Riemannian Geometry

Prof. Zuoqin Wang

Preface

These notes grew out of a graduate course on Riemannian Geometry taught by Prof. Zuoqin Wang. Their primary aim is to provide a systematic and conceptually coherent introduction to Riemannian geometry at the graduate level, emphasizing the intrinsic viewpoint and the deep interplay between local differential invariants and global geometric structure.

Riemannian geometry begins with one of the most profound shifts in the history of mathematics: the realization, articulated by Bernhard Riemann in his 1854 habilitation lecture, that geometry need not be tied to ambient Euclidean space. Instead, geometric structure can be encoded intrinsically through a smoothly varying inner product on the tangent spaces of a manifold. From this simple yet revolutionary idea arises a rich and far-reaching theory.

The Riemannian metric serves as the fundamental object. It defines length, angle, area, and volume; it induces a canonical connection; and from this connection emerges curvature. These notions are local in definition but global in consequence. The central theme of Riemannian geometry is precisely this passage from local differential data to global geometric and topological information.

The structure of these notes reflects this philosophy.

We begin with the basic foundations: Riemannian metrics, induced structures, and the Levi-Civita connection. Curvature tensors are then introduced, not merely as formal constructions but as geometric measurements of deviation from flatness. Particular attention is given to the Riemann curvature tensor, Ricci curvature, and sectional curvature, as they encode different layers of geometric information.

The theory of geodesics follows naturally. Geodesics serve both as geometric analogues of straight lines and as analytical tools. Through variational methods, one studies how curvature governs the behavior of geodesics. This leads to Jacobi fields, conjugate points, and the index form — fundamental tools for understanding global geometry.

Subsequent chapters develop comparison theorems and volume comparison results, which reveal how curvature bounds control topology and global structure. These results culminate in sphere theorems and rigidity phenomena, illustrating the remarkable strength of curvature constraints.

The final part introduces Bochner techniques and spectral considerations, highlighting the deep connections between curvature, analysis, and topology. Here one sees clearly that Riemannian geometry is not merely a branch of differential geometry, but a meeting point of geometry, analysis, and global topology.

Throughout the exposition, the emphasis is on clarity of structure rather than encyclopedic completeness. Computations are carried out when they illuminate geometric meaning; coordinate expressions are used as tools, not as ends in themselves. Many classical results are included with proofs, while some deeper results are presented to indicate broader directions of the theory.

These notes assume familiarity with smooth manifolds and basic differential geometry. A solid foundation in linear algebra and multivariable calculus is essential. Some acquaintance with analysis and topology will be helpful, particularly in the later chapters.

Riemannian geometry is a subject in which simple definitions give rise to profound consequences. It is my hope that these notes convey not only the technical framework but also the structural beauty of the theory.

Any remaining errors are my own.

Xumin Liang

Contents

Preface **i**

1 Riemannian Basics **1**
- 1.1 Introduction . . . 1
- 1.2 The Riemannian Metric . . . 6
- 1.3 The Riemannian Distance . . . 14
- 1.4 The Riemannian Measure . . . 19

2 Connections **27**
- 2.1 The Linear Connection . . . 27
- 2.2 The Levi-Civita Connection . . . 33

3 Curvature **42**
- 3.1 The Curvature Tensor . . . 42
- 3.2 The Riemann Curvature . . . 48
- 3.3 The Ricci and the Sectional Curvature . . . 57
- 3.4 Riemannian Manifolds with Constant Curvatures . . . 63
- 3.5 The Method of Moving Frames . . . 70

4 Geodesics and Variations **77**
- 4.1 Geodesics as Self-parallel Curves (on Manifolds with Connection) . . . 77
- 4.2 Geodesics on Riemannian Manifolds . . . 82
- 4.3 Existence of Shortest Geodesics . . . 88
- 4.4 Completeness . . . 91
- 4.5 Variations of Length and Energy . . . 97

5 Jacobi Fields **109**
- 5.1 Jacobi Fields . . . 109
- 5.2 Immediate Applications of Jacobi Field to Curvature . . . 115
- 5.3 Conjugate Point and Applications . . . 119
- 5.4 The Index Form . . . 124
- 5.5 Cut Locus . . . 131

6 Comparison Theorems **136**
- 6.1 Theorems on Curvature v.s. Topology . . . 136
- 6.2 Rauch Comparison Theorem . . . 142
- 6.3 The Global Hessian and Toponogov Comparison . . . 146

7 Volume Comparison and Applications **154**
- 7.1 The Laplacian and Volume Comparison . . . 154
- 7.2 Applications of the Volume Comparison Theorem . . . 159
- 7.3 The Sphere Theorem . . . 166

8 Bochner and Spectral Theory **173**
8.1 Bochner's Technique and Applications 173
8.2 Spectral Geometry 179

References **186**

Chapter 1

Riemannian Basics

1.1 Introduction

1.1.1 Riemann's Inaugural Lecture

On June 10, 1854, B. Riemann gave one of the most famous job talk in the history of mathematics, with title "On the hypothesis which lie at the foundation of geometry". This talk not only gained a job for him (as a privatdocent at Göttingen University), but also offered jobs for many of us including me: two of our courses, Manifolds and Riemannian geometry, born in this probationary inaugural lecture.

What Riemann did in this talk was trying to develop a higher dimensional intrinsic geometry. It is a very broad and abstract generalization of the intrinsic differential geometry of surfaces in $\mathbb{R}^3$ developed by Gauss[1].

At the beginning of Riemann's talk was a brief "plan of investigation", in which he started with the sentence "geometry presuppose the concept of space". To clear the confusion over non-Euclidean geometry at that time, he proposed to distinguish metric properties from the topological properties of the space. The major part of the talk was divided into three parts. In part one Riemann introduced the conception of manifolds, characterized as locally looks like n-dimensional Euclidean space[2]. Part tow is the major part of the talk, in which Riemann developed the desired intrinsic geometry, started by introducing a positive definite quadratic form (the Riemannian metric)[3] at each point. The crucial question Riemann asked himself in this part was: when does two Riemannian metrics locally isometric? By a dimension counting argument, Riemann argues that there should be a set of $\frac{n(n-1)}{2}$ functions which will determine the metric completely. They are nothing else but sectional curvatures (as a generalization of Gauss curvature for surfaces in $\mathbb{R}^3$) associated to 2-dimensional vector subspaces of the tangent space! Finally in part three, Riemann dealt with possible applications, especially to questions in physics.[4]

[1] In 1827, Gauss published a famous paper "General investigation of curved surfaces", in which he proved his Theorema Egregium ("remarkable theorem" in Latin): the Gauss curvature of a surface can be determined entirely by measuring distances along paths on the surface (intrinsic), and does not depend on how the surface might be embedded in 3-dimensional space (extrinsic).

[2] Riemann's definition of manifold is a very primitive form. Since most of his audience were non-mathematicians (faculty of Göttingen University), Riemann tried his best to make his lecture intelligible to general audience. The modern abstract definition of manifolds as "topological spaces that are Hausdorff, second countable and locally Euclidean" was introduced by H. Weyl in 1912.

[3] In fact, Riemann was also aware of the existence of more general "metrics" that could be used to measure the length of tangent vectors, including the so-called Finsler metric that was developed by Finsler in 1918.

[4] About 60 years later, Einstein used the theory of pseudo-Riemannian manifolds (a generalization of Riemannian manifolds) to developed his general theory of relativity. In particular, his equations for gravitation are constraints on the curvature of spacetime.

1.1.2 Riemannian Geometry for Euclidean Submanifolds: A Quick Survey on Undergraduated Differential Geometry

Before we introduce the abstract conception of Riemannian metric on a smooth manifold, let's start with some basic geometry that we learnt in undergraduate differential geometry course (in a higher dimensional fashion). As one can imagine, differential geometry starts by taking derivative. It turns out that all those important geometric quantities appears by this way.

Curves in $\mathbb{R}^N$

Let $\gamma : I \to \mathbb{R}^N$ be a smooth curve defined on a finite interval $I = [0, T]$. By definition the arc length $s = s(t)$ is given by

$$s(t) = \int_0^t \|\gamma'(\tau)\| \mathrm{d}\tau.$$

Since s is strictly increasing, we may change variable and write γ as

$$\gamma = \gamma(s), \ s \in [0, l],$$

where l is the length of γ.

We start with the unit tangent vector $\gamma'(s)$: since $\|\gamma'(s)\| = 1$, i.e.

$$\langle \gamma'(s), \gamma'(s) \rangle = 1,$$

taking derivative one gets

$$\langle \gamma'', \gamma'(s) \rangle = 0,$$

i.e. $\gamma''(s) \perp \gamma'(s)$. In other words, $\gamma''(s)$ is a normal vector.

By definition,

$$\kappa(s) := \|\gamma''(s)\|$$

is called the **curvature** of γ at $\gamma(s)$, and the vector

$$n(s) := \frac{\gamma''(s)}{\|\gamma''(s)\|}$$

is called the **principal normal** γ at $\gamma(s)$.

Remark. Note that $n(s)$ is again a unit vector. So we may repeat this process. What we will get is the torsion and the binormal. If we continue this process for the binormal, we will get Frenet formula.

The First Fundamental Form

Now let M be an n-dimensional manifold embedded into $\mathbb{R}^N$. For simplicity, we suppose $U \subset \mathbb{R}^n$ is an open set, and suppose

$$\varphi : U \subset \mathbb{R}^n \to \mathbb{R}^N$$

is an injective immersion such that $\varphi(U) = M$ (or a portion of M). In what follows we denote

$$\varphi_j = \frac{\partial \varphi}{\partial x^j}, \ 1 \leqslant j \leqslant n.$$

Then $T_{\varphi(x)}M = \mathrm{span}(\varphi_1, \ldots, \varphi_n)$. Now let $\mu = (\mu^1, \ldots, \mu^n) : I \to U$ be a curve in U, so that $\gamma = \varphi \circ \mu : I \to M$ be a curve in $\mathbb{R}^N$ that sits in M. Then

$$\gamma'(t) = \frac{\mathrm{d}(\varphi \circ \mu)}{\mathrm{d}t} = \sum_{j=1}^n \frac{\mathrm{d}\mu^j}{\mathrm{d}t} \varphi_j(\mu(t))$$

The arc length of γ is again given by

$$s(t) = \int_0^t \|\gamma'(\tau)\|\mathrm{d}\tau$$

If we denote $v^j := \dfrac{\mathrm{d}\mu^j}{\mathrm{d}t}$, $x = \mu(t)$ and

$$g_{jk}(x) := \langle \varphi_j(x), \varphi_k(x) \rangle,$$

then we get

$$\left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2 = \|\gamma'(t)\|^2 = \sum_{j,k=1}^n g_{jk}(\mu(t))v^j v^k.$$

After polarizing, we get a quadratic form

$$\mathrm{I}\left(\sum_j v^j\varphi_j \sum_k w^k\varphi_k\right) := \sum_{j,k=1}^n g_{jk}(x)v^j w^k$$

defined on $T_{\varphi(x)}M$, which is known as the first fundamental form of M.

The Second Fundamental Form

We may continue to calculate the second derivative to get

$$\gamma''(t) = \frac{\mathrm{d}^2(\varphi \circ \mu)}{\mathrm{d}t^2} = \sum_{j=1}^n \frac{\mathrm{d}^2\mu^j}{\mathrm{d}t^2}\varphi_j(\mu(t)) + \sum_{j,k=1}^n \frac{\mathrm{d}\mu^j}{\mathrm{d}t}\frac{\mathrm{d}\mu^k}{\mathrm{d}t}\varphi_{jk}(\mu(t)),$$

where $\varphi_{jk}(x) = \dfrac{\partial^2\varphi}{\partial x^j \partial x^k}(x)$. Note that the first term lies in $T_{\gamma(t)}M$. So when projecting to the **normal plane** $N_{\gamma(t)}M = (T_{\gamma(t)}M)^\perp$, and denoting

$$h_{jk} = \mathrm{Proj}_{N_{\gamma(t)}M}(\varphi_{jk}),$$

one gets

$$\sum_{j,k=1}^n h_{jk}v^j v^k = \mathrm{Proj}_{N_{\gamma(t)}M}(\gamma''(t)).$$

For simplicity let's take arc length parametrization, so that $\sum v^j\varphi_j$ is a unit vector (i.e. $\sum g_{jk}\dfrac{\mathrm{d}\mu^j}{\mathrm{d}s}\dfrac{\mathrm{d}\mu^k}{\mathrm{d}s} = 1$). Then we get

$$\sum_{j,k=1}^n h_{jk}v^j v^k = \kappa(s)\,\mathrm{Proj}_{N_{\gamma(s)}M}(n(s)).$$

After polarizing, the resulting quadratic form

$$\mathrm{II}\left(\sum v^j\varphi_j, \sum w^k\varphi_k\right) := \sum h_{jk}v^j v^k$$

(defined on T_xM with value in N_xM) is known as the **second fundamental form** of M. In the case M is a hypersurface (i.e. $n = N-1$), by fixing an orientation on M one may identify $N_{\gamma(t)}M$ with $\mathbb{R}$, and thus II can be viewed as a real-valued quadratic form.

The Christoffel Symbols

Interesting quantities also appears when we study the tangent component of $\gamma''(t)$. Since $\varphi_{jk}(x) - h_{jk}(x) \in T_{\varphi(x)}M = \mathrm{span}(\varphi_1, \dots, \varphi_n)$, one may write

$$\varphi_{jk}(x) = \sum_{l=1}^{n} \Gamma_{jk}^{l} \varphi_l(x) + h_{jk}(x).$$

Pairing with the vector φ_i, one gets

$$\langle \varphi_{jk}, \varphi_i \rangle = \sum_{l=1}^{n} \Gamma_{jk}^{l} g_{li}.$$

A miracle is that the mysterious coefficients Γ_{jk}^{l} can be calculated via g_{jk}'s: From

$$\partial_k g_{ij} = \langle \varphi_{ik}, \varphi_j \rangle + \langle \varphi_i, \varphi_{jk} \rangle$$

one gets

$$\langle \varphi_{jk}, \varphi_i \rangle = \frac{1}{2}(\partial_k g_{ij} + \partial_j g_{ik} - \partial_i g_{kj}).$$

So if we denote $(g^{ij}) = (g_{ij})^{-1}$, then

$$\Gamma_{jk}^{l} = \sum_i g^{il} \langle \varphi_{jk}, \varphi_i \rangle = \frac{1}{2} \sum_i g^{il} (\partial_k g_{ij} + \partial_j g_{ik} - \partial_i g_{kj}).$$

The functions Γ_{jk}^{l} are known as **Christoffel symbols**. Note that they are determined by the first fundamental form. In summary, we see that

$$\gamma''(t) = \sum_{j=1}^{n} \left(\frac{\mathrm{d}^2 \mu^j}{\mathrm{d}t^2} + \sum_{i,k=1} \Gamma_{ik}^{j} \frac{\mathrm{d}\mu^i}{\mathrm{d}t} \frac{\mathrm{d}\mu^k}{\mathrm{d}t} \right) \varphi_j(\mu(t)) \bmod N_{\gamma(t)}M.$$

The Covariant Derivative and Geodesics

In particular, if γ is parameterized by arc length s (i.e. $\sum g_{jk} \frac{\mathrm{d}\mu^j}{\mathrm{d}s} \frac{\mathrm{d}\mu^k}{\mathrm{d}s} = 1$), then

$$\sum_{j=1}^{n} \left(\frac{\mathrm{d}^2 \mu^j}{\mathrm{d}s^2} + \sum_{i,k=1}^{n} \Gamma_{ik}^{j} \frac{\mathrm{d}\mu^i}{\mathrm{d}s} \frac{\mathrm{d}\mu^k}{\mathrm{d}s} \right) \varphi_j(\mu(s)) = \kappa(s) \, \mathrm{Proj}_{T_{\gamma(t)}M}(n(s)).$$

The length

$$\kappa_g(s) := \left\| \sum_{j=1}^{n} \left(\frac{\mathrm{d}^2 \mu^j}{\mathrm{d}s^2} + \sum_{i,k=1}^{n} \Gamma_{ik}^{j} \frac{\mathrm{d}\mu^i}{\mathrm{d}s} \frac{\mathrm{d}\mu^k}{\mathrm{d}s} \right) \varphi_j(\mu(s)) \right\|$$

is known as the geodesic curvature of γ. If $\kappa_g(s) \equiv 0$, then γ is called a geodesic. They are locally shortest paths (generalizations of straight lines in Euclidean space and great circles in sphere) in M.

More generally, given any vector field $X = \sum X^j(x)\varphi_j(x)$ along γ (which, by definition, is tangent to M everywhere), the same computation yields

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \sum_{j,k=1}^{n} (\partial_k X^j + \Gamma_{ik}^{j} X^i) \frac{\mathrm{d}\mu_k}{\mathrm{d}t} \varphi_j(\mu(t)) \mathrm{mod} N_{\gamma(t)}M,$$

which is known as the **covariant derivative** of the vector field X along γ.

The Riemann Curvature

What about vector fields N that are normal to M? We may calculate the tangential component of the derivative of $N(x)$ in a similar way. For this purpose, we write

$$\partial_i N(x) = \sum_{k=1}^{n} N_i^k(x)\varphi_k(x) \bmod N_{\varphi(x)}M.$$

Start with the equation $\langle N(x), \varphi_j(x)\rangle = 0$. By taking derivative, we get

$$\langle \partial_i N(x), \varphi_j(x)\rangle + \langle N, \varphi_{ij}\rangle = 0,$$

i.e.

$$\sum_k N_i^k g_{kj} = -\langle h_{ij}, n\rangle.$$

It follows

$$N_i^k = -\sum_j \langle h_{ij}, N\rangle g^{kj}$$

and thus we get, for any normal vector field N on M,

$$\partial_i N(x) = -\sum_{j,k} \langle h_{ij}, N\rangle g^{kj}\varphi_k(x) \bmod N_{\varphi(x)}M.$$

Applying this formula to the normal vector fields h_{ij}, we may calculate the tangential component of $\varphi_{ijk} = \partial_i\partial_j\partial_k\varphi$. Since $\varphi_{ij} = \sum \Gamma_{ij}^l \varphi_l + h_{ij}$, we get

$$\varphi_{kij} = \partial_k \varphi_{ij} = \sum_{m=1}^{n} \left(\partial_k \Gamma_{ij}^m + \sum_l \Gamma_{ij}^l \Gamma_{lk}^m - \sum_l \langle h_{ij}, h_{kl}\rangle g^{lm} \right) \varphi_m \bmod N_{\varphi(x)}M.$$

Since $\varphi_{kij} = \varphi_{jik}$, we get

$$\partial_j \Gamma_{ik}^m - \partial_k \Gamma_{ij}^m + \sum_{l=1}^{n} (\Gamma_{ik}^l \Gamma_{lj}^m - \Gamma_{ij}^l \Gamma_{lk}^m) = \sum_{l=1}^{n} (\langle h_{ik}, h_{jl}\rangle - \langle h_{ij}, h_{kl}\rangle) g^{lm}.$$

We define

$$R_{ijk}{}^m := \partial_j \Gamma_{ik}^m - \partial_k \Gamma_{ij}^m + \sum_{l=1}^{n} (\Gamma_{ik}^l \Gamma_{lj}^m - \Gamma_{ij}^l \Gamma_{lk}^m)$$

and let

$$R_{ijkl} := \sum_m g_{lm} R_{ijk}^m,$$

then we get $R_{lijk} = \langle h_{ik}, h_{jl}\rangle - \langle h_{ij}, h_{kl}\rangle$. The $(0,4)$-tensor

$$R\left(\sum X^l\varphi_l, \sum Y^i\varphi_i, \sum Z^j\varphi_j, \sum W^k\varphi_k\right) := \sum R_{lijk} X^l Y^i Z^j W^k$$

on T_xM is called the **Riemann curvature tensor**. It admits many nice symmetry properties from which one can show that the quantity

$$\frac{R(X,Y,X,Y)}{\langle X,X\rangle\langle Y,Y\rangle - \langle X,Y\rangle^2}$$

depends only on the two dimensional plane $\operatorname{span}(X,Y)$. It is known as the **sectional curvature** of M with respect to the plane. By taking a basis there are $\dfrac{n(n-1)}{2}$ such functions, and they are the $\dfrac{n(n-1)}{2}$ functions first studied by Riemann!

1.2 The Riemannian Metric

As we have seen, in Riemannian geometry there will be lots of summations for quantities with many indices. To simplify notions and computations, from now on we will follow the **Einstein summation convention**.

Definition 1.1 (Einstein Summation Convention)**.** If an expression is a product of several terms with indices, and if an index variable appears twice in this expression, once as an upper index in one term and once as a lower index in another term[5], then (unless otherwise stated) the expression is understood to be a summation over all possible values of that index (usually from 1 to the space dimension). For example,

$$a_i b^i := \sum_i a_i b^i, \quad a^{ijkl} b_{il}^m c_j := \sum_{i,j,l} a^{ijkl} b_{il}^m c_j.$$

One should also be aware of how we choose upper and lower indices in this course (trying to meet Einstein summation convention). For example, vector fields are always indexed by lower indices (like $X_1, X_2, \ldots$) while the coefficients of vector fields will be indexed by upper indices (e.g. $a^1 X_1 + a^2 X_2$). Similarly a collection of 1-forms will be indexed by upper indices while the coefficients of their linear combinations will be indexed by lower indices.

1.2.1 The Riemannian Metric

Definition of Riemannian Metric

Let M be a smooth manifold of dimension m, in other words, M is a second countable Hausdorff topological space such that near every point $p \in M$, there is a neighborhood U of p which is diffeomorphic to a domain in $\mathbb{R}^m$. Moreover, if we denote by $\{x^1, \ldots, x^m\}$ the coordinate functions on U, then the tangent space T_pM is spanned by the vectors $\{\partial_1, \ldots, \partial_m\}$, and its dual T_p^*M (the cotangent space) is spanned by $\{\mathrm{d}x^1, \ldots, \mathrm{d}x^m\}$.

Definition 1.2. A **Riemannian metric** g on M is an assignment of an inner product

$$g_p(\cdot, \cdot) = \langle \cdot, \cdot \rangle_p$$

on T_pM for each $p \in M$, such that the assignment depends smoothly on p.

Remark.

(1) As we have seen, the Riemannian metric g is motivated by the first fundamental form of a surface in space. They will be used to measure the length of curves in M.

(2) "Smooth dependence" $\Leftrightarrow$ if X, Y are two smooth vector fields on an open subset $U \subset M$, then $f(p) = \langle X_p, Y_p \rangle_p$ is a smooth function on U.

(3) The Riemannian metric g itself is NOT a **metric** (also known as a **distance function**) on M. Recall that a distance function on M is a continuous function $d : M \times M \to \mathbb{R}$ so that for all $p, q, r \in M$,

- $d(p, q) \geqslant 0$, and $d(p, q) = 0$ if and only if $p = q$.
- $d(p, q) = d(q, p)$.
- $d(p, r) \leqslant d(p, q) + d(q, r)$.

However, we will see soon that g induces a natural distance function d on M, and the topology generated by d on M coincides with its original manifold topology.

[5] Note: an upper index in the denominator will be regarded as a lower index, and vice versa.

Riemannian Metric as a Tensor Field

We may also use the language of tensors. By definition

$$g : \Gamma^\infty(TM) \times \Gamma^\infty(TM) \to \mathcal{C}^\infty(M)$$

defined in the obvious way is $\mathcal{C}^\infty(M)$-bilinear, and thus can be viewed as a $(0,2)$-tensor on M. The remaining conditions of being an inner product (i.e. symmetric and positive definite) at each point now becomes, in the language of tensors, that the $(0,2)$-tensor g is symmetric and positive definite. So we get another description of a Riemannian metric g:

A Riemannian metric g is a smooth symmetric $(0,2)$-tensor field that is positive definite.

We remark that many geometric structures on smooth manifold M are defined as a special tensor field. For example, an almost complex structure on M is a special $(1,1)$-tensor field, a symplectic structure on M is a special $(0,2)$ tensor field, while a Poisson structure on M is a special $(2,0)$ tensor field.

Riemannian Metric via Local Coordinates

One can represent the Riemannian metric g using local coordinates as follows. Let $\{U, x^1, \ldots, x^m\}$ be a coordinate patch. We denote

$$g_{ij}(p) = \langle \partial_i, \partial_j \rangle_p.$$

It is easy to see that the functions g_{ij} have the following properties:

- For all i, j, $g_{ij}(p)$ is smooth in p.
- $g_{ij} = g_{ji}$, so the matrix $(g_{ij}(p))$ is symmetric at any p.
- The matrix $(g_{ij}(p))$ is also positive definite for any p.

Note that although g is intrinsically defined, the functions g_{ij} depend on the choice of coordinate system. If $\{\tilde{x}^1, \ldots, \tilde{x}^n\}$ is another coordinated system on U, then

$$\tilde{\partial}_i = \frac{\partial x^k}{\partial \tilde{x}^i} \partial_k.$$

It follows that

$$\tilde{g}_{ij} := \langle \tilde{\partial}^i, \tilde{\partial}^j \rangle = \frac{\partial x^k}{\partial \tilde{x}^k} g_{kl} \frac{\partial x^l}{\partial \tilde{x}^j}.$$

In other words, the matrices $(\tilde{g}_{ij})$ and (g_{ij}) are related by the matrix equation

$$(\tilde{g}_{ij}) = J^{\mathrm{T}}(g_{ij})J$$

where J is the Jacobian matrix whose (i,j)-element is $\left(\frac{\partial x^i}{\partial \tilde{x}^j}\right)$.

Since for any smooth vector fields $X = X^i\partial_i$ and $Y = Y^j\partial_j$ in U,

$$\langle X_p, Y_p \rangle_p = X^i(p)Y^j(p)\langle \partial_i, \partial_j \rangle_p = g_{ij}(p)X^i(p)Y^j(p),$$

so locally we can write the 2-tensor g as

$$g = g_{ij} \mathrm{d}x^i \otimes \mathrm{d}x^j.$$

The Dual Riemannian Metric on the Cotangent Space

Since each matrix (g_{ij}) is positive definite, it is invertible. We will denote by (g^{ij}) the inverse matrix of (g_{ij}), i.e. they satisfy

$$g_{ij}g^{jk} = \delta_i^k.$$

Then the matrix (g^{ij}) is again positive definite, and we can use it to define a dual inner product structure g^* on T_p^*M for each p. More explicitly, for any 1-forms

$$\omega = \omega_i \mathrm{d}x^i \text{ and } \eta = \eta_i \mathrm{d}x^i$$

on U, we define

$$g^*(\omega, \eta) = \langle \omega, \eta \rangle_p^* := g^{ij}(p)\omega_i(p)\eta_j(p).$$

We will leave as a simple exercise for the reader to check that this definition is independent of the choices of coordinates.

The Musical Isomorphisms

Since g is non-degenerate and bilinear on T_pM, it gives us an isomorphism between T_pM and T_p^*M via[6]

$$\flat : T_pM \to T_p^*M,$$
$$\flat(X_p)(Y_p) := g_p(X_p, Y_p).$$

(Pronunciation of $\flat$: flat.) It is not hard to see that $\flat$ maps smooth vector fields to smooth 1-forms, and gives rise to a vector bundle isomorphism between TM and T^*M.

In local coordinates, if we denote $X = X^i\partial_i$ and take $Y = \partial_j$ for each j, then

$$\flat(X)(\partial_j) = g(X, \partial_j) = g_{ij}X^i,$$

so we conclude

$$\flat(X^i\partial_i) = g_{ij}X^i \mathrm{d}x^j.$$

In other words, $\flat$ "lowers the indices" via g_{ij}, i.e. changes the coefficients from X^i to $X_i := g_{ij}X^i$.

We will denote the inverse map of $\flat$ by

$$\sharp : T_p^*M \to T_pM.$$

Then in local coordinates,

$$\sharp(w_i \mathrm{d}x^i) = g^{ij}w_i\partial_j.$$

So $\sharp$ "raises the indices" via g^{ij}. We will call $\flat$ and $\sharp$ the **musical isomorphisms**[7].

Note that for any 1-form ω and η,

$$g_p(\sharp\omega, \sharp\eta) = g_{ij}g^{ki}\omega_k g^{lj}\eta_l = \delta_j^k \omega_k \eta_l g^{lj} = g^{kl}\omega_k\eta_l = \langle \omega, \eta \rangle_p^*.$$

In other words, the dual inner product $g_p^*(\omega, \eta)$ on T_p^*M we mentioned above can be defined as $g_p(\sharp\omega, \sharp\eta)$, which is a coordinate-free definition of g^*.

Riemannian Metric for Tensors

Given the Riemannian inner product g on T_pM and the induced inner product g^* on T^*M, one may further define a natural inner product $T_l^k(g)$, also denoted by g if there is no ambiguity, on the tensor product space $(T_pM)^{\otimes k} \otimes (T_p^*M)^{\otimes l}$ as follows:

Let $W = (T_pM)^k \times (T_p^*M)^l$ (the Cartesian product). Consider the map $W \times W \to \mathbb{R}$ given by

$$((X_1, \dots, X_k, \omega_1, \dots, \omega_l), (Y_1, \dots, Y_k, \eta_1, \dots, \eta_l)) \mapsto g(X_1, Y_1) \cdots g(X_k, Y_k) g^*(\omega_1, \eta_1) \cdots g^*(\omega_l, \eta_l).$$

[6] Although $\dim T_pM = \dim T_p^*M$, without using a Riemannian metric or some other extra structure, we don't have a natural isomorphism between T_pM and T_p^*M.

[7] In music, the symbol $\flat$ means lower in pitch while the symbol $\flat$ means higher in pitch.

It is a multi-linear map which is linear in each entry. By universality of tensor product, it gives rise to a unique bilinear map

$$(T_pM)^{\otimes k} \otimes (T_p^*M)^{\otimes l} \times (T_pM)^{\otimes k} \otimes (T_p^*M)^{\otimes l} \to \mathbb{R}$$

which can be proven to be an inner product.

This inner product can be characterized by the following property: Suppose $e_1, \ldots, e_m$ is an orthonormal basis of (T_pM, g_p), and $e^1, \ldots, e^m$ its dual basis of (T_p^*M, g_p^*). Then the induced inner product on $(T_pM)^{\otimes k} \otimes (T_p^*M)^{\otimes l}$ is defined so that

$$\{e_{i_1} \otimes \cdots \otimes e_{i_k} \otimes e^{j_1} \otimes \cdots \otimes e^{j_l}\}$$

form an orthonormal basis.

In local coordinates, if $T = T^{i_1 \cdots i_k}_{j_1 \cdots j_l} \partial_{i_1} \otimes \cdots \otimes \partial_{i_k} \otimes \mathrm{d}x^{j_1} \otimes \cdots \otimes \mathrm{d}x^{j_k}$ and likewise for a (k, l)-tensor S, then

$$\langle T, S \rangle = g^{j_1 b_1} \cdots g^{j_l b_l} g_{i_1 a_1} \cdots g_{i_k a_k} T^{i_1 \cdots i_k}_{j_1 \cdots j_l} S^{a_1 \cdots a_k}_{b_1 \cdots b_l}.$$

As an example, we see that the length square of the metric tensor g itself is

$$|g|^2 = \langle g, g \rangle = g^{ik} g^{jl} g_{ij} g_{kl} = \delta^k_j \delta^j_k = m.$$

1.2.2 Riemannian Manifolds

Riemannian Manifolds: Definition and Simplest Example

Let M be a smooth manifold.

Definition 1.3. Let g be Riemannian metric on M. Then we call the pair (M, g) a **Riemannian manifold**. (Sometimes we omit g and say M is a Riemannian manifold.)

Example. The simplest manifold of dimension m is $\mathbb{R}^m$, on which we can endow many Riemannian metrics:

(1) The standard inner product on $\mathbb{R}^m$ defines a **canonical Riemannian metric** g_0 on $\mathbb{R}^m$ via

$$g_0(X, Y) = \sum_i X^i Y^i.$$

Alternatively, this means the matrix (g_{ij}) is the identity matrix:

$$(g_0)_{ij} = \delta_{ij}.$$

In the notion of tensors, we can write

$$g_0 = \mathrm{d}x^1 \otimes \mathrm{d}x^1 + \cdots + \mathrm{d}x^m \otimes \mathrm{d}x^m.$$

(2) More generally, for any positive definite $m \times m$ matrix $A = (a_{ij})$, the formula

$$g_p^A(X_p, Y_p) := X_p^{\mathrm{T}} A Y_p$$

defines a Riemannian metric on $\mathbb{R}^m$ in which case $g^A_{ij} = a_{ij}$. Equivalently,

$$g^A = \sum_{i,j} a_{ij} \mathrm{d}x^i \otimes \mathrm{d}x^j.$$

(3) Since $\mathbb{R}^m$ admits a global coordinate system, one may even describe all possible Riemannian metrics on $\mathbb{R}^m$: Endow the space $\mathrm{Sym}(m)$ of all $m \times m$ symmetric matrices (which is linearly isomorphic to $\mathbb{R}^{m(m+1)/2}$) the standard smooth structure, then the subset $\mathrm{PosSym}(m)$ of all positive definite $m \times m$ matrices is open and thus again a smooth $m \times m$ matrices is open and thus again a smooth manifold. By definition, any smooth map

$$g : \mathbb{R}^m \to \mathrm{PosSym}(m) \subset \mathrm{Sym}(m)$$

defines a Riemannian metric on $\mathbb{R}^m$, and vice versa.

Example. On the torus $\mathbb{T}^m = (S^1)^m$, one has the following flat Riemannian metric

$$g_0 = \mathrm{d}\theta^1 \otimes \mathrm{d}\theta^1 + \cdots + \mathrm{d}\theta^m \otimes \mathrm{d}\theta^m.$$

Example. Consider the upper half plane $\mathbb{H}^2 = \{(x,y) \mid y > 0\}$. On $\mathbb{H}^2$ the Riemannian metric

$$g_{(x,y)} = \frac{1}{y^2}(\mathrm{d}x \otimes \mathrm{d}x + \mathrm{d}y \otimes \mathrm{d}y)$$

is known as the **hyperbolic metric**, and $(\mathbb{H}^2, g)$ is known as the hyperbolic plane.

Constructing New Riemannian Manifolds

There are many ways to construct new Riemannian manifolds from old ones, for example,

(1) Let (M, g_M) and (N, g_N) be two Riemannian manifolds, then $g_M \oplus g_N$ defined by

$$(g_M \oplus g_N)_{(p,q)}((X_p, Y_q), (X'_p, Y'_q)) = (g_M)_p(X_p, X'_p) + (g_N)_q(Y_q, Y'_q)$$

is a Riemannian metric on $M \times N$, whose matrix is simply

$$\begin{pmatrix} (g_M)_{m_1 \times m_1} & 0 \\ 0 & (g_N)_{m_2 \times m_2} \end{pmatrix},$$

where m_1, m_2 are the dimensions of M and N respectively.

Definition 1.4. We will call $(M \times N, g_M \oplus g_N)$ the **product Riemannian manifold** of (M, g_M) and (N, g_N).

For example,

- The Euclidean space $(\mathbb{R}^m, g_0)$ is the Riemannian product of m copies of $(\mathbb{R}, g_0)$.
- The torus $(\mathbb{T}^m, g_0)$ is the Riemannian product of m copies of the standard circle $(S^1, \mathrm{d}\theta \otimes \mathrm{d}\theta)$.
- The hyperbolic plane $(\mathbb{H}, g)$ is NOT a Riemannian products of two 1-dimensional manifolds.

(2) Let (N, g_N) be a Riemannian manifold, and $f : M \to N$ a smooth **immersion**, i.e. $\mathrm{d}f_p : T_pM \to T_{f(p)}N$ is injective for all $p \in M$. Then the "pull-back metric" f^*g_N on M defined by

$$(f^*g_N)_p(X_p, Y_p) = (g_N)_{f(p)}(\mathrm{d}f_p(X_p), \mathrm{d}f_p(Y_p))$$

is a Riemannian metric on M.

Definition 1.5. We call $g_M := f^*g_N$ the **induced metric** or the **pulled-back metric** on M with respect to f, and call $f : (M, g_N) \to (N, g_N)$ an **isometric immersion**. If f is an embedding, then f is called an **isometric embedding**.

(3) Let (N, g_N) be a Riemannian manifold, and $M \subset N$ be an immersed/embedded submanifold. Then the inclusion map $\iota : M \to N$ is an immersion, which defines an induced Riemannian metric on M.

Definition 1.6. We call (M, ι^*g_N) an immersed/embedded **Riemannian submanifold** of (N, g_N). (Usually "Riemannian submanifold" refers to "embedded Riemannian submanifold").

Note that under the identification of T_pM with $\mathrm{d}\iota_p(T_pM) \subset T_pN$, the induced metric $(\iota^*g_N)_p$, viewed as an inner product or a tensor field, is just the restriction of g_N onto the subspace $T_pM \subset T_pN$.

(4) Let (M, g) be any Riemannian manifold, and $u : M \to \mathbb{R}$ an arbitrary smooth function on M. Then $e^{2u}g$ defined by

$$(e^{2u}g)_p(X_p, Y_p) = e^{2u(p)}g_p(X_p, Y_p)$$

is a Riemannian metric on M.

Definition 1.7. We say a Riemannian metric g' on M is **conformal** to g if

$$g' = e^{2u} g$$

for some $u \in \mathcal{C}^\infty(M)$.

By definition, if two Riemannian metrics g' and g are conformal, then when we replace g by g', for each p, all vectors in T_pM are stretched in length by the same constant $e^{u(p)}$, while the angle between any pair of vectors in T_pM keeps the same.

S^2 as a Riemannian submanifold of $\mathbb{R}^3$

Example. Let $M = S^2$ be the unit 2-sphere in $\mathbb{R}^3$. The induced Riemannian metric g (from the canonical Riemannian metric g_0 on $\mathbb{R}^3$) is known as the **round metric**. To calculate g locally, we need to choose a coordinate patch.

For example, we can use cylindrical coordinates θ and z to parameterize S^2,

$$x = \sqrt{1-z^2}\cos\theta, \quad y = \sqrt{1-z^2}\sin\theta, \quad z = z,$$

with $0 < \theta < 2\pi$, $-1 < z < 1$. Then

$$\mathrm{d}x = \frac{-z}{\sqrt{1-z^2}}\cos\theta \mathrm{d}z - \sqrt{1-z^2}\sin\theta\mathrm{d}\theta$$

and

$$\mathrm{d}y = \frac{-z}{\sqrt{1-z^2}}\sin\theta \mathrm{d}z + \sqrt{1-z^2}\cos\theta\mathrm{d}\theta.$$

It follows

$$\begin{aligned} g_{S^2} &= [\mathrm{d}x\otimes\mathrm{d}x + \mathrm{d}y\otimes\mathrm{d}y + \mathrm{d}z\otimes\mathrm{d}z]|_{S^2} \\ &= \frac{z^2}{1-z^2}\mathrm{d}z\otimes\mathrm{d}z + (1-z^2)\mathrm{d}\theta\otimes\mathrm{d}\theta + \mathrm{d}z\otimes\mathrm{d}z \\ &= \frac{1}{1-z^2}\mathrm{d}z\otimes\mathrm{d}z + (1-z^2)\mathrm{d}\theta\otimes\mathrm{d}\theta. \end{aligned}$$

Alternatively, one may use the colatitude $\theta \in (0,\pi)$ and the longitude $\varphi \in (0, 2\pi)$ to parameterize S^2 as

$$x = \sin\theta\cos\varphi, \quad y = \sin\theta\sin\varphi, \quad z = \cos\theta.$$

A similar computation as above will give us

$$g_{S^2} = \mathrm{d}\theta\otimes\mathrm{d}\theta + \sin^2\theta\mathrm{d}\varphi\otimes\mathrm{d}\varphi.$$

Isometries and Local Isometries

Next let's define the notion of "equivalence" in the Riemannian world.

Definition 1.8. Let (M, g_M) and (N, g_N) be two Riemannian manifolds.

(1) If $\varphi : M \to N$ is a local diffeomorphism such that $g_M = \varphi^* g_N$, then we call φ a **local isometry**.

(2) If a local isometry $\varphi : (M, g_M) \to (N, g_N)$ is invertible, then we call φ an **isometry**, in which case we say (M, g_M) and (N, g_N) are **isometric** Riemannian manifolds.

Isometries are crucial in Riemannian geometry since isometric Riemannian manifolds will be viewed as the same. Local isometries are also important in studying local invariants like curvatures.

Remark. A map $\varphi : (M, g_M) \to N, g_N$ is a local isometry if and only if for any $p \in M$, there exists a neighborhood U of p in M so that $\varphi : U \to \varphi(U)$ is an isometry.

Example. For any $m \times m$ positive definite matrix A, $(\mathbb{R}^m, g^A)$ is isometric to $(\mathbb{R}^m, g_0)$. [Can you write down the isometry?]

Example. On the set $M = \mathbb{R}_{>0} \times (0, 2\pi)$, consider the Riemannian metric

$$g = \mathrm{d}r \otimes \mathrm{d}r + r^2 \mathrm{d}\theta \otimes \mathrm{d}\theta.$$

Then the map

$$\varphi : M \to \mathbb{R}^2 - \{(x, 0) \mid x \geqslant 0\}$$
$$(r, \theta) \mapsto (r\cos\theta, r\sin\theta)$$

(where the latter is endowed with the standard Euclidean metric) is an isometry. [Obviously (M, g) is really the polar coordinate system for $\mathbb{R}^2$.]

Example. For the standard metrics on $\mathbb{R}^m$ and T^m: If we regard $\mathbb{T}^m = \mathbb{R}^m/\mathbb{Z}^m$, then the projection $\pi : (\mathbb{R}^m, g_0) \to (\mathbb{T}^m, g_0)$ is a local isometry but not a global isometry.

The Isometry Group

Obviously isometries satisfies the following functorality:

- If $\varphi : (M, g_M) \to (N, g_N)$ is an isometry, the φ^{-1} is an isometry.
- If $\varphi : (M, g_M) \to (N, g_N)$ and $\psi : (N, g_N) \to (P, g_P)$ are two isometries, then the composition $\psi \circ \varphi : (M, g_M) \to (P, g_P)$ is again an isometry.

In particular, if we let

$$\mathrm{Isom}(M, g) = \{\varphi : (M, g) \to (M, g) \mid \varphi \text{ is an isometry}\}.$$

Then $\mathrm{Isom}(M, g)$ is a group. It is a subgroup of the diffeomorphism group

$$\mathrm{Diff}(M) = \{\varphi : M \to M \mid \varphi \text{ is a diffeomorphism}\}.$$

Definition 1.9. We call $\mathrm{Isom}(M, g)$ the **isometry group** of (M, g).

For example,

- The isometry group of $(\mathbb{R}^m, g_0)$ is the Euclidean group $E(m) = O(m) \ltimes \mathbb{R}^m$.
- The isometry group of (S^2, g_{round}) is the orthogonal group $O(3)$.

Remark. A remarkable theorem proved by Myers and Steenrod in 1939 claims

Theorem 1.10 (Myers-Steenrod)**.** Let (M, g) be any Riemannian manifold. Then with respect to the compact open topology, there is a smooth structure on $\mathrm{Isom}(M, g)$ so that $\mathrm{Isom}(M, g)$ is a Lie group, which is compact if M is compact. Moreover, the obvious action of $\mathrm{Isom}(M, g)$ on M is smooth.

On the other hand, as we have learned in the course of smooth manifold, the diffeomorphism group $\mathrm{Diff}(M)$ can be regarded as an "infinite dimensional Lie group" whose Lie algebra is the Lie algebra of all smooth vector fields on M (for simplicity we may assume M is compact). So the isometry group $\mathrm{Isom}(M, g)$ of a Riemannian manifold (M, g), as a Lie group which is finite dimensional and carries a smooth structure, is much nicer than the diffeomorphism group $\mathrm{Diff}(M)$ of the underlying manifold M. What is the Lie algebra of $\mathrm{Isom}(M, g)$? They are nothing else but those vector fields whose flow are isometries, known as **Killing vector fields**.

By the remark above we see that the Riemannian structure is much more **rigid** than the smooth structure. As we know, locally manifolds of the same dimension are always the same. However, this is not the case for Riemannian manifolds: there are rich local geometry in the Riemannian world.

The Existence of Riemannian Metric

The first remarkable theorem in this course is

Theorem 1.11. On any smooth manifold M, there exist (many) Riemannian metrics on any smooth manifold M.

We shall give two proofs of this theorem.

The First Proof. We first take a locally finite covering of M by coordinate patches $\{U_\alpha, x_\alpha^1, \dots, x_\alpha^m\}$. It is clear that one can choose a Riemannian metric g_α on each U_α, e.g. one may take

$$g_\alpha = \sum_i \mathrm{d}x_\alpha^i \otimes \mathrm{d}x_\alpha^i.$$

Let $\{\rho_\alpha\}$ be a partition of unity subordinate the chosen covering $\{U_\alpha\}$. We define

$$g = \sum_\alpha \rho_\alpha g_\alpha.$$

Note that this is in fact a finite sum in the neighborhood of each point. It is positive definite since for any $p \in M$, there always exist some α such that $\rho(\alpha)(p) > 0$. So it is a Riemannian metric on M. □

The Second Proof. According to the famous Whitney embedding theorem, any smooth manifold M of dimension m can be embedded into $\mathbb{R}^{2m+1}$ as a smooth submanifold, and thus each Riemannian metric on $\mathbb{R}^{2m+1}$ will induce a Riemannian submanifold metric on M. □

Remark. One may ask: How large is the space of all Riemannian metrics on a given smooth manifold? Let

$$\mathrm{Riem}(M) = \{g \mid g \text{ is a Riemannian metric on } M\}$$

be the set of all Riemannian metrics on M. Motivated by the first proof, it is easy to see that if g_1, g_2 are two Riemannian metrics on M, so is $ag_1 + bg_2$ for a, $b > 0$. As a consequence, $\mathrm{Riem}(M)$ (as a subset in the finite dimensional vector space of all symmetric $(0,2)$-tensor fields on M) is a positive convex cone.

Of course a natural question is: Given a manifold, can one find a Riemannian metric that is "best" in some sense? This is one of the main targets in Riemannian geometry. In this course we shall define various kind of invariants (curvatures) of Riemannian metrics, and we shall study the relations between these invariants and the topology of the underlying manifold.

Remark. In the second proof we used the Whitney embedding theorem. For Riemannian manifolds (M,g), there is a much stronger embedding theorem proved by the famous Nobel prize (in Economics) winner John Nash in 1956,

Theorem 1.12 (Nash Embedding Theorem)**.** Any m-dimensional Riemannian manifold (M,g) can be isometrically embedded into the standard $(\mathbb{R}^N, g_0)$ as a Riemannian submanifold, where

$$N = \begin{cases} \dfrac{m(3m+1)}{2}, & \text{if } M \text{ is compact,} \\ \dfrac{m(m+1)(3m+11)}{2}, & \text{if } M \text{ is noncompact.} \end{cases}$$

For compact manifolds, the dimension N was lowered by Gromov in 1986 to

$$N = \frac{m^2+5m+6}{2},$$

and then was further lowered by Günther in 1989 to

$$N = \max\left\{\frac{m^2+5m}{2}, \frac{m^2+3m+10}{2}\right\}.$$

In particular, any 2-dimensional smooth Riemannian manifold can be isometrically embedded into $\mathbb{R}^{10}$ (instead of $\mathbb{R}^{17}$ by Nash). It is still not known whether this dimension can be further lowered[8].

1.3 The Riemannian Distance

1.3.1 Length of Curves

As we mentioned, the Riemannian metric g on M itself is not a metric on M. Nevertheless, it can be viewed as an "infinitesimal metric" and it generates a true metric on M which makes M a metric space. To define this metric, we first define the length of a (piecewise smooth) curve in (M, g).

The Length of a Curve

Let $\gamma : [a, b] \to M$ be a smooth immersed **parametric** curve in M. For any t,

$$\dot{\gamma}(t) = \mathrm{d}\gamma\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)$$

is a tangent vector in $T_{\gamma(t)}M$, and thus has a length with respect to g. We shall always assume that the parametrization is **regular**, i.e. $\dot{\gamma}(t) \neq 0$ for all t. As in undergraduate differential geometry course, it is natural to define

Definition 1.13. The **length** of γ is

$$\mathrm{Length}(\gamma) := \int_a^b \|\dot{\gamma}(t)\|_{\gamma(t)} = \int_a^b \sqrt{g_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))}\mathrm{d}t$$

As one can imagine, the length should be a property of the geometric curve and thus should be independent of the choice of different parametrizations:

Lemma 1.14. Length(γ) is independent of the choices of regular parametrizations.

Proof. Let $\gamma_1 : [c, d] \to M$ be another regular parametrization of the same geometric curve as γ. Then there exists a smooth function $t_1 : [a, b] \to [c, d]$ so that $\gamma_1(t_1(t)) = \gamma(t)$. It follows

$$\dot{\gamma}(t) = \dot{\gamma}_1(t_1)\frac{\mathrm{d}t_1}{\mathrm{d}t}(t).$$

Since both parametrizations are regular, we get $\frac{\mathrm{d}t_1}{\mathrm{d}t}(t) \neq 0$ and thus the function $t_1 = t_1(t)$ is either strictly increasing, or strictly decreasing. Now the conclusion follows from the standard change of variable formula,

$$\begin{aligned}\mathrm{Length}(\gamma_1) &= \int_c^d \sqrt{\langle \dot{\gamma}_1(t_1), \dot{\gamma}_1(t_1)\rangle_{\gamma_1(t_1)}}\mathrm{d}t_1 \\ &= \int_a^b \sqrt{\left\langle \dot{\gamma}(t)\left(\frac{\mathrm{d}t_1}{\mathrm{d}t}(t)\right)^{-1}, \dot{\gamma}(t)\left(\frac{\mathrm{d}t_1}{\mathrm{d}t}(t)\right)^{-1}\right\rangle_{\gamma(t)}}\frac{\mathrm{d}t_1}{\mathrm{d}t}(t)\mathrm{d}t = \mathrm{Length}(\gamma),\end{aligned}$$

where for simplicity we assumed t_1 is strictly increasing with respect to t. □

By the same change of variable argument one can prove

Lemma 1.15. Let $\varphi : (M, g_M) \to (N, g_N)$ be a local isometry, and γ be a regular parametric curve in M. Then

$$\mathrm{Length}_M(\gamma) = \mathrm{Length}_N(\varphi(\gamma)).$$

[8] However, if instead of $\mathcal{C}^\infty$ maps (or $\mathcal{C}^r$-maps for $r \geqslant 3$), one only require a "$\mathcal{C}^1$-isometries", i.e. $\mathcal{C}^1$-diffeomorphism $\varphi : M \to N$ with $\varphi^* g_N = g_M$, then Nash showed in 1955 that any (M^m, g) ca be $\mathcal{C}^1$-isometrically embedded into $\mathbb{R}^{2m+1}$. Of course "$\mathcal{C}^1$-isometries" are not natural objects in Riemannian geometry, because as we have seen last time, basic Riemannian geometric quantities like curvatures need third order derivatives of the embedding (second order derivatives of Riemannian metric).

