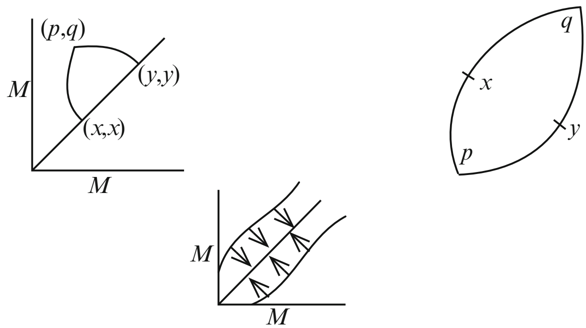


Fig. 12.19 Critical points for diagonal and deformation

Figure 12.19 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 $\leq \frac{1}{\cos\alpha} |(p,q)\,\Delta|$ 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 $|pq| < \delta$, then p and q are joined by a curve $t \mapsto H(p,q,t)$, $0 \leq t \leq 1$, whose length is $\leq \frac{\sqrt{2}}{\cos\alpha}|pq|$. Furthermore, the map $(p,q,t) \mapsto H(p,q,t)$ is continuous. For simplicity, we let $C = \frac{\sqrt{2}}{\cos\alpha}$ in the constructions below.

We now have the first ingredient in our proof.

Corollary 12.6.4. *If $f_0, f_1 : X \to M$ are two continuous maps such that*

$$|f_0(x)\, f_1(x)| < \delta$$

for all $x \in X$, then f_0 and f_1 are homotopy equivalent.

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.,

$$\Delta^k = \left\{ \left(x^0, \ldots, x^k\right) \mid x^0 + \cdots + x^k = 1 \text{ and } x^0, \ldots, x^k \in [0,1] \right\}.$$

The basis vectors $e_i = \left(\delta_i^1, \ldots, \delta_i^k\right)$ are the vertices of the simplex.

Lemma 12.6.5. *Suppose we have $k+1$ points $p_0, \ldots, p_k \in B(p,r) \subset M$. 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 $f(e_i) = p_i$.

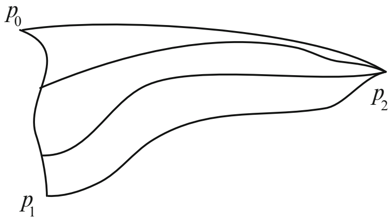


Fig. 12.20 Homotopy construction of a simplex

Proof. Figure 12.20 gives the essential idea of the proof. The construction is by induction on k. For $k = 0$ there is nothing to show.

Assume that the statement holds for k and that we have $k + 2$ points $p_0, \ldots, p_{k+1} \in B(p, r)$. 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 $i = p_0, \ldots, p_k$. We then define

$$\bar{f} : \Delta^{k+1} \to B\left(p, r + 2r \cdot C \cdot \frac{C^{k+1} - 1}{C - 1}\right),$$

$$\bar{f}\left(x^0, \ldots, x^k, x^{k+1}\right) = H\left(f\left(\frac{x^0}{\sum_{i=1}^k x^i}, \ldots, \frac{x^k}{\sum_{i=1}^k x^i}\right), p_{k+1}, x^{k+1}\right).$$

This clearly gives a well-defined continuous map as long as

$$\begin{aligned}
\left|p_{k+1} f\left(\frac{x^0}{\sum_{i=1}^k x^i}, \ldots, \frac{x^k}{\sum_{i=1}^k x^i}\right)\right| &\leq \left|p f\left(\frac{x^0}{\sum_{i=1}^k x^i}, \ldots, \frac{x^k}{\sum_{i=1}^k x^i}\right)\right| + |p p_{k+1}| \\
&\leq \left(2r \cdot C \cdot \frac{C^k - 1}{C - 1} + r\right) + r \\
&= 2r \cdot \frac{C^{k+1} - 1}{C - 1} \\
&< \delta.
\end{aligned}$$