Arc Length Parametrization

Although there are lots of different ways to parameterize a curve, there is one parametrization, the arc-length parametrization defined below, that is best and thus will be extensively used in this course.

To describe the arc-length parametrization, we start with any regular parametric curve $\gamma : [a, b] \to M$. We will call

$$s(t) = \int_a^t \sqrt{\langle \dot\gamma(\tau), \dot\gamma(\tau)\rangle_{\gamma(\tau)}}\mathrm{d}\tau$$

the **arc-length function** of γ. Obviously $s = s(t)$ is a strictly increasing function mapping the interval $[a, b]$ to the interval $[0, \mathrm{Length}(\gamma)]$. We will denote by $t = t(s)$ the inverse function of $s = s(t)$. One can reparameterize γ via the parameter s,

$$\gamma_1(s) = \gamma(t(s)), \qquad 0 \leqslant s \leqslant \mathrm{Length}(\gamma).$$

This is called the **arc-length parametrization**. It has the following nice property

Proposition 1.16. For the arc-length parametrization, $\langle \dot\gamma_1(s), \dot\gamma_1(s)\rangle_{\gamma_1(s)} \equiv 1$.

Proof. At $s = s(t)$, we have $t'(s) = (s'(t))^{-1} = \langle \dot\gamma(t), \dot\gamma(t)\rangle^{-\frac{1}{2}}$. So

$$\langle \dot\gamma_1(s), \dot\gamma_1(s)\rangle_{\gamma_1(s)} = \langle \dot\gamma(t)t'(s), \dot\gamma(t)t'(s)\rangle_{\gamma_1(s(t))} = t'(s)^2\langle \dot\gamma(t), \dot\gamma(t)\rangle_{\gamma(t)} = 1.$$

□

Conversely by the definition of the arc-length function $s(t)$ above, we see that if a parametrization of γ satisfies $\langle \dot\gamma, \dot\gamma\rangle \equiv 1$, then it is simply a translation of the arc-length parametrization. Usually a curve with $\|\dot\gamma\| \equiv 1$ is called a **normal curve**, and a curve with $\|\dot\gamma\| \equiv c$ is called a **curve of constant speed**.

Now suppose (M, g) is a 1-dimensional Riemannian manifold. Then locally near each point, a local coordinate neighborhood is a curve, and the arc-length parameter s is a local coordinate. By definition, $g_{11} = g(\partial_s, \partial_s) = 1$, and thus locally

$$g = \mathrm{d}s \otimes \mathrm{d}s.$$

As an immediate consequence, we get

Corollary 1.17. Any two 1-dimensional Riemannian manifolds are locally isometric.

Remark. Discussions above can be easily extended to piecewise $\mathcal{C}^1$ curves in M: If $\gamma : [a, b] \to M$ is a continuous map, and there exists a partition

$$a = a_0 < a_1 < a_2 < \cdots < a_N = b$$

of $[a, b]$ so that each $\gamma|_{[a_i,a_{i+1}]}$ is a regular $\mathcal{C}^1$ curve, then one can simply define

$$\mathrm{Length}(\gamma) = \sum_{i=0}^{N-1} \mathrm{Length}(\gamma|_{[a_i,a_{i+1}]}).$$

1.3.2 The Riemannian Distance

Now we are ready define a metric (i.e. a distance function) on a connected Riemannian manifold (M, d). For any p, $q \in M$, let

$$\mathcal{C}_{pq} = \{\gamma : [a, b] \to M \mid \gamma \text{ is piecewise smooth and } \gamma(a) = p, \gamma(b) = q\}.$$

Note that the connectedness of M guarantees that the set $\mathcal{C}_{pq}$ is nonempty.

Definition 1.18. The **distance** between p and q on a Riemannian manifold (M, g) is defined to be

$$\mathrm{dist}(p, q) = \inf\{\mathrm{Length}(\gamma) \mid \gamma \in \mathcal{C}_{pq}\}.$$

Obviously for the standard Euclidean space $(\mathbb{R}^m, g_0)$, dist is the Euclidean distance function on $\mathbb{R}^m$.

Let's give another example:

Example. Consider the hyperbolic plane $(\mathbb{H}, g_{\text{hyp}})$, where $g_{\text{hyp}} = \frac{1}{y^2}(\mathrm{d}x \otimes \mathrm{d}x + \mathrm{d}y \otimes \mathrm{d}y)$. Let $p = (0, a)$, and $q = (0, b)$, where $b > a$. Let $\gamma : [0, 1] \to \mathbb{H}$ be a regular curve with $\gamma(0) = p$ and $\gamma(1) = q$. Denote $\gamma(t) = (x(t), y(t))$, then $y(t) > 0$ and

$$L(\gamma) = \int_0^1 \sqrt{\frac{1}{y(t)^2}(x'(t)^2 + y'(t)^2)}\mathrm{d}t \geqslant \int_0^1 \frac{y'(t)}{y(t)}\mathrm{d}t = \log\frac{b}{a},$$

and the equality achieves if $x(t) = 0$ and y is monotonely increasing. It follows

$$\mathrm{dist}(p, q) = \log\frac{b}{a}.$$

In what follows, we will show that d is a metric, and the metric topology coincides with the given manifold topology. The crucial idea behind the proofs is a comparison between the Riemannian metric on a compact set with the Euclidean metric g_0.

The Riemannian Distance is a Distance

Now we prove

Theorem 1.19. For any connected Riemannian manifold (M, g), the distance function dist makes M into a metric space.

Proof. It is easy to check that the function

$$\mathrm{dist} : M \times M \to \mathbb{R}$$

satisfies most axioms for a distance function, e.g. for any $p, q, r \in M$,

(a) $\mathrm{dist}(p, p) = 0$, $\mathrm{dist}(p, q) \geqslant 0$.

(b) $\mathrm{dist}(p, q) = \mathrm{dist}(q, p)$.

(c) $\mathrm{dist}(p, r) \leqslant \mathrm{dist}(p, q) + \mathrm{dist}(q, r)$.

It remains to show that for $p \neq q$, we must have

$$\mathrm{dist}(p, q) > 0.$$

Take a chart (φ, U, V) around q with[9] $p \notin U$, so that

$$\varphi(q) = 0 \in V = B_1(0) \subset \mathbb{R}^m.$$

Then

$$h = (\varphi^{-1})^*(g_U)$$

is a Riemannian metric on V so that (V, h) is isometric to $(U, g|_U)$. Let

$$\lambda = \inf\{\text{the smallest eigenvalue of the matrix } (h_{ij})_x \mid x \in \overline{B_{1/2}(0)}\}.$$

Then for any $x \in \overline{B_{1/2}(0)}$ and any $X \in T_xV$, we have

$$\langle X, X\rangle_h = h_{ij}X^iX^j \geqslant \sum_i \lambda(X^i)^2 = \lambda\langle X, X\rangle_{g_0}.$$

[9]Here we used that any manifold is Hausdorff. For a non-Hausdorff locally Euclidean space like "the line with two origins", the two origins would have distance zero if we define a "Riemannian distance" as above.

For any piecewise smooth curve γ starting from $0 = \varphi(q)$ and ending at some point on $\partial B_{1/2}(0)$, and $\tilde{\gamma}$ be the first portion of the curve γ that sits totally in $B_{1/2}(0)$ (so $\tilde{\gamma}$ is a piecewise smooth curve that starts at q and still ends at some point on $\partial B_{1/2}(0)$), reparameterized with parameters in $[0,1]$. Then

$$\begin{aligned}\mathrm{Length}_h(\gamma) \geqslant \mathrm{Length}_h(\tilde{\gamma}) &= \int_0^1 \sqrt{\langle \dot{\tilde{\gamma}}, \dot{\tilde{\gamma}} \rangle_h} \mathrm{d}t \\ &\geqslant \sqrt{\lambda} \int_0^1 \sqrt{\langle \dot{\tilde{\gamma}}, \dot{\tilde{\gamma}} \rangle_{g_0}} \mathrm{d}t = \sqrt{\lambda} = \mathrm{Length}_{g_0}(\tilde{\gamma}) \geqslant \frac{\sqrt{\lambda}}{2}.\end{aligned}$$

Since any curve from p to q must intersect $\varphi(\partial B_{1/2}(0))$ at some point, we conclude

$$\mathrm{dist}(p,q) \geqslant \frac{\sqrt{\lambda}}{2} > 0,$$

as desired. □

Remark. Obviously if $\varphi : (M, g_M) \to (N, g_N)$ is an isometry, and d_M, d_N are the corresponding distance functions, then $\varphi : (M, d_M) \to (N, d_N)$ is distance preserving in the sense that

$$\mathrm{dist}_N(\varphi(p), \varphi(q)) = \mathrm{dist}_M(p,q), \qquad \forall p, q \in M.$$

It turns out that the converse is also true, again proven by Myers and Steenrod in 1939:

Theorem 1.20 (Myers-Steenrod)**.** Let (M, g_M) and (N, g_N) be Riemannian manifolds, and dist_M, dist_N be the corresponding distance functions. If $\varphi : (M, \mathrm{dist}_M) \to (N, \mathrm{dist}_N)$ is surjective and distance-preserving, then it is an isometry (and in particular it is smooth).

In other words, the concept "isometry" in the world of Riemannian geometry coincides with the concept "isometry" in the world of metric spaces. Their proof was simplified and the theorem was strengthened by Palais in 1959 to

Theorem 1.21 (Palais)**.** The Riemannian distance of a Riemannian manifold determines its structure as a manifold (the smooth structure) and its Riemannian metric.

Remark. One may further ask the following question:

Given a metric d on a smooth manifold M, is it true that d can be realized as the Riemannian distance for some Riemannian metric g?

The answer is no: Consider the taxicab metric on $\mathbb{R}^2$. Then any two points can be connected by infinitely many "shortest curves". On the other hand, as we will see later, on a Riemannian manifold, near any point there is a neighborhood in which any point can be connected to the given point by a unique shortest curve.

Continuity of the Distance Function

We may further study the topology generated by the metric dist. First we prove

Proposition 1.22. For any fixed p, the function

$$f(\cdot) = \mathrm{dist}(\cdot, p)$$

is continuous on M (with respect to the manifold topology).

Proof. Since manifolds are second countable, it is enough to prove sequential continuity of f. As in the proof of the previous theorem we take a coordinate patch (φ, U, V) centered at q with $\varphi(U) = V = B_1(0) \subset \mathbb{R}^m$. Let q_i be a sequence of points that tends to q with respect to the manifold topology, i.e. for any k, there exists $N(k)$ such that for all $i \geqslant N(k)$, $\varphi(q_i) \in B_{1/k}(0)$. We want to prove $f(q_i) \to f(q)$. By triangle inequality (see (c) above), we have

$$|f(q_i) - f(q)| \leqslant \mathrm{dist}(q, q_i).$$

So it suffices to prove $\mathrm{dist}(q, q_i) \to 0$ as $i \to \infty$.

Again we let $h = (\varphi^{-1})^*(g|_U)$ be the induced metric on V. Denote

$$\Lambda = \sup\{\text{the greatest eigenvalue of the matrix } (h_{ij})_x \mid x \in \overline{B_{1/2}(0)}\}.$$

Then for any $x \in \overline{B_{1/2}(0)}$ and any $X \in T_xV$, we have

$$\langle X, X\rangle_h \leqslant \Lambda \langle X, X\rangle_{g_0}.$$

So if we take

$$\tilde{\gamma}_i : [0,1] \to V,$$
$$\tilde{\gamma}_i(t) = t\varphi(q_i)$$

be the "straight line segment" from $0 = \varphi(q)$ to $\varphi(q_i)$, then for $i \geqslant N(k)$,

$$\mathrm{Length}_h(\tilde{\gamma}_i) = \int_0^1 \sqrt{\langle \dot{\tilde{\gamma}}_i, \dot{\tilde{\gamma}}_i\rangle_h} dt \leqslant \sqrt{\Lambda} \int_0^1 \sqrt{\langle \dot{\tilde{\gamma}}_i, \dot{\tilde{\gamma}}_i\rangle_{g_0}} dt = \sqrt{\Lambda}\,\mathrm{Length}_{g_0}(\tilde{\gamma}_i) \leqslant \frac{\sqrt{\Lambda}}{k}.$$

Since $\varphi : (U, g) \to (V, h)$ is an isometry, we conclude that

$$\mathrm{dist}(q, q_i) \leqslant \mathrm{Length}_g(\varphi^{-1} \circ \tilde{\gamma}_i) = \mathrm{Length}_h(\tilde{\gamma}_i) \leqslant \sqrt{\Lambda}/k,$$

for all $i \geqslant N(k)$. This completes the proof. □

Remark. The distance functions $\mathrm{dist}(\cdot, p)$ are among the most important functions (maybe the only natural geometric functions) on a Riemannian manifold. Note the triangle inequality implies that the "distance to p" function $f(q) = \mathrm{dist}(p, q)$ is not only continuous, but also Lipschitz continuous (with respect to the distance). One may further ask: are they smooth? Even in the Euclidean case, it is obvious that $\mathrm{dist}(\cdot, p)$ is not smooth at $x = p$ (this singularity can be eliminated by considering f^2). On the other hand, as we will see later, this function is smooth in a "punctured neighborhood" $U - \{p\}$ near p, but might have singularities at other points.

Example. Consider the round sphere (S^2, g_{round}). Let $p = (0, 0, -1)$ be the south pole. It is easy to see that for $q = (x, y, z)$,

$$f(q) = d(p, q) = \pi(1 - \arccos z) = \pi(1 - \arccos \sqrt{1 - x^2 - y^2})$$

which is not smooth at both the south pole and the north pole. [In fact for any compact Riemannian manifold, the function $d_p(\cdot) = \mathrm{dist}(p, \cdot)$ is not smooth at some point $q \neq p$. Since the restriction of Euclidean distance function $d_{\text{Euc}}(p, \cdot)$ to S^2 is smooth on $S^2 - \{p\}$, we conclude that the restriction of the Euclidean distance function to S^2 is NOT the Riemannian distance for any Riemannian metric on S^2!]

The Metric Topology

As a consequence of the continuity, we prove

Corollary 1.23. The metric topology on M induced by the metric dist coincides with the manifold topology on M.

Proof. The continuity of $f(\cdot) = \mathrm{dist}(p, \cdot)$ implies that any metric open ball is also open in the manifold topology. Conversely, for any (manifold) open neighborhood U of q in M, by shrinking U we may assume U is a coordinate neighborhood and $(\varphi, U, V = B_1(0))$ is a coordinate chart. Repeat the proof of Theorem 1.19, we see any point $p \notin U$ must have distance $d(p, q) \geqslant \sqrt{\lambda}/2$, which is equivalent to say that the metric open ball of radius $\sqrt{\lambda}/2$ sits in U. So the two topologies on M coincide. □

We will call the metric ball

$$B(q, r) = \{p \in M \mid \mathrm{dist}(p, q) < r\}$$

a **geodesic ball** of radius r centered at q. For r small it has the topology of an Euclidean ball, while for r large it may have very complicated topology.

1.4 The Riemannian Measure

1.4.1 The Riemannian Measure

The Riemannian Volume in Tangent Space

Not only a Riemannian metric g (as an infinitesimal distance, i.e. a distance defined in each tangent space) on M gives rise to a canonical metric structure on M, but also it defines a canonical measure structure (or to be more precise, a volume density) on M through an "infinitesimal volume" (i.e. volume defined in each tangent space). The idea is standard: as in multi-variable calculus, to define the volume or integrate a function over M, one simply start with a coordinate chart, using which one can divide M into small coordinate pieces, and then approximate each small piece $\{(x^1, \ldots, x^m) \mid a^i \leqslant x^i \leqslant a^i + h^i\}$ by the parallelepiped in T_pM (where $p = (x^1, \ldots, x^m)$) generated by $(h^1\partial_1, \ldots, h^m\partial_m)$.

Now the problem is reduces to: how do we define a volume of a parallelepiped in a finite dimensional inner product space? Well, one can always define the volume of a unit cube to be 1 (here we use not only the lengths of vectors, but also the angles between vectors), and then use multi-linearity to extend the definition to more general parallelepipeds. So to compute the volume of the parallelepiped generated by $\partial_1, \partial_2, \ldots, \partial_m$, we start with an orthonormal basis $e_1, \ldots, e_m$ of (T_pM, g_p), and define the volume of the parallelotope generated by $e_1, \ldots, e_m$ to be

$$V_p(e_1, e_2, \ldots, e_m) = 1.$$

Then we write $\partial_i = a_i^j e_j$, which implies

$$V_p(\partial_1, \partial_2, \ldots, \partial_m) = |\det(a_i^j)|.$$

For simplicity we denote $A = (a_i^j)$. From the observation

$$g_{ij} = g(\partial_i, \partial_j) = g(a_i^k e_k, a_j^l e_l) = \sum_k a_i^k a_j^k = (AA^{\mathrm{T}})_{ij},$$

We conclude $(g_{ij}) = AA^{\mathrm{T}}$, and thus the "infinitesimal volume" we are calculating is

$$V_p(\partial_1, \partial_2, \ldots, \partial_m) = |\det(a_i^j)| = \sqrt{G},$$

where $G = \det(g_{ij})$.

Remark. Alternatively, one can define $V_p(\partial_1, \partial_2, \ldots, \partial_m)$ as "the length of the vector $\partial_1 \wedge \partial_2 \wedge \cdots \wedge \partial_m$ in the space $\otimes^m T_pM$" (with respect to the induced metric on tensors that we introduced in Section 1.2), and similar computation yields the same result.

Integrals of Compactly Supported Continuous Functions

Now let (M, g) be a Riemannian manifold. We start with a continuous function f with compact support, so that $\mathrm{supp}(f)$ is contained in one chart (φ, U, V). As motivated by the previous computation, we may define

$$\int_M f \mathrm{d}V_g := \int_V (f\sqrt{G}) \circ \varphi^{-1} \mathrm{d}x^1 \cdots \mathrm{d}x^m,$$

where $\mathrm{d}x^1 \cdots \mathrm{d}x^m$ the Lebesgue measure on $\mathbb{R}^m$.

Lemma 1.24. The definition above is independent of the choices of coordinate charts containing $\mathrm{supp}(f)$.

Proof. Let $(\tilde{\varphi}, \tilde{U}, \tilde{V})$ be another coordinate chart containing $\mathrm{supp}(f)$, on which the coordinates are denoted by $y^1, \ldots, y^m$. Then as we have seen in Section 1.2,

$$(g_{ij}) = J^{\mathrm{T}}(\tilde{g}_{kl})J,$$

where $J = \left(\dfrac{\partial y^i}{\partial x^j}\right)$ is the Jacobian of the map $\tilde{\varphi}\circ\varphi^{-1}$. As a consequence, we get

$$\sqrt{G(p)} = \sqrt{\tilde{G}(p)}|\det(J(\varphi(p)))|$$

for $p = \varphi^{-1}(x) = \tilde{\varphi}^{-1}(y)$, and thus by change of variables in $\mathbb{R}^m$,

$$\sqrt{\tilde{G}\circ\tilde{\varphi}^{-1}}\mathrm{d}y^1\cdots\mathrm{d}y^m = \sqrt{\tilde{G}\circ\tilde{\varphi}^{-1}(\tilde{\varphi}\circ\varphi^{-1})}|\det(J)|\mathrm{d}x^1\cdots\mathrm{d}x^m = \sqrt{G\circ\varphi^{-1}}\mathrm{d}x^1\cdots\mathrm{d}x^m.$$

The conclusion follows. □

Of course in general, even if f is compactly supported, one cannot assume that $\mathrm{supp}(f)$ is contained in one single chart. However, one can extend the above definition to general $f\in\mathcal{C}_c(M)$ easily by using partition of unity: Let $\{(\varphi_\alpha, U_\alpha, V_\alpha)\}$ be a system of locally finite coordinate charts that cover M, with local coordinates $\{x^1_\alpha,\ldots,x^m_\alpha\}$ on each U_α, and let $\{\rho_\alpha\}$ be a partition of unity subordinate to the open covering $\{U_\alpha\}$. Then we define

$$\int_M f\mathrm{d}V_g := \sum_\alpha \int_{\varphi_\alpha(U_\alpha)} (f\rho_\alpha\sqrt{G^\alpha})\circ(\varphi_\alpha)^{-1}\mathrm{d}x^1_\alpha\cdots\mathrm{d}x^m_\alpha,$$

Note that by locally finiteness of U_α and compactness of $\mathrm{supp}(f)$, the sum is in fact a finite sum. Moreover, if $\{(\tilde{\varphi}_\beta, \tilde{U}_\beta, \tilde{V}_\beta)\}$ is another atlas, then by Lemma 1.24,

$$\int_{\varphi_\alpha(U_\alpha\cap\tilde{U}_\beta)} (f\rho_\alpha\tilde{\rho}_\beta\sqrt{G^\alpha})\circ(\varphi_\alpha)^{-1}\mathrm{d}x^1_\alpha\cdots\mathrm{d}x^m_\alpha = \int_{\tilde{\varphi}_\beta(U_\alpha\cap\tilde{U}_\beta)} (f\rho_\alpha\tilde{\rho}_\beta\sqrt{G^\beta})\circ(\varphi_\beta)^{-1}\mathrm{d}x^1_\beta\cdots\mathrm{d}x^m_\beta$$

since both sides equal to $\int_M \rho_\alpha\tilde{\rho}_\beta f\mathrm{d}V_g$, which implies

$$\sum_\alpha \int_{\varphi_\alpha(U_\alpha)} (f\rho_\alpha\sqrt{G^\alpha})\circ(\varphi_\alpha)^{-1}\mathrm{d}x^1_\alpha\cdots\mathrm{d}x^m_\alpha = \sum_\beta \int_{\varphi_\beta(U_\beta)} (f\rho_\beta\sqrt{G^\beta})\circ(\varphi_\beta)^{-1}\mathrm{d}x^1_\beta\cdots\mathrm{d}x^m_\beta.$$

In other words, $\int_M f\mathrm{d}V_g$ is well-defined for any $f\in\mathcal{C}_c(M)$.

The Riemannian Measure

Since manifolds are always locally compact and Hausdorff, and since the linear functional

$$\mu:\mathcal{C}_c(M)\to\mathbb{R},$$
$$f\mapsto\mu(f) = \int_M f\mathrm{d}V_g$$

is positive (i.e. $f\geqslant 0$ implies $\mu(f)\geqslant 0$), by Riesz representation theorem, μ gives rise to a unique Radon measure on M. Now one can further extend the integral to more general functions using the standard machinery developed in real analysis:

- first define the (upper) integral of a lower semi-continuous positive function f to be the supremum of integrals of compactly-supported functions that are no more than f,
- then define the (upper) integral of a positive function f as the infimum of the (upper) integral of all lower semi-continuous positive function f that are greater than f,
- a function f is said to be **integrable** if there exists a sequence g_n in $\mathcal{C}_c(M)$ so that the (upper) integrals of the sequence $|g_n - f|$ converge to 0.

As usual we denote the space of integrable functions as $L^1(M,g)$, which by definition is the completion of $\mathcal{C}_c(M)$ with respect to suitable norm.

As usual, for any $1 \leqslant p < \infty$ one can define the L^p norm on $\mathcal{C}_c^\infty$ via

$$\|f\|_{L^p} := \left(\int_M |f|^p \mathrm{d}V_g\right)^{\frac{1}{p}},$$

and define $L^p(M,g)$ to be the completion of $\mathcal{C}_c^\infty$ under the L^p norm. Similarly one can define $L^\infty(M,g)$. It is not hard to extend the theory to complex-valued functions. In the special case $p=2$, one can define an inner product structure on $L^2(M,g)$ by

$$\langle f_1, f_2 \rangle_{L^2} := \int_M f_1 \overline{f}_2 \mathrm{d}V_g$$

which make $L^2(M,g)$ into a Hilbert space.

One can also talk about the volume of any Borel set (or more generally, measurable subsets) A in M, which is defined to be

$$\mathrm{Vol}(A) = \int_M \chi_A \mathrm{d}V_g$$

Remark. In the above definition, we don't assume M to be oriented or compact. What we really get is a volume density, which, on a local chart, can be written as

$$\mathrm{d}V_g = \sqrt{G} \circ \varphi^{-1} \mathrm{d}x^1 \cdots \mathrm{d}x^m.$$

We will call dVol the **Riemannian volume element** (or **volume density**) on (M,g).

Remark. In the special case where M is oriented, then we may choose an orientation-compatible coordinate patch near each point, and define (locally on each chart)

$$\omega_g = \sqrt{G} \mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^m.$$

One can check that ω_g is a well-defined global volume form on M, which is called the **Riemannian volume form** for the oriented Riemannian manifold (M,g).

Remark. Suppose (M,g) is an m-dimensional Riemannian manifold, and S an r-dimensional submanifold of M, where $r < m$. Then the Riemannian submanifold metric $g_S := \iota^* g$ on M gives a natural measure (an r-dimensional volume density) on S. Here are two special cases:

- If $\gamma : I \to M$ is a simple smooth curve, then with respect to the coordinates t (from the parametrization), we have $g_\gamma = g(\partial_t, \partial_t)\mathrm{d}t \otimes \mathrm{d}t = |\dot\gamma(t)|^2 \mathrm{d}t \otimes \mathrm{d}t$, and thus the induced 1-dimensional volume density (i.e. **length density**) on γ is simply $|\dot\gamma|\mathrm{d}t$, which is exactly what we used to calculate the length of γ.
- If M is a smooth manifold with boundary, in which case the boundary ∂M is a smooth submanifold of dimension $m-1$, then one gets a natural Riemannian submanifold metric and thus a volume density on ∂M. In this case the volume density on ∂M is usually called a **surface density** (or **hypersurface density**) and will be denoted by $\mathrm{d}S_g$.

The Change of Variable Formula

By using the standard change of variable formula for the Lebesgue measure in $\mathbb{R}^m$, together with a partition of unity argument, one can easily prove the following

Proposition 1.25 (Change of Variables in Riemannian Setting)**.** Let $\varphi : M \to N$ be a diffeomorphism, and h a Riemannian metric on N. Then

$$\int_M f \circ \varphi \, \mathrm{d}V_{\varphi^* h} = \int_N f \mathrm{d}V_h, \quad \forall f \in L^1(N,h).$$

In particular, we see isometries preserve the Riemannian volume densities.

As another consequence, suppose $\dim M \leqslant \dim N$, $\varphi : M \to N$ is an embedding, and $\iota : \varphi(M) \to N$ is the inclusion map. Let g be a Riemannian metric on M and h be a Riemannian metric on N, then[10]

$$\int_M f \circ \varphi \frac{\mathrm{d}V_{\varphi^* h}}{\mathrm{d}V_g} \mathrm{d}V_g = \int_{\varphi(M)} f \, \mathrm{d}V_{\iota^* h}, \quad \forall f \in L^1(N, h),$$

where $\frac{\mathrm{d}V_{\varphi^* h}}{\mathrm{d}V_g}$ is the Radon-Nikodyn derivative of the two corresponding Riemannian measures on M (which are by definition σ-finite measures). In particular, one may find the area of $\varphi(M)$ (or integrals over $\varphi(M)$) in the target space by doing computations in the source space M. It is a very special case of the so-called **area formula** in geometric measure theory where φ is only supposed to be Lipschitz and need not be injective, and the measures encountered are replaced by the Hausdorff measure.

There is a "dual" version of the area formula above, known as the **co-area formula**, in which, with the help of a map $\varphi : M \to N$ with $\dim M \geqslant \dim N$, one could use integrals over level sets $\varphi^{-1}(q)$ in target space to compute integrals over the source space M. In the very general version of co-area formula in geometric measure theory, φ is only supposed to be a Lipschitz map, and people use the Hausdorff measures. In what follows we will prove a simplest version of co-area formula (for $N = \mathbb{R}$) that is already very useful in Riemannian geometry. To state the theorem, we need the concept of gradient vector fields associated to a function.

The Gradient

Let (M, g) be a Riemannian manifold. For any smooth function f on M, the differential $\mathrm{d}f$ is a smooth 1-form on M. By using the musical isomorphism $\sharp : T^*M \to TM$, we will get a smooth vector field on M:

Definition 1.26. The **gradient vector field** of f is $\nabla f = \sharp(\mathrm{d}f)$.

It is not hard to find out ∇f in local charts: By definition, ∇f is the vector field so that for any vector field $X = X^i \partial_i$,

$$g(\nabla f, X) = \mathrm{d}f(X) = Xf = X^i \partial_i f.$$

It follows that locally

$$\nabla f = g^{ij} \partial_i f \partial_j.$$

In particular, for $g = g_0$ in $\mathbb{R}^m$, we get the ordinary gradient of f.

As in multivariable calculus, the gradient vector field of a function is always perpendicular to its regular level sets:

Lemma 1.27. Suppose f is a smooth function on M and c is regular value of f. Then the gradient vector field ∇f is perpendicular to the level set $f^{-1}(t)$.

Proof. Since c is a regular value, by the regular level set theorem, $f^{-1}(c)$ is a smooth submanifold of M. Let X be a vector field tangent to $f^{-1}(c)$. Then we learned from manifold theory that $Xf = 0$ on $f^{-1}(c)$. It follows

$$g(\nabla f, X) = Xf = 0$$

on $f^{-1}(c)$. So ∇f is perpendicular to $f^{-1}(c)$. □

The Coarea Formula: a Simple Version

Fix a smooth function $u \in \mathcal{C}^\infty(M)$ and let

$$\Omega_t := u^{-1}((-\infty, t)), \quad \Gamma_t := u^{-1}(t).$$

[10] Note that it makes no sense to write an expression like $\int_M f \circ \varphi |\det \mathrm{d}\varphi| \mathrm{d}V_g$ even if φ is a diffeomorphism, since $\mathrm{d}\varphi$ is a linear map between different vector spaces.

For any regular value t of u, Γ_t is a smooth submanifold of dimension $m-1$ in M. By Sard's theorem, critical values of u form a measure zero set in $\mathbb{R}$ (and thus can be ignored in the integration $\int_{\mathbb{R}}$ below). Now we can prove

Theorem 1.28 (The Co-area Formula, a Simple Version)**.** Let (M,g) be a Riemannian manifold. For any regular value t of u, let g_t be the induced Riemannian metric on Γ_t and denote the corresponding Riemannian volume density on Γ_t by $\mathrm{d}S_t$. Then for any integrable function f on M, one has

$$\int_M f|\nabla u|\,\mathrm{d}V_g = \int_{\mathbb{R}}\left(\int_{\Gamma_t} f\,\mathrm{d}S_t\right)\mathrm{d}t.$$

Proof. First note that if we let C be the set of critical points of u, then C is closed. It follows that $M\setminus C$ is an open submanifold in M, and obviously

$$\int_M f|\nabla u|\,\mathrm{d}V_g = \int_{M\setminus C} f|\nabla u|\,\mathrm{d}V_g.$$

So we may replace M by $M\setminus C$ without changing both sides. In other words, we may assume u admits no critical point on M.

No consider the vector field

$$X = \frac{\nabla u}{|\nabla u|^2}$$

on M. By Lemma 1.27, X is perpendicular to $T_q\Gamma_c$ at any $q\in\Gamma_c$ for any c. Let φ_t be the (local) flow generated by X. Then by definition,

$$\frac{\mathrm{d}}{\mathrm{d}t}u(\varphi_t(q)) = \mathrm{d}u(X(\varphi_t(q))) = \langle\nabla u, X\rangle_{\varphi_t(q)} = 1.$$

It follows that if $q\in\Gamma_c$, then $\varphi_t(q)\in\Gamma_{c+t}$ for t small enough. Now we choose a neighborhood A of q in Γ_c so that

$$\psi : (-\varepsilon,\varepsilon)\times A \to M,$$
$$(y,t)\mapsto \varphi_t(y)$$

is a diffeomorphism onto an open subset $U=\psi((-\varepsilon,\varepsilon)\times A)$ in M. By shrinking A if necessary, we may suppose A is a coordinate patch on Γ_c and let $y^1,\dots,y^{m-1}$ be corresponding coordinate functions. Then $\{t,y^1,\dots,y^{m-1}\}$ form a set of coordinate functions on U. With respect to these coordinates, and in view of the facts $\partial_t = X$ and $X\perp\partial_{y^i}$ for all i, the Riemannian metric g has the form

$$g = \langle X,X\rangle \mathrm{d}t\otimes\mathrm{d}t + h_{ij}\mathrm{d}y^i\otimes\mathrm{d}y^j,$$

where $h_{ij}=g(\partial_{y^i},\partial_{y^j})$. Since $\langle X,X\rangle = \frac{1}{|\nabla u|^2}$, the volume density

$$\mathrm{d}V_g = \frac{1}{|\nabla u|}\sqrt{\det(h_{ij})}\mathrm{d}t\mathrm{d}y^1\cdots\mathrm{d}y^{m-1} = \frac{1}{|\nabla u|}\mathrm{d}t\mathrm{d}S_t.$$

So we conclude that for any $\rho\in\mathcal{C}_c(U)$,

$$\int_M \rho f|\nabla u|\mathrm{d}V_g = \int_U \rho f\sqrt{\det G_t}\mathrm{d}t\mathrm{d}y^1\cdots\mathrm{d}y^{m-1} = \int_{c-\varepsilon}^{c+\varepsilon}\left(\int_{\Gamma_t\cap U}\rho f\mathrm{d}S_t\right)\mathrm{d}t.$$

Now the conclusion follows from a standard partition of unity argument. □

As a corollary, we get

Corollary 1.29. Suppose the critical values of u form a closed subset[11] in $\mathbb{R}$, and $\mathrm{Vol}(\Omega_t)<\infty$, then the function $t\mapsto\mathrm{Vol}(\Omega_t)$ is smooth at regular value t, and

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Vol}(\Omega_t) = \int_{\Gamma_t}\frac{1}{|\nabla u|}\mathrm{d}S_t.$$

[11] This condition holds if u is a proper function.

Proof. For any regular t, take $\varepsilon > 0$ so that $(t, t+\varepsilon)$ is free of critical values. By taking $f = \frac{1}{|\nabla u|}$ we get, for $h \in (0, \varepsilon)$,

$$\mathrm{Vol}(\Omega_{t+h}) - \mathrm{Vol}(\Omega_t) = \int_t^{t+h} \left(\int_{\Gamma_t} \frac{1}{|\nabla u|} \mathrm{d}S_t \right) \mathrm{d}t.$$

It follows

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Vol}(\Omega_t) = \lim_{h\to 0} \frac{1}{h} \int_t^{t+h} \left(\int_{\Gamma_t} \frac{1}{|\nabla u|} \mathrm{d}S_t \right) \mathrm{d}t = \int_{\Gamma_t} \frac{1}{|\nabla u|} \mathrm{d}S_t.$$

□

1.4.2 The Laplace-Beltrami Operator

The Divergence of a Vector Field

Let X be a smooth vector field on M. Take a coordinate patch $(U, x^1, \ldots, x^m)$ (which is of course orientable) on M, then the volume element

$$\omega_g = \sqrt{G}\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^m$$

is locally an n-form on U. Of course one may choose other coordinates on U, then the corresponding volume forms are either the same, or differ by a negative sign. As a result, the following definition is independent of the choice of coordinate charts:

Definition 1.30. The **divergence** of X is the function $\mathrm{div}(X)$ on M such that

$$(\mathrm{div}X)\omega_g = \mathrm{d}(\iota(X)\omega_g).$$

Remark. According to Cartan's magic formula, the definition above is equivalent to

$$\mathcal{L}_X(\omega_g) = \mathrm{div}(X)\omega_g,$$

where $\mathcal{L}_X$ is the Lie derivative along the vector field X. This coincides with the geometric definition of divergence in the case of $\mathbb{R}^m$: the divergence of a vector field is the infinitesimal rate of change of the volume element along the vector field.

Let's calculate $\mathrm{div}(X)$ locally. Let $X = X^i\partial_i$, then

$$\begin{aligned}(\mathrm{div}X)\sqrt{G}\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^m &= \mathrm{d}\left(\iota(X^i\partial_i)\sqrt{G}\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^m\right) \\ &= \mathrm{d}\left(\sum_i X^i\sqrt{G}(-1)^{i-1}\mathrm{d}x^1 \wedge \cdots \wedge \widehat{\mathrm{d}x^i} \wedge \cdots \wedge \mathrm{d}x^m\right) \\ &= \partial_i(X^i\sqrt{G})\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^m,\end{aligned}$$

so we conclude

$$\mathrm{div}(X^i\partial_i) = \frac{1}{\sqrt{G}}\partial_i(X^i\sqrt{G}).$$

We may replace X by fX to get

$$\mathrm{div}(fX) = f\mathrm{div}X + (\partial_i f)X^i = f\mathrm{div}X + g(\nabla f, X).$$

In other words,

Corollary 1.31. For any smooth vector field $X \in \Gamma^\infty(TM)$ and any smooth function $f \in \mathcal{C}^\infty(M)$, one has

$$\mathrm{div}(fX) = f\mathrm{div}X + g(\nabla f, X).$$

As an application, we prove

Theorem 1.32 (The Divergence Theorem I)**.** Let X be a smooth vector field with compact support on a Riemannian manifold (M, g), then

$$\int_M \mathrm{div}(X)\mathrm{d}V_g = 0.$$

Proof. First we assume that X is supported in a local chart (φ, U, V) and $X = X^i\partial_i$ with $X^i \in \mathcal{C}_c^\infty(U)$. Then

$$\begin{aligned}\int_M \mathrm{div}(X)\mathrm{d}V_g &= \int_U \frac{1}{\sqrt{G}}\partial_i(X^i\sqrt{G})\mathrm{d}V_g \\ &= \int_{\varphi(U)} \partial_i(X^i\sqrt{G}\circ\varphi^{-1})\mathrm{d}x^1\cdots\mathrm{d}x^m = 0.\end{aligned}$$

The general case follows from partition of unity and Corollary 1.31:

$$\sum_\alpha \rho_\alpha \mathrm{div}(X) = \sum_\alpha \mathrm{div}(\rho_\alpha X) - g(\nabla(\sum_\alpha \rho_\alpha), X) = \sum_\alpha \mathrm{div}(\rho_\alpha X)$$

and thus

$$\int_M \mathrm{div}(X)\mathrm{d}V_g = \int_M \sum \rho_\alpha \mathrm{div}(X)\mathrm{d}V_g = \int_M \sum_\alpha \mathrm{div}(\rho_\alpha X)\mathrm{d}V_g = 0.$$

□

The Laplace-Beltrami Operator

Let (M, g) be a Riemannian manifold.

Definition 1.33. For any smooth function f, we define the **Laplacian** of f to be

$$\Delta f = -\mathrm{div}(\nabla f).$$

Locally, Δf is given by

$$\Delta f = -\mathrm{div}(g^{ij}\partial_i f\partial_j) = -\frac{1}{\sqrt{G}}\partial_i(\sqrt{G}g^{ij}\partial_j f),$$

i.e.

$$\Delta = -\frac{1}{\sqrt{G}}\partial_i(\sqrt{G}g^{ij}\partial_j).$$

We shall call Δ the **Laplace-Beltrami** operator. It is a second order differential operator on M, and is the most important differential operator on Riemannian manifolds. It plays an essential role on the analysis of Riemannian manifolds.

Theorem 1.34 (Green's Formula I)**.** Suppose f and h are smooth function on M and either f or h is compactly supported. Then

$$\int_M f\Delta h\mathrm{d}V_g = \int_M g(\nabla f, \nabla h)\mathrm{d}V_g = \int_M h\Delta f\mathrm{d}V_g.$$

Proof. We have seen

$$\mathrm{div}(fX) = f\mathrm{div}X + g(\nabla f, X).$$

It follows

$$\mathrm{div}(f\nabla h) = -f\Delta h + g(\nabla f, \nabla h).$$

Now the theorem follows from the fact that $f\nabla h$ is compactly supported. □

In particular if M is compact (without boundary), then any smooth function is compactly supported. Replacing h by $\overline{h}$ if they are complex-valued, we can rewrite the above formula as

$$\langle f, \Delta h\rangle_{L^2} = \langle \Delta f, h\rangle_{L^2}.$$

In other words, we get

Corollary 1.35. If M is compact, then Δ is densely defined symmetric operator on $L^2(M, g)$.

As another immediate consequence, we see that Δ is a positive operator:

Corollary 1.36. If M is compact, then $\langle \Delta f, f \rangle_{L^2} \geqslant 0$.

Remark. Both the divergence theorem and the Green's formula can be generalized to the case where M is a **compact Riemannian manifold with boundary**, i.e. M is

- an m-dimensional smooth manifold with boundary
- M is also a compact subset of an m-dimensional Riemannian manifold N
- The Riemannian structure on M coincide with that of N

So ∂M carries

(1) an outward normal vector field ν

(2) an induced Riemannian metric from g_N, and thus a volume density $\mathrm{d}A$.

Then for any smooth vector field X on M and any smooth functions f, h on M,

- (Divergence Theorem II) $\int_M \mathrm{div} \mathrm{d}V_g = \int_{\partial M} g(X, \nu) \mathrm{d}A$,
- (Green's Formula II) $\int_M f \Delta h \mathrm{d}V_g = \int_M g(\nabla f, \nabla h) \mathrm{d}V_g - \int_{\partial M} g(\nu, \nabla h) f \mathrm{d}A$.

Details will be left as an exercise.

Laplacian v.s. Isometry

Why the operator Δ is so important in Riemannian geometry? Since differential operators are local, it is quite obvious that if $\varphi : (M, g_M) \to (N, g_N)$ is a local isometry, then $\psi^*(\Delta_N f) = \Delta_M(\psi^* f)$. Conversely,

Proposition 1.37. A diffeomorphism $\psi : M \to N$ is an isometry between (M, g_M) and (N, g_N) if and only if it commutes with the Beltrami-Laplace operators, i.e.

$$\psi^*(\Delta_N f) = \Delta_M(\psi^* f), \quad \forall f \in \mathcal{C}^\infty(N).$$

Proof. Obviously if φ is an isometry, then it commutes with the Beltrami-Laplace operators. Conversely, suppose the diffeomorphism ψ commutes with the Beltrami-Laplace operators. Take a chart (φ, U, V) on M so that $(\varphi \circ \psi^{-1}, \psi(U), V)$ is a chart on N. Denote the coordinates by $x^1, \ldots, x^m$ and $y^1, \ldots, y^m$ respectively. Then under these coordinates, for $y = \psi(x)$ we have

$$\partial_i^M (f \circ \psi)(x) = \frac{\partial((f \circ \psi) \circ \varphi^{-1})}{\partial x^i}(\varphi(x)) = \frac{\partial(f \circ (\varphi \circ \psi^{-1})^{-1})}{\partial x^i}(\varphi \circ \psi^{-1}(y)) = \partial_i^N f(y)$$

and thus

$$(\partial_i^N \partial_j^N f) \circ \psi = \partial_i^M \partial_j^M (f \circ \psi).$$

On the other hand, we have

$$\begin{aligned}(\psi^* \Delta_N f)(x) = (\Delta_N f)(\psi(x)) &= -\frac{1}{\sqrt{G_N}} \partial_i^N (\sqrt{G_N} g_N^{ij} \partial_j^N f)(\psi(x)) \\ &= -(g_N^{ij} \partial_i^N \partial_j^N f)(\psi(x)) + \cdots.\end{aligned}$$

and

$$\Delta_M(\psi^* f)(x) = -\frac{1}{\sqrt{G}} \partial_i^M (\sqrt{G} g_M^{ij} \partial_j^M (f \circ \psi))(x) = -(g_M^{ij} \partial_i^M \partial_j^M (f \circ \psi))(x) + \cdots.$$

where $\cdots$ represents terms that involve only first order derivatives of f. So by comparing the coefficients of second order terms, we get $g_M^{ij}(x) = g_N^{ij}(\psi(x))$, as desired. ☐

Chapter 2

Connections

2.1 The Linear Connection

Now we will introduce a new structure on (the tangent bundle of) a smooth manifold: the linear connection structure. Roughly speaking, a linear connection is a structure that can be viewed as an abstraction of the concept "parallel" (which is of course one of the most important concepts in geometry): geometrically it starts by giving us a way to connect or identify different tangent spaces over nearby points. When endowed with a linear connection structure on its tangent bundle, a manifold will look infinitesimally like Euclidean space not just smoothly, but as an affine space. In particular, with a linear connection structure, one can

- transport vectors in a "parallel way" along a curve,
- differentiate vector fields as if they were functions on the manifold with values in a fixed vector space (so that "parallel vector fields" have derivative zero).

It turns out that the structure of defining "parallel transport" on a smooth manifold is equivalent to the structure of defining "covariant derivative".

2.1.1 Linear Connections: Many Faces

There are many different ways to define a linear connection on the tangent bundle. Here we mainly focus on two of them which are most useful for us.

Parallel Transports

From its geometric origin, linear connection was used to characterize parallelism. We start by axiomize the structure of parallelism on smooth manifolds. By staring at the parallelism structure in the Euclidean space, it is not hard to see that the first step should be "identifying tangent spaces in Euclidean space $\mathbb{R}^m$ by translations". On a smooth manifold, instead of transporting/translating a vector along a line in a parallel fashion, we will have to transport a vector along a curve since we no longer have the concept of straight lines in general.

How to transport a vector in a parallel way along a curve? If we transport a vector from one point to another point along different paths, do we get the same result (i.e. path-dependence)? One may start with two simple examples:

- For the standard Euclidean space: the simplest way is to identify all $T_x\mathbb{R}^m$ with $\mathbb{R}^m$ via the global coordinate system, and thus transport a vector from one point to another "without changing its direction".
- For the round sphere: According to the hairy ball theorem, there is no way to find a global coordinate system on S^2. However, we can still try to transport (in some reasonable way) the north-pointing vector at a point on the equator to the north pole. You may try to transport it along the longitude directly to the north pole, or first transport it along the equator to the

opposite point and then transport the resulting vector along the longitude to the north pole. Although we have not specified what do we mean by "parallel transport", by looking at this example with the most natural way of transport, you may convince yourself that "parallel transport" depends on the path.

Of course in these two examples, we really used some kind of Riemannian metric structure to get a geometrically reasonable parallel transport. In fact, the theory of connection/covariant derivative as developed by Ricci, Levi-Civita at the turn of 20th century did require the presence of a Riemannian structure, but it was soon realized by many other mathematicians (including E. Cartan, Schouten and Weyl) that such a structure could be defined abstractly without the presence of a Riemannian metric.

The following definition is complicated (you don't need to memorize this definition since we will mainly use the definition of linear connection via covariant derivative below) but quite natural:

Definition 2.1. A **parallel transport structure** on a smooth manifold M is a family

$$\{P^\gamma : \gamma \text{ is a piecewise smooth curve on } M\}$$

that assigns to each piecewise smooth curve $\gamma : [a, b] \to M$ a linear isomorphism

$$P^\gamma : T_{\gamma(a)}M \to T_{\gamma(b)}M$$

so that

(1) Suppose the start point of γ_2 is the endpoint of γ_1, then $P^{\gamma_1\gamma_2} = P^{\gamma_2} \circ P^{\gamma_1}$, where $\gamma_1\gamma_2$ is the curve formed by "first γ_1, then γ_2".

(2) Denote by $-\gamma$ "the curve γ in the opposite direction", then $P^{-\gamma} = (P^\gamma)^{-1}$.

(3) P^γ depends smoothly on γ in the following sense: Suppose U is an open subset in M, X is smooth vector field on U, $\gamma_u(\cdot)$ is a smooth family of curves (i.e. $\Gamma(u, t) = \gamma_u(t)$ is a smooth map from $U \times [0, 1]$ to M) in M with parameter in U such that $\gamma_u(0) = u$, then the map $U \to TM$ given by $u \mapsto P^{\gamma_u}(X(u))$ is smooth.

(4) If γ_1 and γ_2 are in first order osculating, i.e.

$$\gamma_1(0) = \gamma_2(0), \quad \dot\gamma_1(0) = \dot\gamma_2(0),$$

then for any $X_0 \in T_{\gamma_1(0)}$,

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} P^{\gamma_1}_{0,t}(X_0) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} P^{\gamma_2}_{0,t}(X_0),$$

where $P^\gamma_{t_1,t_2}$ represents the parallel transport along the curve $\gamma([t_1, t_2])$ (so that $P^{\gamma_1}_{0,t}(X_0)$ is a curve in TM starting at the point $(\gamma_1(0), X_0)$.).

Given such a parallel transport structure on M, one may define the directional derivative of a vector field Y along a vector field X. To illustrate, we start with the Euclidean case. Let X and Y be two smooth vector fields on $\mathbb{R}^m$. Then at a point x the directional derivative of Y along X is the limit

$$(D_XY)(x) = \lim_{t\to 0} \frac{Y(x + tX) - Y(x)}{t}, \tag{2.1}$$

where both x and X are viewed as vectors in $\mathbb{R}^m$. On a smooth manifold M with $X, Y \in \Gamma^\infty(TM)$, it makes no sense to talk about $Y(p+tX)$ since $p+tX$ is no longer a point on the manifold. However, in the Euclidean case $\gamma(t) = x + tX$ is simply a curve starting at x in the direction X. So on M, one may replace $x+tX$ by a curve $\gamma(t)$ on M with $\gamma(0) = p$ and $\dot\gamma(0) = X_p$, and consider the limit

$$\lim_{t\to 0} \frac{Y(\gamma(t)) - Y(p)}{t}$$

which, however, is still not well-defined since the vectors $Y(\gamma(t))$ and $Y(p)$ belong to different vector spaces. It is at this step that we may use our parallel transport structure to identify two different tangent spaces! As a result, we may define the directional derivative we are looking for to be

$$D_XY(p) = \lim_{t\to t_0} \frac{(P^{\gamma}_{t_0,t})^{-1}(Y(\gamma(t))) - Y(\gamma(t_0))}{t-t_0},$$

where γ is a smooth curve on M such that $\gamma(t_0) = p$ and $\dot\gamma(t_0) = X_p \in T_pM$. In particular, different choices of parallel transport structure may give us different directional derivatives.

Remark. There is an even abstract way to define the connection/parallel transport structure via suitable choices of "horizontal subspaces at each point". It has the advantage that it can be extended to define connections on general fiber bundles. (One may search the word "Ehresmann connection" on internet to get more details.)

Linear Connections as Directional Derivatives

So by attaching to a smooth manifold an extra (and complicated) parallel transport structure, one can define directional derivative of a vector field along another vector field. It turns out that life will be much easier if, instead of axiomizing the parallelism structure, we start by axiomizing the directional derivative (since it has a much simpler algebraic structure).