Moreover, it has the property that

$$\begin{aligned}
\left|p \bar{f}(\cdot)\right| &\leq |p\, p_{k+1}| + \left|p_{k+1} \bar{f}(\cdot)\right| \\
&\leq r + 2r \cdot C \cdot \frac{C^{k+1} - 1}{C - 1}.
\end{aligned}$$

This concludes the induction step. $\square$

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. 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. Thus it suffices to prove

Lemma 12.6.6. *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*

$$\begin{aligned} \mathrm{diam} &\leq D, \\ \mathrm{vol} &\geq v, \\ \sec &\geq -k^2, \end{aligned}$$

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 on $M \cup N$, 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 construct a continuous map $f : M \to N$ as follows. First use the Hausdorff approximation to map all the vertices $\{p_i\} \subset M$ of the triangulation to points $\{q_i\} \subset N$ such that $|p_i q_i| < \varepsilon$. If $(p_{i_0}, \ldots, p_{i_n})$ forms a simplex in the triangulation of M, then by the choice of the triangulation $\{p_{i_0}, \ldots, p_{i_n}\} \subset B(x, \varepsilon)$ for some $x \in M$. Thus $\{q_{i_0}, \ldots, q_{i_n}\} \subset B(q_{i_0}, 4\varepsilon)$. Therefore, if

$$8\varepsilon \frac{C^n - 1}{C - 1} < \delta,$$

then lemma 12.6.5 can be used to define f on the simplex spanned by $(p_{i_0}, \ldots, p_{i_n})$. 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 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 $(p_{i_0}, \ldots, p_{i_n})$. Then we have

$$\begin{aligned} |x f(x)| &\leq |x\, p_{i_0}| + |p_{i_0} f(x)| \\ &\leq 2\varepsilon + \varepsilon + |q_{i_0} f(x)| \end{aligned}$$

$$\begin{aligned} &\leq 3\varepsilon + 4\varepsilon + 8\varepsilon \cdot C \cdot \frac{C^n - 1}{C - 1} \\ &= 7\varepsilon + 8\varepsilon \cdot C \cdot \frac{C^n - 1}{C - 1}. \end{aligned}$$

We can construct $g : N \to M$ in the same manner. This map will, of course, also satisfy

$$|y\, g(y)| \leq 7\varepsilon + 8\varepsilon \cdot C \cdot \frac{C^n - 1}{C - 1}.$$

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:

$$\begin{aligned} |y f \circ g(y)| &\leq |y\, g(y)| + |g(y) f \circ g(y)| \\ &\leq 14\varepsilon + 16\varepsilon \cdot C \cdot \frac{C^n - 1}{C - 1}; \\ |x\, g \circ f(x)| &\leq 14\varepsilon + 16\varepsilon \cdot C \cdot \frac{C^n - 1}{C - 1}. \end{aligned}$$

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

$$14\varepsilon + 16\varepsilon \cdot \frac{C^{n+1} - 1}{C - 1} < \delta,$$

the two spaces are homotopy equivalent. Thus, ε depends in an explicit way on $C = \frac{\sqrt{2}}{\cos\alpha}$ and δ. It is possible, in turn, to estimate α and δ from n, k, v, and D. Thus there is 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 section 11.1.4 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. There are inclusions of classes of closed Riemannian n-manifolds

$$\left\{ \begin{array}{ll} \text{diam} & \leq D \\ \sec & \geq -k^2 \end{array} \right\} \supset \left\{ \begin{array}{ll} \text{diam} & \leq D \\ \text{vol} & \geq v \\ \sec & \geq -k^2 \end{array} \right\} \supset \left\{ \begin{array}{ll} \text{diam} & \leq D \\ \text{vol} & \geq v \\ |\sec| & \leq k^2 \end{array} \right\}$$

with strengthening 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 $\left\{S^3/\mathbb{Z}_p\right\}$ 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

$$\begin{aligned} \mathrm{vol} &= v, \\ 1 \geq \sec &> 0. \end{aligned}$$

The family of lens spaces $\left\{S^3/\mathbb{Z}_p\right\}$ with curvature $= 1$ also shows that the lower volume bound is necessary in both of these 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 [59] and in a more general case in [85]. One can also prove many of the above results for manifolds with certain types of integral curvature bounds, see for instance [91] and [88]. The volume [54] also contains complete discussions of generalizations to the case where one has merely Ricci curvature bounds.