To illustrate the structure behind directional derivative, again we start with the Euclidean case, i.e. the formula (2.1). Let $X = X^i\partial_i$ and $Y = Y^j\partial_j$. Then a simple computation yields

$$D_XY = X^i\partial_i(Y^j)\partial_j.$$

In order to generalize this to vector fields on manifolds, one may take closer look of the dependence of D_XY with X and Y. It is not hard to see

- Fixing Y, the map
$$\begin{aligned} DY : \Gamma^\infty(T\mathbb{R}^m) &\to \Gamma^\infty(T\mathbb{R}^m), \\ X &\mapsto D_XY \end{aligned}$$
is $\mathcal{C}^\infty(M)$-linear with respect to X, i.e.
$$D_{fX}Y = fD_XY, \quad \forall f \in \mathcal{C}^\infty(\mathbb{R}^m). \tag{2.2}$$
- Fixing X, the map
$$\begin{aligned} D_X : \Gamma^\infty(T\mathbb{R}^m)Y &\to \Gamma^\infty(T\mathbb{R}^m), \\ Y &\mapsto D_XY \end{aligned} \tag{2.3}$$
satisfies the Leibniz rule
$$D_X(fY) = fD_XY + (Xf)Y, \quad \forall f \in \mathcal{C}^\infty(\mathbb{R}^m).$$

Now let M be any smooth manifold (so no Riemannian structure is assumed). For any smooth vector fields $X, Y \in \Gamma^\infty(TM)$, we would like to study the "directional derivative" of Y along the direction of X. Here are two natural candidates:

- The Lie derivative
$$\mathcal{L}_XY = [X, Y]$$
is not good for this purpose, since it is not $\mathcal{C}^\infty(M)$-linear (tensorial) in X.
- One may embed M into some Euclidean space $\mathbb{R}^N$ as we did in Section 1.1: For any $X, Y \in \Gamma^\infty(TM)$, one can extend them to vector fields $\tilde X$, $\tilde Y$ on $\mathbb{R}^N$, and then define D_XY to be some kind of "projection" of $D_{\tilde X}\tilde Y$ onto the tangent space of M. One can check that what we get satisfies the conditions (2.2) and (2.3) that we want.

So, with only smooth structure at hand, there are lots of different ways (e.g. arising from different embeddings) to define the directional derivative of Y with respect to X.

Note that both (2.2) and (2.3) are natural in defining the "directional derivative" of Y with respect to X: X represents the direction that we want to take "derivative", so one should have pointwise linearity; Y represents the vector field to be differentiated, so one should have a Leibniz law. With the purpose as a guiding, we define

Definition 2.2. A **linear connection** ∇ on a smooth manifold M is a bilinear map

$$\begin{aligned}\nabla : \Gamma^\infty(TM) \times \Gamma^\infty(TM) &\to \Gamma^\infty(TM),\\ (X,Y) &\mapsto \nabla_X Y\end{aligned}$$

such that for any $X, Y \in \Gamma^\infty(TM)$ and any $f \in \mathcal{C}^\infty(M)$,

(1) $\nabla_{fX} Y = f\nabla_X Y$,

(2) $\nabla_X(fY) = f\nabla_X Y + (Xf)Y$.

The vector field $\nabla_X Y$ is called the **covariant derivative** of Y along X.

Example. Let $M = \mathbb{R}^m$. Then the usual directional derivative

$$\nabla_X Y = \nabla_{X^i\partial_i}(Y^j\partial_j) = X^i\partial_i(Y^j)\partial_j.$$

is a linear connection. More generally, for any choice of m^3 functions $\gamma_{ij}^k \in \mathcal{C}^\infty(\mathbb{R}^m)$,

$$\nabla_X Y := X^i\partial_i(Y^j)\partial_j + X^iY^j\gamma_{ij}^k\partial_k.$$

defines a linear connection on $\mathbb{R}^m$.

One can regard a linear connection as a map

$$\begin{aligned}\nabla : \Gamma^\infty(TM) &\to \Gamma^\infty(T^*M \otimes TM),\\ Y &\mapsto \nabla Y\end{aligned}$$

in the understanding that $\nabla Y(X, \omega) := \omega(\nabla_X Y)$. Then we automatically have $\mathcal{C}^\infty(M)$-linearity on X, and the condition (2.2) of a linear connection becomes

$$\nabla(fY) = \mathrm{d}f \otimes Y + f\nabla Y.$$

More generally, given any vector bundle E over M, one can define a linear connection (or a covariant derivative) over E to be a linear map

$$\nabla : \Gamma^\infty(E) \to \Gamma^\infty(T^*M \otimes E)$$

such that

$$\nabla(fs) = \mathrm{d}f \otimes s + f\nabla s, \quad \forall f \in \mathcal{C}^\infty(M), s \in \Gamma^\infty(E).$$

Remark. In general, if E, F are two vector bundles over a smooth manifold M, then a linear map $P : \Gamma^\infty(E) \to \Gamma^\infty(F)$ is a **differential operator** if $\mathrm{supp}(Pu) \subset \mathrm{supp}(u)$ for all section $u \in \Gamma^\infty(E)$. So a connection in E is a first order differential operator from section of E to sections of $T^*M \otimes E$ (which has an extra property that "its principal symbol is the identity map in $T^*M \otimes E$").

2.1.2 Basic Properties of Linear Connections

Locality of Linear Connections

Now let ∇ be a linear connection on a smooth manifold M. We shall prove that $\nabla_X Y$ depends only on local information of X and Y. Since $\nabla_X Y$ is tensorial in X and is a "derivative" in Y, one immediately gets

Proposition 2.3 (Locality I)**.** For any open subset $U \subset M$, if $X|_U = \tilde{X}|_U$ and $Y|_U = \tilde{Y}_U$, then $\nabla_X Y|_U = \nabla_{\tilde{X}}\tilde{Y}|_U$.

Proposition 2.4 (Locality II)**.** If $X(p) = \tilde{X}(p)$, then $\nabla_X Y(p) = \nabla_{\tilde{X}} Y(p)$.

As a consequence, for any vector $v \in T_pM$ and any vector field $Y \in \Gamma^\infty(TM)$, one can define $\nabla_v Y(p)$, the "directional derivative" of Y at p along the direction v, to be the vector $\nabla_X Y(p)$, where X is any vector field such that $X(p) = v$.

On the other hand side, it is not hard to construct vector fields X, Y, $\overline{Y}$ such that $Y(p) = \overline{Y}(p)$ but $\nabla_X Y(p) \neq \nabla_X \overline{Y}(p)$. However, we have

Proposition 2.5 (Locality III)**.** Let $\gamma : (-\varepsilon, \varepsilon) \to M$ be a smooth curve on M with

$$\gamma(0) = p \text{ and } \dot{\gamma}(0) = v.$$

Suppose X, Y, $\overline{Y}$ are vector fields on M such that $X(p) = v$ and

$$Y(\gamma(t)) = \overline{Y}(\gamma(t)), \quad -\varepsilon < t < \varepsilon.$$

Then

$$\nabla_X Y(p) = \nabla_X \overline{Y}(p).$$

Proof. It suffices to prove that if $Y = 0$ along γ, then $\nabla_v Y(p) = 0$. Pick a local coordinate patch $(U, x^1, \dots, x^m)$ near p such that $x(p) = 0$ and that the geometric curve γ has the defining equation $x^2 = \cdots = x^m = 0$ near p. Then $v = a\partial_1$ for some scalar a, and the condition "$Y = 0$ along γ" means $Y = Y^j \partial_j$ with $Y^j(x^1, 0, \dots, 0) = 0$ for all j. In particular,

$$Y^j(p) = 0 \text{ and } \partial_1 Y^j(p) = 0$$

for all j. It follows

$$\nabla_v Y(p) = \nabla_{a\partial_1} Y^j \partial_j(p) = a(\partial_1(Y^j)(p)\partial_j + Y^j(p)\nabla_{\partial_1}\partial_j) = 0.$$

□

Linear Connections in Local Coordinates: the Christoffel Symbols

Let ∇ be a linear connection on M, and let $(U, x^1, \dots, x^m)$ be a coordinate chart. Since $\nabla_{\partial_i}\partial_j$ is a smooth vector field on U (here we used Locality I), there exists smooth functions Γ_{ij}^k on U such that

$$\nabla_{\partial_i}\partial_j = \Gamma_{ij}^k \partial_k.$$

Definition 2.6. The functions Γ_{ij}^k are called the **Christoffel symbols** of ∇ (with respect to the given chart).

For example, if we consider the linear connection

$$\nabla_X Y := X^i \partial_i(Y^j)\partial_j + X^i Y^j \Gamma_{ij}^k \partial_k.$$

on $\mathbb{R}^m$, then the functions γ_{ij}^k's are exactly the Christoffel symbols. In particular for the canonical linear connection on $\mathbb{R}^m$, the Christoffel symbols are all zero.

Under coordinate change from $(x^1, \dots, x^m)$ to $(\tilde{x}^1, \dots, \tilde{x}^m)$, one has

$$\tilde{\Gamma}_{ij}^k = \frac{\partial \tilde{x}^k}{\partial x^t} \frac{\partial x^r}{\partial \tilde{x}^i} \frac{\partial x^s}{\partial \tilde{x}^j} \Gamma_{rs}^t + \frac{\partial^2 x^r}{\partial \tilde{x}^i \partial \tilde{x}^j} \frac{\partial \tilde{x}^k}{\partial x^r}.$$

The proof will be left as a happy exercise. According to this transformation formula,

- As one can anticipate, the Christoffel symbols do not transform like tensor.
- However, the "bad term" (i.e. the second term) depends only on coordinate change and not on the linear connection. So if we have two linear connections, ∇ and $\overline{\nabla}$, on M, then the difference $\nabla - \overline{\nabla}$ is a tensor in both X and Y (which, of course, can be easily checked via definition).

The Existence of Linear Connections

Note that the sum and the difference of two linear connections will no longer be a linear connection. However, it is easy to check by definition that

Lemma 2.7. If $\nabla^{(1)}, \dots, \nabla^{(k)}$ are linear connections on M, and $f_1, \dots, f_k$ are smooth functions on M such that $f_1 + \cdots + f_k = 1$, then the sum $f_i \nabla^{(i)}$ is also a linear connection on M.

So the set of all linear connections on M is a convex set, although it is not a linear space.

As in the case of Riemannian metrics, one can prove the existence of linear connection in different ways:

Theorem 2.8. There exists (plenty of) linear connections on M.

Sketch of the First Proof. On a chart U_α, one may use $\nabla_X \partial_i \equiv 0$ to define linear connection $\nabla^{(\alpha)}$ (with vanishing Christoffel symbol) (or you may define a linear connection with any prescribed Christoffel symbol if you prefer to). Then take a partition of unity and add them to get a linear connection $\nabla = \rho_\alpha \nabla^{(\alpha)}$ on M. □

Sketch of the Second Proof. Embed M into $\mathbb{R}^N$ and check that the projection

$$\nabla_X Y := \mathrm{Proj}_{TM}(D_{\tilde{X}} \tilde{Y})$$

we mentioned earlier is a linear connection. □

The most interesting linear connection on M can be constructed as follows: On any Riemannian manifold (M, g) we will construct explicitly a unique linear connection called the **Levi-Civita connection**, which has nice properties like "torsion free" and "metric-compatibility". Since there exist plenty of Riemannian metrics on M, there must be plenty of linear connections on M.

From Linear Connection to Parallel Transport and Back

With linear connection at hand, we may define the conception of "parallel":

Definition 2.9. Let M be a smooth manifold with a linear connection ∇. Let $\gamma : [a, b] \to M$ be an embedded smooth curve in M, and X a vector field on M. If

$$\nabla_{\dot{\gamma}(t)} X = 0, \quad \forall t,$$

then we say X is **parallel** along γ.

Example. Let $M = \mathbb{R}^n$ with the standard Euclidean space, with standard linear connection ∇ such that $\nabla_X Y = X(Y^j)\partial_j$. Let γ be any curve and X be a vector field. Then for X to be parallel along γ, we need

$$0 = \nabla_{\dot{\gamma}(t)} X = \dot{\gamma}(t)(X^i)\partial_i = \frac{\mathrm{d}}{\mathrm{d}t}(X^i \circ \gamma)\partial_i.$$

It follows that X is parallel along γ if and only if X^i's are constants on γ, i.e. if and only if X is constant vector field along γ.

Theorem 2.10. For any smooth curve $\gamma : [a, b] \to M$, any $t_0 \in [a, b]$ and any vector $X_0 \in T_{\gamma(t_0)}M$, there exists a unique vector field X defined on γ with $X(\gamma(t_0)) = X_0$, such that X is parallel along γ.

Proof. It is enough to prove the theorem for the case when the curve lies in one coordinate patch, since the general case follows from this local existence/uniqueness and a standard compactness argument. Suppose in a local chart, $X_0 = X_0^j \partial_j|_{\gamma(t_0)}$. To find the parallel vector field $X = X^j \partial_j$, we need to solve the equation

$$0 = \nabla_{\dot{\gamma}(t)}(X^j \partial_j) = \frac{\mathrm{d}X^j(\gamma(t))}{\mathrm{d}t}\partial_j + X^k(\gamma(t))\nabla_{\dot{\gamma}(t)}\partial_k.$$

If we let $f^j(t) = X^j(\gamma(t))$ and let $a_k^j(t)$ be such that $\nabla_{\dot{\gamma}(t)}\partial_k = a_k^j(t)\partial_j$, then we get a system of linear ODEs

$$(f^j)'(t) + f^k(t)a_k^j(t) = 0, \quad 1 \leqslant j \leqslant m$$

with initial conditions $f^j(t_0) = X_0^j$. Now apply the classical existence and uniqueness results for system of linear ODEs. □

Definition 2.11. Let X be the unique parallel vector field along γ with $X(\gamma(t_0)) = X_0$. We will call

$$P^\gamma_{t_0,t} : T_{\gamma(t_0)}M \to T_{\gamma(t)}M,$$
$$X_0 = X(\gamma(t_0)) \mapsto X(\gamma(t))$$

the **parallel transport** from $\gamma(t_0)$ to $\gamma(t)$ along γ.

Proof. The linearity comes from the fact that the solution of a homogeneous linear ODE system depends linearly on initial data (the superposition principle). The map $P^\gamma_{t_0,t}$ is invertible since $P^\gamma_{t_0,t}P^{-\gamma}_{a+b-t,a+b-t_0} = \mathrm{Id}$, where $-\gamma$ is the "opposite curve" $(-\gamma)(s) = \gamma(a+b-s)$. □

One can check that given a linear connection ∇, the maps P^γ defined in this way form a parallel transport structure on M.

Conversely, the linear connection ∇ is determined by its parallel transports:

Proposition 2.12. Let $\gamma : [a, b] \to M$ be a smooth curve on M such that $\gamma(t_0) = p$ and $\dot\gamma(t_0) = X_0 \in T_pM$. Then for any vector field $Y \in \Gamma^\infty(TM)$,

$$\nabla_{X_0}Y(p) = \lim_{t\to t_0} \frac{(P^\gamma_{t_0,t})^{-1}(Y(\gamma(t)) - Y(\gamma(t_0)))}{t - t_0}.$$

Proof. Let $\{e_1, \dots, e_m\}$ be a basis of T_pM. Let $e_i(t) = P^\gamma_{t_0,t}(e_i)$. Then by the previous lemma, $\{e_1(t), \dots, e_m(t)\}$ is a basis of $T_{\gamma(t)}M$. So there exist functions $Y^i(t)$ along curve γ so that $Y(\gamma(t)) = Y^i(t)e_i(t)$. It follows that

$$(P^\gamma_{t_0,t})^{-1}(Y(\gamma(t))) = Y^i(t)e_i.$$

So

$$\lim_{t\to t_0} \frac{(P^\gamma_{t_0,t})^{-1}(Y(\gamma(t)) - Y(p))}{t - t_0} = \lim_{t\to t_0} \frac{Y^i(t)e_i - Y(p)}{t - t_0} = \dot Y^i(t_0)e_i.$$

On the other hand side,

$$\nabla_{X_0}Y(p) = (\nabla_{X_0}Y^i)(p)e_i + Y^i(t_0)\nabla_{X_0}e_i(t_0) = \dot Y^i(t_9)e_i,$$

so the conclusion follows. □

2.2 The Levi-Civita Connection

2.2.1 Induced Linear Connections on Tensors

Linear Connections on the Trivial Line Bundle

Let M be a smooth manifold, and E a vector bundle over M. As we have seen, a linear connection on E is bilinear map

$$\nabla : \Gamma^\infty(TM) \times \Gamma^\infty(E) \to \Gamma^\infty(E),$$
$$(X, s) \mapsto \nabla_X s$$

such that for any $f \in \mathcal{C}^\infty(M)$,

$$\nabla_{fX}s = f\nabla_X s$$

and

$$\nabla_X(fs) = f\nabla_X s + (Xf)s.$$

Again there are numerous choices of linear connections on any vector bundle.

Consider the simplest vector bundle, the trivial line bundle $M \times \mathbb{C}$, which will be regarded as $\otimes^{0,0}TM$ below. Since $\Gamma^\infty(\otimes^{0,0}TM) = \mathcal{C}^\infty(M)$, by definition a linear connection on this bundle is a bilinear map

$$\nabla : \Gamma^\infty(TM) \times \mathcal{C}^\infty(M) \to \mathcal{C}^\infty(M)$$

that satisfies the two conditions above. Since we are only considering "directional derivative of smooth functions", we have an obvious and perfect candidate, namely

$$\begin{aligned}\nabla : \Gamma^\infty(TM) \times \mathcal{C}^\infty &\to \mathcal{C}^\infty(M),\\ (X, f) &\mapsto \nabla_X f := Xf,\end{aligned} \tag{2.4}$$

which obviously satisfies the two conditions, and is canonical in the sense that it depends only on the smooth structure of M.

Although we will use the canonical linear connection ∇ defined by (2.4) for smooth functions, we should point out that there exist many other interesting linear connections. In fact, for any smooth 1-form $\omega \in \Omega^1(M)$, we have

$$\nabla_X(gf) = X(gf) + gf\omega(X) = (Xg)f + g(Xf + f\omega(X)) = (Xg)f + g\nabla_X f,$$

which implies

Lemma 2.13. For any 1-form ω on M,

$$\nabla_X^\omega f := Xf + f\omega(X)$$

is a linear connection on $\otimes^{0,0}TM$.

Equivalently, we can write this connection as $\nabla = \mathrm{d} + \omega$.

The Induced Linear Connection on Cotangent Bundle

Suppose we are given a linear connection ∇ on $\otimes^{1,0}TM = TM$. Together with the canonical linear connection ∇ on $\otimes^{0,0}TM = M \times \mathbb{R}$, next let's try to find a reasonable linear connection on the cotangent bundle T^*M. By definition the covariant derivative we want to construct is a bilinear map

$$\begin{aligned}\nabla : \Gamma^\infty(TM) \times \Gamma^\infty(T^*M) &\to \Gamma^\infty(T^*M),\\ (X, \omega) &\mapsto \nabla_X\omega\end{aligned}$$

with given properties. The idea is simple and natural: we need to apply the pairing between T^*M and TM. Note that the linear connection ∇ on TM gives rise to a parallel transport map $P^\gamma_{0,t} : T_{\gamma(0)}M \to T_{\gamma(t)}M$, and by taking dual one gets a linear isomorphism

$$(P^\gamma_{0,t})^* : T^*_{\gamma(t)}M \to T^*_{\gamma(0)}M.$$

With this map at hand, it is thus natural to define the covariant derivative to be

$$\nabla_X\omega(p) := \lim_{t\to 0} \frac{(P^\gamma_{0,t})^*\omega_{\gamma(t)} - \omega_p}{t}, \tag{2.5}$$

where γ is any curve with $\gamma(0) = p$ and $\dot\gamma(0) = X_p$. To get a clear sense of this formula using ∇ on TM instead of using P^γ, let's pair the 1-form $\nabla_X\omega$ with any vector field Y, to get

$$(\nabla_X\omega)(Y) = \lim_{t\to 0} \frac{(P^\gamma_{0,t})^*\omega_{\gamma(t)}(Y_p) - \omega_p(Y_p)}{t} = \lim_{t\to 0} \frac{\omega_{\gamma(t)}(P^\gamma_{0,t}(Y_p)) - \omega_p(Y_p)}{t}$$

We have

$$\begin{aligned}\omega_{\gamma(t)}(P^\gamma_{0,t}(Y_p)) - \omega_p(Y_p) &= \omega_{\gamma(t)}(P^\gamma_{0,t}(Y_p)) - \omega_{\gamma(t)}(Y_{\gamma(t)}) + \omega_{\gamma(t)}(Y_{\gamma(t)}) + \omega_{\gamma(t)}(Y_{\gamma(t)}) - \omega_p(Y_p)\\ &= -\omega_{\gamma(t)}\left(P^\gamma_{0,t}((P^\gamma_{0,t})^{-1}(Y_{\gamma(t)}) - Y_p)\right) + \omega_{\gamma(t)}(Y_{\gamma(t)}) - \omega_p(Y_p).\end{aligned}$$

So in view of the facts

$$\lim_{t\to 0} \frac{(P^\gamma_{0,t})^{-1}(Y_{\gamma(t)}) - Y_p}{t} = \nabla_X Y$$

and

$$\lim_{t\to 0} \frac{\omega_{\gamma(t)}(Y_{\gamma(t)}) - \omega_p(Y_p)}{t} = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \omega(Y)(\gamma(t)) = \dot\gamma(0)(\omega(Y)) = X(\omega(Y)) = \nabla_X(\omega(Y))$$

we get the desired formula

$$(\nabla_X\omega)(Y) = \nabla_X(\omega(Y)) - \omega(\nabla_X Y). \tag{2.6}$$

Note that although it looks like our "definition formula" (2.5) may depend on the curve γ, the formula (2.6) shows that it is independent of the choice of γ.

Induced Linear Connection for Tensors

One can continue this process. Let

$$(P^\gamma_{0,t})^{(r,s)} : \otimes^{r,s} T_{\gamma(0)}M \to \otimes^{r,s} T_{\gamma(t)}M$$

be the naturally induced linear isomorphism (which equals $P^\gamma_{0,t}$ on tangent components, and equals $((P^\gamma_{0,t})^*)^{-1}$ on cotangent components). Then for any tensor field $T \in \Gamma^\infty(\otimes^{r,s}TM)$, one may naturally define

$$\nabla_X T(p) := \lim_{t\to 0} \frac{((P^\gamma_{0,t})^{(r,s)})^{-1} T_{\gamma(t)} - T_p}{t}, \tag{2.7}$$

where γ is any curve with $\gamma(0) = p$ and $\dot\gamma(0) = X_p$.

After some standard but messy computations as above, one can convert the conceptional definition above to a "computable" formula

$$\begin{aligned}(\nabla_X T)(\omega_1, \dots, \omega_r, Y_1, \dots, Y_s) =& \nabla_X(T(\omega_1, \dots, \omega_r, Y_1, \dots, Y_s)) \\ &- \sum_i T(\omega_1, \dots, \nabla_X\omega_i, \dots, \omega_r, Y_1, \dots, Y_s) \\ &- \sum_j T(\omega_1, \dots, \omega_r, Y_1, \dots, \nabla_X Y_j, \dots, Y_s).\end{aligned}$$

Example. Let ∇ be a linear connection on M, and g be a Riemannian metric which is a $(0,2)$-tensor field on M. Applying the induced linear connection to g we get

$$(\nabla_X g)(Y, Z) = X\langle Y, Z\rangle - \langle \nabla_X Y, Z\rangle - \langle Y, \nabla_X Z\rangle.$$

Parallel Tensors

As in the case of vector fields, the linear connection ∇ on $\otimes^{r,s}TM$ satisfies the three locality properties. We may also talk about parallel tensors:

Definition 2.14. A tensor field T is called **parallel along** γ if $\nabla_{\dot\gamma} T = 0$, and is called **parallel** (in all directions) if $\nabla_X T = 0$ for all $X \in \Gamma^\infty(TM)$.

Example. Under the natural pairing between T_p^*M with T_pM, we may view the identity map $\mathrm{Id} : \Gamma^\infty(TM) \to \Gamma^\infty(TM)$ as a $(1,1)$-tensor via

$$I(\omega, Y) = \omega(Y).$$

It is not surprising that I (which comes from the identity map) is parallel:

$$(\nabla_X I)(\omega, Y) = X(\omega(Y)) - (\nabla_X\omega)(Y) - \omega(\nabla_X Y) = 0,$$

which gives a second explanation of (2.6).

Compatibility of the Induced Linear Connection

Now let M be a smooth manifold, and ∇ a linear connection on (the tangent bundle of) M. As we have seen, ∇ induces linear connections on all tensor bundles $\otimes^{r,s}TM$ over M. It turns out that the induced connections are consistent in the sense that they are compatible with the two natural operations on tensors: the tensor product and the contraction.

To see this, let's consider two examples:

Example. For any $Y \in \Gamma^\infty(TM)$ and $\omega \in \Omega^\infty(T^*M)$, applying ∇ to the $(1,1)$-tensor field $Y \otimes \omega$ we get

$$\begin{aligned}\nabla_X(Y\otimes\omega)(\eta, Z) &= X(\eta(Y)\omega(Z)) - (\nabla_X\eta)(Y)\omega(Z) - \eta(Y)\omega(\nabla_X Z) \\ &= X(\eta(Y))\omega(Z) - (\nabla_X\eta)(Y)\omega(Z) + \eta(Y)X(\omega(Z)) - \eta(Y)\omega(\nabla_X Z) \\ &= ((\nabla_X Y)\otimes\omega)(\eta, Z) + (Y\otimes(\nabla_X\omega))(\eta, Z).\end{aligned}$$

In other words,

$$\nabla_X(Y\otimes\omega) = (\nabla_X Y)\otimes\omega + Y\otimes(\nabla_X\omega).$$

Example. Here is another way to understand the fact $\nabla I = 0$: Let $\mathcal{C}_1^1$ be the contraction map that pairs the first tangent component to the first cotangent component, then

$$X(\omega(Y)) = \nabla_X(\mathcal{C}_1^1(Y \otimes \omega)),$$

and by the previous example,

$$\mathcal{C}_1^1(\nabla_X(Y \otimes \omega)) = \mathcal{C}_1^1((\nabla_X Y) \otimes \omega + Y \otimes \nabla_X \omega) = \omega(\nabla_X Y) + (\nabla_X \omega)(Y).$$

In other words, the fact "the identity map I being parallel" implies the fact "∇ commutes with $\mathcal{C}_1^1$" for $(1,1)$-tensor. Similarly one can show that for an (r,s)-tensor, ∇ commutes with all contraction $\mathcal{C}_j^i$'s.

Now we can state the compatibility of ∇ with the two tensor operations:

Theorem 2.15. Given a linear connection ∇ on TM, the induced linear connection

$$\nabla : \Gamma^\infty(TM) \times \Gamma^\infty(\otimes^{r,s} TM) \to \Gamma^\infty(\otimes^{r,s} TM),$$
$$(X, T) \mapsto \nabla_X T,$$

on tensor bundles $\otimes^{r,s} TM$ above is compatible with the tensor product operation[1]

$$\nabla_X(T_1 \otimes T_2) = (\nabla_X T_1) \otimes T_2 + T_1 \otimes (\nabla_X T_2) \tag{2.8}$$

and commutes with the contractions

$$\mathcal{C}_j^i(\nabla_X T) = \nabla_X \mathcal{C}_j^i(T), \tag{2.9}$$

where $1 \leqslant i \leqslant r$, $1 \leqslant j \leqslant s$, and

$$\mathcal{C}_j^i : \Gamma^\infty(\otimes^{r,s} TM) \to \Gamma^\infty(\otimes^{r-1,s-1} TM)$$

is the contraction map that pairs the i-th vector with the j-th covector.

The proof is merely a simple but messy computation which we will omit. Instead, we will show how do we recover (2.6) using conditions (2.8) and (2.9):

$$\begin{aligned}\nabla_X(\omega(Y)) = \nabla_X(\mathcal{C}_1^1(Y \otimes \omega)) &= \mathcal{C}_1^1(\nabla_X(Y \otimes \omega)) \\ &= \mathcal{C}_1^1(Y \otimes \nabla_X \omega + \nabla_X Y \otimes \omega) = (\nabla_X \omega)(Y) + \omega(\nabla_X Y)\end{aligned}$$

which is another way to write (2.6).

Moreover, by a tedious messy induction argument in the same philosophy, one can even recover (2.7) by using (2.8) and (2.9). In other words, one has

Theorem 2.16. Given any linear connection ∇ on the tangent bundle TM, there is a unique linear connection on all tensor fields that coincides with ∇ on TM, coincides with (2.4) on functions, and satisfies compatibility conditions (2.8) and (2.9) above.

The Hessian of a Function

Let M be a smooth manifold and ∇ a linear connection on M. One may equivalently write the induced linear connections on tensor bundles as maps

$$\nabla : \Gamma^\infty(\otimes^{r,s} TM) \to \Gamma^\infty(T^*M \otimes (\otimes^{r,s} TM)) = \Gamma^\infty(\otimes^{r,s+1} TM)$$

with the understanding that

$$\nabla T(\ldots, X) = (\nabla_X T)(\cdots).$$

[1] This fact has another beautiful explanation: For any X, the covariant derivative operator ∇_X is a derivation on the (graded tensor) algebra of all tensor fields on M!

Then one may iterate ∇ to get

$$\nabla^2 : \Gamma^\infty(\otimes^{r,s}TM) \to \Gamma^\infty(\otimes^{r,s+2}TM)$$

(or even higher order powers) in the understanding that

$$\nabla^2 T(\ldots, X, Y) = (\nabla_Y \nabla T)(\ldots, X) = (\nabla_Y, \nabla_X T)(\cdots) - (\nabla_{\nabla_Y X} T)(\cdots). \tag{2.10}$$

[Note that $\nabla^2 T(\ldots, X, Y) \neq (\nabla_Y \nabla_X T)(\cdots)$ in general.]

In particular, if we take $r = s = 0$, i.e. consider functions $f \in \mathcal{C}^\infty(M)$, we get

$$\nabla^2 f(X, Y) = (\nabla_Y \mathrm{d}f)(X) = YXf - (\nabla_Y X)f.$$

The bilinear form $\nabla^2 f$ is known as the **Hessian** of f with respect to ∇.

Torsion Tensors of a Linear Connection

For a general linear connection ∇, the Hessian is not interesting, since it might be non-symmetric. A natural question is: when will $\nabla^2 f$ symmetric? We calculate:

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

It follows that the vector field

$$\mathcal{T}(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y] \tag{2.11}$$

measures how far $\nabla^2 f$ from being symmetric. A direct computation shows

$$\mathcal{T}(fX, Y) = \mathcal{T}(X, fY) = f\mathcal{T}(X, Y).$$

In other words, $\mathcal{T}$ is really a $(1, 2)$-tensor (where we identify $\mathcal{T}$ with the $(1, 2)$-tensor $\widetilde{T}(\omega, X, Y) := \omega(\mathcal{T}(X, Y))$).

Definition 2.17. For any linear connection ∇ on TM, the map

$$\mathcal{T} : \Gamma^\infty(TM) \times \Gamma^\infty(TM) \to \Gamma^\infty(TM)$$

defined by (2.11) is called the **torsion tensor** of ∇.

Example. Consider the connection ∇ defined on $\mathbb{R}^3$ so that with respect to the standard frame e_1, e_2, e_3,

$$\nabla_{e_i} e_j = e_i \times e_j,$$

where $\times$ is the cross product. Then

$$\mathcal{T}(e_i, e_j) = e_i \times e_j - e_j \times e_i = 2e_i \times e_j.$$

To understand the effect of the torsion, let's parallel transport the vector e_2 along the e_1-axis starting at the origin. Let $X = a(x)e_1 + b(x)e_2 + c(x)e_3$ be the parallel transport of e_2 along the e_1-axis. Then we have

$$0 = \nabla_{e_1} X = a'(x)e_1 + (b'(x) - c(x))e_2 + (c'(x) + b(x))e_3,$$

i.e.

$$a'(x) = 0, \quad b'(x) = c(x), \quad c'(x) = -b(x).$$

Together with the initial condition $a(0) = c(0) = 0$, $b(0) = 1$, we will get

$$X(x) = (\cos x)e_2 - (\sin x)e_3.$$

From this formula one can see that in the presence of a torsion, how the vector e_2 "twist" when we parallel transport it.

Back to the Hessian $\nabla^2 f$. We have seen that for $\nabla^2 f$ to be symmetric for all f, one need the linear connection to have vanishing torsion tensor.

Definition 2.18. If $\mathcal{T} = 0$, we call ∇ a **torsion free** (or **symmetric**) connection.

The name "symmetric connection" comes from local computation: if we write

$$\widetilde{\mathcal{T}} = T_{ij}^k \partial_k \otimes \mathrm{d}x^i \otimes \mathrm{d}x^j,$$

i.e. we let T_{ij}^k to be the functions such that

$$\mathcal{T}(\partial_i, \partial_j) = T_{ij}^k \partial_k,$$

then from

$$\mathcal{T}(\partial_i, \partial_j) = \nabla_{\partial_i} \partial_j - \nabla_{\partial_j} \partial_j - [\partial_i, \partial_j] = \Gamma_{ij}^k \partial_k - \Gamma_{ji}^k \partial_k$$

one gets

$$T_{ij}^k = \Gamma_{ij}^k - \Gamma_{ji}^k.$$

As a consequence,

Corollary 2.19. ∇ is torsion free if and only if $\Gamma_{ij}^k = \Gamma_{ji}^k$ for all i, j.

Remark. In particular, we see that the symmetric-condition "$\Gamma_{ij}^k = \Gamma_{ji}^k$ for all i, j" is independent of the choice of local coordinates.

2.2.2 The Levi-Civita Connection

We know that for any smooth manifolds, there are numerous choices of metric structures and measure structures. But with a Riemannian metric structure g at hand, one can define a unique canonical metric structure and measure structure associated to g. The same phenomena happens for linear connections: Given a Riemannian metric g, there is a unique canonical linear connection associated to g, known as the Levi-Civita connection, which has many nice properties.

Metric Compatible Linear Connection

Let (M, g) be a Riemannian manifold. Before we write down the definition of the Levi-Civita connection, we may ask ourselves a question: what kind of nice properties do we want?

First, we may want the linear connection to be torsion free, since under this condition, the Hessian is symmetric (and in fact as we will see later, there will be many other nice symmetry properties under the torsion free condition). Second, we may want the linear connection to be "compatible with the Riemannian metric g".

Now a natural question is:

Question. When we say a linear connection is "compatible with the Riemannian metric g", what do we really mean?

Let's explore this question from the geometric point of view. With a Riemannian metric g at hand, we get an inner product on each T_pM. So a natural requirement would be.

Answer. We may require that each parallel transport

$$P_{0,t}^{\gamma} : (T_{\gamma(0)}, g_{\gamma(0)}) \to (T_{\gamma(t)}, g_{\gamma(t)})$$

preserves the given inner product structure (i.e. is an isometry between the two inner product spaces).

It turns out that the geometric requirement above is equivalent to an algebraic equation on ∇_X (which is easy to use) and also to an analytic equation on g:

Proposition 2.20. Let ∇ be a linear connection on a Riemannian manifold (M, g). Then the following statements are equivalent:

(1) All the parallel transport $P^\gamma_{0,t} : (T_{\gamma(0)}, g_{\gamma(0)}) \to (T_{\gamma(t)}, g_{\gamma(t)})$ are isometries.

(2) For any smooth vector fields X, Y, $Z \in \Gamma^\infty(TM)$, one has

$$X(\langle Y, Z\rangle) = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle.$$

(3) g is parallel, i.e. $\nabla g = 0$.

Proof. (1)⇒(2): Let ∇ be a linear connection such that $P^\gamma_{0,t}$ are isometries. For any vector fields X, Y, $Z \in \Gamma^\infty(TM)$, and any $p \in M$, take a curve γ such that $\gamma(0) = p$ and $\dot\gamma(0) = X_p$. Take an orthonormal basis $\{e_i\}$ of (T_pM, g_p), and let $e_i(t)$ be the parallel transport of e_i along γ. By assumption, $\{e_i(t)\}$ is an orthonormal basis at $\gamma(t)$. If we denote $Y = Y^i(t)e_i(t)$ and $Z = Z^i(t)e_i(t)$, then

$$\langle Y, Z\rangle = \sum Y^i(t)Z^i(t)$$

along γ. So

$$\nabla_{X_p}\langle Y, Z\rangle = \sum X_p(Y^i(t))Z^i(0) + Y^i(0)X_p(Z^i(t)) = \langle \nabla_{X_p} Y, Z_p\rangle + \langle Y_p, \nabla_{X_p} Z\rangle,$$

i.e. ∇ satisfies the desired equation.

(2)⇒(1): Conversely, suppose ∇ be a linear connection on M such that $X(\langle Y, Z\rangle)$ equals $\langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle$ for all X, Y, Z. Fix any curve γ, let $\{e_i\}$ be an orthonormal basis at $p = \gamma(0)$, and let $e_i(t)$ be the parallel transport of e_i along γ, then

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle e_i(t), e_j(t)\rangle = \dot\gamma(t)(\langle e_i(t), e_j(t)\rangle) = \langle \nabla_{\dot\gamma(t)} e_i(t), e_j(t)\rangle + \langle e_i(t), \nabla_{\dot\gamma(t)} e_j(t)\rangle = 0.$$

It follows that $\{e_i(t)\}$ remains to be an orthonormal basis for $(T_{\gamma(t)}, g_{\gamma(t)})$. So the linear map $P^\gamma_{0,t}$ is an isometry.

(2)⇔(3): Recall that the Riemannian metric g is a $(0, 2)$-tensor, and thus one can take its covariant derivative ∇g, which by definition is given by

$$(\nabla_X g)(Y, Z) = X\langle Y, Z\rangle - \langle \nabla_X Y, Z\rangle - \langle Y, \nabla_X Z\rangle.$$

So the conclusion follows. □

Definition 2.21. We say a linear connection ∇ on a Riemannian manifold (M, g) is **compatible** with g if one (and thus all) of the three equivalent conditions in Proposition 2.20 hold.

Remark. Note that by the geometric condition, if ∇ is a metric-compatible linear connection on (M, g), and if X, Y are vector fields parallel along a curve γ, then $\langle X, Y\rangle$ is a constant on γ.

The Levi-Civita Connection

Finally, we define

Definition 2.22. A connection ∇ is on (M, g) is called a **Levi-Civita connection** (also called a **Riemannian connection**) if it is torsion-free and is compatible with g.

We exemplify two simple examples:

Example. Let $M = \mathbb{R}^m$, equipped with the canonical Riemannian metric g_0, then the canonical linear connection (i.e. the one with all Christoffel symbols $\Gamma^l_{ij} = 0$ under the canonical basis) is a Levi-Civita connection.

Example. Equip $M = S^m$ with the round metric $g = g_{\text{round}}$, i.e. the induced metric from the canonical metric in $\mathbb{R}^{m+1}$. We denote by $\overline{\nabla}$ the canonical (Levi-Civita) connection in $\mathbb{R}^{m+1}$. For any X, $Y \in \Gamma^\infty(TS^m)$, one can extend X, Y to smooth vector fields $\overline{X}$ and $\overline{Y}$ on $\mathbb{R}^{m+1}$. By localities we proved last time, the vector

$$\overline{\nabla}_{\overline{X}}\overline{Y}$$

at any point $p \in S^m$ depends only on the vector $\overline{X}(p) = X(p)$ and the vectors $\overline{X}(q) = X(q)$ for $q \in S^m$ near p. In other words, it is independent of the choice of the extension. So for simplicity we will write $\overline{\nabla}_X Y$ instead of $\overline{\nabla}_{\overline{X}}\overline{Y}$ for points on S^m. It is a vector that is not necessary tangent to S^m. We define $\nabla_X Y$ be the "orthogonal projection" of $\overline{\nabla}_X Y$ onto the tangent space of S^m, i.e.

$$\nabla_X Y := \overline{\nabla}_X Y - \langle \overline{\nabla}_X Y, \vec{n}\rangle \vec{n},$$

where $\vec{n} = (x^1, x^2, \dots, x^{m+1})$ is the unit out normal vector on S^m. Observe that

$$\overline{\nabla}_X \vec{n} = X^i \partial_i (x^j)\partial_j = X.$$

It follows $\langle \overline{\nabla}_X Y, \vec{n}\rangle \vec{n} = -\langle Y, \overline{\nabla}_X \vec{n}\rangle \vec{n} = -\langle X, Y\rangle = \vec{n}$ and thus

$$\nabla_X Y = \overline{\nabla}_X Y + \langle X, Y\rangle \vec{n}.$$

We claim that ∇ is a Levi-Civita connection of (S^m, g_{round}). To prove this, first notice ∇ is bilinear, and $\nabla_{fX} Y = f\nabla_X Y$. Also

$$\nabla_X(fY) = \overline{\nabla}_X(fY) + \langle X, fY\rangle \vec{n} = (Xf)Y + f\overline{\nabla}_X Y + f\langle X, Y\rangle \vec{n} = (Xf)Y + f\nabla_X Y.$$

This connection is torsion free because

$$\nabla_X Y - \nabla_Y X = \overline{\nabla}_X Y + \langle X, Y\rangle \vec{n} - \overline{\nabla}_Y X - \langle Y, X\rangle \vec{n} = \overline{\nabla}_X Y - \overline{\nabla}_Y X = [X, Y].$$

Finally this connection is compatible with the metric g, since

$$X\langle Y, Z\rangle = \langle \overline{\nabla}_X Y, Z\rangle + \langle Y, \overline{\nabla}_X Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle,$$

where we used the fact that Z is perpendicular to $\vec{n}$.

Remark. If (M, g) is a Riemannian manifold, with a Levi-Civita connection ∇^M, and if $(N, \iota^* g)$ is a Riemannian submanifold of (M, g), then we can define a connection on N by the same trick, namely orthogonally project ∇^M onto TN,

$$\nabla^N_X Y := (\nabla^M_{\overline{X}} \overline{Y})^{\mathrm{T}},$$

One can prove that it is the Levi-Civita connection on $(N, \iota^* g)$.

The Fundamental Theorem of Riemannian Geometry

Since any Riemannian manifold can be embedded to the standard Euclidean space isometrically, the arguments in the previous remark implies that on any Riemannian manifold, there exists a Levi-Civita connection! In what follows we will give two direct elementary proofs of this fact, and also prove the uniqueness:

Theorem 2.23 (The Fundamental Theorem of Riemannian Geometry)**.** On any Riemannian manifold (M, g), there is a unique Levi-Civita connection.

First proof (local coordinate). We first prove uniqueness. Let ∇ be a Levi-Civita connection. Pick a coordinate chart and let Γ^k_{ij} be the Christoffel symbols. It is enough to prove that the Γ^k_{ij}'s are determined by g_{ij}'s. The trick already appeared in Section 1.1. First we note that by torsion free property, $\Gamma^k_{ij} = \Gamma^k_{ji}$. Second we calculate

$$\begin{aligned}\partial_i g_{jk} = \partial_i (g(\partial_j, \partial_k)) &= g(\nabla_{\partial_i}\partial_j, \partial_k) + g(\partial_j, \nabla_{\partial_i}\partial_k) \\ &= g(\Gamma^l_{ij}\partial_l \partial_l, \partial_k) + g(\partial_j, \Gamma^l_{ik}\partial_l) = \Gamma^l_{ij} g_{lk} + \Gamma^l_{ik} g_{jl}.\end{aligned}$$

Similarly one can prove

$$\partial_j g_{ki} = \Gamma^l_{jk} g_{li} + \Gamma^l_{ji} \quad \text{and} \quad \partial_k g_{ij} = \Gamma^l_{ki} g_{lj} + \Gamma^l_{kj} g_{il}.$$

So we get

$$\partial_j g_{ki} + \partial_i g_{jk} - \partial_k g_{ij} = 2 g_{lk} \Gamma^l_{ij}.$$

It follows

$$2\Gamma_{ij}^l = g^{lk}(\partial_j g_{ki} + \partial_i g_{jk} - \partial_k g_{ij}). \tag{2.12}$$

This proves the uniqueness. [This is essentially the same as we did in Section 1.1.]

For the existence, we can define locally (for $X = X^i\partial_i$ and $Y = Y^j\partial_j$)

$$\nabla_X Y = X^i(\partial_i Y^j)\partial_j + X^i Y^j \Gamma_{ij}^l \partial_l,$$

where Γ_{ij}^l is the function given by (2.12). By tedious computations one can check that this give a Levi-Civita connection whose Christoffel symbols are the Γ_{ij}^l's. □

Second proof (coordinate free). Again we first prove the uniqueness. Assume the Levi-Civita connection exists. Then (use torsion-free and metric-compatibility in turns)

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

It follows that $\nabla_X Y$ must be the vector satisfying

$$2\langle\nabla_X Y, Z\rangle = X(\langle Y, Z\rangle) - Z(\langle Y, X\rangle) + Y(\langle Z, X\rangle) - \langle Z, [Y, X]\rangle + \langle[Z, Y], X\rangle - \langle Y, [X, Z]\rangle. \tag{2.13}$$

The right hand side is determined by the metric. So the uniqueness is proved. [(2.13) is called the **Koszul** formula, which reduce to (2.13) if we take X, Y, Z to be ∂_i, ∂_j and ∂_l.]

To prove the existence, one "only need" to check that the $\nabla_X Y$ defined by the above formula satisfies all conditions of Levi-Civita connections. □

Chapter 3

Curvature

3.1 The Curvature Tensor

3.1.1 The Curvature Tensor of a Linear Connection

Derivations on the Graded Tensor Algebra

Let M be a smooth manifold endowed with a linear connection ∇. As we have seen last time, ∇ induces a linear connection

$$\nabla : \Gamma^\infty(TM) \times \Gamma^\infty(\otimes^{k,l} TM) \to \Gamma^\infty(\otimes^{k,l} TM)$$

on each tensor bundle $\otimes^{k,l} TM$. Moreover, all these linear connections are compatible as a whole set of connections in the sense that they are compatible with the tensor product operation and the contraction operation for tensors.

Let's take a closer look of the "tensor product compatibility". Denote by $\Gamma^\infty(\otimes^{*,*} TM)$ the graded tensor algebra of all smooth tensor fields on M. Then the tensor product compatibility means that for any smooth vector field $X \in \Gamma^\infty(TM)$, the map

$$\nabla_X : \Gamma^\infty(\otimes^{*,*} TM) \to \Gamma^\infty(\otimes^{*,*} TM)$$

satisfies $\nabla_X(S \otimes T) = \nabla_X S \otimes T + S \otimes \nabla_X T$. In other words, ∇_X is a derivation on $\Gamma^\infty(\otimes^{*,*} TM)$.

Now let $\mathcal{D}$ be the set of all derivations on the tensor algebra $\Gamma^\infty(\otimes^{*,*} TM)$, which are by definition linear maps D such that $D(S \otimes T) = DS \otimes T + S \otimes DT$. A standard fact (which is easy to verify via definition) is that $\mathcal{D}$ is a Lie algebra (with respect to commutator), namely if D_1, D_2 are two derivations, so is their commutator

$$[D_1, D_2] = D_1 \circ D_2 - D_2 \circ D_1.$$

Example. For any smooth vector field X, the Lie derivative $\mathcal{L}_X$ is a derivation on $\Gamma^\infty(\otimes^{*,*} TM)$. Moreover, $\mathcal{L}_X$ satisfies (when acting on any tensor field)

$$\mathcal{L}_{[X,Y]} = \mathcal{L}_X \circ \mathcal{L}_Y - \mathcal{L}_Y \circ \mathcal{L}_X.$$

In other words, the linear map "$X \mapsto \mathcal{L}_X$" is a Lie algebra homomorphism from "the Lie algebra of all smooth vector fields on M" to "the Lie algebra of all derivations on $\Gamma^\infty(\otimes^{*,*} TM)$".

Now consider the linear map

$$\Phi : \Gamma^\infty(TM) \to \mathcal{D}$$
$$X \mapsto \nabla_X.$$

One may ask: Is Φ a Lie algebra homomorphism? In other words, do we have

$$\nabla_{[X,Y]} = \nabla_X \circ \nabla_Y - \nabla_Y \circ \nabla_X?$$

Unfortunately the answer is no in general[1] (as we will see soon). So we are naturally led to study the map $R(X,Y) : \Gamma^\infty(\otimes^{k,l}TM) \to \Gamma^\infty(\otimes^{k,l}TM)$ defined by

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

Let's start with two simple cases:

- First for $k = l = 0$, i.e. $T = f \in \mathcal{C}^\infty(M)$, the map $R(X,Y)$ is zero, since
$$R(X,Y)f = \nabla_X\nabla_Y f - \nabla_Y\nabla_X f - \nabla_{[X,Y]}f = XYf - YXf - [X,Y]f = 0.$$
- Next we study the case $k = 1$, $l = 0$, i.e. $T = \omega$ is a smooth 1-form. It turns out that one can convert $R(X,Y)$ on 1-forms to $R(X,Y)$ on vector fields:

Lemma 3.1. For any 1-form $\omega \in \Omega^1(M)$,

$$(R(X,Y)\omega)(Z) = -\omega(R(X,Y)Z).$$

Proof. Compute by definition. Details left as an exercise. □

In view of the fact that the graded tensor algebra $\Gamma^\infty(\otimes^{*,*}TM)$ is generated by smooth functions, vector fields and 1-forms, together with the fact that $R(X,Y)$ is again a derivation on $\Gamma^\infty(\otimes^{*,*}TM)$, we conclude that to study $R(X,Y)$ on all tensor fields, it is enough to study $R(X,Y)$ on vector fields!

The Curvature Tensor of a Linear Connection

We define

Definition 3.2. Let M be a smooth manifold and ∇ a linear connection on M. We call the map $R : \Gamma^\infty(TM) \times \Gamma^\infty(TM) \times \Gamma^\infty(TM) \to \Gamma^\infty(TM)$ defined by

$$R(X,Y)Z = \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z \tag{3.1}$$

the **curvature tensor** of ∇.

As we explained above, R measures to what extend the map Φ fails to be a Lie algebra homomorphism.

Remark. In many books, the definition of curvature tensor is different from the above formula by a negative sign. Both definitions have their own advantages. So when you open a new book on Riemannian geometry, you should first glance at its definition of the curvature tensor.

Example. For the standard linear connection $\overline{\nabla}$ on $\mathbb{R}^m$, we have

$$\overline{\nabla}_{X^i\partial_i}(Y^j\partial_j) = X^i\partial_i(Y^j)\partial_j.$$

which implies

$$\overline{\nabla}_X\overline{\nabla}_Y Z - \overline{\nabla}_Y\overline{\nabla}_X Z = X^i\partial_i(Y^j\partial_j Z^k)\partial_k - Y^j\partial_j(X^i\partial_i Z^k)\partial_k = \overline{\nabla}_{[X,Y]}Z$$

and thus its curvature tensor $\overline{R} \equiv 0$.

Example. Consider $M = S^m$. Last time we have seen that

$$\nabla_X Y = \overline{\nabla}_X Y + \langle X, Y\rangle\vec{n}$$

is the Levi-Civita connection on S^m, where $\overline{\nabla}$ is the standard connection on $\mathbb{R}^{m+1}$. It follows

$$\begin{aligned}\nabla_X\nabla_Y Z &= \overline{\nabla}_X\nabla_Y Z + \langle X, \nabla_Y Z\rangle\vec{n}\\ &= \overline{\nabla}_X(\overline{\nabla}_Y Z + \langle Y, Z\rangle\vec{n}) + Y\langle X, Z\rangle\vec{n} - \langle\nabla_Y X, Z\rangle\vec{n}\\ &= \overline{\nabla}_X\overline{\nabla}_Y Z + X(\langle Y, Z\rangle)\vec{n} + \langle Y, Z\rangle X + Y(\langle X, Z\rangle)\vec{n} - \langle\nabla_Y X, Z\rangle\vec{n}.\end{aligned}$$

In view of the fact $\overline{R} = 0$, we get

$$\begin{aligned}R(X,Y)Z =& X(\langle Y, Z\rangle)\vec{n} + \langle Y, Z\rangle X + Y(\langle X, Z\rangle)\vec{n} - \langle\nabla_Y X, Z\rangle\vec{n} - Y(\langle X, Z\rangle)\vec{n}\\ &- \langle X, Z\rangle Y - X(\langle Y, Z\rangle)\vec{n} + \langle\nabla_X Y, Z\rangle\vec{n} - \langle[X,Y], Z\rangle\vec{n}\\ =& \langle Y, Z\rangle X - \langle X, Z\rangle Y.\end{aligned}$$

[1]Maybe we should say fortunately the answer is no, otherwise there will be no Riemannian geometry, and the world will be boring.

The Curvature Tensor is a Tensor

Now we prove that R is a tensor of type $(1,3)$, in the sense

$$\widetilde{R}(\omega, X, Y, Z) := \omega(R(X,Y),Z)$$

is really an element in $\Gamma^\infty(\otimes^{1,3}TM)$:

Proposition 3.3. The curvature tensor R is a $(1,3)$-tensor.

Proof. We need to prove

$$R(fX,Y)Z = R(X,fY)Z = R(X,Y)(fZ) = fR(X,Y)Z.$$

Here we only check one of them:

$$\begin{aligned} R(fX,Y)Z &= f\nabla_X\nabla_Y Z - \nabla_Y(f\nabla_X Z) - \nabla_{(fX)Y-Y(fX)}Z \\ &= f(\nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z) - (Yf)\nabla_X Z + (Yf)\nabla_X Z \\ &= fR(X,Y)Z. \end{aligned}$$

The others are similar and are left as happy exercises. □

Locally, we write the tensor R (or $\widetilde{R}$) as[2]

$$R = R^l_{ijk}\mathrm{d}x^i \otimes \mathrm{d}x^j \otimes \mathrm{d}x^k \otimes \partial_l,$$

i.e. if we denote

$$R(\partial_i, \partial_j)\partial_k = R^l_{ijk}\partial_l,$$

then the coefficients R^l_{ijk} are related to the Christoffel symbols by

$$R^l_{ijk} = \partial_i\Gamma^l_{jk} - \partial_j\Gamma^l_{ik} + \Gamma^s_{jk}\Gamma^l_{is} - \Gamma^s_{ik}\Gamma^l_{js},$$

which is a consequence of the fact

$$R^l_{ijk}\partial_l = R(\partial_i,\partial_j)\partial_k = \nabla_{\partial_i}(\Gamma^s_{jk}\partial_s) - \nabla_{\partial_j}(\Gamma^s_{ik}\partial_s).$$

Note that this also implies that the curvature tensor for the standard connection on $\mathbb{R}^m$ is identically zero, since its Christoffel symbols are all zero.

The Commutator of Two Coordinate Covariant Derivatives of a Surface

Consider an embedded 2-dimensional parametric surface in M,

$$\varphi : U \subset \mathbb{R}^2 \to S = \varphi(U) \subset M.$$

Denote the parameters for U by s and t. We let ∂_s and ∂_t be the two coordinate vector fields

$$\partial_s := \mathrm{d}\varphi\left(\frac{\partial}{\partial s}\right) \quad \text{and} \quad \partial_t := \mathrm{d}\varphi\left(\frac{\partial}{\partial t}\right)$$

on the 2-dimensional surface S. Note that by locality, for any smooth vector field Z on surface S, the expression $\nabla_{\partial_s}\nabla_{\partial_t}Z$ make sense and will be thought of as the iterated second order covariant derivative of Z with respect to ∂_s, ∂_t. It turns out that R measures the non-commutativity of such iterated covariant derivatives:

Proposition 3.4. For any smooth vector field Z (defined on the surface S),

$$\nabla_{\partial_s}\nabla_{\partial_t}Z - \nabla_{\partial_t}\nabla_{\partial_s}Z = R(\partial_s,\partial_t)Z.$$

Proof. Take a coordinate chart of M near S so that S is defined by $x^3 = \cdots = x^m = 0$. Then $(s,t,x^3,\ldots,x^m)$ is a local coordinate system on M in a tubular neighborhood of S. Now the conclusion follows from the fact as coordinate vector fields, $[\partial_t, \partial_s] = 0$. □

Remark. Geometrically, R is closely related to "the holonomy along an infinitesimal path which is the boundary of $\varphi((0,\varepsilon)\times(0,\varepsilon))$". Details are left as a term project.

[2] Note that here we are using a "non-standard" order: for the local expression of the $(1,3)$-tensor R, we write the "1"-part (i.e. the vector ∂_l) after the "3"-parts (i.e. the co-vectors). In some books people use different orders like $R(\partial_i,\partial_j)\partial_k = R^l_{kij}\partial_l$. In other words, in writing local expressions of $\widetilde{R}(\omega,X,Y,Z)$, we always want to put the index for ω next to the index for Z. The reason will be clear in next section.

Flat Connection

We are interested in linear connections with vanishing curvature tensor, i.e. with $R = 0$. For example, the standard connection on $\mathbb{R}^m$ satisfies $R = 0$. The nice point for $\mathbb{R}^m$ is that the coordinate vector fields are parallel along any vector fields.

Definition 3.5. Let M be a smooth manifolds with a connection ∇. We say (M, ∇) admits a **local flat frame** everywhere if near any point p, there is a set of vector fields $X_1, \ldots, X_m$ on a neighborhood U of p such that

(1) [frame] $\{X_i(q) \mid 1 \leqslant i \leqslant m\}$ form a basis of T_qM for every $q \in U$.

(2) [flatness] $\nabla_Y X_i = 0$ for all i and for all vector field Y.

It is easy to see that if (M, ∇) admits a local flat frame everywhere, then $R(X, Y)X_i = 0$ for all X, Y, and thus $R \equiv 0$ since R is a tensor. Conversely,

Proposition 3.6. Let M be a smooth manifold, ∇ be a linear connection. Then $R = 0$ if and only if (M, ∇) admits a local flat frame everywhere.

Proof. It remains to prove the "only if" part. Without loss of generality, we may take U to be a coordinate neighborhood and $p = (0, \ldots, 0)$ the origin. We start with any basis $\{v_1, \ldots, v_m\}$ of T_pM and let $X_i(p) = v_i$. We extend X^i to the "line" $\{(a, 0, \ldots, 0)\}$ by parallel transporting the vector $X_i(p)$ along the curve $\gamma_0(t) = (t, 0, \ldots, 0)$. Then we extend further to the "plane" $\{(a, b, 0, \ldots, 0)\}$ by parallel transporting each $X_i(\gamma_0(a))$ along $\gamma_a(t) = (a, t, 0, \ldots, 0)$. Repeating this procedure, we get a set of smooth (why?) vector fields $\{X_1, \ldots, X_m\}$ on the whole of U. By construction, they are a frame. It remains to prove the flatness.

First by construction, we have

- $\nabla_{\partial_1} X_i = 0$ at any point on the line $(a, 0, \ldots, 0)$.
- $\nabla_{\partial_2} X_i = 0$ at any point on the plane $(a, b, 0, \ldots, 0)$.

Moreover, since $R = 0$ and $[\partial_1, \partial_2] = 0$, we get

$$\nabla_{\partial_2} \nabla_{\partial_1} X_i = \nabla_{\partial_1} \nabla_{\partial_2} X_i = 0$$

on the plane $(a, b, 0, \ldots, 0)$. As a consequence, $\nabla_{\partial_1} X_i$ is parallel along each line $\gamma_a(t) = (a, t, 0, \ldots, 0)$, with initial condition $(\nabla_{\partial_1} X_i)(a, 0, \ldots, 0) = 0$. By uniqueness, one must have $\nabla_{\partial_1} X_i = 0$ along each γ_a. In other words, we get

- $\nabla_{\partial_1} X_i = 0$, $\nabla_{\partial_2} X_i = 0$ at any point on the plane $(a, b, 0, \ldots, 0)$.

  By the same argument, we get

- $\nabla_{\partial_1} X_i = 0$, $\nabla_{\partial_2} X_i = 0$, $\nabla_{\partial_3} X_i = 0$ at any point of the form $(a, b, c, 0, \ldots, 0)$.

Continuing this argument, one can see that $\nabla_{\partial_j} X_i = 0$ for all i, j, at all points in U. As a consequence, $X_1, \ldots, X_m$ form a local flat frame. □

As a consequence,

Corollary 3.7. If (M, g) is a Riemannian manifold for which the curvature of the Levi-Civita connection vanishes, then (M, g) is locally isometric to $(\mathbb{R}^m, g_0)$.

Proof. In the proof above, we take $\{v_1, \ldots, v_m\}$ to be an orthonormal basis. Then after parallel transport, the vector fields $X_1, \ldots, X_m$ are orthonormal everywhere. Since ∇ is Levi-Civita connection, it is torsion free. It follows that for any i, j,

$$0 = \nabla_{X_i} X_j - \nabla_{X_j} X_i - [X_i, X_j] = -[X_i, X_j].$$

By Frobenius theorem, there exists a local coordinate chart with $X_i = \partial_i$ for all i. In this chart, we have $g_{ij} = \langle \partial_i, \partial_j \rangle = \delta_{ij}$, and thus locally $g = \sum \mathrm{d}x^i \otimes \mathrm{d}x^i$. □

Definition 3.8. A linear connection ∇ on M is called **flat** if $R = 0$. A Riemannian manifold is called **flat** if the Levi-Civita connection is flat.

3.1.2 Symmetries of the Curvature Tensor for Torsion Free Connection

Curvature Tensor as Commutator of ∇^2 (for Torsion Free Connection)

Regard ∇ as a map of the form

$$\nabla : \Gamma^\infty(\otimes^{k,l}TM) \to \Gamma^\infty(\otimes^{k,l+1}TM).$$

We have studied the composition $\nabla^2 : \Gamma^\infty(\otimes^{0,0}TM) \to \Gamma^\infty(\otimes^{0,2}TM)$ which maps any $f \in \mathcal{C}^\infty(M)$ to its Hessian

$$\nabla^2 f : \Gamma^\infty(TM) \times \Gamma^\infty(TM) \to \mathcal{C}^\infty(M),$$
$$\nabla^2 f(X,Y) = \nabla_Y \nabla_X f - \nabla_{\nabla_Y X} f.$$

In general, for any $T \in \Gamma^\infty(\otimes^{k,l}TM)$, the second covariant derivative map ∇^2 sends T to the $(k, l+2)$-tensor $\nabla^2 T$, which, by definition, equals

$$\nabla^2 T(\dots, X, Y) = (\nabla_Y, \nabla_X T)(\cdots) - (\nabla_{\nabla_Y X} T)(\cdots).$$

For simplicity, we will denote

$$\nabla^2_{X,Y} : \Gamma^\infty(\otimes^{k,l}TM) \to \Gamma^\infty(\otimes^{k,l}TM),$$
$$T \mapsto \nabla^2 T(\dots, X, Y).$$

One should be aware fo the difference between the second covariant derivative and the iterated covariant derivative.

From now on suppose ∇ is torsion free, which as we have seen, is equivalent to the fact that $\nabla^2_{X,Y} f$ is symmetric with respect to the entries X and Y. A natural question is: if ∇ is torsion free, is $\nabla^2_{X,Y} T$ symmetric with respect to X and Y for all T? The answer is no. For example, if $T = Z$ is a vector field,

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

where in the last step we used the fact ∇ is torsion free, i.e. $\nabla_X Y - \nabla_Y X = [X,Y]$. More generally, by exactly the same computation one has

Lemma 3.9 (Ricci Identity)**.** Let ∇ be any torsion free connection, then

$$\nabla^2_{Y,X}(T) - \nabla^2_{X,Y}(T) = R(X,Y)T, \quad \forall T \in \Gamma^\infty(\otimes^{k,l}TM).$$

In conclusion, for a torsion free connection, the operator $R(X,Y)$ (on any tensors) measures the non-commutativity of second covariant derivatives.

Remark. One may ask: What about higher order covariant derivatives? It turns out that under the torsion free assumption, the operator $R(X,Y)$ also measures the non-commutativity of higher covariant derivatives. For example,

$$\begin{aligned}\nabla^3_{Y,Z,X} T - \nabla^3_{Z,Y,X} T &= (\nabla_X \nabla^2 T)(Y,Z) - (\nabla_X \nabla^2 T)(Z,Y) \\ &= \nabla_X \nabla^2_{Y,Z} T - \nabla^2_{\nabla_X Y, Z} T - \nabla^2_{Y, \nabla_X Z} T - \nabla_X \nabla^2_{Z,Y} T + \nabla^2_{\nabla_X Z, Y} T + \nabla^2_{Z, \nabla_X Y} T \\ &= -\nabla_X (R(Y,Z)T) + R(\nabla_X Y, Z)T + R(Y, \nabla_X Z)T.\end{aligned}$$

The First/Algebraic Bianchi Identity

Now we study symmetric of the curvature tensor R. By definition one immediately see that for any linear connection ∇, R admits the following anti-symmetry:

$$R(X,Y)Z = -R(Y,X)Z \tag{3.2}$$

In local coordinates, it can be written as

$$R^l_{ijk} = -R^l_{jik}.$$

It turns out that for torsion free connections, R admits more symmetries.

Proposition 3.10 (The First Bianchi Identity)**.** If ∇ is a torsion-free, then

$$R(X,Y)Z + R(Y,Z)X + R(Z,X)Y = 0. \tag{3.3}$$

Proof. Since for torsion free connection, we have $\nabla_X Y - \nabla_Y X - [X,Y] = 0$, so

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

where in the last step we used the Jacobi identity for vector fields. □

In local coordinates the first Bianchi identity can be written as

$$R^l_{ijk} + R^l_{jki} + R^l_{kij} = 0.$$

The Second/Differential Bianchi Identity

It turns out that not only R has the above cyclic symmetry, but also its covariant derivative ∇R has a similar cyclic symmetry. To understand ∇R, let's go back to the $(1,3)$-tensor $\widetilde{R}$ defined by

$$\widetilde{R}(\omega, X, Y, Z) = \omega(R(X,Y)Z).$$

It follows by definition that $\nabla\widetilde{R}$ is the $(1,4)$-tensor given by

$$\begin{aligned}
(\nabla\widetilde{R})(\omega, X, Y, Z, W) &= (\nabla_W\widetilde{R})(\omega, X, Y, Z) \\
&= \nabla_W(\omega(R(X,Y)Z)) - (\nabla_W\omega)(R(X,Y)Z) - \omega(R(\nabla_W X, Y)Z) \\
&\quad - \omega(R(X,\nabla_W Y)Z) - \omega(R(X,Y)\nabla_W Z) \\
&= \omega(\nabla_W(R(X,Y)Z) - R(\nabla_W X, Y)Z - R(X,\nabla_W Y)Z - R(X,Y)\nabla_W Z).
\end{aligned}$$

So it is reasonable to define $\nabla_W R$ as

$$(\nabla_W R)(X,Y,Z) := \nabla_W(R(X,Y)Z) - R(\nabla_W X, Y)Z - R(X,\nabla_W Y)Z - R(X,Y)\nabla_W Z.$$

Proposition 3.11 (The Second Bianchi Identity)**.** Suppose ∇ is torsion free, then

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

Proof. By definition,

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

Using the torsion-freeness and (3.2), one can simplify the middle two columns to

$$R([X,Z],Y)W + R([Y,X],Z)W + R([Y,X],Z)W.$$

Now expand each R using its definition, the whole expression becomes a summation of 27 terms, the first 9 terms being

$$\begin{aligned}
&+\nabla_X\nabla_Y\nabla_Z W - \nabla_X\nabla_Z\nabla_Y W - \nabla_X\nabla_{[Y,Z]}W \\
&+\nabla_{[X,Z]}\nabla_Y W - \nabla_Y\nabla_{[X,Z]}W - \nabla_{[[X,Z],Y]}W \\
&-\nabla_Y\nabla_Z\nabla_X W + \nabla_Z\nabla_Y\nabla_X W + \nabla_{[Y,Z]}\nabla_X W,
\end{aligned}$$

the second and third 9 terms are similar to the first 9 terms above: one just replace X, Y, Z by Y, Z, X and Z, X, Y respectively. It is not hard to check that all those expressions containing three ∇'s (12 terms in total) cancel out trivially, all those expressions containing two ∇'s (also 12 terms in total) cancel out by using the fact $[X,Y]=-[Y,X]$, and the remaining three terms

$$\nabla_{[[X,Z],Y]}W+\nabla_{[[Y,X],Z]}W+\nabla_{[[Z,Y],X]}W=0$$

in view of the Jacobi identity. □

In local coordinates we can write $\nabla_{\partial_n}R=R^l_{ijk;n}\mathrm{d}x^i\otimes\mathrm{d}x^j\otimes\mathrm{d}x^k\otimes\partial_l$. Then the second Bianchi identity can be written as

$$R^l_{ijk;n}+R^l_{jnk;i}+R^l_{nik;j}=0.$$

Remark. If we denote $\sum_{\bigcirc X,Y,Z}$ to be the cyclic sum over X, Y, Z, then the first and second Bianchi identities can be written as

$$\sum_{\bigcirc X,Y,Z}R(X,Y)Z=0\quad\text{and}\quad\sum_{\bigcirc X,Y,Z}(\nabla_XR)(Y,Z,W)=0.$$

More generally, if ∇ is not torsion free, then in terms of the torsion tensor $\mathcal{T}$,

$$\sum_{\bigcirc X,Y,Z}R(X,Y)Z=\sum_{\bigcirc X,Y,Z}((\nabla_X\mathcal{T})(Y,Z)+\mathcal{T}(\mathcal{T}(X,Y),Z))$$

and

$$\sum_{\bigcirc X,Y,Z}(\nabla_XR)(Y,Z,W)+\sum_{\bigcirc X,Y,Z}R(\mathcal{T}(X,Y),Z)W=0.$$

3.2 The Riemann Curvature

3.2.1 The Riemann Curvature Tensor

The Riemann Curvature Tensor of Type $(0,4)$

Given any linear connection ∇ on M, one gets a type $(1,3)$ curvature tensor R

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

which measures the non-commutativity of "second order/iterated covariant derivatives". Locally one may write R as

$$R=R^l_{ijk}\mathrm{d}x^i\otimes\mathrm{d}x^j\otimes\mathrm{d}x^k\otimes\partial_l.$$

Now suppose (M,g) is a Riemannian manifold and ∇ is the Levi-Civita connection. By using the Riemannian metric g (via the musical isomorphism) one can convert the $(1,3)$-tensor R to a $(0,4)$-tensor $Rm\in\Gamma^\infty(\otimes^{0,4}TM)$ defined by

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

Definition 3.12. We call Rm the **Riemann curvature tensor** of (M,g).

Locally if we write

$$Rm=R_{ijkl}\mathrm{d}x^i\otimes\mathrm{d}x^j\otimes\mathrm{d}x^k\otimes\mathrm{d}x^l,$$

then

$$R_{ijkl}=Rm(\partial_i,\partial_j,\partial_k,\partial_l)=-g(R^m_{ijk}\partial_m,\partial_l)=-g_{ml}R^m_{ijk}.$$

In other words, the Riemannian metric "lower one of the index".

Example. For S^m (equipped with the standard round metric), we have seen

$$R(X,Y)Z = \langle Y,Z\rangle X - \langle X,Z\rangle Y.$$

Thus the Riemann curvature tensor is

$$Rm(X,Y,Z,W) = -\langle Y,Z\rangle\langle X,W\rangle + \langle X,Z\rangle\langle Y,W\rangle.$$

Now we introduce the **Kulkarni-Nomizu product** $\owedge$ that converts 2 symmetric $(0,2)$-tensors T_1 and T_2 into one $(0,4)$-tensor $T_1 \owedge T_2$ defined by

$$\begin{aligned}(T_1 \owedge T_2)(X,Y,Z,W) :=& T_1(X,Z)T_2(Y,W) + T_1(Y,W)T_2(X,Z)\\ &-T_1(X,W)T_2(Y,Z) - T_1(Y,Z)T_2(X,W).\end{aligned}$$

As a result, we get a very brief expression for the Riemann curvature tensor of S^m,

$$Rm = \frac{1}{2} g \owedge g.$$

Symmetries of Rm

By definition the $(1,3)$-tensor R admits the anti-symmetry

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

Moreover, if ∇ is torsion free, then the curvature tensor R admits two more cyclic symmetry, namely the first Bianchi identity

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

and the second Bianchi identity

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

Obviously, one can convert the symmetries of R to symmetries of Rm, namely

$$Rm(X,Y,Z,W) + Rm(Y,X,Z,W) = 0, \tag{3.4}$$

the first Bianchi identity

$$Rm(X,Y,Z,W) + Rm(Y,Z,X,W) + Rm(Z,X,Y,W) = 0, \tag{3.5}$$

and the second Bianchi identity

$$(\nabla_X Rm)(Y,Z,W,V) + (\nabla_Y Rm)(Z,X,W,V) + (\nabla_Z Rm)(X,Y,W,V) = 0, \tag{3.6}$$

or in local coordinates as

$$\begin{aligned}R_{ijkl} + R_{jikl} &= 0,\\ R_{ijkl} + R_{jkil} + R_{kijl} &= 0,\\ R_{ijkl;n} + R_{jnkl;i} + R_{nikl;j} &= 0.\end{aligned}$$

where we denote $R_{ijkl;n} = (\nabla_{\partial_n} R)(\partial_i, \partial_j, \partial_k, \partial_l)$.

By starting at the Riemann curvature tensor Rm of the standard S^m, we may find more (anti-)symmetries than the ones we have seen, e.g. one can exchange Z with W to get a negative sign, or even exchange X, Y with Z, W. In fact these two (anti-)symmetries are consequences of metric compatibility, and thus hold for any Riemannian manifold:

Proposition 3.13. The Riemann curvature tensor Rm satisfies

$$Rm(X,Y,Z,W) = -Rm(X,Y,W,Z) \tag{3.7}$$

and

$$Rm(X,Y,Z,W) = Rm(Z,W,X,Y). \tag{3.8}$$

Proof. For simplicity, we denote $f = \langle Z, Z\rangle$, then by metric compatibility,

$$\langle \nabla_X Z, Z\rangle = Xf - \langle Z, \nabla_X Z\rangle,$$

in other words,

$$\langle \nabla_X Z, Z\rangle = \frac{1}{2} Xf.$$

It follows

$$\langle \nabla_X \nabla_Y Z, Z\rangle = X\langle \nabla_Y Z, Z\rangle - \langle \nabla_Y Z, \nabla_X Z\rangle = \frac{1}{2} X(Yf) - \langle \nabla_Y Z, \nabla_X Z\rangle.$$

So

$$\begin{aligned} -Rm(X, Y, Z, Z) &= \langle R(X, Y)Z, Z\rangle \\ &= \langle \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z, Z\rangle \\ &= \frac{1}{2} X(Yf) - \frac{1}{2} Y(Xf) - \frac{1}{2}[X, Y]f \\ &= 0. \end{aligned}$$

As a consequence, we get

$$\begin{aligned} &Rm(X, Y, Z, W) + Rm(X, Y, W, Z) \\ =&Rm(X, Y, Z + W, Z + W) - Rm(X, Y, Z, Z) - Rm(X, Y, W, W) \\ =&0, \end{aligned}$$

which implies (3.7).

The equation (3.8) is a consequence of (3.7) first one together with (3.4) and (3.5). In fact, by the first Bianchi identity (3.5) one has

$$\begin{aligned} Rm(X, Y, Z, W) + Rm(Y, Z, X, W) + Rm(Z, X, Y, W) = 0, \\ Rm(Y, Z, W, X) + Rm(Z, W, Y, X) + Rm(W, Y, Z, X) = 0, \\ Rm(Z, W, X, Y) + Rm(W, X, Z, Y) + Rm(X, Z, W, Y) = 0, \\ Rm(W, X, Y, Z) + Rm(X, Y, W, Z) + Rm(Y, W, X, Z) = 0, \end{aligned}$$

Adding these equations and using (3.4) and (3.7), we get

$$Rm(Z, X, Y, W) + Rm(W, Y, Z, X) = 0,$$

which is equivalent to (3.8). □

By using (3.8), we may rewrite the second Bianchi identity (3.6) as

$$(\nabla_U Rm)(Y, Z, V, W) + (\nabla_V Rm)(Y, Z, W, U) + (\nabla_W Rm)(Y, Z, U, V) = 0, \qquad (3.6')$$

In local coordinates, the identities (3.7), (3.8) and (3.6') become

$$R_{ijkl} = -R_{ijlk}, R_{ijkl} = R_{klij}, \quad \text{and} \quad R_{ijkl;n} + R_{ijln;k} + R_{ijmk;l} = 0.$$

The Curvature Operator $\mathcal{R}$

According to (3.4) and (3.7), the Riemann curvature tensor Rm can be considered as acting on two bi-vectors $X \wedge Y$ and $Z \wedge W$ instead of acting on four vectors X, Y, Z, W. In other words, we may write Rm as

$$\widetilde{Rm} : \Lambda^2(TM) \times \Lambda^2(TM) \to \mathcal{C}^\infty(M).$$

Since the Riemannian metric on M induces an inner product on each $\Lambda^2(T_pM)$, one may convert the tensor Rm into an operator $\mathcal{R} : \Lambda^2(TM) \to \Lambda^2(TM)$ such that

$$\langle \mathcal{R}(X \wedge Y), Z \wedge W\rangle = \widetilde{Rm}(X \wedge Y, Z \wedge W) := Rm(X, Y, Z, W).$$

Moreover, the symmetry equation (3.8) implies that $\mathcal{R}$ is a self-adjoint operator on each $\Lambda^2(T_pM)$. The operator $\mathcal{R}$ is called the **curvature operator**.

3.2.2 Decomposition of the Riemann Curvature Tensor

Some Tensor Algebra: Symmetric Tensors

Let V be any vector space. Recall that $\Lambda^2 V^* \subset \otimes^2 V^*$ represents the space of anti-symmetric 2-tensors on V, while $S^2 V^* \subset \otimes^2 V^*$ represents the space of symmetric 2-tensors on V. Any 2-tensor T on V can be decomposed uniquely as the summation of a symmetric 2-tensor and an anti-symmetric 2-tensor as

$$T(u,v) = \frac{T(u,v)+T(v,u)}{2} + \frac{T(u,v)-T(v,u)}{2}.$$

If $\dim V = m$, then we have

$$\dim \Lambda^2 V^* = \frac{m(m-1)}{2}, \quad \text{and} \quad \dim S^2 V^* = \frac{m(m+1)}{2}.$$

Note that by definition, $S^2(\Lambda^2 V^*)$ contains $(0,4)$-tensors that are symmetric with respect to $(1,2) \leftrightarrow (3,4)$ and anti-symmetric with respect to $1 \leftrightarrow 2$ and $3 \leftrightarrow 4$, i.e.

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

For example, one can easily check that for any two symmetric $(0,2)$-tensor $S,T \in S^2(V^*)$, their Kulkarni-Nomizu product $S \owedge T \in S^2(\Lambda^2 V^*)$. The set $S^2(\Lambda^2 V^*)$ is a vector space of dimension

$$\dim S^2(\Lambda^2 V^*) = \frac{m(m-1)(m^2-m+2)}{8}. \tag{3.9}$$

Moreover, the space of 4-forms, $\Lambda^4 V^*$, is a subspace of $S^2(\Lambda^2 V^*)$ with dimension

$$\dim \Lambda^4 V^* = \binom{m}{4}. \tag{3.10}$$

Let $\alpha, \beta \in \Lambda^2 V^*$ be any two linear 2-forms, both viewed as skew-symmetric 2-tensors on V. Define the **symmetric product** of α and β to be the $(0,4)$-tensor

$$(\alpha \odot \beta)(X,Y,Z,W) := \alpha(X,Y)\beta(Z,W) + \alpha(Z,W)\beta(X,Y). \tag{3.11}$$

In local coordinates, one can write

$$(\alpha \odot \beta)_{ijkl} = \alpha_{ij}\beta_{kl} + \alpha_{kl}\beta_{ij}.$$

Obviously each $\alpha \odot \beta$ is in $S^2(\Lambda^2 V^*)$. It turns out that these $(0,4)$-tensors generates the whole space $S^2(\Lambda^2 V^*)$:

Lemma 3.14. Any element in $S^2(\Lambda^2 V^*)$ can be written as a linear combination of elements of the form $\alpha \odot \beta$.

Proof. Let $e^1, \ldots, e^m$ be a basis of V^*, then

$$E^1 = e^1 \wedge e^2, E^2 = e^1 \wedge e^3, \ldots, E^{m(m-1)/2} = e^{m-1} \wedge e^m$$

is a basis of $\Lambda^2 V^*$, and $E^i \odot E^j$ $(i \leqslant j)$ are linearly independent in $S^2(\Lambda^2 V^*)$ (check). By dimension counting, we see these elements form a basis of $S^2(\Lambda^2 V^*)$. □

Some Tensor Algebra: Curvature-like Tensors

To explore the cyclic symmetry that arise in the first Bianchi identity, we define

Definition 3.15. The **Bianchi symmetrization** of any $T \in S^2(\Lambda^2 V^*)$ is the 4-tensor

$$bT(X,Y,Z,W) = \frac{1}{3}(T(X,Y,Z,W) + T(Y,Z,X,W) + T(Z,X,Y,W)). \tag{3.12}$$

So the first Bianchi identity for Rm now becomes the simple equation $b(Rm) = 0$.

Definition 3.16. If $T \in S^2(\Lambda^2 V^*)$ and $b(T) = 0$, we call T a **curvature-like** tensor. The set of all curvature-like tensors is denoted by $\mathscr{C}$.

Example. For any $S, T \in S^2V^*$, we have $b(S \owedge T) = 0$ since

$$\begin{aligned} 3b(S \owedge T)_{ijkl} =& S_{ik}T_{jl} + S_{jl}T_{ik} - S_{il}T_{jk} - S_{jk}T_{il} + S_{jk}T_{li} + S_{li}T_{jk} \\ & - S_{ji}T_{lk} - S_{lk}T_{ji} + S_{lk}T_{ij} + S_{ij}T_{lk} - S_{lj}T_{ik} - S_{ik}T_{lj} \\ =& 0. \end{aligned}$$

As a result, $S \owedge T$ is a curvature-like tensor.

Note that by definition, curvature-like tensors are exactly those tensors satisfying all the algebraic symmetries that Rm admit, namely (3.4), (3.7), (3.8) and (3.5).

To study the set $\mathscr{C}$ of all curvature-like tensors, we first study the image of the Bianchi symmetrization b. Note that for any $\alpha, \beta \in \Lambda^2(V^*)$,

$$\begin{aligned} 3(b(\alpha \odot \beta))_{ijkl} &= \alpha_{ij}\beta_{kl} + \alpha_{kl}\beta_{ij} + \alpha_{jk}\beta_{il} + \alpha_{il}\beta_{jk} + \alpha_{ki}\beta_{jl} + \alpha_{jl}\beta_{ki} \\ &= \alpha_{ij}\beta_{kl} - \alpha_{ik}\beta_{jl} + \alpha_{il}\beta_{jk} + \alpha_{jk}\beta_{il} - \alpha_{jl}\beta_{ik} + \alpha_{kl}\beta_{ij} \\ &= \frac{4!}{2!2!}\frac{1}{4!}\sum_{\pi \in S_4}((\alpha \otimes \beta)^{\pi})_{ijkl} \\ &= (\alpha \wedge \beta)_{ijkl}. \end{aligned}$$

In other words, we have

Lemma 3.17. For any $\alpha, \beta \in \Lambda^2(V^*)$, $b(\alpha \odot \beta) = \dfrac{1}{3}\alpha \wedge \beta$.

In view of Lemma 3.14, we conclude that the map b has image

$$\mathrm{im}(b) = \Lambda^4 V^* \subset \Lambda^2(V^*).$$

In particular, for any $T \in S^2(\Lambda^2 V^*)$ we have $bT \in S^2(\Lambda^2 V^*)$. By definition it is straightforward to check

Lemma 3.18. For any $T \in S^2(\Lambda^2 V^*)$, one has $b(b(T)) = T$.

So the Bianchi symmetrization map b, as a linear map $b : S^2(\Lambda^2 V^*) \to S^2(\Lambda^2 V^*)$, is a projection. It follows from the standard linear algebra that

$$S^2(\Lambda^2 V^*) = \ker(b) \oplus \mathrm{im}(b) = \mathscr{C} \oplus \Lambda^4 V^*.$$

As a consequence, $\mathscr{C}$ is a vector space of dimension

$$\dim \mathscr{C} = \frac{m(m-1)(m^2-m+2)}{8} - \binom{m}{4} = \frac{1}{12}m^2(m^2-1). \tag{3.13}$$

Some Tensor Algebra: Metric Contractions

Now suppose the vector space V is endowed with an inner product $g(\cdot,\cdot) = \langle \cdot, \cdot \rangle$, so that one can identify V^* with V using the musical isomorphisms $\flat$ and $\sharp$. In particular, for any $(0,4)$-tensor T and for any vectors X, Y, Z, the linear map

$$T(Z, X, \cdot, Y) : V \to \mathbb{R},$$

is in V^* and thus be identified with a vector $\sharp T(Z, X, \cdot, Y) \in V$.

Definition 3.19. For any $(0,4)$-tensor T on an inner product space $(V, \langle \cdot, \cdot \rangle)$, the **Ricci contraction** $c(T)$ of T is the following $(0,2)$-tensor:[3]

$$c(T)(X, Y) := \mathrm{tr}(Z \mapsto \sharp T(Z, X, \cdot, Y)). \tag{3.14}$$

For a Riemannian manifold, we call $Rc := c(Rm)$ its **Ricci curvature tensor**.

[3] One can define metric contractions between other pairs of indices. However, for curvature-like tensors, what one get are either Ricci contraction, or zero.

It turns out that $c(T)$ is symmetric if T is curvature-like:

Lemma 3.20. If $T \in \mathscr{C}$, then $c(T)(X,Y) = c(T)(Y,X)$.

Proof. Fix X and Y. Let $K : V \to V$ and $\widetilde{K} : V \to V$ be the maps

$$K(Z) = \sharp T(Z, X, \cdot, Y) \quad \text{and} \quad \widetilde{K}(Z) = \sharp T(\cdot, X, Z, Y).$$

Then for any $Z, W \in V$,

$$\langle K(Z), W\rangle = \langle \sharp T(Z, X, \cdot, Y), W\rangle = T(Z, X, W, Y) = \langle Z, \sharp T(\cdot, X, W, Y)\rangle = \langle Z, \widetilde{K}(W)\rangle.$$

So $\widetilde{K}$ is the transpose of K, and thus they have the same trace. But by definition $\mathrm{tr}(K) = c(T)(X,Y)$, while $\mathrm{tr}(\widetilde{K}) = c(T)(Y,X)$ (since $T \in \mathscr{C} \subset S^2(\Lambda^2 V^*)$). □

Similarly one can define the trace of a $(0,2)$-tensor T using the metric: the map

$$T(X, \cdot) : V \to \mathbb{R}$$

is an element in V^*, and thus can be identified with a vector $\sharp T(X, \cdot)$ in V.

Definition 3.21. The **trace** of a $(0,2)$-tensor T on $(V, \langle \cdot, \cdot \rangle)$ is

$$\mathrm{tr}(T) := \mathrm{trace}(X \mapsto \sharp T(X, \cdot)). \tag{3.15}$$

For a Riemannian manifold (M,g), we call $S(g) = \mathrm{tr}(Rc)$ the **scalar curvature**.

Example. For the metric tensor g, by definition $\sharp g(X, \cdot) = X$ and thus $\mathrm{tr}(g) = m$.

Locally if $v^1, \dots, v^n$ be a basis of V^*, and if we denote $g^{pq} = g^*(v^p, v^q)$ (where g^* is the dual metric on V^*), then one can check [as exercise]

$$c(T)_{ij} = g^{pq} T_{ipjq} \quad \text{and} \quad \mathrm{tr}(T) = g^{ij} T_{ij}.$$

Note that $\langle T, g\rangle = T_{ij} g_{kl} g^{ik} g^{jl} = T_{ij} \delta^i_l g^{jl} = T_{ij} g^{ij}$, one has

$$\mathrm{tr}(T) = \langle T, g\rangle. \tag{3.16}$$

The Dual of the Ricci Contraction

Now let's study the Ricci contraction map

$$\begin{aligned} c : \mathscr{C} &\to S^2 V^*, \\ S &\mapsto c(S) \end{aligned}$$

which maps a curvature-like tensor to a symmetric 2-tensor. We also have a map

$$\begin{aligned} \Psi : S^2 V^* &\to \mathscr{C}, \\ T &\mapsto T \owedge g. \end{aligned} \tag{3.17}$$

that maps any symmetric 2-tensor to a curvature-like tensor. It turns out with respect to the induced metrics on the spaces of tensors that we learned in Section 1.2, these two maps are almost adjoint to each other:

Lemma 3.22. For any $S \in \mathscr{C}$ and any $T \in S^2 V^*$, one has

$$\langle S, \Psi(T)\rangle = 4\langle c(S), T\rangle.$$

Proof. We calculate using an orthonormal basis of V, so that $g_{ij} = g^{ij} = \delta_{ij}$:

$$\begin{aligned}\langle S, T \owedge g\rangle &= \sum S_{ijkl}(T_{ik}g_{jl} + T_{jl}g_{ik} - T_{jk}g_{il} - T_{il}g_{jk}) \\ &= \sum (S_{ijkl}T_{ik} + S_{ijik}T_{jk} - S_{ijki}T_{jk} - S_{ijjk}T_{ik}) \\ &= \sum 4S_{ijkj}T_{ik} \\ &= 4\sum (c(S))_{ik}T_{ik} \\ &= 4\langle c(S), T\rangle.\end{aligned}$$

where each summation is over all indices that appeared. ☐

In particular, if $S_1 = \Psi(T_1) \in \mathrm{im}(\Psi)$, $S_2 \in \ker(c)$, then

$$\langle S_1, S_2\rangle = \langle \Psi(T_1), S_2\rangle = 4\langle T_1, c(S_2)\rangle = 0.$$

So we get

Corollary 3.23. $\mathrm{im}(\Psi) \perp \ker(c)$.

Similarly, we may calculate $c \circ \Psi$ via an orthonormal basis: for any $T \in S^2V^*$,

$$\begin{aligned}c(T \owedge g)_{ij} &= g^{pq}(T_{ij}g_{pq} - T_{pj}g_{iq} - T_{iq}g_{pj} + T_{pq}g_{ij}) \\ &= mT_{ij} - T_{ij} - T_{ij} + \mathrm{tr}(T)g_{ij} \\ &= (m-2)T_{ij} + \mathrm{tr}(T)g_{ij}.\end{aligned}$$

In other words, we have

Lemma 3.24. For any symmetric 2-tensor $T \in S^2V^*$,

$$c(\Psi(T)) = (m-2)T + \mathrm{tr}(T)g.$$

Together with (3.16), we get

$$|\Psi(T)|^2 = \langle \Psi(T), \Psi(T)\rangle = 4\langle c(\Psi(T)), T\rangle = 4(m-2)|T|^2 + 4(\mathrm{tr}(T))^2. \tag{3.18}$$

Decomposition of a Curvature-like Tensor via the Metric: Step 1

Now suppose $m > 2$. We first prove

Proposition 3.25. The map Ψ is injective for $m > 2$, and is bijective for $m = 3$.

Proof. Suppose $m \geqslant 2$ and $\Psi(T) = 0$. Then by (3.18), $T = 0$, so Ψ is injective. For $m = 3$ it is bijective since $\dim S^2V^* = \dim \mathscr{C} = 6$. ☐

As a result, the Ricci contraction map $c : \mathscr{C} \to S^2V^*$ is surjective, and thus

$$\dim \mathrm{im}(\Psi) + \dim \ker(c) = \dim S^2V^* + \dim \ker(c) = \dim \mathrm{im}(c) + \dim \ker(c) = \dim \mathscr{C}.$$

So by Corollary 3.23, we really have an orthogonal decomposition

$$\mathscr{C} = \ker(c) \oplus \mathrm{im}(\Psi).$$

In particular, for any curvature-like tensor $T \in \mathscr{C}$, there is a curvature-like tensor $W \in \ker(c)$ and a symmetric 2-tensor $A \in S^2V^*$ so that

$$T = W + A \owedge g,$$

and the decomposition is orthogonal. To find out A, we apply c to both sides to get

$$c(T) = c(\Psi(A)) = (m-2)A + \mathrm{tr}(A)g.$$

To find out $\mathrm{tr}(A)$, we continue to take $\mathrm{tr}(A)$ for both sides,

$$\mathrm{tr}(c(T)) = (m-2)\mathrm{tr}(A) + m\mathrm{tr}(A) = 2(m-1)\mathrm{tr}(A).$$

So we get

$$A = \frac{1}{m-2}\left(c(T) - \frac{\mathrm{tr}(c(T))}{2(m-1)}g\right). \tag{3.19}$$

For a Riemannian manifold (M, g), we have

$$c(Rm) = Rc \quad \text{and} \quad \mathrm{tr}(c(Rm)) = S.$$

In this case, the tensors W and A have their own names:

Definition 3.26. For a Riemannian manifold (M, g), we call

$$A = \frac{1}{m-2}\left(Rc - \frac{S}{2(m-1)}g\right)$$

the **Schouten tensor** of (M, g), and call

$$W := Rm - A \owedge g$$

the **Weyl curvature tensor** (or conformal curvature tensor) of (M, g).

Both Weyl curvature tensor and the Schouten tensor play very important roles in conformal geometry. For example, we will show that the Weyl tensor is invariant under conformal transformations. Note that by Corollary 3.23 and Proposition 3.25, the Weyl curvature tensor vanishes for $m = 3$:

Proposition 3.27. If $m = 3$, then $W = 0$.

Decomposition of a Curvature-like Tensor via the Metric: Step 2

We may continue to decompose any $R \in S^2V^*$ into orthogonal ones via the map $\mathrm{tr} : S^2V^* \to \mathbb{R}$. Again we need the dual of tr, which, in view of (3.16), is simply

$$\mathrm{tr}^* : \mathbb{R} \to S^2V^*,$$
$$t \mapsto tg.$$

By repeating the same arguments, again we get an orthogonal decomposition

$$S^2V^* = \ker(\mathrm{tr}) \oplus \mathrm{im}(\mathrm{tr}^*).$$

So there exist $E \in \ker(\mathrm{tr})$ and $t \in \mathbb{R}$ so that R can be decomposed orthogonally to

$$R = E + tg,$$

Applying the map tr to both sides we get $\mathrm{tr}(R) = t\mathrm{tr}(g) = mt$, and thus

$$t = \frac{\mathrm{tr}(R)}{m}.$$

Note that if $R = Rc$ is the Ricci curvature tensor of (M, g), then $t = \dfrac{S}{m}$.

Definition 3.28. For a Riemannian manifold (M, g), we call

$$E := Rc - \frac{S}{m}g$$

the **traceless Ricci tensor** of (M, g).

We point out that the two decompositions are compatible, in the sense that

$$\mathrm{im}(\Psi) = \Psi(S^2V^*) = \Psi(\ker(\mathrm{tr})) \oplus \Psi(\mathrm{im}(\mathrm{tr}^*))$$

is an orthogonal decomposition of $\mathrm{im}(\Psi)$, since for $t \in \mathbb{R}$ and any $E \in \ker(\mathrm{tr})$,

$$\langle \Psi(tg), \Psi(E) \rangle = \langle c(\Psi(tg)), E \rangle = \langle (2m-2)(tg), E \rangle = 2(m-2)t\mathrm{tr}(E) = 0.$$

Thus we end up with an orthogonal decomposition

$$\mathscr{C} = \ker(c) \oplus \Psi(\ker(\mathrm{tr})) \oplus \Psi(\mathrm{im}(\mathrm{tr}^*)),$$

so that any $T \in \mathscr{C}$ can be decomposed into three curvature-like tensors which are orthogonal to each other: the "Weyl part" W that lies in $\ker(c)$, $E \owedge g$ for a trace-less symmetric 2-tensor E, and a multiple of $g \owedge g$.

Remark. A more algebraic way to understand the two decompositions: Consider the natural action of $O(m)$ on V that preserves $\langle \cdot, \cdot \rangle$, which induces natural actions of $O(m)$ on S^2V^* and on $\mathscr{C}$ that preserve the induced inner products.

- For the case of S^2V^*, since g (and thus the 1-dimensional space $\mathbb{R}g$) is invariant under the $O(m)$-action, one may decompose $S^2V^* = \mathbb{R}g \oplus (\mathbb{R}g)^\perp$. This decomposition can be explained via the $O(m)$-invariant map tr as above, and thus $(\mathbb{R}g)^\perp = \ker(\mathrm{tr})$ consists of all traceless symmetric $(0,2)$-tensors.
- Similarly since c is $O(m)$-equivalent, it induce a decomposition of $\mathscr{C}$ into $\ker(c) \oplus (\ker(c))^\perp$, and we have seen $(\ker(c))^\perp = \Psi(S^2V^*)$ since c is surjective.

Can we decompose further? The answer is no, since one can prove the $O(m)$-action on $\ker(\mathrm{tr})$ and on $\ker(c)$ are transitive (i.e. they are irreducible representations of $O(m)$).

Decomposition of the Riemann Curvature Tensor

Now suppose (M, g) be a Riemannian manifold. First we have a decomposition

$$Rm = W + A \owedge g.$$

We may continue to decompose the Schouten tensor A: By definition formula (3.19), the traceless part of A equals the traceless part of $\frac{Rc}{m-2}$, which is $\frac{E}{m-2}$. Since

$$\frac{1}{m-2}\frac{S}{m}g - \frac{1}{m-2}\frac{S}{2(m-1)}g = \frac{S}{2m(m-1)}g,$$

we get the orthogonal decomposition of Schouten tensor:

$$A = \frac{E}{m-2} + \frac{S}{2m(m-1)}g.$$

So we end up with an orthogonal decomposition

$$Rm = W + \frac{1}{m-2}E \owedge g + \frac{S}{2m(m-1)}g \owedge g \tag{3.20}$$

Our final goal in this section is to prove

Theorem 3.29. For any Riemannian manifold, the (pointwise) norm squares of the Riemann/Ricci/Weyl curvature tensors and the scalar curvature are related by

$$|Rm|^2 = |W|^2 + \frac{4}{m-2}|Rc|^2 - \frac{2}{(m-1)(m-2)}S^2.$$

Proof. In view of (3.18) and the fact $\operatorname{tr}(g) = |g|^2 = m$, we have

$$|E \owedge g|^2 = 4(m-2)|E|^2,$$
$$|g \owedge g|^2 = 4(m-2)m + 4m^2 = 8m(m-1).$$

Since the decomposition (3.20) is orthogonal, we get

$$|Rm|^2 = |W|^2 + \frac{4}{m-2}|E|^2 + \frac{2}{m(m-1)}S^2.$$

Finally, we use

$$\begin{aligned} |E|^2 &= \left\langle Rc - \frac{S}{m}g, Rc - \frac{S}{m}g \right\rangle \\ &= |Rc|^2 - \frac{2S}{m}\langle Rc, g\rangle + \frac{S^2}{m^2}|g|^2 \\ &= |Rc|^2 - \frac{2S^2}{m} + \frac{S^2}{m} \\ &= |Rc|^2 - \frac{S^2}{m} \end{aligned}$$

to get

$$|Rm|^2 = |W|^2 + \frac{4}{m-2}|Rc|^2 - \frac{2}{(m-1)(m-2)}S^2.$$

□

3.3 The Ricci and the Sectional Curvature

3.3.1 The Ricci and the Sectional Curvature

The Ricci curvature of a Riemannian Manifold

We start with some simple algebra. Let $B : V \times V \to \mathbb{R}$ be a symmetric bilinear form defined on a vector space V. Then we can assign to it a quadratic form

$$Q : V \to \mathbb{R},$$
$$Q(v) := B(v, v).$$

Conversely, we can recover the symmetric bilinear form B from its quadratic form Q via the polarization formula

$$2B(u, v) = Q(u + v) - Q(u) - Q(v).$$

There is also a more succinct way to recover B from Q is via

$$2B(u, v) = [Q(u + tv)]'(0),$$

where $'$ refers to t-derivative.

Recall that the Ricci curvature tensor Ric is the contraction of the Riemann curvature tensor Rm,

$$Rc(X, Y) = c(Rm)(X, Y) = \operatorname{tr}(Z \mapsto \sharp Rm(Z, X, \cdot, Y)).$$

It is a symmetric $(0, 2)$-tensor field on M. In local coordinates one has

$$Rc_{ij} = g^{pq}R_{ipjq}$$

Applying the previous trick to Rc, we see that to study the $(0, 2)$-tensor Rc, it is enough to study the real-valued function $Rc(X, X)$ defined on TM, which is easier to handle. In view of the fact $Rc(\lambda X, \lambda X) = \lambda^2 Rc(X, X)$, we may simplify a bit further by studying $Rc(X, X)$ only for unit-length vector fields $X \in \Gamma^\infty(SM)$:

Definition 3.30. For any unit tangent vector $X_p \in S_pM \subset T_pM$, we call

$$Ric(X_p) = Rc(X_p, X_p)$$

the **Ricci curvature** of M at p in the direction of X_p.

So the Ricci curvature function Ric is not a function on M, but a function on the unit sphere bundle $SM \subset TM$ (one can think of the Ricci curvature as a function defined on one-dimensional subspace of T_pM). It encodes all information of the tensor Rc via

$$Rc(X_p, Y_p) = \frac{1}{2}[\|X_p + Y_p\|^2 Ric(\widehat{X_p + Y_p}) - \|X_p\|^2 Ric(\widehat{X_p}) - \|Y_p\|^2 Ric(\widehat{Y_p})].$$

where $Y_p \neq -X_p$, and we denoted $\widehat{X} = X/\|X\|$.

Reduce a Curvature-like Tensor to its Bi-quadratic from

Now we move from symmetric $(0,2)$-tensor to "very symmetric" $(0,4)$-tensors that we studied last time, namely, curvature-like tensors T. We let

$$Q(X, Y) := T(X, Y, X, Y)$$

be the bi-quadratic form associated to T. It turns out that one can recover T from Q in the same spirit above:

Lemma 3.31. Let T be a curvature-like tensor, and let

$$f_{X,Y,Z,W}(t) = Q(X + tZ, Y + tW) - t^2(Q(X, W) + Q(Z, Y)).$$

Then $(f_{X,Y,Z,W} - f_{Y,X,Z,W})''(0) = 12T(X, Y, Z, W)$.