12.7 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 [51], by Abresch-Meyer, Colding, Greene, and Zhu in [54], by Cheeger in [28], and by Karcher in [32]. 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.

12.8 Exercises

EXERCISE 12.8.1. Let (M, g) be a closed simply connected positively curved manifold. Show that if M contains a totally geodesic closed hypersurface, 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. This also shows that it suffices to assume that $H^1(M, \mathbb{Z}_2) = 0$.

EXERCISE 12.8.2. Let (M, g) be a complete noncompact manifold with $\sec \geq 0$ and soul $S \subset M$.

(1) Show that if $X \in TS$ and $V \in T^{\perp}S$, then $\sec(X, V) = 0$.
(2) Show that if the soul has codimension 1 and trivial normal bundle, then $(M, g) = \left(S \times \mathbb{R}, g_S + dr^2\right)$.
(3) Show that if the soul has codimension 1 and nontrivial normal bundle, then a double cover splits as in (2).

EXERCISE 12.8.3. Show that the solution to

$$\begin{aligned} \ddot{\zeta} &= -(1+\varepsilon)\,\zeta, \\ \zeta(0) &= \psi(0) > 0, \\ \dot{\zeta}(0) &= \dot{\psi}(0) > 0. \end{aligned}$$

is given by

$$\zeta(t) = \sqrt{(\psi(0))^2 + \frac{\left(\dot{\psi}(0)\right)^2}{1+\varepsilon}} \cdot \sin\left(\sqrt{1+\varepsilon}\cdot t + \arctan\left(\frac{\psi(0)\cdot\sqrt{1+\varepsilon}}{\dot{\psi}(0)}\right)\right).$$

EXERCISE 12.8.4. 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 $\sec \geq k$.

EXERCISE 12.8.5. Let (M, g) be a complete Riemannian manifold. Show that if all sectional curvatures on $B(p, 2R)$ are $\geq k$, then Toponogov's comparison theorem holds for hinges and triangles in $B(p, R)$.

EXERCISE 12.8.6. Let (M, g) be a complete Riemannian manifold with $\sec \geq k$. Consider $f(x) = |xq| - |xp|$ and assume that both $|xq|$ and $|xp|$ are smooth at x. Show that Toponogov's theorem can be used to bound $|\nabla f|\,|_x$ from below in terms of the distance from x to a segment from q to p.

EXERCISE 12.8.7 (HEINTZE AND KARCHER). Let $c \subset (M, g)$ be a geodesic in a Riemannian n-manifold with $\sec \geq -k^2$. Let $T(c, R)$ be the normal tube around c of radius R, i.e., the set of points in M that can be joined to c by a segment of length $\leq R$ that is perpendicular to c. The last condition is superfluous when c is a closed geodesic, but if it is a loop or a segment, then not all points in M within distance R of c will belong to this tube. On this tube introduce coordinates (r, s, θ), where r denotes the distance to c, s is the arclength parameter on c, and $\theta = \left(\theta^1, \ldots, \theta^{n-2}\right)$ are spherical coordinates normal to c. These give adapted coordinates for the distance r to c. 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

$$\mathrm{vol} T(c, R) \leq f(n, k, R, L(c)),$$

for some continuous function f depending on dimension, lower curvature bound, radius, and length of c. Moreover, as $L(c) \to 0, f \to 0$. Use this estimate to prove lemmas 11.4.9 and 12.6.2. This shows that Toponogov's theorem is not needed for the latter result.