Proof. Obviously $f_{X,Y,Z,W}$ is a polynomial of degree 4 in t, whose quadratic coefficients equals

$$T(Z, W, X, Y) + T(Z, Y, X, W) + T(X, W, Z, Y) + T(X, Y, Z, W).$$

which, by using the symmetries for curvature-like tensors, equals

$$2T(X, Y, Z, W) + 2T(Z, Y, X, W).$$

Similarly the quadratic coefficient of $f_{Y,X,Z,W}$ equals

$$2T(Y, X, Z, W) + 2T(Z, X, Y, W).$$

So the quadratic coefficient of $f_{X,Y,Z,W}(t) - f(Y, X, Z, W)(t)$ is

$$2T(X, Y, Z, W) + 2T(Z, Y, X, W) - 2T(Y, X, Z, W) - 2T(Z, X, Y, W),$$

which, after applying the first Bianchi identity, equals $6T(X, Y, Z, W)$. □

Remark. One may explicitly write down a "pure algebraic polarization formula" of $T(X, Y, Z, W)$ in terms of the bi-quadratic form Q, which is quite lengthy.

The Sectional Curvature

As a result, to study the curvature tensor Rm of a Riemannian manifold (M, g), it is enough to study the associated bi-quadratic form, namely, $Rm(X, Y, X, Y)$.

Again, by using some simple algebra, we can simplify a bit further.

Lemma 3.32. For any $T \in \otimes^2(\Lambda^2 V^*)$ and any $X, Y \in V$, if we denote $X' = aX + bY$, $Y' = cX + dY$, then

$$T(X', Y', X', Y') = (ad - bc)^2 T(X, Y, X, Y).$$

Proof. This follows from a very simple computation:

$$\begin{aligned}T(X', Y', X', Y') &= T(aX + bY, cX + dY, aX + bY, cX + dY)\\ &= (ad - bc)T(X, Y, aX + bY, cX + dY)\\ &= (ad - bc)^2 T(X, Y, X, Y).\end{aligned}$$

□

Now suppose (M, g) is a Riemannian manifold. Recall that $\frac{1}{2}g \owedge g$ is a curvature-like tensor, such that

$$\frac{1}{2}g \owedge g(X_p, Y_p, X_p, Y_p) = \langle X_p, X_p\rangle\langle Y_p, Y_p\rangle - \langle X_p, Y_p\rangle^2,$$

which is nothing else but the square of the area of the parallelogram with sides X_p, Y_p. Applying the previous lemma to Rm and $\frac{1}{2}g \owedge g$, we immediately get

Proposition 3.33. The quantity

$$K(X_p, Y_p) := \frac{Rm(X_p, Y_p, X_p, Y_p)}{\langle X_p, X_p\rangle\langle Y_p, Y_p\rangle - \langle X_p, Y_p\rangle^2}$$

depends only on the two dimensional plane $\Pi_p = \mathrm{span}(X_p, Y_p) \subset T_pM$, i.e. it is independent of the choices of basis $\{X_p, Y_p\}$ of Π_p.

Definition 3.34. We will call

$$K(\Pi_p) = K(X_p, Y_p)$$

the **sectional curvature** of (M, g) at p with respect to the plane Π_p.

Note that the Ricci curvature that we studied above are closely related to the sectional curvatures: If X_p is a unit vector in T_pM, we may extend X_p to an orthonormal basis $\{e_1 = X_p, e_2, \dots, e_m\}$ of T_pM. As a result,

$$Ric(X_p) = \sum_{i\geqslant 2} Rm(e_i, e_1, e_i, e_1) = \sum_{i\geqslant 2} K(e_i, e_1). \tag{3.21}$$

In other words, the Ricci curvature $Ric(X_p)$ is "the sum of sectional curvatures" for a specially chosen set of $m - 1$ pairwise orthogonal planes containing X_p.

Example. Here are three basic examples:

(1) For the Euclidean space $(\mathbb{R}^m, g_0)$, one has $Rm = 0$, so

$$K(\Pi_p) \equiv 0 \quad \text{and} \quad Ric(X_p) \equiv 0.$$

(2) For the unit sphere (S^m, g_{round}), one has $Rm = \frac{1}{2}g \owedge g$, so

$$K(\Pi_p) \equiv 1 \quad \text{and} \quad Ric(X_p) \equiv m - 1.$$

One may also prove the conclusion by calculating the Christoffel symbols.

(3) For the hyperbolic space $(H^m, g_{\text{hyperbolic}})$, one can prove (as exercise)

$$K(\Pi_p) \equiv -1 \quad \text{and} \quad Ric(X_p) \equiv -(m - 1).$$

Remark. One may give a conceptional proof of the fact that these three spaces have constant sectional curvature: For $(\mathbb{R}^m, g_0)$, the isometry group $E(m) = O(m) \ltimes \mathbb{R}^m$ acts transitively on the set $Gr_2(T\mathbb{R}^m) = \{\Pi_p \mid p \in \mathbb{R}^m, \Pi_p \subset T_pM \text{ is a plane}\}$. Since the sectional curvature is invariant under the action of the isometric group, we must have $K(\Pi_p) = K(\Pi'_q)$, i.e. the sectional curvature is a constant. The same phenomena occurs for (S^m, g_{round}) (with isometry group $O(m + 1)$) and for $(H^m, g_{\text{hyperbolic}})$ (with isometry group $O^+(m, 1)$).

Remark. We may compare the sectional curvature K with the curvature operator $\mathcal{R}_p : \Lambda^2(T_pM) \to \Lambda^2(T_pM)$. By definition, we have

$$K(X_p, Y_p) = \frac{\langle \mathcal{R}_p(X_p \wedge Y_p), X_p \wedge Y_p \rangle}{\langle X_p, X_p \rangle \langle Y_p, Y_p \rangle - \langle X_p, Y_p \rangle^2}.$$

As a consequence, the curvature operator $\mathcal{R}$ determines the sectional curvature K.

3.3.2 Basic Properties of Sectional Curvatures

Sectional Curvature for Low Dimensional Manifolds

Let (M, g) be a Riemannian manifold. By definition, the sectional curvature K is NOT a function on M, but instead a function on the Grassmannian bundle

$$Gr_2(TM) = \{(p, \Pi_p) \mid p \in M, \Pi_p \subset T_pM \text{ is 2-dimensional}\}.$$

(1) If M has dimension $m = 1$, obviously $R \equiv 0$ and it makes no sense to talk about sectional curvature (so again as we have seen in Section 1.3, essentially there is no Riemannian geometry in dimensional one).

(2) If M has dimension $m = 2$, i.e. is a surface, then one can regard the sectional curvature K as a function defined on M in the natural way. Moreover, in view of the equation (3.21), for any direction X_p we have $Ric(X_p) = K(p)$. In other words, for surfaces the sectional curvature, the Ricci curvature and the scalar curvature are all the same. One can show that in this case K is really the Gauss curvature in undergraduate differential geometry course.

(3) If M has dimension $m \geqslant 3$, by using the exponential map that we will learn later, for each 2-dimensional plane $\Pi_p \in T_pM$, locally one gets a 2-dimensional submanifold S_p near p in M whose tangent space at p is Π_p, and the sectional curvature $K(\Pi_p)$ is nothing else but the Gauss curvature of S_p (endowed with the subspace Riemannian metric) at p. Thus the sectional curvature is a natural generalization of the Gauss curvature to higher dimensional Riemannian manifolds.

(4) In general, the sectional curvatures encodes more information than the Ricci curvatures. However, if M has dimension $m = 3$, then according to the equation (3.21), for an orthonormal basis e_1, e_2, e_3.

$$\begin{aligned} Ric(e_1) &= K(e_1, e_2) + K(e_1, e_3), \\ Ric(e_2) &= K(e_1, e_2) + K(e_2, e_3), \\ Ric(e_3) &= K(e_1, e_3) + K(e_2, e_3). \end{aligned}$$

As a result, in this case the Ricci curvatures determine the sectional curvatures: For each plane Π_p, one just start with an orthonormal basis $\{e_1, e_2\}$ of Π_p, extend it to an orthonormal basis $\{e_1, e_2, e_3\}$ of T_pM and then solve the above system of equations to get

$$2K(\Pi_p) = Ric(e_1) + Ric(e_2) - Ric(e_3).$$

Riemannian Manifolds with Curvature Bounds

Unlike algebraic quantities like curvature tensors, the sectional/Ricci/scalar curvatures are real-valued functions. Since we may compare real numbers, we can define

Definition 3.35. Let (M, g) be a Riemannian manifold. We say (M, g) has

(1) constant sectional curvature c if $K(\Pi_p) = c$ for all p and all planes $\Pi_p \subset T_pM$.

(2) constant Ricci curvature c if $Ric(X_p) = c$ for all p and all vectors $X_p \in S_pM$.

(3) positive sectional curvature if $K(\Pi_p) > 0$ for all p and all planes $\Pi_p \subset T_pM$.

Similarly, we may define

- (M, g) has negative/nonpositive/nonnegative sectional curvature if $K(\Pi_p)$ is negative/nonpositive/nonnegative for all p and all planes $\Pi_p \subset T_pM$.
- (M, g) has positive/negative/nonpositive/nonnegative Ricci curvature if $Ric(X_p)$ is positive/negative/nonpositive/nonnegative for all p and all $X_p \in S_pM$.
- (M, g) has sectional curvature $K \geqslant c$ or $K \leqslant c$ if $K(\Pi_p) \geqslant c$ or $K(\Pi_p) \leqslant c$ for all p and all planes $\Pi_p \subset T_pM$.
- (M, g) has Ricci curvature $Ric \geqslant c$ or $Ric \leqslant c$ if $Ric(X_p) \geqslant c$ or $Ric(X_p) \leqslant c$ for all p and all vectors $X_p \in S_pM$.

More generally, given two Riemannian metrics g_1 and g_2 on any smooth manifold M, with sectional curvature functions K_{g_1}, K_{g_2} and Ricci curvature functions Ric_{g_1}, Ric_{g_2}, we may compare K_{g_1} and K_{g_2} as functions on $Gr_2(TM)$, and compare Ric_{g_1} and Ric_{g_2} as functions on SM.

Example. We can write down the change of sectional curvature under scaling of the metric: if we scale a Riemannian metric g to λg, where λ is a positive constant, then

(1) by the Koszul formula, the Levi-Civita connection $\nabla_X Y$ remains unchanged,

(2) it follows that the $(1, 3)$-curvature tensor R remains unchanged,

(3) as a result, the Riemann curvature tensor is changed to $Rm_{\lambda g} = \lambda Rm_g$,

(4) and thus the Ricci curvature tensor remains unchanged: $Rc_{\lambda g} = Rc_g$,

(5) but the two sectional curvatures are related by $K_{\lambda g} = \lambda^{-1} K_g$,

(6) and it follows that $Ric_{\lambda g} = \lambda^{-1} Ric_g$ (no conflict with (4) since unit vectors changed)

(7) and thus $S_{\lambda g} = \lambda^{-1} S_g$.

In particular, for each c one gets a Riemannian manifold with constant sectional curvature c, namely the space $\left(S^m, \frac{1}{c} g_{\text{round}}\right)$ if $c > 0$, and $\left(H^m, \frac{1}{-c} g_{\text{hyperbolic}}\right)$ if $c < 0$.

Manifolds with curvature bounds will be one of the major themes of this course.

Remark. Since the curvature operator $\mathcal{R}_p$ is a symmetric operator on a real vector space, it has real eigenvalues.

Definition 3.36. We say (M, g) is a Riemannian manifold with **positive curvature operator** if all eigenvalues of $\mathcal{R}_p$ are positive.

Obviously if (M, g) has positive curvature operator, then it has positive sectional curvature. However, the converse is not true:

- It was proven by C. Bohm and B. Wilking in 2008 that manifolds with positive curvature operators are space forms, i.e. are complete Riemannian manifolds with constant sectional curvature.
- On the other hands, there exist Riemannian manifolds with positive sectional curvature which are not constant (e.g. the complex projective space $\mathbb{CP}^m$ endowed with the Fubini-Study metric).

So for such Riemannian manifolds the curvature operator is not positive. The secret is: the space $\Lambda^2(T_pM)$ is a vector space that contains elements of the form $u_1 \wedge v_1 + u_2 \wedge v_2$ which do not correspond to any 2-dimensional plane in T_pM. After all, the Grassmannian $Gr_2(T_pM)$ is a smooth manifold of dimension $2(m-2)$, while the space of bi-vectors $\Lambda^2(T_pM)$ has dimension $\binom{m}{2}$.

Riemannian Manifolds with Isotropic Sectional Curvature at a Point

Finally study the following question: At a given point, when will the sectional curvature be independent of the choice of $\Pi_p \subset T_pM$?

Proposition 3.37. Let (M, g) be a Riemannian manifold and $p \in M$. The following are equivalent:

(1) $K(\Pi_p) = c$ for all $\Pi_p \subset T_pM$.

(2) $Rm_p = \frac{c}{2} g_p \owedge g_p$.

(3) $R_p(X_p, Y_p)Z_p = c(\langle Y_p, Z_p\rangle X_p - \langle X_p, Z_p\rangle Y_p)$ for any X_p, Y_p, $Z_p \in T_pM$.

(4) $\mathcal{R}_p = c\mathrm{Id}$ on $\Lambda^2 T_pM$.

(5) The Weyl curvature tensor $W_p = 0$ and Ricci curvature tensor $Rc_p = (m-1)cg_p$.

Proof. (1)⇔(2): According to Lemma 3.31, if T is a curvature-like tensor, then

$$T \equiv 0 \Leftrightarrow T(X, Y, X, Y) = 0, \ \forall X, Y.$$

Apply this to the curvature-like tensor $T = Rm_p - \frac{c}{2} g_p \owedge g_p$, we see

$$Rm_p = \frac{c}{2} g_p \owedge g_p \Leftrightarrow K(\Pi_p) = c, \ \forall \Pi_p \subset T_pM.$$

(2)⇔(3) and (4)⇒(1) are obvious.

(2)⇒(4): Take an orthonormal basis $\{e_i\}$ of T_pM, then $\{e_i \wedge e_j \mid i < j\}$ is a basis of $\Lambda^2 T_pM$. On this basis,

$$\langle \mathcal{R}_p(e_i \wedge e_j), e_k \wedge e_l\rangle = Rm_p(e_i, e_j, e_k, e_l) = c\delta_{ik}\delta_{jl},$$

where in the last step we used the fact $i < j$, $k < l$. As a result, we see

$$\mathcal{R}_p(e_i \wedge e_j) = ce_i \wedge e_j, \ \forall i < j.$$

In other words, $\mathcal{R}_p = c\mathrm{Id}$ for all $e_i \wedge e_j$. Since $\mathcal{R}_p$ is linear, $\mathcal{R}_p = c\mathrm{Id}$ on $\Lambda^2 T_pM$.

(2)⇒(5): We start with the unique orthogonal decomposition

$$Rm_p = W_p + \frac{1}{m-2} E_p \owedge g_p + \frac{S(p)}{2m(m-1)} g_p \owedge g_p,$$

where $E_p = Rc_p - \frac{S(p)}{m} g_p$ is the traceless Ricci tensor. If (2) holds, then by the uniqueness of the decomposition,

$$W_p = 0, \quad E_p = 0 \quad \text{and} \quad S(p) = m(m-1)c.$$

As a result,

$$Rc_p = (m-1)cg_p.$$

(5)⇒(2): Conversely if $W_p = 0$ and $Rc_p = (m-1)cg_p$, then

$$S(p) = \mathrm{tr}(Rc_p) = m(m-1)c.$$

So $E_p = 0$ and thus

$$Rm_p = \frac{c}{2} g_p \owedge g_p.$$

This completes the proof. □

Recall from Section 3.1 that a Riemannian manifold is flat if the $(1,3)$-curvature tensor $R = 0$. As a consequence,

Corollary 3.38. A Riemannian manifold (M, g) is flat manifold if and only if its sectional curvatures are identically zero.

Similarly by using the polarization formula for Ricci curvature tensor, we may also easily get a Ricci curvature version:

Proposition 3.39. Let (M, g) be a Riemannian manifold and $p \in M$. Then $Ric(X_p) = c$ for all $X_p \in S_pM$ if and only if $Rc_p = cg_p$.

3.4 Riemannian Manifolds with Constant Curvatures

On any smooth manifold there are numerous different Riemannian metrics, most of which are not interesting to us. Today we will briefly discuss some results on very special Riemannian metrics, namely Riemannian metrics with constant curvatures (including sectional, Ricci, scalar and Einstein curvature).

3.4.1 Schur's Theorem: From Fiber Constant to Constant

As we have seen, the sectional curvature and the Ricci curvature are functions not defined on M itself, but defined on some fiber bundles over M, namely the Grassmannian 2-plane bundle $Gr_2(TM)$ and the sphere bundle SM. Before we study Riemannian manifolds with constant sectional or Ricci curvatures, let's first study an "intermediate" case, namely Riemannian manifolds whose sectional or Ricci curvatures are fiber-wise constant. It turns out that for connected Riemannian manifolds of dimension $m \geqslant 3$, fiber-wise constant sectional/Ricci curvature will force the Riemannian manifold to have globally constant sectional/Ricci curvature. This result was first established by German mathematician F. Schur in 1886 (for the sectional curvature case).

The Contracted Bianchi Identity

The main tool in the proof of Schur's theorem is the second Bianchi identity

$$(\nabla Rm)(U,V,X,Y,Z) + (\nabla Rm)(U,V,Y,Z,X) + (\nabla Rm)(U,V,Z,X,Y) = 0.$$

We will first prove the stronger Ricci curvature version of Schur's theorem, for which what we need is

Proposition 3.40 (The Contracted Bianchi Identity)**.** For any Riemannian manifold,

$$\nabla S = 2c_{1,3}\nabla Rc,$$

where $c_{1,3}$ is the metric contraction in the first and third entry.

Proof. Since the metric contractions commute with[4] ∇, we may apply metric contractions to the Bianchi identity. Contracting the first and the third entries then contract the second and fourth entries:

$$0 = \sum_{\bigcirc 3,4,5} (\nabla Rm) \Rightarrow 0 = c_{2,4}c_{1,3} \sum_{\bigcirc 3,4,5} (\nabla Rm).$$

Note that

$$c_{V,Y}c_{U,X}(\nabla Rm)(U,V,X,Y,Z) = \nabla(c_{V,Y}c_{U,X}Rm)(U,V,X,Y,Z) = (\nabla S)(Z),$$

while

$$\begin{aligned} c_{V,Y}c_{U,X}(\nabla Rm)(U,V,Y,Z,X) &= -c_{U,X}(\nabla c_{V,Y}Rm)(U,V,Z,Y,X) \\ &= -c_{U,X}(\nabla Rc)(U,Z,X) \end{aligned}$$

and

$$\begin{aligned} c_{V,Y}c_{U,X}(\nabla Rm)(U,V,Z,X,Y) &= -c_{V,Y}\nabla(c_{U,X}Rm)(U,V,X,Z,Y) \\ &= -c_{V,Y}(\nabla Rc)(V,Z,Y) \end{aligned}$$

So we arrived at

$$\nabla S = 2c_{1,3}\nabla Rc,$$

which completes the proof. □

In local coordinates, the **contracted Bianchi identity** can be written as

$$\partial_k S = 2g^{ij}Rc_{ik;j}.$$

[4]We note that the metric compatibility implies $\nabla g^* = 0$, where $g^* = g^{ij}\partial_i\partial_j$ is the "dual of g". This in turn implies that the musical isomorphisms commute with the covariant derivative ∇. Now consider the **metric contraction** $c_{i,j}$ that contracts the ith entry with the jth entry of a $(0,k)$ tensor T, where $1 \leqslant i \neq j \leqslant k$. Since $c_{i,j}$ can be written as a composition of the standard contraction C^i_j with a musical isomorphism, and since ∇ commutes with C^i_j, one can prove that ∇ also commutes with the metric contraction $c_{i,j}$.

Schur's Theorem

Now we are ready to prove

Theorem 3.41 (Schur)**.** Let (M, g) be a connected Riemannian manifold of dimension $m \geqslant 3$.

(1) If $Ric(X_p) = f(p)$ depends only on p, then (M, g) has constant Ricci curvature.

(2) If $K(\Pi_p) = f(p)$ depends only on p, then (M, g) has constant sectional curvature.

Proof. (1) Under the assumption we have $Rc_p = f(p)g_p$. It follows

$$S(p) = \mathrm{tr}(Rc_p) = f(p)\mathrm{tr}(g_p) = mf(p).$$

So by the contracted Bianchi identity and the fact $\nabla g = 0$ (which implies $g_{ij;k} = 0$),

$$m\partial_k f = \partial_k S = 2g^{ij}Rc_{ik;j} = 2g^{ij}(fg)_{ik;j} = 2g^{ij}(\partial_j f)g_{ik} = 2\partial_k f.$$

It follows that $\partial_k f = 0$ for any k and thus f is a constant.

(2) If $K(\Pi_p) = f(p)$, then

$$Ric(X_p) = (m-1)f(p).$$

So by (1), f is constant.[5] □

Remark. Obviously the theorem fails in dimension 2, in which case the sectional/Ricci curvature is always a function on M but need not be a constant.

3.4.2 Riemannian Manifolds with Constant Curvatures

Manifolds with Constant Sectional Curvatures

Now we study Riemannian manifolds with **constant sectional curvatures**, i.e. $K(\Pi_p) = k$ for all $p \in M$ and all $\Pi_p \in T_pM$. According to what we have proved last time, (M, g) has constant curvature k if and only if

$$Rm = \frac{k}{2} g \owedge g,$$

which is also equivalent to the fact "(M, g) has Weyl curvature tensor $W = 0$ and Ricci curvature tensor $Rc = (m-1)kg$".

We have constructed, for any constant k, a simple Riemannian manifold which has constant sectional curvature k, namely

(a) $\left(S^m, \frac{1}{k}g_{\mathrm{round}}\right)$ if $k > 0$,

(b) $(\mathbb{R}^m, g_0)$ if $k = 0$,

(c) $\left(H^m, -\frac{1}{k}g_{\mathrm{hyperbolic}}\right)$ if $k < 0$.

Of course there are many other constant sectional curvature manifolds, e.g.

- Any open subset in a Riemannian manifold of constant sectional curvature is again a Riemannian manifold of constant sectional curvature. To make our lives easier, we will exclude such examples[6] by studying only connected complete [i.e. when endowed with the Riemannian distance d, (M, d) is complete as a metric space] Riemannian manifolds of constant sectional curvature, which are known as **space forms**.

[5] We will give another direct proof of this fact using moving frames next time.

[6] Unfortunately, it is not true that any constant sectional curvature Riemannian manifold is an open submanifold of a complete constant sectional curvature Riemannian manifold. For example, one may start with $S^2 \setminus \{N, S\}$ be the standard sphere with the north/south poles removed, and consider its universal covering (which is topologically $\mathbb{R}^2$ with pull-back Riemannian metric). According to the Killing-Hopf theorem below, the metric can't be complete. Incomplete Riemannian manifolds are far from well-understood.

- If (M, g) has constant sectional curvature, $\pi : M \to N$ is a smooth normal [i.e. the Deck transformation group acts freely on each fiber] covering map and g is invariant under all its Deck transformations, then $(N, \pi_* g)$ has constant sectional curvature. Since universal cover is always normal, by this way we can easily construct
  - a constant positive sectional curvature metric on the real projective space $\mathbb{RP}^m = S^m/\mathbb{Z}_2$ and on the Lens space $L(p, q)$.
  - a flat metric (constant curvature zero metric) on the torus $\mathbb{T}^m = \mathbb{R}^m/\mathbb{Z}^m$, [in local coordinates $\theta^1, \dots, \theta^m$ on $\mathbb{T}^m = S^1 \times \cdots \times S^1$, the flat metric has the form $g = d\theta^1 \otimes d\theta^1 + \cdots + d\theta^m \otimes d\theta^m$.]
  - a constant negative sectional curvature metric on any closed orientable surface Σ_g of genus $g \geqslant 2$.
- Conversely if (M, g) has constant sectional curvature, and $\pi : \widetilde{M} \to M$ is a smooth covering, then $(\widetilde{M}, \pi^* g)$ has constant sectional curvature.

So it is reasonable to focus first on simply connected complete Riemannian manifolds of constant sectional curvature, and the examples (a), (b), (c) above are all simply connected. It turns out that they are the only ones, both locally (without completeness assumption) and globally (under the assumption of completeness):

Theorem 3.42 (Riemann)**.** Let (M, g) be a Riemannian manifold with constant sectional curvature k, then any point $p \in M$ has a neighborhood that is isometric to an open subset of (a) or (b) or (c).

Theorem 3.43 (Killing-Hopf)**.** Let (M, g) be a complete Riemannian manifold of constant sectional curvature k, then the Riemannian universal cover of (M, g) is either (a) or (b) or (c) above (depending on the sign of k).

We will postpone the proofs of both theorems to later.

According to Killing-Hopf theorem, any complete Riemannian manifold of constant sectional curvature is the quotient of one of the three canonical examples above by a group (which is a subgroup of the corresponding isometry group) that acts freely and properly discontinuously. As a result, only very few smooth manifolds can admit a constant sectional curvature metric. For example, there is no constant sectional curvature metric on $S^2 \times S^1$ since its universal cover is $S^2 \times \mathbb{R}$, which is not one of the above three.

In fact, we have seen all complete even dimensional Riemannian manifolds which has positive sectional curvature:

Corollary 3.44. If (M, g) is a compact Riemannian manifolds of even dimension $m = 2k$, and g has constant sectional curvature 1, then (M, g) is isometric to either (S^m, g_{round}) or its quotient $(\mathbb{RP}^m, g)$.

Proof. Since M is compact, (M, g) is complete. By Killing-Hopf theorem, (M, g) is the quotient of (S^m, g_{round}) by a subgroup

$$\Gamma \subset \text{Iso}(S^m, g_{\text{round}}) = O(m + 1)$$

which acts on S^m freely and properly discontinuously.

Now let $\gamma \in \Gamma$. If γ has an eigenvalue 1, then γ fixes a point in S^m (which is the unit eigenvector of γ) and thus by freeness of the action, $\gamma = \text{Id}$.

As a consequence, for $\text{Id} \neq \gamma \in \Gamma$, 1 is not an eigenvalue of γ. We may consider its square matrix γ^2, which is again an element in Γ. Since m is even, $\gamma^2 \in SO(m + 1)$ must has eigenvalue 1 and thus $\gamma^2 = \text{Id}$. It follows that all eigenvalues of γ are -1 and thus $\gamma = -\text{Id}$.

So we must have $\Gamma = \{\text{Id}\}$ or $\Gamma = \{+\text{Id}\}$, and the conclusion follows. □

Remark. The result fails in odd dimension, since as we have mentioned, all lens space admit constant sectional curvature Riemannian metric.

Spaces with Constant Ricci Curvatures: Einstein Manifolds

Now let's turn to Riemannian manifolds with constant Ricci curvature, i.e. satisfying $Ric(X_p) = k$ for any $p \in M$ and any $X_p \in S_pM$. It turns out that such manifolds play important roles in Einstein's general theory of relativity (in a slightly different framework, i.e. pseudo-Riemannian geometry): The Einstein field equation (which, together with the geodesic equation that we will discuss later, form the core of the mathematical formulation of general relativity) has the form

$$Rc - \frac{1}{2}Sg + \Lambda g = \kappa T,$$

where S is the scalar curvature function, Λ is known as the cosmological constant[7], κ is the Einstein gravitational constant, and T is the so-called stress-energy tensor. In the case of vacuum where $T = 0$, the equation becomes

$$Rc = \left(\frac{S}{2} - \Lambda\right) g.$$

According to Schur's theorem, the function $\frac{S}{2} - \Lambda$ must be a constant, and thus (M, g) has constant Ricci curvature.

Definition 3.45. We say a Riemannian manifold (M, g) is an **Einstein manifold** if there exists a constant λ such that

$$Rc = \lambda g.$$

Einstein manifolds with $\lambda = 0$ are known as **Ricci-flat** manifolds.

Obviously, if (M, g) has constant sectional curvature k, then (M, g) is an Einstein manifold since

$$Rc = c(Rm) = (m - 1)kg.$$

Since $\mathrm{tr}(Rc) = S$ (the scalar curvature) and $\mathrm{tr}(g) = m$ (the dimension of M), we conclude that the constant λ for an Einstein manifold must be

$$\lambda = \frac{S}{m}.$$

Since the traceless Ricci tensor $E = Rc - \frac{S}{m}g$, we conclude

Corollary 3.46. (M, g) is an Einstein manifold if and only if $E = 0$.

In particular, if (M, g) is an Einstein manifold and $W = 0$, then (M, g) has constant sectional curvature. Since $W = 0$ in dimension 3, it follows

Proposition 3.47. For $m = 2$ or 3, (M, g) is Einstein if and only if (M, g) has constant sectional curvature.

So to find an Einstein manifold that is not of constant sectional curvature, one must look at manifolds of dimension at least 4. To discover "which manifold admits an Einstein metric and which does not" is still a very active research topic today. Here is an example:

Example. Let $M = S^m \times S^m$ [or more generally the product of two m-dimensional Riemannian manifolds that have the same constant sectional curvature], equipped with the product metric

$$g = \pi_1^* g_{S^m} + \pi_2^* g_{S^m}.$$

Note that S^m has constant curvature 1, so that it is Einstein and

$$Rc(g_{S^m}) = (m - 1)g_{S^m}.$$

It follows

$$Rc(g) = \pi_1^* Rc(g_{S^m}) + \pi_2^* Rc(g_{S^m}) = (m - 1)\pi_1^* g_{S^m} + (m - 1)\pi_2^* g_{S^m} = (m - 1)g.$$

In other words, (M, g) is an Einstein manifold.

[7]According to Gamow, Einstein regard the introduction of the cosmological term as "the biggest blunder of his life". So we name these manifolds as Einstein manifolds as a punishment (joke).

On the other hand, (M, g) is not of constant sectional curvature for $m > 1$. This can be proved by using Killing-Hopf theorem, or by direct computation: for $(p, q) \in S^m \times S^m$, if we let e_1, e_2 be linearly independent vectors T_pS^m and let $e_3 \in T_qS^m$, then

$$K(\mathrm{d}\iota_q^1(e_1), \mathrm{d}\iota_q^1(e_2)) = 1, \quad K(\mathrm{d}\iota_q^1(e_1), \mathrm{d}\iota_p^1(e_3)) = 0,$$

where $\iota_q : S^m \to S^m \times S^m$ is the embedding that maps p to (p, q), while $\iota_p : S^m \to S^m \times S^m$ is the embedding that maps q to (p, q).

Remark. Again there are many topological restrictions for a smooth manifold to admit an Einstein metric. For example, by using the famous Chern-Gauss-Bonnet theorem, which has the following form for orientable closed 4-manifolds,

$$\chi(M) = \frac{1}{32\pi^2} \int_M (|Rm|^2 - 4|Rc|^2 + S^2)\mathrm{d}v,$$

where $\chi(M)$ is the Euler characteristic of M, one can easily prove:

Theorem 3.48 (Berger)**.** If M is an orientable closed 4-manifold and M admits an Einstein metric g, then $\chi(M) \geqslant 0$, and the equality holds if and only if g is flat.

In particular, since $\chi(S^3 \times S^1) = 0$ and $S^3 \times S^1$ admits no flat metric (by Killing-Hopf), we conclude that $S^3 \times S^1$ admits no Einstein metric.

The above result was further strengthened by Thorpe and Hitchin as follows:

Theorem 3.49 (Hitchin-Thorpe Inequality)**.** If M is an orientable closed 4-manifold and M admits an Einstein metric g, then

$$\chi(M) \geqslant \frac{3}{2}|\tau(M)|,$$

where $\tau(M)$ is the signature of M. Moreover, if the equality holds, then the Einstein metric is a Ricci flat metric.

Riemannian Manifolds with Constant Scalar Curvature

Now we turn to Riemannian manifolds with constant scalar curvature. There are many such examples, e.g.

- any Einstein manifold has constant scalar curvature,
- the product of two Riemannian manifolds with constant scalar curvatures is again a Riemannian manifold with constant scalar curvature since [as exercise]

Proposition 3.50. Let S_i $(i = 1, 2)$ be the scalar curvature of (M_i, g_i), and S the scalar curvature of $(M_1 \times M_2, \pi_1^* g_1 + \pi_2^* g_2)$. Then $S(p, q) = S_1(p) + S_2(q)$.

For simplicity, we only consider compact manifolds. It turns out that there always exist lots of constant scalar curvature metrics on any compact manifold M.

(1) Let's start with the case of dimension 2, i.e. S is a compact surface. Note that in this case all curvatures are the same. As we have seen, since the universal covering of S is either S^2 or $\mathbb{R}^2$ or D (the unit disc), M admits a constant curvature metric. However, even in this case, one can say a lot more: Suppose M is orientable (so that it has a Riemann surface structure), then by lifting to the universal covering and using the famous uniformization theorem in complex analysis, one can prove

Theorem. For any Riemannian metric g on an compact orientable surface S, there is $u \in \mathcal{C}^\infty(S)$ so that $(M, \mathrm{e}^{2u}g)$ has constant curvature.

[Note that according to the Gauss-Bonnet theorem, in this case the sign of the constant curvature depends on the topology of M.]

(2) In dimension $m \geqslant 3$, unlike the sectional curvature or the Ricci curvature, the scalar curvature encodes relatively few information. For example, Kazdan and Warner solved the prescribed scalar curvature problem and get

Theorem (Kazdan-Warner)**.** For any compact smooth manifold M of dimension $m \geqslant 3$,exactly one of the following will happen:

(a) For any $f \in \mathcal{C}^\infty(M)$, there exists a Riemannian metric g on M whose scalar curvature function is f.

(b) A function $f \in \mathcal{C}^\infty(M)$ is the scalar curvature of some Riemannian metric on M if and only if either $f \equiv 0$, or f is negative somewhere.

(c) A function $f \in \mathcal{C}^\infty(M)$ is the scalar curvature of some Riemannian metric on M if and only if f is negative somewhere.

As consequences, we immediately get

- any compact manifold of dimension $m \geqslant 3$ admits a Riemannian metric whose scalar curvature is any given negative constant.
- For any compact manifold M, by Proposition 3.50 the manifold $M \times S^m$ (where $m \geqslant 2$) admits a positive scalar curvature metric, and thus any function on $M \times S^m$ can be realized as its scalar curvature.

On the other hand, Atiyah-Singer index theorem gives topological obstructions for the existence of positive scalar curvature metric. It is also known that any torus $\mathbb{T}^m$ admits no positive scalar curvature metric [Schoen-Yau for $3 \leqslant m \leqslant 7$ (related to the positive mass theorem), Gromov-Lawson for general m].

(3) Since for surfaces, one can always find a constant scalar curvature metric in any given conformal class, it is natural to ask whether the same result holds in higher dimension [a classical problem in geometric analysis]:

The Yamabe Problem. Given a compact[8] Riemannian manifold (M, g) of dimension $m \geqslant 3$, is there $u \in \mathcal{C}^\infty(M)$ so that $(M, \mathrm{e}^{2u}g)$ has constant scalar curvature?

By computing the scalar curvature under conformal change, the problem is reduced to finding a positive solution to the partial differential equation

$$\frac{4(m-1)}{m-2}\Delta\varphi + S\varphi = \lambda\varphi^{\frac{m+2}{m-2}}.$$

Although there is a gap in Yamabe's origin solution (as pointed out by Trudinger), by combining the works of Yamabe (1960), Trudinger (1968), Aubin (1976) and Schoen (1984), the answer is **YES**. In particular, for any compact manifold of dimension $m \geqslant 3$, there exist lots of constant scalar curvature metrics.

Riemannian Manifolds with $W = 0$ (Locally Conformally Flat Manifolds)

Finally, we return to constant sectional curvature metrics, but in the framework of conformal geometry. Of course globally there are topological restrictions for the existence (by Killing-Hopf theorem), so we restrict ourselves to the local setting, i.e. study Riemannian manifolds (M, g) so that near each point p, there is an open neighborhood U and a smooth function $u \in \mathcal{C}^\infty(U)$ so that the metric $(U, \mathrm{e}^{2u}g)$ has constant sectional curvature. According to Theorem 3.42, if $(U, \mathrm{e}^{2u}g)$ has constant sectional curvature, then by shrinking U if needed, the manifold $(U, \mathrm{e}^{2u}g)$ is isometric to an open subset of one of the three canonical constant sectional curvature spaces, namely the

[8]One may pose the same problem for complete noncompact manifolds. The answer is no in general, and counterexamples were constructed by Jin (1988). The problem of finding conditions under which the problem has a solution is still a topic of research today.

Euclidean space or the sphere or the hyperbolic space (with scaled metric). Let's take a closer look of these three metrics. For $\mathbb{R}^m$ and H^m, we have

$$g_0 = \mathrm{d}x^1 \otimes \mathrm{d}x^1 + \cdots + \mathrm{d}x^m \otimes \mathrm{d}x^m \qquad \text{for } \mathbb{R}^m$$
$$g_{\text{hyperbolic}} = \frac{1}{(x^m)^2}(\mathrm{d}x^1 \otimes \mathrm{d}x^1 + \cdots + \mathrm{d}x^m \otimes \mathrm{d}x^m) \quad \text{for } H^m$$

while for S^m, we may use steregraphic coordinates $(t^1, \dots, t^m)$ [so that any point which is not the north pole on S^m can be written as $\left(\frac{2t^1}{1+|t|^2}, \dots, \frac{2t^m}{1+|t|^2}, \frac{1-|t|^2}{1+|t|^2}\right)$] to get [as exercise]

$$g_{\text{round}} = \frac{4}{(1+|t|^2)^2}(\mathrm{d}t^1 \otimes \mathrm{d}t^1 + \cdots + \mathrm{d}t^m \otimes \mathrm{d}t^m) \;\; \text{for } S^m.$$

So in all these three cases, after multiplying a conformal factor one gets a flat metric. In other words, "locally conformally constant sectional curvature" in the same as "locally conformally flat":

Definition 3.51. We say a Riemannian manifold (M, g) is **locally conformally flat** if for any $p \in M$, there is a neighborhood U of p and a smooth function $u \in \mathcal{C}^\infty(U)$ so that the metric $\overline{g} = \mathrm{e}^{2u} g$ is flat in U.

Example. We give a couple examples:

- Any constant sectional curvature space is locally conformally flat (as just explained).
- Any surface (with any real analytic Riemannian metric) is locally conformally flat:

  Theorem (Gauss, 1822)**.** On any surface with real analytic Riemannian metric g, there exists coordinates x, y and smooth function $u = u(x, y)$ so that

  $$g = \mathrm{e}^{2u}(\mathrm{d}x \otimes \mathrm{d}x + \mathrm{d}y \otimes \mathrm{d}y).$$

  [Such coordinates are known as isothermal coordinates.]

To find out necessary conditions for a metric g to be locally conformally flat, we need to find local quantities that are invariant under a conformal change of metric.

Lemma 3.52. If $\overline{g} = \mathrm{e}^{2u} g$, then $\overline{W} = \mathrm{e}^{2u} W$.

Proof. In Problem Sheet 2, we will see $\overline{Rm} = \mathrm{e}^{2u}(Rm - g \owedge T)$ for some symmetric 2-tensor T. Decomposing both $\overline{Rm}$ and Rm into Weyl and non-Weyl parts, we get

$$\overline{W} - \mathrm{e}^{2u} W = g \owedge \widetilde{T},$$

for some symmetric 2-tensor $\widetilde{T}$. So $\overline{W} - \mathrm{e}^{2u} W \in \ker(c) \cap \operatorname{im}(\Psi)$. By Corollary 3.23, we must have $\overline{W} - \mathrm{e}^{2u} W = 0$. ☐

Since any flat metric has Weyl curvature tensor $W = 0$, we conclude

Corollary 3.53. If (M, g) is locally conformally flat, then $W = 0$.

It turns out that the condition is also necessary for $m \geqslant 4$:

Theorem (Weyl-Schouten)**.** Let (M, g) be a Riemannian manifold.

(1) For $m \geqslant 4$, (M, g) is locally conformally flat if and only if $W = 0$.

(2) For $m = 3$, (M, g) is locally conformally flat if and only if [A is the Schouten tensor]

$$(\nabla_X A)(Y, Z) - (\nabla_Y A)(X, Z) = 0, \quad \forall X, Y, Z \in \Gamma^\infty(TM).$$

[The proof will be left as part of an exercises.]

For example, by calculating the Weyl curvature tensor one can easily prove that when endowed with the standard product metric, both $S^2 \times S^2$ and $S^2 \times \mathbb{R}^2$ are not locally conformally flat, while $S^{m_1} \times H^{m_2}$ is locally conformally flat.

3.5 The Method of Moving Frames

In Riemannian geometry, one frequently encounters with heavy computations (especially for those problems related to curvatures). There are three different methods to do these calculations: the invariant method via global vector fields and tensor fields, the local method via carefully chosen coordinate charts (under the help of Einstein summation convention), and E. Cartan's method of moving frames via calculus of differential forms. Now we will give a brief introduction to the method of moving frames where the use of differential forms is emphasized [when compared with tensor fields, differential forms have the advantage that they can be pulled-back via smooth maps, and we have the powerful tool of exterior derivative].

3.5.1 Cartan's Method of Moving Frames

The Connection 1-forms for a Linear Connection in a Local Frame

Let M be a smooth manifold and ∇ a linear connection on M. We can regard ∇ (acting on vector fields) as a linear map

$$\nabla : \Gamma^\infty(TM) \to \Gamma^\infty(TM \otimes T^*M).$$

So if $\{e_1, \dots, e_m\}$ is a **local frame** [i.e. for each $p \in U$, $e_1(p), \dots, e_m(p)$ form a basis of T_pM] of TM defined on an open set $U \subset M$, then one can find a set of one forms $\{\theta_i^j\}_{1 \leqslant i,j \leqslant m}$ defined on U so that $\nabla_X e_i = \theta_i^j(X) e_j$ for all $X \in \Gamma^\infty(TM)$, i.e.

$$\nabla e_i = e_j \otimes \theta_i^j. \tag{3.22}$$

These θ_i^j's are known as **connection 1-forms** of ∇ with respect to the local frame $\{e_i\}$, which are only locally defined.

Moreover, if we choose another local frame $\{\widetilde{e}_1, \dots, \widetilde{e}_m\}$ on $\widetilde{U}$, and $\widetilde{e}_i = f_i^j e_j$ on $U \cap \widetilde{U}$, then $e_j = (f^{-1})_j^i \widetilde{e}_k$ (where f^{-1} is the inverse of the matrix $f = (f_i^j)$) and thus

$$\begin{aligned}\widetilde{e}_l \otimes \widetilde{\theta}_i^l = \nabla(f_i^j e_j) &= f_i^j \nabla e_j + e_j \otimes \mathrm{d} f_i^j \\ &= f_i^j e_k \otimes \theta_j^k + (f^{-1})_j^l \widetilde{e}_l \otimes \mathrm{d} f_i^j \\ &= \widetilde{e}_l \otimes (f^{-1})_k^l \theta_j^k f_i^j + (f^{-1})_j^l \mathrm{d} f_i^j,\end{aligned}$$

so we end up with

$$\widetilde{\theta}_i^l = (f^{-1})_k^l \theta_j^k f_i^j + (f^{-1})_j^l \mathrm{d} f_i^j,$$

on $U \cap \widetilde{U}$, which can be written in brief as

$$\widetilde{\theta} = f^{-1} \theta f + f^{-1} \mathrm{d} f, \tag{3.23}$$

where $\widetilde{\theta}$ and θ are understood as $m \times m$ matrices whose entries are 1-forms, while f and f^{-1} are invertible $m \times m$ matrices[9] whose entries are functions (and thus one can not exchange their positions in the product above).

To develop Riemannian geometry via differential forms only, let's first derive the dual formula for covariant derivative of differential forms via these connection 1-forms. We denote by $\{\omega^1, \dots, \omega^m\}$ the local **dual co-frame** [i.e. $\omega^i(e_j) = \delta_j^i$ for all i, j] of T^*M defined on U to the given local frame $\{e_1, \dots, e_m\}$. Then we have

$$(\nabla_X \omega^i)(e_j) = X(\omega^i(e_j)) - \omega^i(\nabla_X e_j) = -\omega^i(\theta_j^k(X) e_k) = -\theta_j^i(X).$$

It follows that the linear connection ∇ acting on one forms, viewed as a map

$$\nabla : \Gamma^\infty(T^*M) \to \Gamma^\infty(T^*M \otimes T^*M),$$

can be expressed in terms of the co-frame and the connection 1-forms as

$$\nabla \omega^i = -\omega^j \otimes \theta_j^i. \tag{3.24}$$

[9] So one may regard f as a map from $U \cap \widetilde{U}$ to the general linear group $GL(m)$. If we are in the setting of Riemannian manifold and we are only using local orthonormal frames, then the group encountered is $O(m)$ instead. The method of moving frame works in a more general setting, and there is always such a Lie group behind the theory that plays an important role.

The Connection 1-forms: Torsion Freeness and Metric Compatibility

Now suppose the linear connection ∇ is torsion free. Then

$$\begin{aligned} d\omega^i(X,Y) &= X(\omega^i(Y)) - Y(\omega^i(X)) - \omega^i([X,Y]) \\ &= X(\omega^i(Y)) - Y(\omega^i(X)) - \omega^i(\nabla_X Y - \nabla_Y X) \\ &= (\nabla_X \omega^i)(Y) - (\nabla_Y \omega^i)(X) \\ &= -\omega^j(Y)\theta^i_j(X) + \omega^j(X)\theta^i_j(Y). \end{aligned}$$

So the torsion free condition for a linear connection can be written, in terms of the dual co-frame and the connection 1-forms, as

$$d\omega^i = \omega^j \otimes \theta^i_j - \theta^i_j \otimes \omega_j = \omega_j \wedge \theta^i_j. \tag{3.25}$$

which can be written in brief as $d\omega = -\theta \wedge \omega$.

Next suppose there is a Riemannian metric g on M, and the connection ∇ is metric compatible. To encode the information of the metric into our consideration, it is reasonable to choose an orthonormal frame $\{e_1, \dots, e_m\}$ instead of a general frame. Then

$$0 = \langle \nabla e_i, e_j \rangle + \langle e_i, \nabla e_j \rangle = \langle e_k \otimes \theta^j_i, e_j \rangle + \langle e_i, e_k \otimes \theta^k_j \rangle = \theta^j_i + \theta^j_i,$$

So the metric compatibility of ∇ becomes: for any orthonormal frame, the connection 1-forms satisfy

$$\theta^j_i + \theta^j_i = 0 \tag{3.26}$$

i.e. the matrix of connection 1-forms is anti-symmetric.

Cartan's Formulation of Riemannian Geometry

It turns out that one can develop Riemannian geometry starting with local frames and connection 1-forms (i.e. via the differential 1-forms ω^i, θ^i_j) instead of the Riemannian metric g and its Levi-Civita connection [since one can recover the Riemannian metric g from the local orthonormal co-frame $\{\omega^i\}$, and then recover the Levi-Civita connection ∇ from its connection 1-forms θ^j_i]. We start with a simple lemma:

Lemma 3.54. Suppose $\omega^1, \dots, \omega^s \in \Lambda^1 V^*$ $(s \leqslant m = \dim V)$ are linearly independent.

(1) If $\eta^1, \dots, \eta^s \in \Lambda^1 V^*$ and $\sum \eta^i \wedge \omega^i = 0$, then there exist uniquely determined real numbers A^i_j $(1 \leqslant i, j \leqslant s)$ with $A^i_j = A^j_i$ such that $\eta^i = A^i_j \omega^j$.

(2) If $s = m$, and a collection of linear 1-forms $\theta^i_j \in \Lambda^1 V^*$ $(1 \leqslant i, j \leqslant m)$ satisfy

$$\omega^j \wedge \theta^i_j \quad \text{and} \quad \theta^i_j + \theta^j_i = 0,$$

then $\theta^i_j = 0$.

Proof. (1) Obviously, $\eta^i \in \operatorname{span}\{\omega^1, \dots, \omega^s\}$. Write $\eta^i = A^i_j \omega^j$. Then

$$\sum \eta^i \wedge \omega^i = \sum_{i<j} (A^i_j - A^j_i)\omega^i \wedge \omega^j$$

and the conclusion follows.

(2) Write $\theta^i_j = a^i_{jk}\omega^k$. Then the two conditions becomes

$$a^i_{jk} - a^i_{kj} = 0 \quad \text{and} \quad a^i_{jk} + a^j_{ik} = 0.$$

Thus

$$a^i_{jk} = a^i_{kj} = -a^k_{ij} = -a^k_{ji} = a^j_{ki} = a^j_{ik} = -a^i_{jk}$$

and the conclusion follows. □

Now we state the fundamental theorem of Riemannian geometry [i.e. the existence and uniqueness of Levi-Civita connection] in the language of connection 1-forms:

Theorem 3.55 (E. Cartan)**.** Let $\omega^1, \dots, \omega^m \in \Omega^1(U)$ be a collection of 1-forms on an open set $U \subset M$ that are linearly independent at each point. Then there exists a unique collection of 1-forms, $\theta^i_j \in \Omega^1(U)$ $(1 \leqslant i, j \leqslant m)$, so that

$$\mathrm{d}\omega^i = \omega^i \wedge \theta^i_j \quad \text{and} \quad \theta^i_j + \theta^j_i = 0.$$

[These equations are known as Cartan's structural equations.]

Proof. Uniqueness follows from Lemma 3.54 (2). For the existence, one just start with the Riemannian metric $g = \sum \omega^i \otimes \omega^i$ (so that the dual frame $\{e_i\}$ of $\{\omega^i\}$ is an orthonormal basis for each point in U) and take θ^i_j to be the collection 1-forms for the Levi-Civita connection of this metric. □

Remark. How to get from local to global? To glue, one need the connection 1-forms to satisfy the change of frame formula (2) for any orthogonla transformation f.

The Curvature 2-form

We start with any linear connection on a smooth manifold M. Suppose we are given a local co-frame $\{\omega^i\}$ and the corresponding connection 1-forms θ^i_j. We may express the curvature using differential forms (in terms of the connection 1-forms) as follows. By definition

$$\begin{aligned} R(X,Y)e_i &= \nabla_X \nabla_Y e_i - \nabla_Y \nabla_X e_i - \nabla_{[X,Y]} e_i \\ &= \nabla_X(\theta^j_i(Y)e_j) - \nabla_Y(\theta^j_i(X)e_j) - \theta^j_i([X,Y])e_j \\ &= X(\theta^j_i(Y))e_j + \theta^j_i(Y)\theta^k_j(X)e_k - Y(\theta^j_i(X))e_j - \theta^j_i(X)\theta^k_j(Y)e_k - \theta^j_i([X,Y])e_j \\ &= (\mathrm{d}\theta^j_i)(X,Y)e_j + \theta^j_k \wedge \theta^k_i(X,Y)e_j. \end{aligned}$$

As a consequence, if we denote $R(e_k, e_l)e_i = R^j_{kli} e_j$, then we get

$$\mathrm{d}\theta^j_i + \theta^j_k \wedge \theta^k_i = R^j_{kli}\omega^k \otimes \omega^l = \frac{1}{2} R^j_{kli}\omega^k \wedge \omega^l. \tag{3.27}$$

We shall denote

$$\Omega^j_i = \frac{1}{2} R^j_{kli}\omega^k \wedge \omega^l,$$

and call it the **curvature 2-form**, which can be expressed in terms of θ^j_i's as

$$\Omega^j_i = \mathrm{d}\theta^j_i + \theta^j_k \wedge \theta^k_i. \tag{3.28}$$

The formula can be taken as definition of curvature (for given connection 1-forms) and is usually written in brief as

$$\Omega = \mathrm{d}\theta + \theta \wedge \theta,$$

where Ω is regarded as an $m \times m$ matrix whose entries are 2-forms.

Unlike the connection 1-forms, given a linear connection, the curvature 2-form is independent of the choice of co-frame and thus is globally defined. To see this, we use the frame transformation formula for connection 1-forms above to get

$$\begin{aligned} \widetilde{\Omega} =& \mathrm{d}\widetilde{\theta} + \widetilde{\theta} \wedge \widetilde{\theta} \\ =& (\mathrm{d}f^{-1}) \wedge \theta f + f^{-1}(\mathrm{d}\theta) f - f^{-1}\theta \wedge \mathrm{d}f + (\mathrm{d}f^{-1}) \wedge \mathrm{d}f \\ &+ f^{-1}\theta \wedge \theta f + f^{-1}\theta \wedge \mathrm{d}f + f^{-1}\mathrm{d}f \wedge f^{-1}\theta f + f^{-1}\mathrm{d}f \wedge f^{-1}\mathrm{d}f. \end{aligned}$$

In view of the fact $\mathrm{d}f^{-1} = -f^{-1}(\mathrm{d}f)f^{-1}$, we get

$$\widetilde{\Omega} = f^{-1}(\mathrm{d}\theta + \theta \wedge \theta)f = f^{-1}\Omega f,$$

which is equivalent to say Ω is independent of the choice of frames.

Now suppose (M, g) is a Riemannian manifold. Then we may start with orthonormal co-frame $\{\omega^i\}$, and we have Cartan's structural equations, which implies

$$\Omega_i^j = -\Omega_j^i.$$

We may also express the curvature 2-form Ω_i^j using $R_{ijkl} := Rm(e_i, e_j, e_k, e_l)$ as

$$\Omega_j^i = \frac{1}{2} R_{klj}^i \omega^k \wedge \omega^l = -\frac{1}{2} R_{klji} \omega^k \wedge \omega^l = \frac{1}{2} R_{ijkl} \omega^k \wedge \omega^l.$$

Remark. More generally, one can develop the theory of linear connections on vector bundles (or principal bundles) via moving frames, as follows. Let E be a rank r vector bundle over M, and $\{e_1, \ldots, e_r\}$ a local frame of E. Then one can either define a linear connection

$$\nabla : \Gamma^\infty(E) \to \Gamma^\infty(E \otimes T^*M)$$

via axioms that we mentioned earlier, or via connection 1-forms θ_i^j $(1 \leqslant i, j \leqslant r)$ that are locally defined such that

$$\nabla e_i = e_j \otimes \theta_i^j.$$

As we calculated above, the matrix θ transform under change of basis as

$$\widetilde{\theta} = f^{-1} \theta f + f^{-1} \mathrm{d} f.$$

One can further define the curvature 2-form to be

$$\Omega = \mathrm{d}\theta + \theta \wedge \theta.$$

3.5.2 Applications to Riemannian Geometry

Calculating Curvatures

As the first application, we use moving frames to calculate the curvature of a Riemannian manifold (M, g). Let $\{e_1, \ldots, e_m\}$ be a local orthonormal frame of (M, g). By definition the sectional curvature of the plane spanned by $\{e_i, e_j\}$ is

$$K(e_i, e_j) = Rm(e_i, e_j, e_i, e_j) = R_{ijij} = \Omega_i^j(e_j, e_i).$$

Theorem 3.56. (M, g) has constant sectional curvature c at $p \in M$ if and only if for any local orthonormal frame $\{e_i\}$, at p we have

$$\Omega_j^i = c\omega^i \wedge \omega^j. \tag{3.29}$$

Proof. Suppose (3.29) holds at p for any orthonormal frame. Let Π_p be any two dimensional plane in T_pM. Choose an orthonormal basis $\{e_1, e_2\}$ of Π_p, extend it to an orthonormal frame and denote by $\omega^1, \ldots, \omega^m$ the dual co-frame. Then

$$K(\Pi_p) = K(e_1, e_2) = c\Omega_1^2(e_2, e_1) = c\omega^2 \wedge \omega^1(e_2, e_1) = c.$$

Conversely suppose (M, g) has constant sectional curvature c at p, then with respect to any orthonormal frame,

$$R_{ijkl} = \frac{c}{2} g \owedge g(e_i, e_j, e_k, e_l) = c(\delta_{ik}\delta_{jl} - \delta_{jk}\delta_{il})$$

at p and thus the conclusion follows. □

Example. Consider the upper half space $\mathbb{H}^m$ with the hyperbolic metric

$$g_{\text{hyperbolic}} = \frac{1}{(x^m)^2} (\mathrm{d}x^1 \otimes \mathrm{d}x^1 + \cdots + \mathrm{d}x^m \otimes \mathrm{d}x^m).$$

With the orthonormal frame $\{e_i = x^m \partial_i\}$ and its dual co-frame $\left\{\omega^i = \frac{1}{x^m}\mathrm{d}x^i\right\}$,

$$\omega^j \wedge \theta^i_j = \mathrm{d}\omega^i = -\frac{1}{(x^m)^2}\mathrm{d}x^m \wedge \mathrm{d}x^i = -\omega^m \wedge \omega^i.$$

Observe that for the given co-frame $\{\omega^1, \dots, \omega^m\}$,

$$\theta^i_j = 0,\ (i, j < m) \quad \text{and} \quad \theta^i_m = -\theta^m_i = -\omega^i,\ (i < m)$$

is a solution and thus has to be the unique solution. So we get, for $i, j < m$,

$$\Omega^i_j = \mathrm{d}\theta^i_j + \theta^i_k \wedge \theta^k_j = \theta^i_m \wedge \theta^m_j = -\omega^i \wedge \omega^j$$

and for $i < m$

$$\Omega^i_m = \mathrm{d}\theta^i_m + \theta^i_k \wedge \theta^k_m = -\mathrm{d}\omega^i = -\omega^i \wedge \omega^m.$$

It follows from Theorem 3.55 that the hyperbolic space has constant curvature -1.

Proving the Bianchi Identities

We may also prove the Bianchi identities via moving frame. For the first Bianchi identity, we just take exterior derivative:

$$\begin{aligned} 0 = \mathrm{d}^2\omega^i &= \mathrm{d}\omega^j \wedge \theta^i_j - \omega^j \wedge \mathrm{d}\theta^i_j \\ &= \omega^k \wedge \theta^j_k \wedge \theta^i_j - \omega^j \wedge (\Omega^i_j - \theta^i_k \wedge \theta^k_j) \\ &= -\omega^j \wedge \Omega^i_j \\ &= -\frac{1}{2} R^i_{klj}\omega^j \wedge \omega^k \wedge \omega^l \\ &= -\frac{1}{2}\sum_{j<k<l}(R^i_{klj} + R^i_{ljk} + R^i_{jkl})\omega^j \wedge \omega^k \wedge \omega^l. \end{aligned}$$

As a consequence, we get for distinct k, l, j's,

$$R^i_{klj} + R^i_{ljk} + R^i_{jkl} = 0.$$

If two or three of k, l, j's are the same, then the first Bianchi identity trivial.

Similarly by taking exterior derivative of $\Omega = \mathrm{d}\theta + \theta \wedge \theta$, we get

$$\mathrm{d}\Omega = \mathrm{d}\theta \wedge \theta - \theta \wedge \mathrm{d}\theta = \Omega \wedge \theta - \theta \wedge \Omega. \tag{3.30}$$

One can prove that in local frames, together with the first Bianchi identity, the expression above is equivalent to the second Bianchi identity. In fact, we can give a very quick proof of the sectional curvature version of Schur's theorem via (3.30):

Alternative Proof of Theorem 3.41 (2). Suppose (M, g) has sectional curvature $K(\Pi_p) = f(p)$ for some $f \in \mathcal{C}^\infty(M)$. By Theorem 3.56, $\Omega^i_j = f(p)\omega^i \wedge \omega^j$. So

$$\begin{aligned} \mathrm{d}f \wedge \omega^i \wedge \omega^j + f\mathrm{d}\omega^i \wedge \omega^j - f\omega^i \wedge \mathrm{d}\omega^j = \mathrm{d}\Omega^i_j &= \Omega^i_k \wedge \theta^k_j - \theta^i_k \wedge \Omega^k_j \\ &= -f\omega^i \wedge \omega^k \wedge \theta^j_k - f\theta^i_k \wedge \omega^k \wedge \omega^j \\ &= -f\omega^i \wedge \mathrm{d}\omega^j + f\mathrm{d}\omega^i \wedge \omega^j. \end{aligned}$$

It follows $\mathrm{d}f \wedge \omega^i \wedge \omega^j = 0$ for all i, j, and, since $m \geqslant 3$, $\mathrm{d}f = 0$, i.e. f is constant. □

Geometry of Riemannian Submanifolds via Moving Frame

Let $(\overline{M}, \overline{g})$ be a Riemannian manifold of dimension m, and $\iota : S \hookrightarrow \overline{M}$ a smooth submanifold of dimension s endowed with the submanifold metric $g = \iota^*\overline{g}$. For simplicity make the following index convention:

- $1 \leqslant A, B, \ldots \leqslant m$,
- $1 \leqslant i, j, \ldots \leqslant s$,
- $s + 1 \leqslant \alpha, \beta, \ldots \leqslant m$.

As usual, we denote by NS the normal bundle of S in M.

We have three different ways to develop the Riemannian geometry of S. Here we take the moving frame approach. So let's start with a special local orthonormal frame $\{\overline{e}_1, \ldots, \overline{e}_m\}$ of $(\overline{M}, \overline{g})$ with the property that $\overline{e}_i = \mathrm{d}\iota(e_i)$ on S for $1 \leqslant i \leqslant s$ and $\{e_1, \ldots, e_s\}$ form a local orthonormal frame of S. Denote by $\{\overline{\varpi}^1, \ldots, \overline{\varpi}^m\}$ the dual co-frame of $\{\overline{e}_1, \ldots, \overline{e}_m\}$. Then by definition,

$$\iota^*\overline{\varpi}^\alpha = 0. \tag{3.31}$$

Let $\overline{\theta}_B^A$ the connection 1-forms of $(\overline{M}, \overline{g})$ corresponding to the local frame $\{\overline{e}_A\}$. Then Cartan's structural equations of $\overline{M}$ reads

$$\overline{\theta}_B^A + \overline{\theta}_A^B = 0 \quad \text{and} \quad \mathrm{d}\overline{\varpi}^A = \overline{\varpi}^B \wedge \overline{\theta}_B^A.$$

It follows that as 1-forms on S, $\omega^i := \iota^*\overline{\varpi}^i$ and $\theta_j^i := \iota^*\overline{\theta}_j^i$ satisfy (here we used (3.31))

$$\theta_j^i + \theta_i^j = 0 \quad \text{and} \quad \mathrm{d}\omega^i = \omega^j \wedge \theta_j^i.$$

By uniqueness in Theorem 3.56, θ_j^i's are the connection 1-forms on S associate with the co-frame $\{\omega^1, \ldots, \omega^m\}$. [This proves the remark on page 40.]

We may also study connection 1-forms with indices α's. Using (3.31) twice we get

$$0 = \mathrm{d}\iota^*\overline{\varpi}^\alpha = \iota^*\overline{\varpi}^A \wedge \iota^*\overline{\theta}_A^\alpha = \omega^i \wedge \iota^*\overline{\theta}_i^\alpha.$$

Thus by Lemma 3.54 (1), there exist uniquely determined functions h_{ij}^α such that

$$h_{ij}^\alpha = h_{ji}^\alpha \quad \text{and} \quad \iota^*\overline{\theta}_i^\alpha = h_{ij}^\alpha \omega^j.$$

Definition 3.57. We call the map $\mathrm{II} : \Gamma^\infty(TS) \times \Gamma^\infty(TS) \to \Gamma^\infty(NS)$ defined by

$$\mathrm{II}(X, Y) = h_{ij}^\alpha \omega^i(X)\omega^j(Y)\overline{e}_\alpha$$

the **second fundamental form** of (S, g) as a Riemannian submanifold of $(\overline{M}, \overline{g})$.

Note that the fact $h_{ij}^\alpha = h_{ji}^\alpha$ implies $\mathrm{II}(X, Y) = \mathrm{II}(Y, X)$. We may write

$$\mathrm{II} = h_{ij}^\alpha \omega^i \otimes \omega^j \otimes \overline{e}_\alpha.$$

To see the formula above is independent of the choices of frames, let's reveal the true face of $\mathrm{II}(X, Y)$ by expressing it in the invariant formulation. We will use $\overline{\nabla}$ and ∇ to denote the Levi-Civita connections for $(\overline{M}, \overline{g})$ and (S, g) respectively. For $X, Y \in \Gamma^\infty(TS)$, we denote $\overline{X} = \mathrm{d}\iota(X)$ and $\overline{Y} = \mathrm{d}\iota(Y)$. Note that if $X = X^i e_i$, then $\overline{X} = X^i\overline{e}_i$. So on S, we have

$$\begin{aligned}
\overline{\nabla}_{\overline{Y}}\overline{X} - \mathrm{d}\iota(\nabla_Y X) &= \overline{Y}(X^i)\overline{e}_i + X^i\overline{\theta}_i^A(\overline{Y})\overline{e}_A - \mathrm{d}\iota(Y(X^i)e_i - X^i\theta_i^j(Y)e_j) \\
&= X^i\overline{\theta}_i^\alpha(\overline{Y})\overline{e}_\alpha \\
&= \omega^i(X)\iota^*\overline{\theta}_i^\alpha(Y)\overline{e}_\alpha \\
&= h_{ij}^\alpha\omega^i(X)\omega^j(Y)\overline{e}_\alpha.
\end{aligned}$$

In other words, for any vector field X, Y tangent to S, we have

$$\mathrm{II}(X, Y) = \overline{\nabla}_{\overline{Y}}\overline{X} - \mathrm{d}\iota(\nabla_Y X).$$

In view of the fact $\nabla_Y X$ is the tangential component of $\overline{\nabla}_{\overline{Y}}\overline{X}$, we conclude that $\mathrm{II}(X, Y)$ is really the normal component of $\overline{\nabla}_{\overline{Y}}\overline{X}$.

Example. According to the example on page 40-41, for the unit sphere S^m viewed as a Riemannian submanifold of $\mathbb{R}^{m+1}$, we have

$$\mathrm{II}(X,Y) = -\langle X,Y\rangle \vec{n}.$$

The second fundamental form is closely related to the curvature 2-form of (S,g): If we pull back $\overline{\Omega}_j^i = \mathrm{d}\overline{\theta}_j^i + \overline{\theta}_A^i \wedge \overline{\theta}_j^A$ to S compare with $\Omega_j^i = \mathrm{d}\theta_j^i + \theta_k^i \wedge \theta_j^k$, we get

$$\Omega_j^i = \iota^*\overline{\Omega}_j^i - \iota^*\overline{\theta}_\alpha^i \wedge \iota^*\overline{\theta}_j^\alpha = \iota^*\overline{\Omega}_j^i + \sum_\alpha \iota^*\overline{\theta}_i^\alpha \wedge \iota^*\overline{\theta}_j^\alpha = \iota^*\overline{\Omega}_j^i + \sum_\alpha h_{ik}^\alpha h_{jl}^\alpha \omega^k \wedge \omega^l.$$

As a consequence,

$$R_{ijkl} = \overline{R}_{ijkl} + (h_{ik}^\alpha h_{jl}^\alpha - h_{il}^\alpha h_{jk}^\alpha),$$

which is known as **Gauss equation**.

Example. In the case S is a hypersurface in (M,g), i.e. has co-dimension 1, then one may pair the second fundamental form with $\overline{e}_m$ and thus for each $p \in S$, regard II_p as a symmetric quadratic form on T_pS. With the help of the Riemannian metric, one can convert this symmetric quadratic form into a symmetric operator on T_pM, which is known as the **shape operator**. The eigenvalues of the shape operator are known as the **principal curvatures** of S at p. Its trace and the determinant are known as the **mean curvature** and the **Gauss curvature** of S at p.

In particular, if S is a 2-dimensional surface isometrically embedded in $\mathbb{R}^3$, the only sectional curvature is $R_{1212} = h_{11}h_{22} - h_{12}^2$, which is exactly the Gauss curvature of S. As a consequence, we get Gauss Theorem Egregium: The Gauss curvature [which is defined by the second fundamental form which is extrinsic] is in fact intrinsic [since the sectional curvature depends only on the Riemannian metric and thus is intrinsic].

We say S is a **totally geodesic submanifold** if $\mathrm{II} = 0$, i.e. $h_{ij}^\alpha = 0$ for all i, j, α. From Gauss equation one gets [there is still an issue here that we will explain later.]

Theorem 3.58. Let S be a totally geodesic 2-dimensional submanifold of M with $T_pS = \Pi_p$. Then the sectional curvature $K(\Pi_p)$ of M is the Gauss curvature of S.

Chapter 4

Geodesics and Variations

4.1 Geodesics as Self-parallel Curves (on Manifolds with Connection)

Now we turn to the next topic in this course: geodesic, which is a generalization of the notion of straight line in the Euclidean space. As we know, a line in $\mathbb{R}^m$ is both a curve "with constant direction", and a curve that "minimize distances between any two points on it". As a result, we will have two ways to define geodesics on Riemannian manifolds, which, as we will see, are equivalent. On the other hand, for the first method (i.e. regard geodesics as curves "with constant directions"), what we need is the existence of a covariant derivative instead of a Riemannian metric structure, and as a result, it works for any smooth manifold with a linear connection. So today we will introduce the first method, i.e. focus on "non-metric properties" of geodesics.

4.1.1 Geodesics on Manifolds with Linear Connections

Geodesics for Manifolds with Linear Connections

Let M be a smooth manifold. To define a geodesic as a "curve with constant direction", what we need is a structure that can be used to compare tangent vectors at different points along a curve, i.e. a parallel transport, or equivalently, a linear connection. So we let ∇ be a linear connection on M. Now suppose $\gamma : [a, b] \to M$ is a smooth curve in M. Then "γ is a geodesic" means that the tangent vector field $\dot\gamma$ is "unchanged" along γ (under parallel transport), i.e. is covariantly constant along γ:

Definition 4.1. We say γ is a **geodesic** if $\dot\gamma$ is parallel along γ, i.e.

$$\nabla_{\dot\gamma(t)}\dot\gamma = 0, \quad \forall t.$$

In local coordinates, if we write $\gamma(t) = (x^1(t), \dots, x^m(t))$, then

$$\dot\gamma(t) = \mathrm{d}\gamma\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) = \dot x^i(t)\partial_i.$$

Now suppose $X = X^i\partial_i$ is a smooth vector field near γ [If X is only defined on γ, then we need to extend it to a smooth vector field in a neighborhood of γ. By locality of ∇, the extension will not affect the computation below]. If we denote $f^i(t) = X^i(\gamma(t))$, then

$$\nabla_{\dot\gamma(t)}X^i = \dot\gamma(t)X^i = \frac{\mathrm{d}}{\mathrm{d}t}(X^i \circ \gamma) = \dot f^i(t)$$

[i.e. the covariant derivative of any function along γ is its t-derivative] and thus

$$(\nabla_{\dot\gamma}X)|_{\gamma(t)} = (\dot\gamma(t)X^i)\partial_i + \Gamma^k_{ij}\dot x^i(t)f^j(t)\partial_k = \dot f^k(t)\partial_k + \Gamma^k_{ij}\dot x^i(t)f^j(t)\partial_k.$$

As a result, the condition $\nabla_{\dot\gamma}X = 0$, i.e. "X is parallel along γ" becomes

$$\dot f^k(t) + \Gamma^k_{ij}(\gamma(t))\dot x^i(t) f^j(t) = 0, \quad \forall k.$$

Apply this to the vector field $X = \dot\gamma$, we see γ is a geodesic if and only if locally its coordinate functions satisfy the following system of second order ODEs

$$\ddot x^k(t) + \dot x^i(t)\dot x^j(t)\Gamma^k_{ij} = 0, \quad 1 \leqslant k \leqslant m. \tag{4.1}$$

Remark. A natural question is:

Question: is a re-parametrization of a geodesics still a geodesic?

Suppose γ is a geodesic and $\dot\gamma \neq 0$ (otherwise γ is constant), and $\widetilde\gamma(s) = \gamma(t(s))$ is a regular re-parametrization of γ, then

$$\begin{aligned}\nabla_{\dot{\widetilde\gamma}(s)}\dot{\widetilde\gamma}(s) &= \nabla_{\dot{\widetilde\gamma}(s)}(t'(s)\dot\gamma(t(s))) \\ &= \dot\gamma(t(s)) + (t'(s))^2\nabla_{\dot\gamma(t(s))}\dot\gamma(t(s)) = t''(s)\dot\gamma(t(s)).\end{aligned}$$

So $\widetilde\gamma$ is also a geodesic if and only if $t'' = 0$, i.e. $t(s) = as + b$ for some constants a and b. So the answer to the above question is:

Answer: A re-parametrization of a geodesics is still a geodesic if and only if the re-parametrization is linear.

Basic Examples

Example. Let $M = \mathbb{R}^m$, equipped with standard linear connection ∇ such that $\nabla_X Y = X(Y^j)\partial_j$, or equivalently, $\Gamma^k_{ij} = 0$. Let γ be any curve and X be a vector field. Then for X to be parallel along γ, we need $\dot f^k(t) = 0$ for all k, i.e. if and only if X^i's are constants on γ [so X is a constant vector field in $\mathbb{R}^m$ along γ in the usual sense].

In particular, the geodesic equations in $\mathbb{R}^m$ above become

$$\ddot x^k(t) = 0, \quad 1 \leqslant k \leqslant m.$$

The solution to the system are linear functions, i.e. $x^k(t) = a_k t + b_k$ for some constants a_k, b_k. As a consequence, γ is a geodesic if and only if it is the straight line in the direction $\vec a = \langle a_1, \dots, a_m\rangle$ that passes the point $(b_1, \dots, b_m)$.

Example. Consider $M = S^m$ the m-sphere, equipped with the Levi-Civita connection. For any $p \in S^m$, regarded as a unit vector $p = \vec u \in \mathbb{R}^{m+1}$, and for any unit tangent vector $\vec w \in T_pS^m$, we let

$$\gamma(t) = (\cos t)\vec u + (\sin t)\vec w.$$

be the great circle in S^m passing p in the direction of $\vec w$. Since the Levi-Civita connection on S^m is given by $\nabla_X Y = \overline\nabla_X Y + \langle X, Y\rangle \vec n$, where $\overline\nabla$ is the Levi-Civita connection for $\mathbb{R}^{m+1}$, i.e. with $\overline\Gamma^k_{ij} = 0$. So

$$\nabla_{\dot\gamma}\dot\gamma = \overline\nabla_{\dot\gamma}\dot\gamma + \langle\dot\gamma, \dot\gamma\rangle\vec n = \ddot\gamma + \vec n.$$

But at the point $\gamma(t)$, one has $\vec n = \gamma(t)$, and $\ddot\gamma(t) = -\gamma(t)$. So we get

$$\nabla_{\dot\gamma}\dot\gamma = 0.$$

In other words, any great circle on S^m is a geodesic. [By uniqueness below, up to linear re-parametrizations they are essentially the only geodesics on S^m.]

The Existence, Uniqueness and Smoothness

To find a geodesic is equivalent to solve the system of second order ODEs (4.1). By introducing $y^i = \dot{x}^i$, we may convert it to a system of first order ODEs (with more variables and more equations)

$$\begin{cases} \dot{x}^k = y^k, \\ \dot{y}^k = -\Gamma_{ij}^k y^i y^j, \end{cases} \quad 1 \leqslant k \leqslant m.$$

So suppose we want to find a geodesic with $\gamma(t_0) = p = (p^1, \dots, p^m)$ and $\dot{\gamma}(t_0) = X_p = X^i \partial_i \in T_pM$, then we need to solve the above system with initial condition $x(t_0) = (x^1(t_0), \dots, x^m(t_0)) = p$, $y(t_0) = (y^1(t_0), \dots, y^m(t_0)) = X_p$. According to the fundamental theorem for system of first order ODEs,

- **Existence:** For any $t_0 \in \mathbb{R}$ and any $(p, X_p) \in TM$, there is an open interval $I \ni t_0$ and open set $\mathcal{U} \ni (p, X_p)$ so that for any $(q, X_q) \in \mathcal{U}$, the system has a smooth solution $\gamma_{q,X_q}(t)$ in $t \in I$ with initial condition $x(t_0) = q$, $y(t_0) = X_q$.
- **Smooth Dependence:** The solution above, viewed as a map $\Upsilon(t, q, X_q) = \gamma_{q,X_q}(t)$, is a smooth map from $I \times \mathcal{U}$ to M.
- **Uniqueness:** If (x_1, y_1) is a solution of the system on an interval $I_1 \ni t_0$, (x_2, y_2) is a solution of the system on an interval $I_2 \ni t_0$, both with initial condition (p, X_p) at t_0, then $(x_1, y_1) = (x_2, y_2)$ on $I_1 \cap I_2$.

As a consequence, we conclude

Theorem 4.2. For any $p \in M$ and any $X_p \in T_pM$, there exists an $\varepsilon > 0$ and a unique geodesic $\gamma = \gamma_{p,X_p}$ defined for $|t| < \varepsilon$ such that $\gamma(0) = p$ and $\dot{\gamma}(0) = X_p$. Moreover, the map $\gamma(t; p, X_p) = \gamma_{p,X_p}(t)$ depends smoothly on (t, p, X_p).

Note that by uniqueness, for any $(p, X_p) \in TM$, there is a maximal interval $J_{p,X_p} \subset \mathbb{R}$ on which a geodesic γ with $\gamma(0) = p$ and $\dot{\gamma}(0) = X_p$ exists. Note that by the "linear re-parametrization remark" above,

$$J_{p,tX_p} = \frac{1}{t} J_{p,X_p}.$$

If $J_{p,X_p} = \mathbb{R}$ for all $(p, X_p) \in TM$, then we say (M, ∇) is **geodesically complete**.

Remark. The dependence of the maximal interval J on the initial data (p, X_p) is not continuous: for example, one can consider in the punctured plane $\mathbb{R}^2 - \{(0,0)\}$. Then the geodesic starting at $(-1, 0)$ in the direction $\langle 1, 0 \rangle$ has maximal existence interval $(-\infty, 1)$, while the geodesic starting at $(-1, 0)$ in any other direction has maximal existence interval $\mathbb{R}$.

It is not hard to see that if M is compact, then it must be geodesically complete. We will see later that for Riemannian manifolds, (M, g) is geodesically complete if and only if as a metric space, (M, dist) is complete.

4.1.2 The Exponential Map and Normal Coordinates

The Exponential Map

Let M be a smooth manifold endowed with a linear connection ∇. Consider

$$\mathcal{E} = \{(p, X_p) \mid \gamma_{p,X_p}(t) \text{ is defined on an interval containing } [0,1]\}.$$

[So by definition $\mathcal{E} = TM$ if and only if (M, g) is geodesically complete.]

By existence and smoothness above, for any $(p, X_p) \in TM$ there is $\varepsilon_0 > 0$ and an open neighborhood $\mathcal{U}$ of (p, X_p) so that for any $(q, X_q) \in \mathcal{U}$, the maximal existence interval J_{q,X_q} of γ_{q,X_q} contains the interval $(-\varepsilon_0, \varepsilon_0)$. As a result,

$$J_{q,\varepsilon_0 X_q/2} \supset (-2, 2),$$

So $\mathcal{E}$ contains a neighborhood of the zero section M in TM. Note that $\mathcal{E} \cap T_pM$ is always a star-like subset in T_pM for any p.

Definition 4.3. The **exponential map** is defined to be

$$\begin{aligned}\exp : \mathcal{E} &\to M,\\ (p, X_p) &\mapsto \exp_p(X_p) := \gamma_{p,X_p}(1).\end{aligned}$$

Example. For $(\mathbb{R}^m, g_0)$, we can identify each $T_p\mathbb{R}^m$ with $\mathbb{R}^m$. Then $\exp_p(X_p) = p + X_p$.

Example. For $(S^1, \mathrm{d}\theta \otimes \mathrm{d}\theta)$, we can identify T_eS^1 with $\mathbb{R}^1$. Then $\exp_e(X_p) = \mathrm{e}^{\mathrm{i}X_p}$.

Remark. Let $M = G$ be a Lie group, endowed with the Levi-Civita connection of the bi-invariant metric on G, then $\exp_e$ coincides with the exponential map $\exp : \mathfrak{g} \to G$ in Lie theory. In particular, if G is a matrix Lie group, then

$$\exp_e(A) = I + A + \frac{A^2}{2!} + \cdots + \frac{A^k}{k!} + \cdots.$$

The smoothness of $\Upsilon(t; p, X_p)$ implies that the exponential map is smooth. In particular, for each $p \in M$, the map

$$\exp_p : T_pM \cap \mathcal{E} \to M$$

is smooth. By definition $\exp_p$ maps $0 \in T_pM$ to $p \in M$. As in Lie theory, we also have the following useful lemma:

Lemma 4.4. For any $p \in M$, if we identify $T_0(T_pM)$ with T_pM, then

$$(\mathrm{d}\exp_p)_0 = \mathrm{Id}|_{T_pM} : T_pM \to T_pM.$$

Proof. For any $X_p \in T_0(T_pM) = T_pM$,

$$(\mathrm{d}\exp_p)_0(X_p) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \exp_p(tX_p) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \gamma(1; p, tX_p) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \gamma(t; p, X_p) = X_p.$$

□

So by the inverse function theorem, we immediately get

Corollary 4.5. For any $p \in M$, there exists a neighborhood V of 0 in T_pM and a neighborhood U of p in M so that $\exp_p : V \to U$ is a diffeomorphism.

Normal Neighborhoods and Normal Coordinates

So for any $p \in M$, there exists a neighborhood $U \subset M$ of p and a neighborhood $\widetilde{V} \subset T_pM$ of 0 so that the exponential map $\exp_p : \widetilde{V} \to U$ is a diffeomorphism. By fixing a basis $\{e_i\}$ of T_pM, we may identify $\widetilde{V}$ with an open subset V of $\mathbb{R}^m$, and as a result, the triple $(\exp_p^{-1}, U, V)$ form a local chart of M near p.

Definition 4.6. If $\widetilde{V}$ is star-like, then we call U a **normal neighborhood** of p, call the local chart $(\exp_p^{-1}, U, V)$ a **normal chart** on M, and call the coordinate system $\{U; x^1, \dots, x^m\}$ a **normal coordinate system** centered at p.

By definition, the normal coordinate system centered at p has the nice characterizing property that any geodesic starting at p is given in such coordinates by

$$\gamma : x(t) = (tv^1, tv^2, \dots, tv^m),$$

where $(v^1, \dots, v^m)$ is the direction of the geodesics. Moreover, we have

Lemma 4.7. Let $\{U; x^1, \dots, x^m\}$ be a normal coordinate system centered at p. Then for all $\vec{v} \in \mathbb{R}^m$ and all $1 \leqslant k \leqslant m$, $\Gamma_{ij}^k(p)v^iv^j = 0$. [In particular, if the linear connection ∇ is torsion free, then $\Gamma_{ij}^k(p) = 0$ for all i, j, k.]

Proof. Put the parametric equation $x(t) = (tv^1, tv^2, \dots, tv^m)$ of a geodesic into the geodesic equation, we get for $1 \leqslant k \leqslant m$,

$$0 = \ddot{x}^k(t) + \Gamma_{ij}^k(\gamma(t))\dot{x}^i(t)\dot{x}^j(t) = \Gamma_{ij}^k(\gamma(t))v^iv^j.$$

Letting $t = 0$, we get $\Gamma_{ij}^k(p)v^iv^j = 0$ for all $\vec{v}$ and for any $1 \leqslant k \leqslant m$. □

Normal Convex Neighborhoods

We may go a lot further.

Theorem 4.8 (Whitehead)**.** For any smooth manifold M with a linear connection, any p has a neighborhood U such that U is a normal neighborhood for any $q \in U$.

Let's explain the meaning before we prove the theorem. For any q, $q' \in U$, since U is a normal neighborhood of q, there is a vector $X_{q\to q'} \in T_qM$ so that

$$\gamma_{q,q'}(t) := \exp_q(tX_{q\to q'})$$

is a geodesic from $q = \gamma(0)$ to $q' = \gamma(1)$ that lies entirely in U. Such an open set is called a **convex normal neighborhood** of p. So Whitehead theorem claims that any p admits a normal convex neighborhood. As a consequence, we can prove

Corollary 4.9. Any smooth manifold M admits a good covering.

Proof. Endow with M a linear connection ∇. Then by Whitehead theorem, each $p \in M$ admits a normal convex neighborhood U_p. Because each normal convex neighborhood is contractible [since it is diffeomorphic to a star-like subset in a vector space], and because arbitrary intersection of normal convex neighborhoods is still a normal convex neighborhood, they form a good covering of M. □

Proof of Whitehead Theorem

Proof. **Step 1:** There is a neighborhood U of p such that for any $q \in U$, there is a normal chart $(\exp_q^{-1}, U_q, V_q)$ with $U_q \supset U$.

Take a neighborhoods U_1 of $p \in M$ and a neighborhood $\widetilde{U}_1$ of $(p, 0) \in TM$ over U_1 [i.e. $\pi(\widetilde{U}_1) = U_1$, where $\pi : TM \to M$ is the bundle projection] such that for each $q \in U_1$,

- $\widetilde{U}_1$ is fiberwise star-like, i.e. $V_q = \widetilde{U}_1 \cap T_qM$ is star-like in T_qM,
- the exponential map $\exp_q : V_q \to U_q$ is a diffeomorphism.

Consider the map

$$\begin{aligned}\Psi : \widetilde{U}_1 &\to M \times M,\\ (q, X_q) &\mapsto (q, \exp_q(X_q)).\end{aligned}$$

The Jacobian of Ψ at $(p, 0)$ is $\begin{pmatrix} I & 0 \\ I & I \end{pmatrix}$. So Ψ is a local diffeomorphism, i.e. it maps a smaller neighborhood $\mathcal{U}_1 \subset \widetilde{U}_1$ diffeomorphically onto a neighborhood of (p, p) in $M \times M$. In particular, one may find a neighborhood U of p in M so that $U \times U \subset \Psi(\mathcal{U}_1)$. By construction, $\Psi^{-1}(U \times U) \cap T_qM \subset \mathcal{U}_1 \cap T_qM \subset V_q$ and thus

$$U \subset \exp_q(\Psi^{-1}(U \times U) \cap T_qM) \subset U_q.$$

Step 2: U can be chosen to be normal with respect to any $q \in U$.

We fix a normal chart (φ, U_0, V_0) centered at p, with normal coordinates $x^1, \dots, x^m$, where for simplicity we denote $\varphi = \exp_p^{-1}$. Apply Lemma 4.7 and shrink U_0 if necessary, we may assume that the matrix $(\delta_{ij} - \sum_k \Gamma_{ij}^k x^k)$ is "positive" at each point in U_0, i.e. such that $(\delta_{ij} - \sum_k \Gamma_{ij}^k x^k) v^i v^j \geqslant 0$ for all $\vec{v} \in \mathbb{R}^m$ and all $q \in U_0$. We may assume U_q we get in Step 1 are all inside U_0.

Now we endow T_pM with any inner product, and shrink U we get in Step 1 so that $\varphi(U)$ is a ball of radius δ. By Step 1, for any q, $q' \in U$, there is a vector $X_{q\to q'} \in T_qM$ with $\exp_q(X_q^{q'}) = q'$. Since V_q is star-like, the curve $\gamma_{q,q'}(t) := \exp_q(tX_q^{q'})$ is a geodesic from $q = \gamma(0)$ to $q' = \gamma(1)$ that lies in U_q. It remains to prove that $\gamma_{q,q'}(t)$ $(0 \leqslant t \leqslant 1)$ lies in U.

Since the geodesic $\gamma_{q,q'}$ lies in U_0, we work on its parametric equations $x^i = x^i(t)$. Consider the function $f(t) = \sum_i (x^i(t))^2$. Then

$$\begin{aligned}\ddot{f}(t) &= 2\sum_i [(\dot{x}^k(t))^2 + \ddot{x}^k(t)x^k(t)] \\ &= 2\sum_k [(\dot{x}^k(t))^2 - \Gamma_{ij}^k \dot{x}^i(t)\dot{x}^j(t)x^k(t)] \\ &= 2[\delta_{ij} - \sum_k \Gamma_{ij}^k x^k(t)]_{\gamma(t)}\dot{x}^i(t)\dot{x}^j(t) \geqslant 0.\end{aligned}$$

As a consequence, f is convex and thus $f(t) \leqslant \max\{f(0), f(1)\}$ for $0 \leqslant t \leqslant 1$. Since q, $q' \in U$, we have $f(0)$, $f(1) \leqslant \delta^2$. So the geodesic $\gamma_{q,q'}$ is inside U. □

4.2 Geodesics on Riemannian Manifolds

After defining geodesics as "self-parallel curves" on any smooth manifold with linear connection, today we will put the Riemannian metric structure into this picture and study what do we gain with this new structure (for the geodesics as self-parallel curves and as integral curves, for the exponential map, and for the normal coordinates etc).

4.2.1 Geodesics as Integral Curves

"Speed" of a Geodesics

Let (M, g) be a Riemannian manifold, and $\gamma : [a, b] \to M$ a smooth curve in M. Recall that γ is a geodesics if and only if it is self-parallel, i.e. $\nabla_{\dot{\gamma}}\dot{\gamma} = 0$. By metric compatibility,

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\dot{\gamma}, \dot{\gamma}\rangle = \nabla_{\dot{\gamma}}\langle\dot{\gamma}, \dot{\gamma}\rangle = \langle\nabla_{\dot{\gamma}}\dot{\gamma}, \dot{\gamma}\rangle + \langle\dot{\gamma}, \nabla_{\dot{\gamma}}\dot{\gamma}\rangle = 0.$$

As a result, we get

Proposition 4.10. If γ is a geodesic on a Riemannian manifold, then $|\dot{\gamma}|$ must be a constant for all t.

Note that this also implies that a re-parametrization of a geodesic is again a geodesic if and only if the re-parametrization is a linear re-parametrization.

In particular, on a Riemannian manifold one can always re-parameterize a geodesic so that its "speed" is 1:

Definition 4.11. We will call a geodesics γ on a Riemannian manifold satisfying $|\dot{\gamma}(t)| = 1$ a **normal geodesics**.

Of course given any geodesic, the corresponding normal geodesic is nothing else but the arc-length re-parametrization of the given geodesic.

Geodesics as Integral Curves at the Presence of Metric

Last time by introducing $y^i = \dot{x}^i$ we converted the system of second order ODEs for a geodesic to a system of first order ODEs

$$\begin{cases}\dot{x}^k = y^k, \\ \dot{y}^k = -\Gamma_{ij}^k y^i y^j,\end{cases} \qquad 1 \leqslant k \leqslant m$$

using which we get the existence, smooth dependence and uniqueness of geodesics. In other words, the problem of finding a local geodesic is equivalent to finding the integral curve of the vector field

$$\widetilde{X} = y^k \frac{\partial}{\partial x^k} - \Gamma_{ij}^k y^i y^j \frac{\partial}{\partial y^k}.$$

Although one can show that the vector field $\widetilde{X}$ defined above is really globally defined (i.e. independent of the choice of coordinates), its geometric meaning is not that obvious.

It turns out that if one transfer from the tangent bundle to the cotangent bundle, then there is a geometrically important vector field whose integral curves give geodesics on M. Recall that given any coordinate chart $(U, x^1, \ldots, x^m)$ on M, any 1-form ω can be expressed locally on U as $\omega = \xi_i dx^i$ and as a result, one gets a coordinate chart $(T^*U, x^1, \ldots, x^m, \xi_1, \ldots, \xi_m)$ for the cotangent bundle T^*M.

Now given a Riemannian metric g on M, i.e. an inner product on each tangent space, one gets a dual inner product on each cotangent space. Consider the smooth function defined on $T^*M \setminus \{0\}$ by

$$f(x, \xi) = \frac{1}{2}|\xi|_x^2 = \frac{1}{2}g^{ij}(x)\xi_i\xi_j.$$

Definition 4.12. The **Hamiltonian vector field** of f is

$$H_f = \sum \frac{\partial f}{\partial \xi_i}\frac{\partial}{\partial x^i} - \frac{\partial f}{\partial x^i}\frac{\partial}{\partial \xi_i}.$$

It is a vector field on $T^*M \setminus \{0\}$ which preserves f (and thus preserves $|\xi|_x$),

$$H_f(f) = 0.$$

As a consequence, it defines a vector field on each level of f, and in particular on the cosphere bundle

$$S^*M = \{(x, \xi) \mid \|\xi\|_x = 1\}.$$

By definition the integral curves of H_f are the curves $\Gamma = \Gamma(t)$ such that

$$\dot{\Gamma}(t) = H_f(\Gamma(t)).$$

More precisely, if we denote

$$\Gamma(t) = (x^1, \ldots, x^m(t), \xi_1(t), \ldots, \xi_m(t)),$$

then any integral curve of H_f satisfies the following **Hamilton equations**

$$\begin{cases} \dot{x}^k = \dfrac{\partial f}{\partial \xi_k}, \\ \dot{\xi}_k = -\dfrac{\partial f}{\partial x^k}. \end{cases}$$

The flow generated by H_f on S^*M is called the **geodesic flow** of (M, g), which is very important in studying Riemannian manifolds. Now we prove

Theorem 4.13. Any integral curve of H_f on S^*M, when projected onto M, is a normal geodesic in M. Conversely, any normal geodesic in M arises in this way.

Proof. Let $\Gamma(t) = (x^1(t), \ldots, x^m(t), \xi_1(t), \ldots, \xi_m(t))$ be an integral curve of H_f, then the Hamilton equations become

$$\dot{x}^k = \frac{\partial f}{\partial \xi_k} = \frac{1}{2}g^{ij}\delta_{ik}\xi_j + \frac{1}{2}g^{ij}\xi_i\delta_{jk} = g^{kj}\xi_j$$

$$\dot{\xi}_k = -\frac{\partial f}{\partial x^k} = -\frac{1}{2}\frac{\partial g^{ij}}{\partial x^k}\xi_i\xi_j$$

From the first equation we get $\xi_k = g_{lk}\dot{x}^l$. Put this into the second equation, we have

$$\frac{\partial g_{lk}}{\partial x^i}\dot{x}^i\dot{x}^l + g_{lk}\ddot{x}^l = -\frac{1}{2}\frac{\partial g^{ij}}{\partial x^k}g_{li}\dot{x}^l g_{nj}\dot{x}^n.$$

Note that

$$-\frac{\partial g^{ij}}{\partial x_k}g_{li}g_{nj} = g^{ij}\frac{\partial g_{li}}{\partial x^k}g_{nj} = \frac{\partial g_{nl}}{\partial x^k},$$

the equation becomes

$$g_{lk}\ddot{x}^l = -\frac{\partial g_{lk}}{\partial x^i}\dot{x}^i\dot{x}^l + \frac{1}{2}\frac{\partial g_{nl}}{\partial x^k}\dot{x}^l\dot{x}^n = -\frac{\partial g_{jk}}{\partial x^i}\dot{x}^i\dot{x}^j + \frac{1}{2}\frac{\partial g_{ij}}{\partial x^k}\dot{x}^i\dot{x}^j.$$

In other words,

$$\ddot{x}^l = g^{kl}\left(-\frac{\partial g_{jk}}{\partial x^i}\dot{x}^i\dot{x}^j + \frac{1}{2}\frac{\partial g_{ji}}{\partial x^k}\dot{x}^i\dot{x}^j\right) = -\frac{1}{2}g^{kl}\left(\frac{\partial g_{jk}}{\partial x^i}\dot{x}^i\dot{x}^j + \frac{\partial g_{ik}}{\partial x^j}\dot{x}^j\dot{x}^i - \frac{\partial g_{ji}}{\partial x^k}\dot{x}^i\dot{x}^j\right),$$

which is exactly the geodesic equation since

$$\Gamma^l_{ij} = \frac{1}{2}g^{kl}(\partial_j g_{ki} + \partial_i g_{jk} - \partial_k g_{ij}).$$

So the projected curve $\gamma(t) = (x^1(t), \dots, x^m(t))$ is a geodesic on M. It is normal since

$$g_{kl}\dot{x}^k\dot{x}^l = g_{kl}g^{kj}g^{li}\xi_j\xi_i = g^{ij}\xi_j\xi_i = 1.$$

Conversely, for any geodesic $\gamma(t) = (x^1(t), \dots, x^m(t))$, we let $\xi_k = g_{lk}\dot{x}^l$. Then the above computations shows that $\Gamma(t) = (x^1(t), \dots, x^m(t), \xi_1(t), \dots, \xi_m(t))$ is an integral curve of H_f in S^*M. □

Remark. The function $|\xi|^2$ is the **symbol** of the Laplace-Beltrami operator Δ_g. So the geodesic flow is also closely related to spectral geometry.

Remark. As a consequence, (M, g) is geodesically complete if and only if the vector field H_f on S^*M is complete. Note that if M is compact, then S^*M is compact, and thus any smooth vector field on S^*M is complete. As a result, any compact Riemannian manifold is geodesically complete.

4.2.2 The Exponential Map at the Presence of Metric

The Injectivity Radius

Now let's turn to the exponential map and figure out what do we gain with g. For a Riemannian manifold, by definition the point $\exp_p(X_p)$ is the end point of the geodesic segment that starts at p in the direction of X_p whose length equals $|X_p|$.

In general, the map $\exp : \mathcal{E}_p \cap T_pM \to M$ is not a global diffeomorphism, even if it may be defined everywhere in T_pM. For example, on the round sphere S^m, $\exp_p$ is a diffeomorphism from any ball $B_r(0) \subset T_pM$ of radius $r < \pi$ to an open region in S^m, but it fails to be injective on the ball $B_r(0)$ with $r > \pi$.

Definition 4.14. The **injectivity radius** of Riemannian manifold (M, g) at $p \in M$ is

$$\mathrm{inj}_p(M, g) := \sup\{r \mid \exp_p \text{ is a diffeomorphism on } B_r(0) \subset T_pM\},$$

and the **injectivity radius** of (M, g) is

$$\mathrm{inj}(M, g) := \inf\{\mathrm{inj}_p(M, g) \mid p \in M\}.$$

Example. $\mathrm{inj}(S^m, g_{S^m}) = \pi$.

Remark. If M is compact, then of course

$$0 < \mathrm{inj}(M, g) \leqslant \mathrm{diam}(M, g),$$

where $\mathrm{diam}(M, g) = \sup_{p,q\in M} d(p, q)$ is the diameter of (M, g). But for noncompact manifolds M, we may have $\mathrm{inj}(M, g) = 0$ or $+\infty$. [But for any p, we always have $\mathrm{inj}_p(M, g) > 0$.]

For any $\rho < \mathrm{inj}_p(M, g)$, we have $B_\rho(0) \subset T_pM \cap \mathcal{E}$, where $B_\rho(0)$ is the ball of radius ρ in (T_pM, g_p) centered at 0.

Definition 4.15. We will call $B(p,\rho) = \exp_p(B_\rho(0))$ the **geodesic ball** of radius ρ centered at p in M, and its boundary $S(p,\rho) = \partial B(p,\rho)$ the **geodesic sphere** of radius ρ centered at p in M.

Now let γ be any normal geodesic starting at p. Then for $\rho < \mathrm{inj}_p(M,g)$, we have $\gamma((0,\rho)) \subset B(p,\rho)$ and $\exp_p^{-1}(\gamma((0,\rho)))$ is the line segment in $B_\rho(0) \subset T_pM$ starting at 0 in the direction $\dot\gamma$ whose length is ρ. As a consequence, the geodesics starting at p of lengths less than $\mathrm{inj}_p(M,g)$ are exactly the images under $\exp_p$ of line segments starting at 0 of lengths no more than $\mathrm{inj}_p(M,g)$. In particular,

Corollary 4.16. Suppose $p \in M$ and $\rho < \mathrm{inj}_p(M,g)$. Then for any $q = \exp_p(X_p) \in B(p,\rho)$, the curve $\gamma(t) = \exp_p(tX_p)$ is the unique normal geodesic connecting p to q whose length is less than ρ.

Remark. No matter how close p and q are to each other, one might be able to find other geodesics connecting p to q whose length is longer. To see this, one can look at cylinders or torus, in which case one can always find infinitely many geodesics connecting two arbitrary given points p and q.