EXERCISE 12.8.8. 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 for $k > 2$ there are two k-dimensional bundles.

EXERCISE 12.8.9. Use Toponogov's theorem to show that b_c is convex when $\sec \geq 0$.

Bibliography

1. M. Amann, L. Kennard, *On a Generalized Conjecture of Hopf with Symmetry*, arXiv:1402.7255 (2014)
2. M.T. Anderson, Short geodesics and gravitational instantons. J. Differ. Geom. **31**, 265–275 (1990)
3. M.T. Anderson, Metrics of positive Ricci curvature with large diameter. Manuscripta Math. **68**, 405–415 (1990)
4. M.T. Anderson, J. Cheeger, C^{α}-compactness for manifolds with Ricci curvature and injectivity radius bounded below. J. Differ. Geom. **35**(2), 265–281 (1992)
5. C.S. Aravinda, F.T. Farrell, Non-positivity: curvature vs. curvature operator. Proc. Am. Math. Soc. **133**(1), 191–192 (2005)
6. V.I. Arnol'd, *Mathematical Methods of Classical Mechanics* (Springer, New York, 1989)
7. W. Ballmann, Nonpositively curved manifolds of higher rank. Ann. Math. **122**, 597–609 (1985)
8. W. Ballmann, *Spaces of Nonpositive Curvature* (Birkhäuser, Basel 1995)
9. W. Ballmann, P. Eberlein, Fundamental groups of manifolds of nonpositive curvature. J. Differ. Geom. **25**, 1–22 (1987)
10. W. Ballmann, V. Schroeder, M. Gromov, *Manifolds of Nonpositive Curvature* (Birkhäuser, Boston, 1985)
11. M. Berger, *Riemannian Geometry During the Second Half of the Twentieth Century*, University Lecture series, AMS 17
12. A.L. Besse, *Einstein Manifolds* (Springer, Berlin-Heidelberg, 1978)
13. A.L. Besse, *Manifolds All of Whose Geodesics are Closed* (Springer, Berlin-Heidelberg, 1987)
14. R.L. Bishop, R.J. Crittenden, *Geometry of Manifolds* (Academic Press, New York, 1964)
15. R.L. Bishop, S.I. Goldberg, *Tensor Analysis on Manifolds* (Dover, New York, 1980)
16. C. Böhm, B. Wilking, Manifolds with positive curvature operators are space forms. Ann. Math. (2) **167**(3), 1079–1097 (2008)
17. A. Borel, *Seminar on Transformation Groups*, Annals of Math. Studies 46 (Princeton University Press, Princeton, 1960)
18. R. Bott, L.W. Tu, *Differential Forms in Algebraic Topology* (Springer, New York 1982)
19. S. Brendle, R. Schoen, Classification of manifolds with weakly 1/4-pinched curvatures. Acta Math. **200**(1), 1–13 (2008)
20. S. Brendle, R. Schoen, Manifolds with 1/4-pinched curvature are space forms. J. Am. Math. Soc. **22**(1), 287–307 (2009)
21. K. Burns, R. Spatzier, On topological Tits buildings and their classification. IHES Publ. Math. **65**, 5–34 (1987)