Gauss Lemma

Last time we should that the exponential $(\mathrm{d}\exp_p)_0 = \mathrm{Id}$. Now let $(p, X_p) \in \mathcal{E}$. By definition, $\exp_p$ maps the point $X_p \in T_pM$ to the point $\exp_p(X_p) \in M$. In general, the differential $\mathrm{d}\exp_p$ at X_p is no longer the identity map Id. [In fact, if $(\mathrm{d}\exp_p)_{X_p} = \mathrm{Id}$ for all p and X_p, then $\exp_p$ is an isometry from (T_pM, g_p) to (M,g) and thus (M,g) is flat.] However, we can prove that $\exp_p$ is always a "radial isometry":

Lemma 4.17 (Gauss lemma)**.** Let (M,g) be a Riemannian manifold and $(p,X_p) \in \mathcal{E}$. Then for any $Y_p \in T_pM = T_{X_p}(T_pM)$, we have

$$\langle (\mathrm{d}\exp_p)_{X_p} X_p, (\mathrm{d}\exp_p)_{X_p} Y_p \rangle_{\exp_p(X_p)} = \langle X_p, Y_p \rangle_p.$$

Proof. Without loss of generality, we may assume X_p, $Y_p \neq 0$. By linearity, it's enough to check the lemma for $Y_p = X_p$ and $Y_p \perp X_p$.
Case 1: $Y_p = X_p$. If we denote $\gamma(t) = \exp(tX_p)$, then $X_p = \dot\gamma(0)$ and

$$(\mathrm{d}\exp_p)_{X_p} X_p = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=1} \exp_p(tX_p) = \dot\gamma(1).$$

Since geodesics are always of constant speed, we conclude

$$(\mathrm{d}\exp_p)_{X_p} X_p, (\mathrm{d}\exp_p)_{X_p} X_p \rangle = \langle \dot\gamma(1), \dot\gamma(1) \rangle = \langle \dot\gamma(0), \dot\gamma(0) \rangle = \langle X_p, X_p \rangle.$$

Case 2: $Y_p \perp X_p$. Under this condition one can find a curve $\gamma_1(s)$ in the sphere of radius $|X_p|$ in T_pM with $\gamma_1(0) = X_p$ and $\dot\gamma_1(0) = Y_p$. Since $(p, X_p) \in \mathcal{E}$, we see that there exists $\varepsilon > 0$ so that for all $0 < t < 1$ and $-\varepsilon < s < \varepsilon$,

$$(p, t\gamma_1(s)) \in \mathcal{E}.$$

Let $A = \{(t,s) \mid 0 < t < 1, -\varepsilon < s < \varepsilon\}$ and consider the smooth map

$$f : A \to M,$$
$$(t,s) \mapsto f(t,s) := \exp_p(t\gamma_1(s)).$$

As usual we denote $f_t = \mathrm{d}f\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)$ and $f_s = \mathrm{d}f\left(\frac{\mathrm{d}}{\mathrm{d}s}\right)$. Then by definition

$$f_t(1,0) = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=1} \exp_p(tX_p) = (\mathrm{d}\exp_p)_{X_p} X_p,$$
$$f_s(1,0) = \frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0} \exp_p(\gamma_1(s)) = (\mathrm{d}\exp_p)_{X_p} Y_p$$

and thus

$$\langle (\mathrm{d}\exp_p)_{X_p} X_p, (\mathrm{d}\exp_p)_{X_p} Y_p \rangle = \langle f_t(1,0), f_s(1,0) \rangle.$$

On the other hand, we have

- for each fixed s_0, $f(t,s_0)$ is a geodesic with tangent vector field f_t. So
$$\nabla_{f_t} f_t = 0.$$
- since ∇ is torsion free, $\nabla_{f_s} f_t - \nabla_{f_t} f_s = [f_s, f_t] = \mathrm{d}f([\partial_s, \partial_t]) = 0$ and thus
$$\nabla_{f_s} f_t = \nabla_{f_t} f_s.$$
- since γ_1 lies in the sphere of radius $|X_p|$, the length
$$|f_t| = |\gamma_1(s)| = |X_p|$$
is a constant.

As a consequence of these three facts,
$$\frac{\partial}{\partial t}\langle f_s, f_t\rangle = \langle \nabla_{f_t} f_s, f_t\rangle + \langle f_s, \nabla_{f_t} f_t\rangle = \langle \nabla_{f_s} f_t, f_t\rangle = \frac{1}{2}\nabla_{f_s}\langle f_t, f_t\rangle = 0,$$
i.e. $\langle f_t, f_s\rangle$ is independent of t. Since
$$\lim_{h\to 0} f_s(h,0) = \lim_{h\to 0} \frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0} \exp_p(h\gamma_1(s)) = \lim_{t\to 0} \mathrm{d}(\exp_p)_{hX_p}(hY_p) = 0,$$
We conclude $\langle f_t(1,0), f_s(1,0)\rangle = 0$, which proves the lemma. □

Geometrically, Gauss lemma implies

Corollary 4.18 (The Geometric Gauss Lemma)**.** For any $\rho < \mathrm{inj}_p(M,g)$ and any $q \in S(p,\rho)$, the shortest geodesic connecting p to q is orthogonal to $S(p,\rho)$.

Local Shortest Curves are Geodesics

As a consequence of Gauss lemma, we may strengthen Corollary 4.16 to

Theorem 4.19. Suppose $p \in M$ and $\delta < \mathrm{inj}_p(M,g)$. Then for any $q = \exp_p(X_p) \in B(p,\delta)$, the geodesic $\gamma(t) = \exp_p(tX_p)$ $(0 \leqslant t \leqslant 1)$ is the only piecewise smooth curve connecting p and q with length $d(p,q)$.

Proof. Let $\sigma : [0,1] \to M$ be any piecewise smooth curve with $\sigma(0) = p$, $\sigma(1) = q$, and parameterized with constant speed. We want to show $L(\sigma) \geqslant d(p,q)$, with equality holds if and only if $\sigma = \gamma$.

Without loss of generality, we may assume $p \notin \sigma([0,1])$ [otherwise we may take $t_0 = \sup\{t \mid \sigma(t) = p\}$ and consider the curve $\sigma|_{[t_0,1]}$ instead] and assume $\sigma((0,1)) \subset B(p,\delta)$ [otherwise we may take $t_1 = \inf\{t \mid \sigma(t) \in S(p,\delta)\}$ and consider the curve $\sigma|_{[0,t_1]}$ instead]. As a result, there exists unit vectors $w(t) \in S_pM$ and real numbers $r(t) \in (0,\delta]$ such that
$$\sigma(t) = \exp_p(r(t)w(t)).$$
It follows
$$\dot\sigma(t) = (\mathrm{d}\exp_p)_{r(t)w(t)}(r'(t)w(t) + r(t)\dot w(t)).$$
Note that $w(t) \in S_pM$ for all t implies $w(t) \perp \dot w(t)$. So by Gauss lemma,
$$(\mathrm{d}\exp_p)_{r(t)w(t)}(r'(t)w(t)) \perp (\mathrm{d}\exp_p)_{r(t)w(t)}(r(t)\dot w(t))$$
and thus
$$|\dot\sigma(t)|^2 \geqslant \langle (\mathrm{d}\exp_p)_{r(t)w(t)}(r'(t), w(t)), (\mathrm{d}\exp_p)_{r(t)w(t)}(r'(t)w(t))\rangle = |r'(t)|^2$$

So if we denote $b = \text{Length}(\sigma)$, then $b = |\dot\sigma(t)|$ at all smooth points t of σ and thus

$$\begin{aligned} b = \text{Length}(\sigma) = \int_0^1 |\dot\sigma(1)|\mathrm{d}t &= \frac{1}{b}\int_0^1 |\dot\sigma(t)|^2 \mathrm{d}t \\ &\geqslant \frac{1}{b}\int_0^1 |r'(t)|^2 \mathrm{d}t \\ &\geqslant \frac{1}{b}\left(\int_0^1 |r'(t)|\mathrm{d}t\right)^2 \geqslant \frac{1}{b}\left(\int_0^1 r'(t)\mathrm{d}t\right)^2 \geqslant \frac{\delta^2}{b}, \end{aligned}$$

where we used Cauchy-Schwartz inequality and the fact $r(1) \leqslant \delta$. It follows that $b \geqslant \delta$ as desired. Moreover, if the equality holds, then $\dot w = 0$ and $|r'(t)|$ is constant, which implies that σ is precisely the geodesic $\gamma(t) = \exp_p(tX_p)$. □

Riemannian Metric Tensor in Riemannian Normal Coordinate System

Now we turn to normal coordinate systems for Riemannian manifolds. Since the Levi-Civita connection is torsion-free, we have seen that with respect to any normal coordinate system centered at p,

$$\Gamma_{ij}^k(p) = 0, \quad 1 \leqslant i, j, k \leqslant m.$$

So what do we gain from the metric? Recall that behind a normal coordinate system $(\exp_p^{-1}, U, V)$ there hides an identification between $\widetilde V = \exp_p^{-1}(U) \subset T_pM$ and $V \subset \mathbb{R}^m$, which is realized after a choice of a basis e_i of T_pM. For a Riemannian manifold (M, g), we will always identify $\widetilde V = \exp_p^{-1}(U) \subset T_pM$ and an open subset $V \subset \mathbb{R}^m$ by choosing an orthonormal basis $\{e_1, \dots, e_m\}$ of T_pM, and call the resulting normal coordinate system a **Riemannian normal coordinate system** at p.

With a Riemannian normal coordinate system at hand, we can prove the following stronger result [c.f. (2.12)]:

Lemma 4.20. Let (M, g) be a Riemannian manifold, and $\{U; x^1, \dots, x^m\}$ be a Riemannian normal coordinate system centered at p. Then

(1) For all $1 \leqslant i, j \leqslant m$, $g_{ij}(p) = \delta_{ij}$.

(2) For all $1 \leqslant i, j, k \leqslant m$, $\partial_k g_{ij}(p) = 0$.

(3) $G(p) = 1$ and $\partial_i G(p) = 0$ for all $1 \leqslant i \leqslant m$, where $G = \det(g_{ij})$.

Proof. (1) By definition of Riemannian normal coordinate system, we have $\partial_i|_p = \mathrm{d}(\exp_p)_0 e_i = e_i$, which implies $g_{ij}(p) = \delta_{ij}$ since $\{e_i\}$ is chosen to be orthonormal.

(2) By metric compatibility, we have

$$\partial_k g_{ij}(p) = \langle \nabla_{\partial_k}\partial_i, \partial j\rangle(p) + \langle \partial_i, \nabla_{\partial_k}\partial_j\rangle(p) = \Gamma_{ki}^l(p) g_{lj}(p) + \Gamma_{kj}^l(p) g_{li}(p)$$

and thus the conclusion follows from the fact $\Gamma_{ij}^k(p) = 0$.

(3) This is a direct consequence of (1), (2) and the definition of determinant. □

Remark. As a result, in a Riemannian normal coordinate centered at p, we have

$$g_{ij} = \delta_{ij} + \mathcal{O}(|x|^2) \quad \text{and} \quad \det(g_{ij}) = 1 + \mathcal{O}(|x|^2)$$

near p. In fact, as we will see later, what hides in $\mathcal{O}(x^2)$ are the curvature information of (M, g) at p: the Riemannian curvature for g_{ij}, and the Ricci curvature for $\det(g_{ij})$.

In Riemannian normal coordinate system centered at p, many differential operators have very simple expressions at p. As a result, it can simplify computations a lot. For example, given any smooth vector field $X = X^i\partial_i$, we have defined its divergence to be $\operatorname{div} X = \dfrac{1}{\sqrt{G}}\partial_i(\sqrt{G}X^i)$.

By Lemma 4.20 (3), we have $\partial_i(\sqrt{G})(p) = 0$. So it follows that in a given Riemannian normal coordinate system centered at p,

$$\operatorname{div} X(p) = \sum_i \partial_i X^i(p).$$

As a result, the Laplacian Δf at p also has a very simple expression,

$$\Delta f(p) = -\operatorname{div} \nabla f(p) = -\partial_i^2 f(p).$$

Similarly the Hessian $\nabla^2 f$ of f, in the Riemannian normal coordinates, becomes

$$(\nabla^2 f)(\partial_i, \partial_j)(p) = \partial_j \partial_i f(p) - (\nabla_{\partial_j} \partial_i) f(p) = \partial_j \partial_i f(p).$$

In particular, we see that at each p,

$$\operatorname{tr}(\nabla^2 f)(p) = g^{ij}(p)(\nabla^2 f)(\partial_i, \partial_j)(p) = g^{ij}(p)\partial_i \partial_j f(p) = \partial_i^2 f(p).$$

So we proved

Proposition 4.21. For any $f \in \mathcal{C}^\infty(M)$, $\Delta f = -\operatorname{tr}(\nabla^2 f)$.

This formula can be viewed as a second definition of the Laplace operator Δ.

Strongly Convex Neighborhood

Finally we take a look at Whitehead's theorem for Riemannian manifolds. We may carefully check the proof of Whitehead's theorem last time: in 2 steps we choose the convex normal neighborhood U carefully so that in the normal coordinate system, $\exp_p(U)$ is a ball in $\mathbb{R}^m$. In current setting if we use Riemannian normal coordinate system, then that means U is a small geodesic ball centered at p. Also in step 1 we may choose $\widetilde{U}_1$ carefully so that each V_q is a ball in (T_qM, g_q) instead of only a star-like subset in T_qM, which means each U_q is a geodesic ball in the construction. In view of Theorem 4.19, we conclude that for such a geodesic ball U, any two points $q_1, q_2 \in U$ can be connected by a unique geodesic γ of length $d(q_1, q_2)$, and this minimizing geodesic γ lies in U.

Such a neighborhood is called **strongly convex** or **geodesically convex**. So we get

Theorem 4.22 (Whitehead)**.** Let (M, g) be a Riemannian manifold, then for any $p \in M$ there exists $\rho > 0$ so that the geodesic ball $B(p, \rho)$ is strongly convex.

4.3 Existence of Shortest Geodesics

By using ODEs (or equivalently, vector fields on tangent or cotangent bundles), we have proved local existence of geodesics as well as the local length-minimizing property. In what follows we turn to global aspects of geodesics.

4.3.1 Length Minimizing Curves on Compact Riemannian Manifolds

Length Minimizing Curves are Geodesics

Last time we showed that near any point p, there is a neighborhood U so that for any two points $q_1, q_2 \in U$, there is a unique normal geodesics γ in U that connects q_1 and q_2, and moreover, γ is the only shortest curve connecting q_1 and q_2. As a consequence, one gets

Proposition 4.23. Let p, q be two points on (M, g). If $\gamma : [0, l] \to M$ is a piecewise smooth curve that connects $p = \gamma(0)$ and $q = \gamma(l)$ and is parameterized by arc-length [so $l = \text{Length}(\gamma)$], and if $l = d(p, q)$, then γ is smooth and is a geodesics.

Proof. First note that by definition of the Riemannian distance function d, there is no piecewise smooth curve connecting p and q of length less than l.

By compactness of $\gamma([0,l])$, there is $\varepsilon > 0$ so that any $t_1, t_2 \in [0,l]$ with $|t_1 - t_2| < \varepsilon$, there is a unique arc-length parameterized length-minimizing curve γ_{t_1,t_2} connecting $\gamma(t_1)$ and $\gamma(t_2)$ which is a normal geodesic. If $\gamma_{t_1,t_2} \neq \gamma|_{[t_1,t_2]}$, then γ_{t_1,t_2} is strictly shorter than $\gamma|_{[t_1,t_2]}$, and we may replace $\gamma|_{[t_1,t_2]}$ by γ_{t_1,t_2} to get a piecewise smooth curve which is shorter than γ, a contradiction. So for any $t_1, t_2 \in [0,l]$ with $|t_1 - t_2| < \varepsilon$, $\gamma|_{[t_1,t_2]}$ is smooth and satisfies the geodesic equation. It follows that γ is smooth and satisfies the geodesic equation on the whole $[0,l]$. □

So the shortest curve between any two points on a connected Riemannian manifold must be a geodesic. Here are two natural subsequent questions:

> **Question 1:** Given $p, q \in M$, does there exist a smooth curve of length $d(p,q)$ between p and q [which then becomes a shortest geodesic]?
>
> **Question 2:** Given $p, q \in M$, is the length-minimizing curve the only geodesics connecting p and q?

By studying very simple examples, it is quite obvious that the answers to both questions are **NO**. However, as usual a simple "no" is not a satisfied answer. In the next couple lectures we will give more in-depth answer to these two questions. For example, under which condition the answer to **Question 1** is yes? For **Question 2**, how does a geodesic change from length minimizing to non-minimizing?

Length Minimizing Curves on Compact Riemannian Manifolds

For the remaining of this lecture, we focus on **Question 1**, or more generally, on conditions for the existence of various "global shortest geodesics". So let's start with a simple counterexample to **Question 1**:

Example. Let $M = \mathbb{R}^2 \setminus \{0\}$ be the punctured plane, equipped with standard Euclidean flat metric. Then there is no smooth curve of length 2 connecting the points $(-1,0)$ and $(1,0)$.

One can easily find the issue in this example: there are many curves of length $2+\varepsilon$ connecting $(-1,0)$ and $(1,0)$, but their "limit curve" does not exist as a curve in M because of the puncture. In fact, this is always the case for those examples that **Question 1** has answer "NO": Given any connected Riemannian manifold (M,g) and any $p, q \in M$, by definition there exist piecewise smooth curves γ_ε of length no more than $d(p,q) + \varepsilon$. By using the technique in the proof of Proposition 4.23, one can even assume these curves to be "piecewise geodesics" [i.e. piecewise smooth with each piece a geodesic]. But these curves cannot converge to a piecewise smooth curve in M: If they converge (in the uniform convergence topology) to a piecewise smooth curve in M, then by the lower semi-continuity of the length functional [c.f. Exercise 1.1], the limit curve must has length $d(p,q)$ and thus by Proposition 4.23, must be a shortest geodesic connecting p and q.

Now suppose M is compact. As one can imagine, in this case such a sequence will converge, and thus give us a YES to **Question 1**. The key observation is:

Lemma 4.24. Let $\gamma_i : [0,1] \to M$ be a family of piecewise smooth curve parameterized with constant speed [i.e. with $|\dot{\gamma}_i(t)| = \text{Length}(\gamma_i)$ at all smooth points t of γ_i], such that $\text{Length}(\gamma_i) < L$ for some constant L, then the family $\{\gamma_i\}$ is equicontinuous.

Proof. For any $x_0 \in [0,1]$ and any $\varepsilon > 0$, if we take $\delta = \dfrac{\varepsilon}{L}$, then for $|x - x_0| < \delta$,

$$d(\gamma_i(x), \gamma_i(x_0)) = |x - x_0| \text{Length}(\gamma_i) < \varepsilon, \quad \forall i,$$

and the conclusion follows. □

Now we prove

Theorem 4.25. If (M,g) is a compact connected Riemannian manifold, then for any p, q on M, there is a geodesic of length $d(p,q)$ connecting p and q.

Proof. Let γ_i be a sequence of piecewise smooth curves with

$$\gamma_i(0) = p, \quad \gamma_i(1) = q, \quad \text{Length}(\gamma_i) < d(p,q) + \frac{1}{i},$$

parameterized with "constant speed". By Lemma 4.24, $\{\gamma_i\}$ is equicontinuous. It is also pointwise precompact since M is a compact metric space. Applying Arzela-Ascoli theorem we know γ_i has a subsequence that converges to a continuous map $\gamma : [0,1] \to M$. To get a piecewise smooth curve out of γ, we fix $\varepsilon_0 < \text{inj}(M,g)$, and fix N such that $\frac{1}{N} < \frac{\varepsilon_0}{d(p,q)+1}$. Split each γ_i into N pieces, $\gamma_i^j = \gamma_i|_{[\frac{j}{N},\frac{j+1}{N}]}$. Then γ_i^j converges to $\gamma^j = \gamma|_{[\frac{j}{N},\frac{j+1}{N}]}$. Denote $p_j = \gamma\left(\frac{j}{N}\right)$. Then

$$d(p_j, p_{j+1}) \leqslant \liminf_{i\to\infty} \text{Length}(\gamma_i^j) \leqslant \frac{d(p,q)}{N} < \varepsilon_0.$$

Let $\widetilde{\gamma}$ be the piecewise geodesic obtained by connecting each p_j to p_{j+1} by the shortest geodesic (which exists since $\varepsilon_0 < \text{inj}(M,g)$). Since by definition $p_0 = p$, $p_N = q$ and

$$\text{Length}(\widetilde{\gamma}) = \sum_{j=0}^{N-1} d(p_j, p_{j+1}) \leqslant d(p,q),$$

we conclude from Proposition 4.23 that $\widetilde{\gamma}$ is the shortest geodesic from p to q (whose length is $d(p,q)$). □

Length Minimizing Curves in Given Path-homotopy Class

With a little bit more work, one can find geodesics between p and q that are not absolutely length minimizing, but only "relatively length minimizing":

Theorem 4.26. Let (M,g) be a compact connected Riemannian manifold, and p, q are two points in M. Then in each path-homotopy class of curves γ with $\gamma(0) = p$, $\gamma(1) = q$, there is a length-minimizing curve and the curve is a geodesic.

Proof. Let l_0 be the infimum of length of all piecewise smooth curves in the given path homotopy class, which is positive (at least $d(p,q)$) since $p \neq q$. Again take a sequence of piecewise smooth curves γ_i in the given path homotopy class so that $\text{Length}(\gamma_i) < l_0 + \frac{1}{i}$. By Arzela-Ascoli theorem as above, γ_i has a convergent subsequence whose limit is a continuous curve γ. We take ε_0 small so that each geodesic ball $B(p, 2\varepsilon_0)$ is strongly convex (as in Whitehead theorem). Again we may divide each γ_i into N pieces, and let $p_j = \gamma\left(\frac{j}{N}\right)$. Then we still have $d(p_j, p_{j+1}) < \varepsilon_0$. As a result, $\gamma|_{[\frac{j}{N},\frac{j+1}{N}]}$ is path-homotopic to the shortest geodesic connecting p_j to p_{j+1}. So if we let $\widetilde{\gamma}$ be the piecewise geodesic obtained by connecting each p_j to p_{j+1} by shortest geodesic, then $\widetilde{\gamma}$ is path homotopic to γ and has shortest length in the given homotopy class. Finally, by Proposition 4.23, each sufficiently small part in $\widetilde{\gamma}$ must be geodesic. So $\widetilde{\gamma}$ is a geodesic. □

Remark. It was proved by Serre in 1951 that in any compact Riemannian manifold, there are infinitely many geodesics joining any pair of points. [Note that for the sphere, the geodesics could contain a whole great circle which repeat many times.]

Length Minimizing Closed Curves in Given Free Homotopy Class

One may also apply the same argument to the case $p = q$, i.e. consider closed curves with base point p. There are two issues:

(1) If the homotopy class is trivial, then "the shortest curve" is a single point and thus is not interesting.

(2) In each non-trivial homotopy class of curves with base point p, by the same argument one gets:

Proposition 4.27. There is a shortest curve $\gamma : [0,1] \to M$ with $\gamma(0) = \gamma(1) = p$ in the given homotopy class with base point p, and it is a geodesic.

However, in general γ is not smooth at the point p (i.e. $\dot{\gamma}(0) \neq \dot{\gamma}(1)$). Such a curve is called a **geodesic loop** with base point p.

Note that although a geodesic loop γ is a closed curve on M, it is not closed if we take an "upstairs" point of view: the integral curve in S^*M that corresponds to γ is not a closed curve in S^*M. In applications those geodesics that are "not only closed on M, but also closed on S^*M" are more important:

Definition 4.28. We say a geodesic $\gamma : [0,1] \to M$ is a **closed geodesic** if $\gamma(0) = \gamma(1)$ and $\dot{\gamma}(0) = \dot{\gamma}(1)$.

So closed geodesics are projections of closed integral curves in S^*M to M, and they can also be regarded as smooth maps $\gamma : S^1 \to M$ that satisfies the geodesic equation for all $t \in S^1$. Here are some simple examples:

- Any geodesic on S^m is a closed geodesic.
- On the standard cylinder $S^1 \times \mathbb{R}$, a geodesic is either a closed geodesic ("horizontal circles") or a non-self-intersecting geodesic that "goes to infinity" in both direction. [Similar for the standard torus $S^1 \times S^1$].
- For the one-sheet hyperboloid $x^2 + y^2 - z^2 = 1$, there is a unique closed geodesic (the circle with $z = 0$), and many geodesic loops based at points not on the closed geodesic, as well as many "unbounded geodesics".

As motivated by the last example, to get a closed geodesic one cannot use curves in the "homotopy class with base point p" any more. Instead, one should look at the free homotopy class of closed curves. By adjusting the proofs above, one has

Theorem 4.29. Let (M, g) be a compact connected Riemannian manifold which is not simply connected. Then in each free homotopy class, there is a length-minimizing curve and the curve is a geodesic.

Remark. The theorem fails for non-compact connected Riemannian manifold (even if we add completeness assumption).

Remark. For compact simply-connected Riemannian manifold, one can also prove the existence of a closed geodesic: It was proved by Birkhoff for Riemannian 2-spheres (with any Riemannian metric), and later by Lusternik-Fet for any compact simply connected Riemannian manifold. It was further proved by Gromoll-Meyer in 1971 that for simply connected closed manifolds whose cohomology ring $H^*(M;\mathbb{Q})$ is not generated by a single element, there are always infinitely many closed geodesics.

4.4 Completeness

4.4.1 The Hopf-Rinow Theorem

The Hopf-Rinow Theorem and Consequences

Last time we proved that on any compact Riemannian manifold, any nontrivial path-homotopy class contains a shortest curve and that curve must be a geodesics. Today we will study geodesics in a wider class of Riemannian manifolds, namely, complete Riemannian manifolds, and prove the existence of shortest geodesics in each non-trivial path-homotopy class in such manifolds.

We will first prove the existence of shortest geodesics [i.e. a geodesic of length $d(p,q)$] between any two given points p, q on any complete Riemannian manifold. This is the second part of a well-known theorem proved by Hopf and Rinow in 1931. Recall that a Riemannian manifold (M, g) is called geodesically complete if the maximal defining interval of any geodesic on M is $\mathbb{R}$. On the other hand, any Riemannian (M, g) admits a Riemannian metric structure given by

$$d(p, q) = \inf\{L(\gamma) \mid \gamma \text{ is a piecewise smooth curve connection } p \text{ to } q\},$$

and thus we can talk about the completeness of d: a metric space is complete if any Cauchy sequence in it converges.

Now we state Hopf-Rinow theorem, which contains two parts: the first part claims that for Riemannian manifolds, the two notions of completeness coincide; while the second part claims the existences of shortest geodesic on such manifolds.

Theorem 4.30 (Hopf-Rinow)**.** Let (M, g) be a connected Riemannian manifold.

(Part I) The following statements are equivalent:

(1) (M, d) is a complete metric space.

(2) (M, g) is geodesically complete.

(3) There exists $p \in M$, so that $\exp_p$ is defined for all $X_p \in T_pM$.

(4) [Heine-Borel Property] Any bounded closed subset in M is compact.

(Part II) Moreover, each of the previous statements implies

(5) for any $p, q \in M$, there exists a geodesic of length $d(p, q)$ connecting p and q.

Definition 4.31. A connected Riemannian manifold (M, g) satisfying any of (1)-(4) is called a **complete Riemannian manifold**.

Remark. Property (5) is NOT enough to guarantee that (M, g) is complete. For example, the open unit ball $B_1(0)$ in $(\mathbb{R}^m, g_0)$ satisfies (5), but is not complete.

Remark. For a general metric space, condition (1) does NOT imply condition (4): any infinite dimensional Banach or Hilbert space like L^2 is a counterexample. So as metric spaces, Riemannian manifolds are special (and nice) metric spaces.

We list a couple immediate consequences of Hopf-Rinow theorem. Since any compact metric space is complete, we get another proof of

Corollary 4.32. Any compact Riemannian manifold is geodesically complete.

Since any two points can be connected by a geodesic,

Corollary 4.33. If (M, g) is complete and connected, then for any $p \in M$, the exponential map $\exp_p : T_pM \to M$ is surjective.

Since the Heine-Borel property is inherited by closed subsets, we have [warning: although any closed subspace of a complete metric space is complete, one cannot apply (1) here since "the induced metric on a submanifold S in the metric space (M, d)" is not the same as "the Riemannian distance generated by the induced Riemannian metric on $S \subset (M, g)$"]

Corollary 4.34. Any closed submanifold of a complete Riemannian manifold, when endowed with the induced Riemannian metric, is complete.

Proof of "Hopf-Rinow Theorem, Part II"

We first prove Part II of Hopf-Rinow theorem. More precisely, we prove (2)⇒(5), or equivalently, its local version, namely (3)⇒(5'), where

(5') for any $q \in M$, there exists a geodesic of length $d(p, q)$ connecting p and q.

Denote $r = d(p,q)$. We have already seen that there exists $0 < \delta < r$ so that the exponential map $\exp_p$ is a diffeomorphism from $B_\delta(0) \in T_pM$ to $B(p,\delta) \in M$. Note that the geodesic sphere $S(p,\delta) = \exp_p(S_\delta(0))$ is compact. Since the distance function is continuous [c.f. Section 1.3], there exists $p_0 \in S(p,\delta)$ so that

$$d(p_0,q) = \inf_{p' S(p,\delta)} d(p',q).$$

Let γ be the normal geodesic from p to p_0. By (3), γ is defined over $\mathbb{R}$. Let

$$A = \{s \in [\delta,\gamma] \mid d(\gamma(s),q) = r - s\}.$$

We will show $\sup A = r$, which implies $\gamma(r) = q$.

To prove this, we first notice that $\delta \in A$, since

$$r = d(p,q) = \inf_{p' \in S(p,\delta)} (d(p,p') + d(p',q)) = \delta + \inf_{p' \in S(p,\delta)} d(p',q) = \delta + d(\gamma(\delta),q).$$

So A is nonempty.

Secondly, it's easy to see that A is closed, since the function

$$f(s) = d(\gamma(s),q) - r + s$$

is continuous and $A = f^{-1}(0) \cap [\delta, r]$.

Now let $s_0 = \sup A$. Since A is nonempty and closed, $s_0 \in A$. Suppose $s_0 < r$. Then by repeating the previous argument, we know that there exists $0 < \delta' < r - s_0$ and $p_0' \in S(\gamma(s_0),\delta')$ so that

$$d(p_0',q) = \min_{p' \in S(\gamma(s_0),\delta')} d(p',q) = d(\gamma(s_0),q) - \delta'.$$

Since $s_0 \in A$, we get

$$d(p_0',q) = r - s_0 - \delta'.$$

So by the triangle inequality,

$$d(p_0',p) \geqslant d(p,q) - d(p_0',q) = r - (r - s_0 - \delta') = s_0 + \delta'.$$

On the other hand, the curve $\widetilde{\gamma}$ by connecting p to $\gamma(s_0)$ along γ and then connecting $\gamma(s_0)$ to p_0' by the "radial" minimal geodesic has length exactly $s_0+\delta'$. So $\widetilde{\gamma}$, with the arc-length parametrization, must be a geodesic. Obviously $\widetilde{\gamma}$ has to coincide with γ. In other words, $p_0' = \gamma(s_0+\delta')$. As a consequence,

$$d(\gamma(s_0+\delta'),q) = r - (s_0+\delta'),$$

i.e. $s_0 + \delta' \in A$. This conflicts with the fact that $s_0 = \sup A$.

Proof of "Hopf-Rinow Theorem, Part I"

Having proved (3)⇒(5'), now we prove Part I of Hopf-Rinow's theorem by

$$(4) \Rightarrow (1) \Rightarrow (2) \Rightarrow (3) \quad \text{and} \quad (3) + (5') \Rightarrow (4).$$

(4)⇒(1): This is a standard result in general topology: Let p_i be any Cauchy sequence, then the set $\{p_i\}$ is contained in bounded ball B whose closure is compact by (4). It follows that p_i has a subsequence that converges to some p_0. But p_i is a Cauchy sequence, so the entire sequence $p_i \to p_i$.

(1)⇒(2): Let γ be any normal geodesic on M. By the existence and uniqueness theorem, the maximal defining interval of γ must be an open interval (a,b). If $b < \infty$, then we can take a sequence $s_i \to b-$. In particular, s_i is a Cauchy sequence in $\mathbb{R}$. But γ is a normal geodesic, so

$$d(\gamma(s_i),\gamma(s_j)) \leqslant |s_i - s_j|.$$

As a consequence, $\gamma(s_i)$ is a Cauchy sequence in (M,d). If follows that there exists a point $p \in M$ so that $\gamma(s_i) \to p$.

Since $\mathcal{E}$ is open and $(p, 0) \in \mathcal{E}$, there exists $\varepsilon > 0$ so that $(q, Y_q) \in \mathcal{E}$ for any q with $d(q, p) < \varepsilon$ and any $Y_q \in T_qM$ with $|Y_q| < 2\varepsilon$. So if we take i large enough so that $b - s_i < \frac{\varepsilon}{2}$ and thus $d(\gamma(s_i), p) < \frac{\varepsilon}{2}$, then $\gamma(t; \gamma_{s_i}, \varepsilon\dot\gamma(s_i))$ is defined for $t \in [0, 1]$. In other words, the geodesic $\gamma_1(t) = \gamma(t; \gamma(s_i), \dot\gamma(s_i))$ is well defined for $0 < t < \varepsilon$. Since γ_1 coincides with γ at s_i, they must be the same. In particular, γ can be defined for all $t < s_i + \frac{\varepsilon}{2}$, which exceeds the upper bound b, a contradiction.

Similarly by considering the "reverse geodesic" one also has $a = -\infty$. So any normal geodesic on M, and thus any geodesic on M, has defining interval $\mathbb{R}$.
(2)⇒(3): This is obvious. ((M, g) is geodesically complete ⇔ $\mathcal{E} = TM$.)
(3)+(5')⇒(4): Let $K \subset M$ be a bounded closed set. Then there exists a constant $C > 0$ so that $d(p, k) < C$ for all $k \in K$. According to (3) and (5'), $K \subset \exp_p(\overline{B_C(0)})$, where $\overline{B_C(0)}$ is the **closed** ball of radius C in T_pM, which is compact in T_pM. Since $\exp_p$ is smooth, $\exp_p(\overline{B_C(0)})$ is also compact. Thus K, as a closed subset of a compact set, is compact.

4.4.2 Geodesics and Riemann Covering Map

Lifting to the Riemannian Covering

Next let's turn to prove the existence of length minimizing geodesics in any path-homotopy class of curves connecting p to q. The idea is to straightforward: instead of working on piecewise smooth curves in M connecting given points p and q that lies in a given path-homotopy classes, we will move to the universal covering $\pi : \widetilde{M} \to M$ of M and work on piecewise smooth curves starting with a fixed $\widetilde{p} \in \pi^{-1}(p)$ and ends at the point $\widetilde{q} \in \pi^{-1}(q)$ so that any curve connecting $\widetilde{p}$ and $\widetilde{q}$ projects to a curve in the given path-homotopy classes, and then we can apply the second part of Hopf-Rinow theorem.

For the argument mentioned above to work, we need a couple ingredients. First, we need to lift the complete metric g on M to a complete metric on its universal covering $\widetilde{M}$. Recall

- Let M, N be connected smooth manifolds. A smooth map $f : M \to N$ is said to be a smooth **covering map** if
  (1) for any $q \in N$, there is a neighborhood V of q in N and open subsets U_α of M so that $f^{-1}(V) = \bigcup_\alpha U_\alpha$,
  (2) for each α, $f : U_\alpha \to V$ is a diffeomorphism,
  (3) these U_α's are disjoint.

  As is well known, if $f : M \to N$ is a covering map, then
  - $\dim M = \dim N$ and f is surjective,
  - fix any $p_\alpha \in f^{-1}(q)$, any path (and path homotopy) starts at q in N admits a unique lifting to M that starts at p_α,
  - moreover, if N is simply connected, then f is a global diffeomorphism.
- If (M, g_M) and (N, g_N) are Riemannian manifolds, then a smooth covering map $\pi : M \to N$ is called a **Riemannian covering map** if $\pi^* g_N = g_M$. Note:
  - given any smooth covering map $\pi : M \to N$ and any Riemannian metric on N, one may pullback that metric to M to make the covering map a Riemannian covering [c.f. Exercise 1.3].
  - any Riemannian covering map is a local isometry.
- We also need some standard properties of local isometries. Let $f : (M, g) \to (N, h)$ be a local isometry, then

– M and N have "the same" Riemannian metrics at corresponding points, and thus the same Levi-Civita connections and the same Riemannian curvature at corresponding points [c.f. Exercise 2.1].
– in particular, f maps geodesics into geodesics, and if f is a Riemannian covering, then the lifting of a geodesic is a geodesic.
– for any piecewise smooth curve γ in M, one has $|\dot{\gamma}|_{\gamma(t)} = |\mathrm{d}f_{\gamma(t)}(\dot{\gamma})|_{f(\gamma(t))}$ and thus $L(\gamma) = L(f(\gamma))$.

Now we prove

Proposition 4.35. Let (M, g) be a complete Riemannian manifold, and $\pi : \widetilde{M} \to M$ be a smooth covering map. Then $(\widetilde{M}, \pi^* g)$ is complete.

Proof. For any $\widetilde{p} \in \widetilde{M}$ and any $\widetilde{v} \in T_{\widetilde{p}}\widetilde{M}$, we denote $p = \pi(\widetilde{p})$ and $v = \mathrm{d}\pi_{\widetilde{p}}(\widetilde{v})$. Then by definition of completeness, there is a geodesic $\gamma : \mathbb{R} \to M$ with $\gamma(0) = p$ and $\dot{\gamma}(0) = v$. By the path-lifting property for covering space, there is a unique lifting $\widetilde{\gamma} : \mathbb{R} \to \widetilde{M}$ with $\widetilde{\gamma}(0) = \widetilde{p}$, which is a geodesic since it is the lifting of a geodesic. Moreover, since $\pi : (\widetilde{M}, \pi^* g) \to (M, g)$ is a local isometry and since $\pi \circ \widetilde{\gamma} = \gamma$, we get

$$\dot{\widetilde{\gamma}}(0) = (\mathrm{d}\pi_{\widetilde{p}})^{-1}(\dot{\gamma}(0)) = (\mathrm{d}\pi_{\widetilde{p}})^{-1}(v) = \widetilde{v}.$$

So the geodesic starts at $\widetilde{p}$ in the direction $\widetilde{v}$ is defined over $\mathbb{R}$. □

Length Minimizing Curves in Given Path-homotopy Class

As a consequence, we can extend Theorem 4.26 to complete Riemannian manifolds.

Theorem 4.36. Let (M, g) be a complete connected Riemannian manifold, and p, q are two points in M. Then in each path-homotopy class of curves γ with $\gamma(0) = p$, $\gamma(1) = q$, there is a length-minimizing piecewise smooth curve and it is a geodesic.

Proof. Consider the universal covering $\pi : \widetilde{M} \to M$. Equip $\widetilde{M}$ with the covering Riemannian metric $\pi^* g$. Given any path $\sigma : [0, 1] \to M$ connecting p and q in the given homotopy class, and given any $\widetilde{p} \in \pi^{-1}(p)$, there is a unique lifting $\widetilde{\sigma} : [0, 1] \to \widetilde{M}$ of σ with $\widetilde{\sigma}(0) = \widetilde{p}$. Since $(\widetilde{M}, \pi^* g)$ is complete, by Hopf-Rinow theorem, there is a minimizing geodesic $\widetilde{\gamma}$ from $\widetilde{p}$ to $\widetilde{q} := \widetilde{\sigma}(1)$. Since π is a local isometry, the projection $\gamma = \pi \circ \widetilde{\gamma}$ is a geodesic in M with $\gamma(0) = p$, $\gamma(1) = q$. Since $\widetilde{M}$ is simply connected, $\widetilde{\gamma}$ is path-homotopic to $\widetilde{\sigma}$ and thus γ is path-homotopic to σ.

Finally suppose σ_1 be any piecewise smooth curve in M from p to q in the given path homotopy class, then its lifting $\widetilde{\sigma}_1$ in $\widetilde{M}$ with starting point $\widetilde{\sigma}_1(0) = \widetilde{p}$ must ends at $\widetilde{q}$, and thus by our choice of $\widetilde{\gamma}$,

$$L(\gamma) = \mathrm{Length}(\widetilde{\gamma}) \leqslant \mathrm{Length}(\widetilde{\sigma}_1) = L(\sigma_1).$$

So γ is the shortest curve in the given path homotopy class. □

The Theorem of Ambrose

In Proposition 4.35 we start with a complete Riemannian manifold downstairs and a smooth covering map [topological information], and end with a complete Riemannian structure upstairs so that the map is a local isometry [geometric information]. It turns out that theorem has an "inverse", i.e. given an upstairs complete Riemannian manifolds and a local isometry f [geometric information], the downstairs metric must be complete and the map is a covering map [topological information]:

Theorem 4.37 (Ambrose)**.** Let (M, g) and (N, h) be connected Riemannian manifold, and $f : (M, g) \to (N, h)$ a local isometry. Suppose (M, g) is complete, then f is a smooth covering map, and (N, h) is complete.

Note that "$f : (M, g) \to (N, h)$ a local isometry and (N, h) is complete" is not enough to guarantee f to be a covering map. We give an immediate consequence of Ambrose's theorem, which will be used later in studying structures of Riemannian manifolds of non-positive sectional curvature:

Corollary 4.38. Let (M, g) be a connected Riemannian manifold, and $p \in M$. If $\exp_p : T_pM \to M$ is a local diffeomorphism everywhere, then $\exp_p$ is a covering map.

Proof. Note that the condition implies $\exp_p$ is defined on the whole T_pM, and thus by Hopf-Rinow, (M, g) is complete. Endow T_pM with the metric $\overline{g} = (\exp_p)^* g$, then

- $\exp_p : (T_pM, \overline{g}) \to (M, g)$ is a local isometry.
- For any $v \in T_pM$, the curve $\gamma(t) = tv$ is a geodesic in $(T_pM, \overline{g})$ since its image $\exp_p(tv)$ is a geodesic in (M, g). In other words, $\exp_0 : T_0(T_pM) \to T_pM$ is defined on the whole $T_0(T_pM)$, and thus by Hopf-Rinow, $(T_pM, \overline{g})$ is complete.

So by Ambrose theorem, $\exp_p$ is a covering map. □

Proof of Ambrose's Theorem

We will prove this in four steps.

Step 1 "Lift" geodesics in N to geodesics in M:

Lemma 4.39. Under the conditions of the theorem, given any geodesic $\gamma : [a, b] \to N$ and any $p \in f^{-1}(\gamma(a))$, we can "lift" γ to a geodesic $\overline{\gamma} : [a, b] \to M$ so that $\gamma(t) = f(\overline{\gamma}(t))$ and $\overline{\gamma}(a) = p$. Moreover, the lift is unique.

Proof. Since $\mathrm{d}f_p : T_pM \to T_{\gamma(a)}N$ is a linear isometry, one can find a unique $X_p \in T_pM$ so that

$$\mathrm{d}f_p(X_p) = \dot{\gamma}(a).$$

Let $\overline{\gamma} : [a, b] \to M$ be the geodesic in M with (used completeness here)

$$\overline{\gamma}(a) = p, \ \dot{\overline{\gamma}}(a) = X_p.$$

Then $f \circ \overline{\gamma}$ is a geodesic in N with same initial conditions as γ, and thus $\gamma = f \circ \overline{\gamma}$. The uniqueness of lift also follows directly from the fact that $\mathrm{d}f_p : T_pM \to T_{\gamma(a)}N$ is a linear isometry. □

Step 2 (N, h) is complete.

Fix any point $p \in N$ which lies in the image of f. For any geodesic γ starting at p, by Lemma 4.39, we can lift γ to a geodesic $\overline{\gamma}$ starting at any $\overline{p} \in f^{-1}(p)$. Since (M, g) is complete, $\overline{\gamma}$ is a geodesic defined for all t. It follows that $\gamma = f \circ \overline{\gamma}$ is a geodesic defined for all t. So by Hopf-Rinow theorem, (N, h) is complete.

Step 3 f is surjective.

Fix any point $p \in N$ which lies in the image of f. For any $q \in N$, since (N, h) is complete, there is minimizing geodesic γ from p to q. By Lemma 4.39, we can lift γ to a geodesic $\overline{\gamma}$ starting at any $\overline{\gamma} \in f^{-1}(p)$. It follows that $q \in f^(\overline{\gamma}) \subset \mathrm{Im}(f)$.

Step 4 Verify covering properties.

Step 4.1 Construct V and U_α: Fix any $q \in N$, we may assume $f^{-1}(q) = \{p_\alpha\}_{\alpha \in I}$. Take δ small enough so that $V = B(q, \delta)$ is a normal geodesic ball. For each α, let

$$U_\alpha = B(p_\alpha, \delta) \subset M.$$

We note that each point in U_α can be connected to p_α through a unique minimizing geodesic of length less than δ: if there exists $p' \in U_\alpha$ that can be connected to p_α by two geodesics γ, γ' of lengths less that δ starting at p_α, then $f(\gamma)$ and $f(\gamma')$ are geodesic in N from q to $f(p')$ of lengths less than δ, and thus we must have $f(\gamma) = f(\gamma') =: \widetilde{\gamma}$ and thus $\dot{\gamma}(0) = (\mathrm{d}f_{p_\alpha})^{-1}(\dot{\widetilde{\gamma}}(0)) = \dot{\gamma}'(0)$.

Step 4.2 $f^{-1}(V) = \bigcup_\alpha U_\alpha$: For any $p \in f^{-1}(V)$, let $\gamma : [0, 1] \to N$ be the minimal geodesic in V connecting $f(p)$ to q, and $\overline{\gamma}$ its lift starting at p. Then $f(\overline{\gamma}(1)) = \gamma(1) = q$, so there exists α so that $\overline{\gamma}(1) = p_\alpha$. Note $L(\overline{\gamma}) = L(\gamma) < \delta$, we conclude that $p \in U_\alpha$. So $f^{-1}(V) \subset \bigcup_\alpha U_\alpha$.

Conversely, for any point $p \in U_\alpha$, there is a minimal geodesic $\overline{\gamma} : [0, 1] \to M$ connecting p_α to p with length $< \delta$. It follows that $\gamma = f \circ \overline{\gamma}$ is a geodesic starting from q with length $< \delta$. So $f(p) = f(\overline{\gamma}(1)) = \gamma(1) \in V$, and thus $U_\alpha \subset f^{-1}(V)$.

Step 4.3 $f : U_\alpha \to V$ is diffeomorphism: Since local isometry maps geodesics into geodesics, and geodesics are determined by initial values, we have

$$\exp_q[\mathrm{d}f_{p_\alpha}(X_\alpha)] = f(\exp_{p_\alpha}(X_\alpha))$$

for any $X_\alpha \in T_{p_\alpha}M$. Moreover, when restricted to balls of radius δ, both $\exp_q$ and $\exp_{p_\alpha}$ are diffeomorphisms. Since $\mathrm{d}f$ is an linear isomorphism which is also a diffeomorphism on the whole $T_{p_\alpha}M$, we conclude that

$$f = \exp_q \circ \mathrm{d}f_{p_\alpha} \circ \exp_{p_\alpha}^{-1}$$

when restricted to balls of radius δ, so in particular $f : U_\alpha \to V$ is a diffeomorphism.
Step 4.4 For $\alpha \neq \beta$, $U_\alpha \cap U_\beta = \varnothing$: Suppose there exists $p \in U_\alpha \cap U_\beta$ and $\alpha \neq \beta$. Let $\overline{\gamma}_\alpha$ and $\overline{\gamma}_\beta$ be the minimal geodesic from p to p_α and p_β respectively. Then $f(\overline{\gamma}_\alpha)$ and $f(\overline{\gamma}_\beta)$ are minimal geodesics in N, both from $f(p)$ to q. It follows that $f(\overline{\gamma}_\alpha) = f(\overline{\gamma}_\beta)$, and both $\overline{\gamma}_\alpha$ and $\overline{\gamma}_\beta$ are lifts of $f(\overline{\gamma}_\alpha)$ from p. By uniqueness of lift, $\overline{\gamma}_\alpha = \overline{\gamma}_\beta$ and thus $p_\alpha = p_\beta$.

4.5 Variations of Length and Energy

Although we defined geodesics as "self-parallel" curves, in the last sections we have seen that on Riemannian manifolds, geodesics are closely related to "length minimizing" curves:

- (Section 4.2) on any Riemannian manifold, in a small neighborhood of any point, geodesics are precisely the shortest curves connecting endpoints.
- (Section 4.3 and 4.4) on any complete Riemannian manifold, in each path-homotopy class, there exists a length minimizing curve and it is a geodesic.

On the other hand, we also know the existence of geodesics which are not length minimizing in the given path homotopy class [e.g. closed geodesics on S^m]. In what follows we take a closer look at the relation between geodesics and the length functional.

4.5.1 Geodesics as Critical Points of Energy Functional

The Euler-Lagrange Equation

For any p, $q \in M$, consider

$$\mathcal{C}_{pq} = \{\gamma : [a,b] \to M \mid \gamma \text{ is piecewise smooth and } \gamma(a) = p,\ \gamma(b) = q\}.$$

One may ask: what property distinguish geodesics in $\mathcal{C}_{pq}$ from other curves? One of the answers should be "length-minimizing", at least locally. Now let's attack this problem by studying the length functional directly.

Recall that the length of a piecewise smooth curve $\gamma : [a,b] \to (M,g)$ is

$$L(\gamma) = \text{Length}(\gamma) = \int_a^b |\dot{\gamma}(t)|\mathrm{d}t.$$

To find the minimum of such a functional, for simplicity let's first assume that γ is inside a coordinate patch, and thus is given by a vector-valued function $x(t) = (x^1(t), \ldots, x^m(t))$. Consider a very general question in variational analysis:

Given a smooth function $f = f(t, x, \dot{x})$, find all the minimizer of the functional

$$I(x) = \int_a^b f(t, x(t), \dot{x}(t))\mathrm{d}t$$

in the set of all smooth curves $x(t) = (x^1(t), \ldots, x^m(t))$ with fixed endpoints $x(a) = p$, $x(b) = q$.

Since this "space of smooth curves" is huge (namely, of "infinitely dimensional"), one cannot apply usual methods in calculus to find the minimizer. Fortunately, there is a new branch of mathematics called variational calculus that is invented to handle such problems. The idea is:

convert the one "infinitely dimensional" problem [in which we have infinitely many directions to move] to infinitely many "one-dimensional problems" [in which we fix one direction to move]. Here is how it works in this example: Since we are studying the functional on curves with fixed endpoints, we may fix any smooth map $y(t) = (y^1(t), \dots, y^m(t))$ with $y(a) = y(b) = 0$ and consider the corresponding one-parameter family of curves of the form $x(t) + \varepsilon y(t)$. So if $x = x(t)$ is a minimizer of I, then we must have

$$\begin{aligned} 0 = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Big|_{\varepsilon=0} I(x+\varepsilon y) &= \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Big|_{\varepsilon=0} \int_a^b f(t, x+\varepsilon y, \dot{x}+\varepsilon \dot{y})\mathrm{d}t \\ &= \int_a^b \left(\frac{\partial f}{\partial x^k}(t,x,\dot{x})y^k + \frac{\partial f}{\partial \dot{x}^k}(t,x,\dot{x})\dot{y}^k \right) \mathrm{d}t \\ &= \int_a^b \left(\frac{\partial f}{\partial x^k}(t,x,\dot{x}) - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial f}{\partial \dot{x}^k}(t,x,\dot{x}) \right) y^k \mathrm{d}t. \end{aligned}$$

As a result, we see that if x is a minimizer (or a critical point) of I, then

$$\frac{\partial f}{\partial x^k}(t,x,\dot{x}) = \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial f}{\partial \dot{x}^k}(t,x,\dot{x}), \quad 1 \leqslant k \leqslant m,$$

which is known as the **Euler-Lagrange equation** for the functional I.