P. Petersen, *Riemannian Geometry*, Graduate Texts in Mathematics 171,
DOI 10.1007/978-3-319-26654-1

22. E. Calabi, P. Hartman, On the smoothness of isometries. Duke Math. J. **37**, 741–750 (1970)
23. M.P. do Carmo, *Riemannian Geometry* (Birkhäuser, Boston, 1993)
24. I. Chavel, *Riemannian Geometry, A Modern Introduction* (Cambridge University Press, New York, 1995)
25. J. Cheeger, *Comparison and Finiteness Theorems for Riemannian Manifolds*, Ph. D. thesis, Princeton University, 1967
26. J. Cheeger, Finiteness theorems for Riemannian manifolds. Am. J. Math. **92**, 61–75 (1970)
27. J. Cheeger, Pinching theorems for a certain class of Riemannian manifolds. Am. J. Math. **92**, 807–834 (1970)
28. J. Cheeger et al., *Geometric Topology: Recent Developments*, LNM 1504 (Springer, Berlin-Heidelberg, 1991)
29. J. Cheeger, T.H. Colding, On the structure of space with Ricci curvature bounded below. J. Differ. Geom. **46**, 406–480 (1997)
30. J. Cheeger, D.G. Ebin, *Comparison Theorems in Riemannian Geometry* (North-Holland/Elsevier, New York, 1975)
31. J. Cheeger, D. Gromoll, The splitting theorem for manifolds of nonnegative Ricci curvature. J. Differ. Geom. **6**, 119–128 (1971)
32. S.S. Chern, (ed.), *Global Geometry and Analysis*, 2nd edition, MAA Studies 27 (Mathematical Assocoation of America, Washington, 1989)
33. F.H. Clarke et al. *Nonsmooth Analysis and Control Theory* (Springer, New York, 1998)
34. T.H. Colding, Ricci curvature and volume convergence. Ann. Math. **145**, 477–501 (1997)
35. R. Courant, D. Hilbert, *Methods of Mathematical Physics*, vol. II (Wiley Interscience, New York, 1962)
36. X. Dai, P. Petersen, G. Wei, Integral pinching theorems. Manuscripta Math. **101**(2), 143–152 (2000)
37. D. DeTurk, J. Kazdan, Some regularity theorems in Riemannian geometry. Ann. sci. Éc. Norm. Sup. **14**, 249–260 (1981)
38. P. Eberlein, *Geometry of Nonpositively Curved Manifolds* (The University of Chicago Press, Chicago 1996)
39. R. Edwards, R. Kirby, Deformations of spaces of embeddings. Ann. Math. **93**, 63–88 (1971)
40. L.P. Eisenhart, *Riemannian Geometry* (Princeton University Press, Princeton, 1997)
41. J.H. Eschenburg, E. Heintze, An elementary proof of the Cheeger-Gromoll splitting theorem. Ann. Glob. Anal. Geom. **2**, 141–151 (1984)
42. F. Fang, X. Rong, Homeomorphism classification of positively curved manifolds with almost maximal symmetry rank. Math. Ann. **332**, 81–101 (2005)
43. T. Farrell, L. Jones, Negatively curved manifolds with exotic smooth structures. J. AMS **2**, 899–908 (1989)
44. 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, pp. 143–238
45. F. Galaz-Garcia, C. Searle, Non-negatively curved 5-manifolds with almost maximal symmetry rank. Geom. Topol. **18**(3), 1397–1435 (2014)
46. S. Gallot, Isoperimetric inequalities based on integral norms of Ricci curvature. Astérisque **157-158**, 191–216 (1988) pp
47. S. Gallot, D. Hulin, J. Lafontaine, *Riemannian Geometry* (Springer, Berlin-Heidelberg, 1987)
48. S. Gallot, D. Meyer, Opérateur de courbure et laplacien des formes différentielles d'une variété riemannianne. J. Math. Pures. Appl. **54**, 259–284 (1975)
49. G. Gibbons, S. Hawking, Gravitational multi-instantons. Phys. Lett. B **78**, 430–432 (1978)
50. D. Gilbarg, N.S. Trudinger, *Elliptic Partial Differential Equations of Second-Order*, 2nd edn. (Springer, Berlin-Heidelberg, 1983)
51. R. Greene, S.T. Yau (eds.), Proceedings of Symposia in Pure Mathematics, vol. 54, 3 (1994)
52. M. Gromov, Manifolds of negative curvature. J. Differ. Geom. **12**, 223–230 (1978)
53. M. Gromov, *Metric Structures for Riemannian and Non-Riemannian Spaces* (Birkhäuser, Boston 1999)

54. K. Grove, P. Petersen (eds.), *Comparison Geometry*, vol. 30 (MSRI publications, New York; Cambridge University Press, Cambridge, 1997)
55. M. Gromov, W. Thurston, Pinching constants for hyperbolic manifolds. Invest. Math. **89**, 1–12 (1987)
56. K. Grove, K. Shiohama, A generalized sphere theorem. Ann. Math. (2) **106**(2), 201–211 (1977)
57. K. Grove, B. Wilking, A knot characterization and 1–connected nonnegatively curved 4–manifolds with circle symmetry. Geom. Topol. **18**(5), 3091–3110 (2014)
58. K. Grove, H. Karcher, E. Ruh, Group actions and curvature. Invest. Math. **23**, 31–48 (1974)
59. K. Grove, P. Petersen, J.-Y. Wu, Geometric finiteness theorems via controlled topology. Invest. Math. **99**, 205–213 (1990); *Erratum* Invest. Math. **104**, 221–222 (1991)
60. P. Hajłasz, P. Koskela, Sobolev Met Poincaré, vol. 688 (Memoirs of the AMS, New York, 2000)
61. R.S. Hamilton, The formation of singularities in the Ricci flow. Surveys in Differential Geometry, vol. 2 (International Press, Somerville, 1995), pp. 7–136
62. S. Helgason, *Differential Geometry, Lie Groups and Symmetric spaces* (Academic Press, New York-London, 1962)
63. W.-Y. Hsiang, B. Kleiner, On the topology of positively curved 4-manifolds with symmetry. J. Differ. Geom. **29**, 615–621 (1989)
64. M. Ise, M. Takeuchi, *Lie Groups I, II*, Translations of Mathematical Monographs, vol. 85 (AMS, Providence, 1991)
65. J. Jost, *Riemannian Geometry and Geometric Analysis* (Springer, Berlin-Heidelberg, 1995)
66. J. Jost, H. Karcher, Geometrische Methoden zur Gewinnung von a-priori-Schranken für harmonische Abbildungen. Manuf. Math. **19**, 27–77 (1982)
67. L. Kennard, On the Hopf conjecture with symmetry. Geom. Topol. **17**(1), 563–593 (2013)
68. L. Kennard, Positively curved Riemannian metrics with logarithmic symmetry rank bounds. Commentarii Mathematici Helvetici **89**(4), 937–962 (2014)
69. W. Klingenberg, *Riemannian Geometry*, 2nd edn. (Walter de Gruyter & Co., Berlin, 1995)
70. S. Kobayashi, K. Nomizu, *Foundations of Differential Geometry*, vols. I, II (Wiley-Interscience, New York, 1963)
71. H.B. Lawson Jr., M.-L. Michelsohn, *Spin Geometry* (Princeton University Press, Princeton, 1989)
72. J.M. Lee, *Introduction to Smooth Manifolds* (Springer, New York, 2003)
73. M.J. Micallef, J.D. Moore, Minimal 2-spheres and the topology of manifolds with positive curvature on totally isotropic 2-planes. Ann. Math. **127**, 199–227 (1988)
74. R.K. Miller, A.N. Michel, *Ordinary Differential Equations* (Academic Press, New York-London, 1982)
75. J.W. Milnor, *Morse Theory* (Princeton University Press, Princeton, 1963)
76. J.W. Milnor, J.D. Stasheff, *Characteristic Classes* (Princeton University Press, Princeton, 1974)
77. C.W. Misner, K.S. Thorne, J.A. Wheeler, *Gravitation* (Freeman, New York, 1973)
78. N. Mok, The uniformization theorem for compact Kähler manifolds of nonnegative holomorphic bi-sectional curvature. J. Differ. Geom. **27**, 179–214 (1988)
79. D. Montgomery, L. Zippin, *Topological Transformation Groups* (Wiley-Interscience, New York, 1955)
80. B. O'Neill, *Semi-Riemannian Geometry* (Academic Press, New York-London, 1983)
81. Y. Otsu, On manifolds of positive Ricci curvature with large diameters. Math. Z. **206**, 255–264 (1991)
82. M. Özaydın, G. Walschap, Vector bundles with no soul. PAMS **120**, 565–567 (1994)
83. R. Palais, A global formulation of the Lie theory of transformation groups. Mem. Am. Math. Soc. **22**, 123 (1957)
84. P. Pambuccian, T. Zamfirescu, Paolo Pizzetti: The forgotten originator of triangle comparison geometry. Historia Mathematica **38**, 415–422 (2011)
85. G. Perel'man, *Alexandrov's spaces with curvatures bounded from below II*, preprint

86. S. Peters, Cheeger's finiteness theorem for diffeomorphism classes of manifolds. J. Reine Angew. Math. **349**, 77–82 (1984)
87. P. Petersen, T. Tao, Classification of almost quarter-pinched manifolds. Proc. Am. Math. Soc. **137**(7), 2437–2440 (2009)
88. P. Petersen, G. Wei, Relative volume comparison with integral Ricci curvature bounds. Geom. Funct. Anal. **7**, 1031–1045 (1997)
89. P. Petersen, *Classical Differential Geometry*, http://www.math.ucla.edu/~petersen/DGnotes.pdf
90. P. Petersen, *Manifold Theory*, http://www.math.ucla.edu/~petersen/manifolds.pdf
91. P. Petersen, S. Shteingold, G. Wei, Comparison geometry with integral curvature bounds. Geom. Funct. Anal. **7**, 1011–1030 (1997)
92. G. de Rham, *Differentiable Manifolds* (Springer, Berlin-Heidelberg, 1984)
93. X. Rong, A bochner theorem and applications. Duke Math. J. **91**, 381–392 (1998)
94. E. A. Ruh, Almost flat manifolds. J. Differ. Geom. **17**(1), 1–14 (1982)
95. Y. Shen, S.-h. Zhu, *A sphere theorem for 3-manifolds with positive Ricci curvature and large diameter*, preprint, Dartmouth
96. E.H. Spanier, *Algebraic Topology* (Springer, New York-Berlin-Heidelberg, 1966)
97. M. Spivak, *A Comprehensive Introduction to Differential Geometry*, vol. I-V (Publish or Perish, Wilmington, 1979)
98. S. Tachibana, A theorem on Riemannian manifolds of positive curvature operator. Proc. Japan Acad. **50**, 301–302 (1974)
99. M.E. Taylor, *Partial Differential Equations*, vol. I-III (Springer, New York, 1996)
100. Q.M. Wang, Isoparametric functions on Riemannian manifolds. I. Math. Ann. **277**(4), 639–646 (1987)
101. F.W. Warner, *Foundations of Differentiable Manifolds and Lie Groups* (Springer, New York, 1983)
102. M. Weiss, Curvature and finite domination. Proc. Am. Math. Soc. **124**(2), 615–622 (1996)
103. B. Wilking, On fundamental groups of manifolds of nonnegative curvature. Differ. Geom. Appl. **13**(2), 129–165 (2000)
104. B. Wilking, Index parity of closed geodesics and rigidity of Hopf fibrations. Invent. Math. **144**(2), 281–295 (2001)
105. B. Wilking, Torus actions on manifolds of positive curvature. Acta Math. **191**, 259–297 (2003)
106. J. Wolf, *Spaces of Constant Curvature* (Publish or Perish, Wilmington, 1984)
107. H.H. Wu, *The Bochner Technique in Differential Geometry*, Mathematical Reports vol. 3, part 2 (Harwood Academic Publishers, London, 1988)

Index

P. Petersen, *Riemannian Geometry*, Graduate Texts in Mathematics 171,
DOI 10.1007/978-3-319-26654-1