Arc Length v.s. Energy

We may apply Euler-Lagrange equation above to the function

$$f(t, x(t), \dot{x}(t)) = (\langle \dot{x}(t), \dot{x}(t)\rangle_{x(t)})^{\frac{1}{2}} = (g_{ij}(x(t))\dot{x}^i(t)\dot{x}^j(t))^{\frac{1}{2}}.$$

However, since there is a square root, the computation could be a bit messy. It turn out that there is a small trick that can simplify the computation a lot: instead of the length functional, we can work on the **energy functional**:

$$E(\gamma) = \frac{1}{2}\int_a^b |\dot{\gamma}(t)|^2 \mathrm{d}t.$$

By the Cauchy-Schwartz inequality, for each piecewise smooth curve γ,

$$L(\gamma)^2 = \left(\int_a^b |\dot{\gamma}(t)|\mathrm{d}t\right)^2 \leqslant \left(\int_a^b 1^2 \mathrm{d}t\right)\left(\int_a^b |\dot{\gamma}(t)|^2 \mathrm{d}t\right) = 2(b-a)E(\gamma),$$

with equality holds if and only if $|\dot{\gamma}(t)| \equiv$ constant. In particular, we see that although the length functional $L(\gamma)$ is independent of the choice of parametrizations, the energy functional does depend on the parametrizations (and on the length of the interval $[a,b]$): among different parametrizations of γ on fixed $[a,b]$, $E(\gamma)$ is minimized on the "constant speed parametrization". As a consequence we can prove

Proposition 4.40. A curve $\gamma : [a,b] \to M$ in $\mathcal{C}_{pq}$ minimize the energy functional $E(\gamma)$ if and only if it has constant speed and minimize the length functional $L(\gamma)$.

Proof. Suppose $\gamma : [a,b] \to M$ minimize $E(\gamma)$ but there exists $\gamma' \in \mathcal{C}_{pq}$ such that $L(\gamma') < L(\gamma)$, then for the "constant speed re-parametrization" $\widetilde{\gamma} : [a,b] \to M$ of γ',

$$E(\widetilde{\gamma}) = \frac{1}{2(b-a)}L(\widetilde{\gamma})^2 = \frac{1}{2(b-a)}L(\gamma')^2 < \frac{1}{2(b-a)}L(\gamma)^2 \leqslant E(\gamma'),$$

which is a contradiction. So any minimizer of $E(\gamma)$ must also minimize $L(\gamma)$.

Conversely, if $\gamma : [a,b] \to M$ has constant speed and minimize $L(\gamma)$, but there is another $\gamma' : [a,b] \to M$ in $\mathcal{C}_{pq}$ with $E(\gamma') < E(\gamma)$, then

$$L(\gamma') \leqslant \sqrt{2(b-a)E(\gamma')} < \sqrt{2(b-a)E(\gamma)} = L(\gamma),$$

a contradiction. □

Since any piecewise smooth curve can be reparameterized to have constant speed, to minimize $L(\gamma)$, it is enough to minimize $E(\gamma)$ whose integrand is much simpler. Applying Euler-Lagrange equation to

$$f(t, x(t), \dot{x}(t)) = g_{ij}(x(t))\dot{x}^i(t)\dot{x}^j(t)$$

we get, for $1 \leqslant k \leqslant m$,

$$\frac{\partial g_{ij}}{\partial x^k}\dot{x}^i\dot{x}^j = \frac{\mathrm{d}}{\mathrm{d}t}(g_{kj}\dot{x}^j) + \frac{\mathrm{d}}{\mathrm{d}t}(g_{ik}\dot{x}^i) = 2\frac{\partial g_{kj}}{\partial x^i}\dot{x}^i\dot{x}^j + 2g_{kj}\ddot{x}^j,$$

which, as we have seen in Section 4.2, is equivalent to the geodesic equation

$$\ddot{x}^k + \Gamma_{ij}^k\dot{x}^i\dot{x}^j = 0, \qquad 1 \leqslant k \leqslant m.$$

Amazingly enough, by this way we get not only all the geodesics that are length minimizing curves in $\mathcal{C}_{pq}$ and all the geodesics that are the length minimizing curves in each path homotopy class of $\mathcal{C}_{pq}$[1], but in fact we get ALL the geodesics connecting p and q:

Theorem 4.41. For a Riemannian manifold (M, g), a curve $\gamma : [a, b] \to M$ in $\mathcal{C}_{pq}$ is a geodesic if and only if it satisfies the Euler-Lagrange equation of the energy functional $E(\gamma)$.

This gives another proof of the fact that any length minimizing curve is a geodesic, and also explains why there exist geodesics that are not length minimizing even in the given path-homotopy class: those curves are only "critical points" of E which need not be minimizing among all near-by curves. If one need to find the minimizing geodesics, then as usual one can further calculate the second order derivative $\left.\frac{\mathrm{d}^2}{\mathrm{d}\varepsilon^2}\right|_{\varepsilon=0} E(x + \varepsilon y)$ in a coordinate system, using which one can show that geodesics are always length-minimizing locally in a neighborhood.

4.5.2 Formulas for the First and Second Variations

The calculations above are thought-provoking but has the shortcoming that they are performed in a chart. In what follows we take a global way to calculate the first and second derivatives, and also study variations which could be more general (without fixing endpoints) or more restrictive (with geodesic variation).

Variations

For simplicity we start with smooth variations of a smooth curve:

Definition 4.42. Let $\gamma : [a, b] \to M$ be a smooth curve, and $\varepsilon > 0$.

(1) A smooth **variation** of γ is a smooth map $f : [a, b] \times (-\varepsilon, \varepsilon) \to M$ so that

$$f(t, 0) = \gamma(t)$$

for all $t \in [a, b]$. In what follows, we will also denote $\gamma_s(t) = f(t, s)$.

(2) A variation f is called **proper** if for every $s \in (-\varepsilon, \varepsilon)$,

$$\gamma_s(a) = \gamma(a) \quad \text{and} \quad \gamma_s(b) = \gamma(b).$$

(3) A variation is called a **geodesic variation** if each γ_s is a geodesic.

For simplicity, we denote $R = [a, b] \times (-\varepsilon, \varepsilon)$. Let $f : R \to M$ be a smooth variation of γ. Then $E = f^*TM$ is a vector bundle over R, on which we have an induced linear connection $\tilde{\nabla} = f^*\nabla$ (where ∇ is the Levi-Civita connection on (M, g)). To be rigorous, in what follows we

[1] Although these curves are not length minimizing in $\mathcal{C}_{pq}$, they are in fact length minimizing among "nearby curves", namely among curves of the form $x + \varepsilon y$ in the computation above for ε small enough, since these curves are in the same path-homotopy class.

will calculate via $\widetilde{\nabla}$, and refer to the appendix of this section for the definition and properties of $\widetilde{\nabla}$.

The variation f gives rise to two natural sections of E, namely,

$$f_s(t,s) := (\mathrm{d}f)_{t,s}\left(\frac{\partial}{\partial s}\right) \in T_{f(t,s)}M = E_{t,s}$$

and

$$f_t(t,s) := (\mathrm{d}f)_{t,s}\left(\frac{\partial}{\partial t}\right) \in T_{f(t,s)}M = E_{t,s},$$

where $\frac{\partial}{\partial s}$ and $\frac{\partial}{\partial t}$ are the coordinate vector fields on R. Note that by definition,

$$f_t(t_0,s_0) = \dot{\gamma}_{s_0}(t_0).$$

We are mainly interested in the restriction of the sections f_s and f_t to $s=0$, which are in fact "vector fields along γ". Obviously, we have $\dot{\gamma}(t) = f_t(t,0) = (\mathrm{d}f)_{t,0}\left(\frac{\partial}{\partial t}\right)$.

Definition 4.43. We will call

$$V(t) := f_s(t,0) = (\mathrm{d}f)_{t,0}\left(\frac{\partial}{\partial s}\right)$$

the **variation field** of the variation f.

Note that if γ is an embedded curve and f is an embedding, then both $\dot{\gamma}(t)$ and $V(t)$ can be regarded as vector fields on M along γ in a natural way, and the computations below can be carried out via ∇ instead of $\widetilde{\nabla}$.

The First Variation Formula of Energy for Smooth Variations

Now we compute the variation of E along given variation (without fixing endpoints): Let $f(t,s)$ be a smooth variation of a smooth curve $\gamma : [a,b] \to M$. By Proposition 4.60 and Proposition 4.62, the derivative of $E(\gamma_s)$ is

$$\frac{\mathrm{d}}{\mathrm{d}s}E(\gamma_s) = \frac{1}{2}\int_a^b \langle \dot{\gamma}_s(t), \dot{\gamma}_s(t)\rangle \mathrm{d}t = \int_a^b \langle \widetilde{\nabla}_{\partial/\partial s} f_t, f_t\rangle \mathrm{d}t = \int_a^b \langle \widetilde{\nabla}_{\partial/\partial t} f_s, f_t\rangle \mathrm{d}t.$$

Applying metric compatibility (i.e. Proposition 4.60) again, we get

$$\int_a^b \langle \widetilde{\nabla}_{\partial/\partial t} f_s, f_t\rangle \mathrm{d}t = \int_a^b \frac{\partial}{\partial t}\langle f_s, f_t\rangle \mathrm{d}t - \int_a^b \langle f_s, \widetilde{\nabla}_{\partial/\partial t} f_t\rangle \mathrm{d}t = \langle f_s, f_t\rangle\Big|_{t=a}^{t=b} - \int_a^b \langle f_s, \widetilde{\nabla}_{\partial/\partial t} f_t\rangle \mathrm{d}t.$$

So we get

Theorem 4.44 (The First Variation of Energy)**.** Given any smooth variation $f(t,s)$ of a smooth curve $\gamma : [a,b] \to M$,

$$\frac{\mathrm{d}}{\mathrm{d}s}E(\gamma_s) = \int_a^b \langle \widetilde{\nabla}_{\partial/\partial t} f_s, f_t\rangle \mathrm{d}t = \langle f_s(t,s), f_t(t,s)\rangle\Big|_{t=a}^{t=b} - \int_a^b \langle f_s, \widetilde{\nabla}_{\partial/\partial t} f_t\rangle \mathrm{d}t.$$

In particular,

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0} E(\gamma_s) = -\int_a^b \langle V(t), \nabla_{\dot{\gamma}(t)}\dot{\gamma}\rangle \mathrm{d}t - \langle V(a), \dot{\gamma}(a)\rangle + \langle V(b), \dot{\gamma}(b)\rangle.$$

In particular, if f is a proper smooth variation, then

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0} E(\gamma_s) = -\int_a^b \langle V(t), \nabla_{\dot{\gamma}(t)}\dot{\gamma}\rangle \mathrm{d}t.$$

Again we see that γ is a geodesic (i.e. $\nabla_{\dot{\gamma}}\dot{\gamma} = 0$) if and only if γ is a critical point of the energy functional E among all proper variations.

The First Variation Formula of Length for Smooth Variations

Use the same way, one can calculate the first variation of the length. A trick to simplify the computation is the following observation:

$$\frac{\partial}{\partial s}|\dot{\gamma}_s(t)| = \frac{\partial}{\partial s}\langle f_t, f_t\rangle^{\frac{1}{2}} = \frac{1}{2}\frac{1}{|f_t|}\frac{\partial}{\partial s}\langle f_t, f_t\rangle = \frac{1}{|f_t|}\langle \widetilde{\nabla}_{\partial/\partial t} f_s, f_t\rangle = \left\langle \widetilde{\nabla}_{\partial/\partial t} f_s, \frac{f_t}{|f_t|}\right\rangle.$$

Then following the same computation, one gets

Theorem 4.45 (The First Variation of Length)**.** Let $f(t,s)$ be a smooth variation of a smooth curve γ. Then

$$\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} L(\gamma_s) = -\int_a^b \left\langle V(t), \nabla_{\dot{\gamma}(t)}\frac{\dot{\gamma}}{|\dot{\gamma}|}\right\rangle \mathrm{d}t - \left\langle V(a), \frac{\dot{\gamma}(a)}{|\dot{\gamma}(a)|}\right\rangle + \left\langle V(b), \frac{\dot{\gamma}(b)}{|\dot{\gamma}(b)|}\right\rangle.$$

As an application we prove

Proposition 4.46. Let S be a closed submanifold of (M,g). Suppose γ is a geodesic from $p \notin S$ to $q \in S$ with $L(\gamma) = d(p,S)$. Then γ is perpendicular to S.

Proof. For any $v \in T_qS$, take a curve $\sigma : (-\varepsilon,\varepsilon) \to S$ with $\sigma(0) = q$ and $\dot{\sigma}(0) = v$. Let γ_s be a variation of γ with $\gamma_s(0) = p$ and $\gamma_s(l) = \sigma(s)$, where $l = L(\gamma)$. Then $V(a) = 0$ and $V(b) = v$, and by the first variation formula,

$$0 = \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} L(\gamma_s) = \langle v, \dot{\gamma}(l)\rangle,$$

thus the conclusion follows. □

Piecewise Smooth Curve

More generally, one can consider piecewise smooth curves $\gamma : [a,b] \to M$, i.e. there exists a subdivision

$$a = t_0 < t_1 < t_2 < \cdots < t_k < t_{k+1} = b$$

such that γ is smooth on each interval $[t_i, t_{i+1}]$. We shall consider **piecewise smooth variations** of γ, which is a continuous map $f : [a,b]\times(-\varepsilon,\varepsilon) \to M$ so that f is smooth on each $[t_i,t_{i+1}]\times(-\varepsilon,\varepsilon)$ for each i. Note that this implies

- for each $s \in (-\varepsilon,\varepsilon)$, the curve $t \mapsto \gamma_s(t) = f(t,s)$ is piecewise smooth,
- for each $t \in [a,b]$, the curve $s \mapsto f(t,s)$ is smooth (so f_s is well-defined at t_i's).

Applying the previous theorems to each $[t_i, t_{i+1}] \times (-\varepsilon,\varepsilon)$, we get

Corollary 4.47. Let f be a piecewise smooth variation of curve γ. Then

$$\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} E(\gamma_s) = -\int_a^b \langle V(t), \nabla_{\dot{\gamma}}\dot{\gamma}\rangle \mathrm{d}t - \langle V(a), \dot{\gamma}(a)\rangle + \langle V(b), \dot{\gamma}(b)\rangle - \sum_{i=1}^k \langle V(t_i), \dot{\gamma}(t_i^+) - \dot{\gamma}(t_i^-)\rangle$$

and

$$\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} L(\gamma_s) = -\int_a^b \left\langle V(t), \nabla_{\dot{\gamma}}\frac{\dot{\gamma}(t)}{|\dot{\gamma}|}\right\rangle \mathrm{d}t - \left\langle V(a), \frac{\dot{\gamma}(a)}{|\dot{\gamma}(a)|}\right\rangle + \left\langle V(b), \frac{\dot{\gamma}(b)}{|\dot{\gamma}(b)|}\right\rangle - \sum_{i=1}^k \left\langle V(t_i), \frac{\dot{\gamma}(t_i^+)}{|\dot{\gamma}(t_i^+)|} - \frac{\dot{\gamma}(t_i^-)}{|\dot{\gamma}(t_i^-)|}\right\rangle.$$

The local computations above imply that among smooth curves, geodesics are critical points of the energy functional. A natural question is: If a curve is piecewise smooth, can it be a critical point of the energy functional? Of course for γ be a critical point of the energy functional, it must be a geodesic when restricted to any subinterval where it is smooth, or in other words, it must be "piecewise geodesic".

Corollary 4.48. If a piecewise smooth curve γ is a critical point of the energy functional among proper variations, then it is $\mathcal{C}^1$ and thus a geodesic.

Proof. We can first choose proper variations with variation fields satisfying $V(t_i) = 0$ and deduce that $\nabla_{\dot\gamma}\dot\gamma = 0$ at any smooth point of γ. In particular, the first term in the right hand of the first variation formula vanishes. As a consequence, we have

$$\sum_{i=1}^{k}\langle V(t_i), \dot\gamma(t_i^+) - \dot\gamma(t_i^-)\rangle = 0$$

for any variation field V. Then for each i we can consider all variation fields so that $V(t_j) = 0$ for all $j \neq i$, and conclude that

$$\langle V(t_i), \dot\gamma(t_i^+) - \dot\gamma(t_i^-) = 0$$

for any $V(t_i) \in T_{\gamma(t_i)}$. It follows that $\dot\gamma(t_i^+) = \dot\gamma(t_i^-)$ and thus γ is $\mathcal{C}^1$. □

Piecewise Smooth Curve

Finally we compute the second variation of energy. As in calculus, the second variation is mainly used near critical points, i.e.. near geodesics. So we let $\gamma : [a, b] \to M$ be a geodesic, and $f(t, s)$ be a smooth variation of γ. According to Theorem 4.44, Proposition 4.60 and Proposition 4.62,

$$\begin{aligned}\frac{\mathrm{d}^2}{\mathrm{d}s^2}E(\gamma_s) &= \int_a^b \langle\widetilde\nabla_{\partial/\partial t} f_s, f_t\rangle \mathrm{d}t\\ &= \int_a^b \langle\widetilde\nabla_{\partial/\partial s}\widetilde\nabla_{\partial/\partial t} f_s, f_t\rangle \mathrm{d}t + \int_a^b \langle\widetilde\nabla_{\partial/\partial t} f_s, \widetilde\nabla_{\partial/\partial s} f_t\rangle \mathrm{d}t\\ &= \int_a^b \left\langle \widetilde R\left(\frac{\partial}{\partial s}, \frac{\partial}{\partial t}\right) f_s, f_t\right\rangle \mathrm{d}t + \int_a^b \langle\widetilde\nabla_{\partial/\partial t}\widetilde\nabla_{\partial/\partial s} f_s, f_t\rangle \mathrm{d}t + \int_a^b \langle\widetilde\nabla_{\partial/\partial t} f_s, \widetilde\nabla_{\partial/\partial t} f_s\rangle \mathrm{d}t\end{aligned}$$

There are two $\widetilde\nabla_{\partial/\partial t}$ in the above formula. We may either apply Proposition 4.60 to the first one to get

$$\begin{aligned}&\frac{\mathrm{d}^2}{\mathrm{d}s^2}E(\gamma_s)\\ =&\int_a^b \frac{\partial}{\partial t}\langle\widetilde\nabla_{\partial/\partial s} f_s, f_t\rangle \mathrm{d}t + \int_a^b \left\langle \widetilde R\left(\frac{\partial}{\partial s}, \frac{\partial}{\partial t}\right) f_s, f_t\right\rangle - \langle\widetilde\nabla_{\partial/\partial s} f_s, \widetilde\nabla_{\partial/\partial t} f_t\rangle + \langle\widetilde\nabla_{\partial/\partial t} f_s, \widetilde\nabla_{\partial/\partial t} f_s\rangle \mathrm{d}t\\ =&\langle\widetilde\nabla_{\partial/\partial s} f_s, f_t\rangle\Big|_{(a,s)}^{(b,s)} + \int_a^b \left\langle \widetilde R\left(\frac{\partial}{\partial s}, \frac{\partial}{\partial t}\right) f_s, f_t\right\rangle - \langle\widetilde\nabla_{\partial/\partial s} f_s, \widetilde\nabla_{\partial/\partial t} f_t\rangle + \langle\widetilde\nabla_{\partial/\partial t} f_s, \widetilde\nabla_{\partial/\partial t} f_s\rangle \mathrm{d}t\end{aligned}$$

or apply Proposition 4.60 to both to get

$$\begin{aligned}\frac{\mathrm{d}^2}{\mathrm{d}s^2}E(\gamma_s) =& \int_a^b \frac{\partial}{\partial t}(\langle\widetilde\nabla_{\partial/\partial s} f_s, f_t\rangle + \langle f_s, \widetilde\nabla_{\partial/\partial t} f_s\rangle)\mathrm{d}t\\ &+ \int_a^b \left(\left\langle \widetilde R\left(\frac{\partial}{\partial s}, \frac{\partial}{\partial t}\right) f_s, f_t\right\rangle - \langle\widetilde\nabla_{\partial/\partial s} f_s, \widetilde\nabla_{\partial/\partial t} f_t\rangle - \langle f_s, \widetilde\nabla_{\partial/\partial t}\widetilde\nabla_{\partial/\partial t} f_s\rangle\right)\mathrm{d}t\\ =&(\langle\widetilde\nabla_{\partial/\partial s} f_s, f_t\rangle + \langle f_s, \widetilde\nabla_{\partial/\partial t} f_s\rangle)\Big|_{(a,s)}^{(b,s)}\\ &+ \int_a^b \left(\left\langle \widetilde R\left(\frac{\partial}{\partial s}, \frac{\partial}{\partial t}\right) f_s, f_t\right\rangle - \langle\widetilde\nabla_{\partial/\partial s} f_s, \widetilde\nabla_{\partial/\partial t} f_t\rangle - \langle f_s, \widetilde\nabla_{\partial/\partial t}\widetilde\nabla_{\partial/\partial t} f_s\rangle\right)\mathrm{d}t.\end{aligned}$$

Note that by Proposition 4.62 (a),

$$\left\langle \widetilde R\left(\frac{\partial}{\partial s}, \frac{\partial}{\partial t}\right) f_s, f_t\right\rangle\Big|_{(t,0)} = \langle R(V(t), \dot\gamma(t))V(t), \dot\gamma(t)\rangle = \langle V, R(\dot\gamma, V)\dot\gamma\rangle(t).$$

So by letting $s = 0$ in both formula, we get

Theorem 4.49 (The Second Variation of Energy)**.** Let $f(t,s)$ be a smooth variation of a geodesic $\gamma : [a,b] \to M$, then

$$\begin{aligned}\left.\frac{\mathrm{d}^2}{\mathrm{d}s^2}\right|_{s=0} E(\gamma_s) &= \langle \widetilde{\nabla}_{\partial/\partial s} f_s, \dot\gamma\rangle\Big|_a^b + \int_a^b (\langle R(\dot\gamma, V)\dot\gamma, V\rangle + \langle \nabla_{\dot\gamma} V, \nabla_{\dot\gamma} V)\mathrm{d}t \\ &= (\langle \widetilde{\nabla}_{\partial/\partial s} f_s, \dot\gamma + \langle V, \widetilde{\nabla}_{\dot\gamma} V)\Big|_a^b - \int_a^b \langle V, \nabla_{\dot\gamma}\nabla_{\dot\gamma} V - R(\dot\gamma, V)\dot\gamma(t)\rangle \mathrm{d}t.\end{aligned}$$

In particular, if the variation is proper, then $V(a) = V(b) = 0$ and

$$(\widetilde{\nabla}_{\partial/\partial s} f_s)\Big|_{(a,s)} = \nabla_{V(a)} f_s(a,s) = 0, \qquad (\widetilde{\nabla}_{\partial/\partial s} f_s)\Big|_{(b,s)} = \nabla_{V(b)} f_s(b,s) = 0$$

so we get

$$\begin{aligned}\left.\frac{\mathrm{d}^2}{\mathrm{d}s^2}\right|_{s=0} E(\gamma_s) &= \int_a^b (\langle R(\dot\gamma, V)\dot\gamma, V\rangle + \langle \nabla_{\dot\gamma} V, \nabla_{\dot\gamma} V\rangle)\mathrm{d}t \\ &= \int_a^b \langle V(t), -\nabla_{\dot\gamma}\nabla_{\dot\gamma} V(t) + R(\dot\gamma, V)\dot\gamma(t)\rangle \mathrm{d}t.\end{aligned}$$

Note that

$$\langle R(\dot\gamma, V)\dot\gamma, V\rangle = -Rm(\dot\gamma, V, \dot\gamma, V) = -K(\dot\gamma, V)\mathrm{Area}(\dot\gamma, V),$$

so we see that if (M, g) has non-positive sectional curvature, then for any proper variation of any geodesics,

$$\left.\frac{\mathrm{d}^2}{\mathrm{d}s^2}\right|_{s=0} E(\gamma_s) \geqslant 0.$$

As a result, any geodesic in non-positive sectional curvature space is locally minimizing among nearby curves.

4.5.3 Appendix: The Induced Connection

The Pullback Bundle

Let M, N be two smooth manifolds, ∇ a connection on M and $\varphi : N \to M$ a smooth map. Then we may pullback the tangent bundle $\pi : TM \to M$ over M to a vector bundle $\pi' : E = \varphi^*(TM) \to N$ over N [known as the **pullback bundle**] , where

$$E = \{(x, v) \mid x \in N, v \in TM, \varphi(x) = \pi(v)\} \subset N \times TM.$$

In other words, we just set the fiber E_x of $\pi' : E \to N$ to be the vector space $T_{\varphi(x)}M$. Denote by $\widetilde{\varphi} : E \to TM$ the induced bundle map that maps $(x, v) \in E$ to $v \in TM$. Then we have the following commutative diagram

$$\begin{array}{ccccccc} & & (x,v) & \longmapsto & v & & \\ & & \cap & & \cap & & \\ (x,v) & \in & E & \xrightarrow{\ \widetilde{\varphi}\ } & TM & \ni & v \\ \big\downarrow & & \big\downarrow{\scriptstyle \pi'} & & \big\downarrow{\scriptstyle \pi} & & \big\downarrow \\ x & \in & N & \xrightarrow[\ \varphi\ ]{} & M & \ni & \pi(v) \\ & & \cup & & \cup & & \\ & & x & \longmapsto & \varphi(x) & & \end{array}$$

In this construction, there are two natural ways to obtain sections on E:

- For any section $V \in \Gamma^\infty(TM)$, one can define an assignment

$$\begin{aligned}\widetilde{\varphi}^* V : N &\to E \\ x &\mapsto (x, V_{\varphi(x)}).\end{aligned}$$

  By this way, any smooth vector field $V \in \Gamma^\infty(TM)$ gives rise to a smooth section $\widetilde{\varphi}^* V \in \Gamma^\infty(E)$ of E.

  [Note: if E_i is a local frame of TM near $\varphi(x)$, then $\widetilde{\varphi}^* E_i$ is local frame of E near x.]

- For any section $X \in \Gamma^\infty(TN)$, one can define an assignment

$$\begin{aligned}\mathrm{d}\varphi(X) : N &\to E \\ x &\mapsto (x, (\mathrm{d}\varphi)_x(X_x)).\end{aligned}$$

  By this way, any smooth vector field $X \in \Gamma^\infty(TN)$ gives rise to a smooth section $\mathrm{d}\varphi(X) \in \Gamma^\infty(E)$ of E.

The two constructions are related as follows: For $X \in \Gamma^\infty(TN)$, $V \in \Gamma^\infty(TM)$,

$$\mathrm{d}\varphi(X) = \widetilde{\varphi}^* V \Leftrightarrow \mathrm{d}\varphi_x(X_x) = V_{\varphi(x)} \Leftrightarrow X, V \text{ are } \varphi\text{-related.}$$

In manifold theory, we have seen that if X, V and Y, W are φ-related, then $[X, Y]$ and $[V, W]$ is φ-related. So we get

Proposition 4.50. If $\mathrm{d}\varphi(X) = \widetilde{\varphi}^* V$, $\mathrm{d}\varphi(Y) = \widetilde{\varphi}^* W$, then $\mathrm{d}\varphi([X, Y]) = \widetilde{\varphi}^*([V, W])$.

To extend this proposition to more general vector fields on N, we let E_i be a local frame of TM near $\varphi(x)$, then $\widetilde{\varphi}^* E_i$ is a local frame of E near x.

Proposition 4.51. If X, $Y \in \Gamma^\infty(TN)$ and $\mathrm{d}\varphi(X) = X^i \widetilde{\varphi}^* E_i$, $\mathrm{d}\varphi(Y) = Y^j \widetilde{\varphi}^* E_j$, then

$$\mathrm{d}\varphi([X, Y]) = X(Y^j)\widetilde{\varphi}^* E_j - Y(X^i)\widetilde{\varphi}^* E_i + X^i Y^j \widetilde{\varphi}^*([E_i, E_j]).$$

Proof. For any $x \in N$ and any $f \in \mathcal{C}^\infty(M)$,

$$Y(\varphi^* f)(x) = (\mathrm{d}\varphi_x)(Y_x) f = Y^j(x)(\widetilde{\varphi}^* E_j)_x(f) = Y^j(x)(E_j)_{\varphi(x)}(f) = (Y^j \varphi^*(E_j f))(x),$$

so $Y(\varphi^* f) = Y^j \varphi^*(E_j f)$ and thus

$$Y^j X_x(\varphi^*(E_j f)) - X^i Y_x(\varphi^*(E_i f)) = X^i Y^j \varphi^*(E_i E_j f - E_j E_i f) = X^i Y^j \varphi^*([E_i, E_j] f).$$

It follows that as vectors in $T_{\varphi(x)} M$ acting on $f \in \mathcal{C}^\infty(M)$,

$$\begin{aligned}(\mathrm{d}\varphi[X, Y])_{\varphi(x)} f &= [X, Y]_x(\varphi^* f) \\ &= X_x(Y \varphi^* f) - Y_x(X \varphi^* f) \\ &= X_x(Y^j \varphi^*(E_j f)) - Y_x(X^i \varphi^*(E_i f)) \\ &= X_x(Y^j)\varphi^*(E_j f) + Y^j X_x(\varphi^*(E_j f)) - Y_x(X^i)\varphi^*(E_i f) - X^i Y_x(\varphi^*(E_i f)) \\ &= X_x(Y^j)\varphi^*(E_j f) - Y_x(X^i)\varphi^*(E_i f) + X^i(x) Y^j(x) \varphi^*([E_i, E_j] f).\end{aligned}$$

So we get, as sections in $\Gamma^\infty(E)$,

$$\mathrm{d}\varphi([X, Y]) = X(Y^j)\widetilde{\varphi}^* E_j - Y(X^i)\widetilde{\varphi}^* E_i + X^i Y^j \widetilde{\varphi}^*([E_i, E_j]).$$

□

The Induced Connection on the Pullback Bundle

Since each fiber $E_x = T_{\varphi(x)}M$, one may transplant structures on TM to E. For example, if (M, g) is a Riemannian manifold, then the metric structure on TM gives rise to a metric structure on the pullback bundle E in the natural way, namely one just endow each fiber $E_x = T_{\varphi(x)}M$ with the inner product $g_{\varphi(x)}$.

Although it is not that obvious, we may also transplant linear connections on TM to E:

Proposition 4.52. Given any linear connection ∇ on TM, there exists a unique linear connection $\widetilde{\nabla}$ on E satisfying

$$\widetilde{\nabla}_u(\widetilde{\varphi}^*V) = \widetilde{\varphi}^*(\nabla_{(\mathrm{d}\varphi)_x u}V) \tag{4.2}$$

for any $x \in N$, $u \in T_xN$ and $V \in \Gamma^\infty(TM)$.

Proof. We first prove the uniqueness. Assume $\widetilde{\nabla}$ exists. For any $x_0 \in N$ and any local frame $\{E_i\}_{i=1}^m$ around $\varphi(x_0)$, $\{\widetilde{\varphi}^*E_i\}_{i=1}^m$ is a local frame around x_0. Thus for any $s \in \Gamma^\infty(E)$, we may write

$$s(x) = s^i(x)\widetilde{\varphi}^*E_i(x)$$

for x near x_0. It follows from Leibniz rule and (4.2) that for any $u \in T_xN$,

$$\widetilde{\nabla}_u s = u(s^i)\widetilde{\varphi}^*E_i(x) + s^i(x)\widetilde{\varphi}^*(\nabla_{(\mathrm{d}\varphi)_x u}E_i). \tag{4.3}$$

So $\widetilde{\nabla}$ is determined by ∇ uniquely.

As for the existence, it is sufficient to show (4.3) is independent of the choice of the local frame $\{E_i\}$ [and thus gives rise to a map $\widetilde{\nabla} : \Gamma^\infty(TN) \times \Gamma^\infty(E) \to \Gamma^\infty(E)$ via linearity in $\Gamma^\infty(TN)$] and defines a linear connection on E. Let $E_i' = f_i^j E_j$ be another frame field around $\varphi(x)$. Write $s = \overline{s}^i\widetilde{\varphi}^*E_i'$, then $s^i = \overline{s}^j(\varphi^* f_j^i)$ and thus

$$\begin{aligned} u(s^i)\widetilde{\varphi}^*E_i + s^i\widetilde{\varphi}^*\nabla_{(\mathrm{d}\varphi)_x u}E_i &= u(\overline{s}^j)(\varphi^* f_j^i)\widetilde{\varphi}^*(E_i) + \overline{s}^j u(\varphi^* f_j^i)\widetilde{\varphi}^*(E_i) + \overline{s}^j(\varphi^* f_j^i)\varphi^*\nabla_{(\mathrm{d}\varphi)_x u}E_i \\ &= u(\overline{s}^j)\widetilde{\varphi}^*E_j' + \overline{s}^j\widetilde{\varphi}^*(\nabla_{(\mathrm{d}\varphi)_x u}E_j'). \end{aligned}$$

Thus (4.3) is independent of the choices of the frame field. The verification of linearity and Leibniz's rule is straightforward. □

Definition 4.53. For any smooth map $\varphi : N \to M$ and any linear connection ∇ on M, the unique linear connection $\widetilde{\nabla}$ on $E = \varphi^*(TM)$ defined above is called the **induced connection** of ∇ by φ on E.

To compute $\widetilde{\nabla}_X s$, one usually fix a local frame E_i of TM near $\varphi(x)$, and expend both $\mathrm{d}\varphi(X)$ and s in the induced local frame $\widetilde{\varphi}^*E_i$. Then we may rewrite (4.3) as

Lemma 4.54. Fix a local frame E_i of TM near $\varphi(x)$. Suppose $X \in \Gamma^\infty(TN)$ satisfies $\mathrm{d}\varphi(X) = X^i\widetilde{\varphi}^*E_i$, then for any section $s = s^k\widetilde{\varphi}^*E_k$,

$$\widetilde{\nabla}_X s = X(s^k)\widetilde{\varphi}^*E_k + X^i s^k\widetilde{\varphi}^*E_k + X^i s^k\widetilde{\varphi}^*(\nabla_{E_i}E_k).$$

To get a better understanding, we point out two extreme cases:

- If φ is a diffeomorphism, then E is isomorphic to TN, and $\widetilde{\nabla}$ is obtained by "transplanting everything from M to N via φ in the obvious way".
- If φ is a constant map: Suppose $\varphi(x) \equiv y$ for all $x \in N$, then E is the trivial bundle $N \times T_yM$. In this case any $s \in \Gamma^\infty(E)$ can be written as $s = s^i(x)e_i$, where e_i is a basis of T_yM, and

$$\widetilde{\nabla}_X s = (Xs^i)e_i$$

(which is independent of ∇).

Basic Properties of the Induced Connection

Let $\varphi : N \to M$ be smooth, $E = \varphi^*(TM)$ the pullback bundle, ∇ a linear connection on M, and $\widetilde{\nabla}$ the induced linear connection on E. It is not surprising that the induced connection $\widetilde{\nabla}$ inherits many nice properties from ∇.

First we study the relation between the curvature tensor

$$\widetilde{R}(X,Y)s := \widetilde{\nabla}_X\widetilde{\nabla}_Y s - \widetilde{\nabla}_Y\widetilde{\nabla}_X s - \widetilde{\nabla}_{[X,Y]}s$$

of the induced connection and the curvature tensor of the original linear connection:

Proposition 4.55. Let ∇ be a linear connection on M, $\widetilde{\nabla}$ its induced connection on E. Then for any $X, Y \in \Gamma^\infty(TN)$ and any $s \in \Gamma^\infty(E)$, we have

$$(\widetilde{R}(X,Y)s)\Big|_x = R((\mathrm{d}\varphi)_x X_x, (\mathrm{d}\varphi)_x Y_x)\widetilde{\varphi}(s|_x), \quad \forall x \in N. \tag{4.4}$$

Proof. For simplicity, we take a local frame E_i of TM near $\varphi(x)$ such that $[E_i, E_j] = 0$ [e.g. take the coordinate vector fields]. We can simplify further by assuming $[X,Y]_x = 0$ since both sides of (4.4) depends only on X_x and Y_x, again write $\mathrm{d}\varphi(X) = X^i\widetilde{\varphi}^*E_i$, $\mathrm{d}\varphi(Y) = Y^j\widetilde{\varphi}^*E_j$ and $s = s^k\widetilde{\varphi}^*E_k$. Then by Lemma 4.54,

$$\begin{aligned}\widetilde{\nabla}_X\widetilde{\nabla}_Y s &= \widetilde{\nabla}_X(Y(s^k)\widetilde{\varphi}^*E_k + Y^j s^k\widetilde{\varphi}^*(\nabla_{E_j}E_k))\\ &= X(Y(s^k))\widetilde{\varphi}^*E_k + X^iY(s^k)\widetilde{\varphi}^*(\nabla_{E_i}E_k)\\ &+ X(Y^j s^k)\widetilde{\varphi}^*(\nabla_{E_j}E_k) + X^iY^js^k\widetilde{\varphi}^*(\nabla_{E_i}\nabla_{E_j}E_k).\end{aligned}$$

Evaluating at the point x, and using the facts $[X,Y]_x = 0$ (which, together with the fact $[E_i, E_j] = 0$ and Proposition 4.51 implies $X_x(Y^j) = Y_x(X^j)$) we get

$$\begin{aligned}(\widetilde{R}(X,Y)_s)|_x = (\widetilde{\nabla}_X\widetilde{\nabla}_Y s - \widetilde{\nabla}_Y\widetilde{\nabla}_X s)|_x &= X^iY^js^k\widetilde{\varphi}^*(\nabla_{E_i}\nabla_{E_j}E_k - \nabla_{E_j}\nabla_{E_i}E_k)|_x\\ &= R((\mathrm{d}\varphi)_xX_x, (\mathrm{d}\varphi)_xY_x)\widetilde{\varphi}(s|_x).\end{aligned}$$

□

Next let's turn to torsion-freeness. Although it makes no sense to talk about torsion-freeness for a connection on a general vector bundle, since we can't exchange X and s in the expression $\nabla_X s$. However, as we have seen, every smooth vector field $Y \in \Gamma^\infty(TN)$ gives rise to a smooth section $\mathrm{d}\varphi(Y) \in \Gamma^\infty(E)$, so we could restrict ourselves to such sections and thus talk about "partial-torsion-freeness" for the induced linear connection:

Proposition 4.56. If ∇ is torsion free, then for any X, $Y \in \Gamma^\infty(TN)$, we have

$$\widetilde{\nabla}_X\mathrm{d}\varphi(Y) - \widetilde{\nabla}_Y\mathrm{d}\varphi(X) - \mathrm{d}\varphi([X,Y]) = 0.$$

Proof. Fix a local frame E_i of TM near $\varphi(x)$, then $\widetilde{\varphi}^*E_i$ is a local frame of E. Write $\mathrm{d}\varphi(X) = X^i\widetilde{\varphi}^*E_i$ and $\mathrm{d}\varphi(Y) = Y^i\widetilde{\varphi}^*E_i$ near x. Then by Lemma 4.54,

$$\widetilde{\nabla}_X\mathrm{d}\varphi(Y) = X(Y^j)\widetilde{\varphi}^*E_j + X^iY^j\widetilde{\varphi}^*(\nabla_{E_i}E_j),$$

which implies

$$\begin{aligned}\widetilde{\nabla}_X\mathrm{d}\varphi(Y) - \widetilde{\nabla}_Y\mathrm{d}\varphi(X) &= X(Y^j)\widetilde{\varphi}^*E_j + X^iY^j\widetilde{\varphi}^*(\nabla_{E_i}E_j) - Y(X^i)\widetilde{\varphi}^*E_i - X^iY^j\widetilde{\varphi}^*(\nabla_{E_j}E_i)\\ &= X(Y^j)\widetilde{\varphi}^*E_j - Y(X^i)\widetilde{\varphi}^*E_i + X^iY^j\widetilde{\varphi}^*([E_i,E_j]).\end{aligned}$$

Now the conclusion follows from Proposition 4.51. □

Finally the metric compatibility. As we have mentioned, any Riemannian metric on M induces a metric structure on the pullback bundle E.

Proposition 4.57. If g is a Riemannian metric on M, and ∇ is compatible with g, then the induced connection $\widetilde{\nabla}$ is compatible with the induced metric on E, i.e.

$$X\langle s_1, s_2\rangle = \langle \widetilde{\nabla}_X s_1, s_2\rangle + \langle s_1, \widetilde{\nabla}_X s_s\rangle$$

for any $X \in \Gamma^\infty(TN)$ and any s_1, $s_2 \in \Gamma^\infty(E)$.

Proof. It is enough to prove this property at a point x. Take a local orthonormal frame E_i of TM around $\varphi(x)$. Write $s_j = s_j^i E_i$ for $j = 1, 2$ and denote $u = X_x$, then

$$\begin{aligned}\langle \widetilde{\nabla}_u s_1, s_2\rangle + \langle s_1, \widetilde{\nabla}_u s_2\rangle &= u(s_1^i)s_2^j(x)\delta_{ij} + s_1^i(x)s_2^j(x)\langle \nabla_{(\mathrm{d}\varphi)_x u}E_i, E_j\rangle \\ &\quad + s_1^i(x)u(s_2^j)\delta_{ij} + s_1^i(x)s_2^j(x)\langle E_i, \nabla_{(\mathrm{d}\varphi)_x u}E_j\rangle \\ &= \sum_{i=1}^m u(s_1^i s_2^i) = u\langle s_1, s_2\rangle.\end{aligned}$$

□

The Use of the Induced Connection

Why should we study the induced connection? Because it provides us the correct language to perform and explain computations when we are handling vector fields associated to maps. Here are two applications. Let M be a smooth manifold with a linear connection ∇.

(1) Let $\gamma : [a, b] \to M$ be a smooth curve. We defined the concept of geodesic via $\nabla_{\dot\gamma}\dot\gamma = 0$. However, in this definition we vaguely assumed that γ is an embedded curve, otherwise $\dot\gamma$ need not be "a vector field along γ". On the other hand, from the differential equations of a geodesic, obviously we allow geodesics to be non-embedded curves (like the constant geodesics or geodesics with self-intersections). The correct way is to explain the expression $\nabla_{\dot\gamma}\dot\gamma$ via the induced connection in the case $\gamma : [a, b] \to M$ is not an embedding. We will use $\frac{\mathrm{d}}{\mathrm{d}t}$ to represent the canonical coordinate vector field on $[a, b] \subset \mathbb{R}$. We first extend the concept of parallel vector fields to sections of $\gamma^*(TM)$ which are not necessary vector fields on M:

Definition 4.58. Let $\gamma : [a, b] \to M$ be a smooth curve.

(a) We say $X : [a, b] \to TM$ is a **smooth vector field along** γ if X is smooth and $X(t) \in T_{\gamma(t)}M$ for all t. In other words, if X is a smooth section of $E = \gamma^*(TM)$.

(b) Let X be a smooth vector field along γ. We say X is **parallel** along γ if

$$\widetilde{\nabla}_{\mathrm{d}/\mathrm{d}t}\widetilde{\gamma}^* X = 0, \quad \forall t.$$

Note that $\dot\gamma$ is always a smooth vector field along γ. So we may define

Definition 4.59. We say a smooth map $\gamma : [a, b] \to M$ is a **geodesic** if $\dot\gamma$ is parallel along γ in the sense of Definition 4.58, i.e.

$$\widetilde{\nabla}_{\mathrm{d}/\mathrm{d}t}\widetilde{\gamma}^*\dot\gamma = 0, \quad \forall t.$$

Of course if γ is an embedded curve, then these definitions reduce to the old definitions that we are familiar with.

Applying Proposition 4.57 to this setting, we get

Proposition 4.60. If ∇ is a metric-compatible linear connection, then for any smooth vector fields V, W along γ,

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle V, W\rangle = \langle \widetilde{\nabla}_{\mathrm{d}/\mathrm{d}t}V, W\rangle + \langle V, \widetilde{\nabla}_{\mathrm{d}/\mathrm{d}t}W\rangle.$$

(2) Now we turn to variations of a curve. Let

$$f(t,s) : R = [a,b] \times (-\varepsilon, \varepsilon) \to M$$

be a smooth map which is a variation of a smooth curve γ. Then

$$f_s(t,s) = \mathrm{d}f_{t,s}\left(\frac{\partial}{\partial s}\right) \in T_{f(t,s)}M = E_{t,s}$$

and

$$f_t(t,s) = \mathrm{d}f_{t,s}\left(\frac{\partial}{\partial t}\right) \in T_{f(t,s)}M = E_{t,s}$$

are sections of the induced bundle $E = f^*TM$ over R. By definition we have $f(t,0) = \gamma(t)$ and

$$f_t(t,0) = \dot{\gamma}.$$

Definition 4.61. We will call

$$V(t) = f_s(t,0)$$

the **variation field** of f along γ. [It is a section on $\widetilde{\gamma}^*(TM)$.]

Applying Proposition 4.55, Proposition 4.56 to this setting, we get

Proposition 4.62. Let $f(t,s)$ be any smooth variation. Then

(a) for any (t_0, s_0),

$$\left(\widetilde{R}\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial s} f_t\right)\right)\Big|_{(t_0,s_0)} = R(f_t(t_0,s_0), f_s(t_0,s_0))f_t(t_0,s_0).$$

In particular,

$$\left(\widetilde{R}\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial s} f_t\right)\right)\Big|_{(t,0)} = R(\dot{\gamma}(t), V(t))\dot{\gamma}(t).$$

(b) If ∇ is torsion free, then

$$\widetilde{\nabla}_{\partial/\partial s} f_t = \widetilde{\nabla}_{\partial/\partial t} f_s.$$

In what follows we will mainly use our old notations ∇, in the understanding that if one expression makes no sense, one should explain it in the language of induced connection $\widetilde{\nabla}$.

Chapter 5

Jacobi Fields

5.1 Jacobi Fields

As we have seen, in the second variational formula the curvature term appears. As a result, the formula will play a crucial role in studying the relation between curvature and topology of Riemannian manifolds. Usually the first step will be: start with a geodesic, and take a special variation (e.g. a geodesic variation, sometimes with one endpoint fixed). Thus the variation field of a geodesic variation will be very important for the remaining of this course.

5.1.1 Definition of the Jacobi Field

The Jacobi Field

Let γ be a geodesic in (M, g). Suppose $f : [a, b] \times (-\varepsilon, \varepsilon) \to M$ is a geodesic variation of γ, i.e. each curve

$$\gamma_s = f(\cdot, s)$$

is a geodesic in M. Then for any s,

$$\widetilde{\nabla}_{\partial/\partial t} f_t = \widetilde{\nabla}_{\partial/\partial t} \dot{\gamma}_s = 0.$$

As a consequence,

$$\widetilde{\nabla}_{\partial/\partial t}\widetilde{\nabla}_{\partial/\partial t} f_s = \widetilde{\nabla}_{\partial/\partial t}\widetilde{\nabla}_{\partial/\partial s} f_t = \widetilde{\nabla}_{\partial/\partial t}\widetilde{\nabla}_{\partial/\partial s} f_t - \widetilde{\nabla}_{\partial/\partial s}\widetilde{\nabla}_{\partial/\partial t} f_t = \widetilde{R}\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial s}\right) f_t.$$

Taking $s = 0$, we see that the variation field V of any geodesic variation satisfies

$$\nabla_{\dot{\gamma}}\nabla_{\dot{\gamma}} V = R(\dot{\gamma}, V)\dot{\gamma}. \tag{5.1}$$

Definition 5.1. Let X be a smooth vector field X along a geodesic γ. We call X a **Jacobi field** along γ if the equation (5.1) holds.

Remark. Let γ be a geodesic. There are two trivial Jacobi fields along γ:

- Obviously $X = \dot{\gamma}$ is a Jacobi field. It is the variation field of $f(t, s) = \gamma(t + s)$.
- $X = t\dot{\gamma}$ is a Jacobi field since

  $$\nabla_{\dot{\gamma}}\nabla_{\dot{\gamma}}(t\dot{\gamma}) = \nabla_{\dot{\gamma}}(\dot{\gamma} + t\nabla_{\dot{\gamma}}\dot{\gamma}) = 0$$

  and $R(\dot{\gamma}, t\dot{\gamma})\dot{\gamma} = 0$. It is the variation field of $f(t, s) = \gamma(t + st)$.
- But $X = t^2\dot{\gamma}$ is NOT a Jacobi field since

  $$\nabla_{\dot{\gamma}}\nabla_{\dot{\gamma}}(t^2\dot{\gamma}) = \nabla_{\dot{\gamma}}(2t\dot{\gamma}) = 2\dot{\gamma} \neq 0.$$

It is not amazing that $t^2\dot\gamma$ is no longer a Jacobi field along γ:

Lemma 5.2. Let X be a Jacobi field along γ, then $f(t) = \langle X, \dot\gamma\rangle$ is a linear function.

Proof. According to the Jacobi field equation,

$$f''(t) = \frac{\mathrm{d}^2}{\mathrm{d}t^2}\langle X, \dot\gamma\rangle = \langle \nabla_{\dot\gamma}\nabla_{\dot\gamma} X, \dot\gamma\rangle = \langle R(\dot\gamma, X)\dot\gamma, \dot\gamma\rangle = 0.$$

It follows that $\langle X, \dot\gamma\rangle$ is a linear function along γ. □

Existence and Uniqueness of Jacobi Field

So the variation field of any geodesic variation is a Jacobi fields. As a result, the second variation formula for a geodesic variation is very simple. We will show that conversely, any Jacobi field on γ can be realized as the variation field of some geodesic variation of γ. Before we prove it, we need some basic properties of Jacobi fields.

Let's take a closer look of the equation for Jacobi fields. Since it is a differential equation, it is enough to study the equation in a coordinate chart. Although one may work on a general frame, to simply the computation one may use a special frame that are parallel along γ [so that the covariant derivatives of the frame are as simple as possible]. So we start with an orthonormal basis $\{e_1, \ldots, e_m\}$ of T_pM, with $e_1 = \dot\gamma(a)$, where $p = \gamma(a)$. Let

$$e_i(t) := \text{the parallel transport of } e_i \text{ along } \gamma, \quad 1 \leqslant i \leqslant m.$$

According to Proposition 2.20,

$$\langle e_i(t), e_j(t)\rangle_{\gamma(t)} = \langle e_i, e_j\rangle_{\gamma(a)} = \delta_{ij}.$$

In other words, we get an orthonormal frame $\{e_1(t), \ldots, e_m(t)\}$ along γ with $e_1(t) = \dot\gamma(t)$, and this frame is parallel along γ, i.e.

$$\nabla_{\dot\gamma(t)} e_k(t) = 0, \quad 1 \leqslant k \leqslant m.$$

Let X be a Jacobi field along γ, then with respect to this orthonormal frame we can write $X = X^i(t)e_i(t)$, and we get

$$\nabla_{\dot\gamma} X = \dot X^i(t)e_i(t) \text{ and } \nabla_{\dot\gamma}\nabla_{\dot\gamma} X = \ddot X^i(t)e_i(t).$$

It follows that the Jacobi field equation becomes

$$\ddot X^i(t)e_i(t) - X^i(t)R^j_{1i1}e_j(t) = 0.$$

So we arrived at a system of second order homogeneous ODEs,

$$\ddot X^i(t) - X^j(t)R^i_{1j1} = 0, \quad 1 \leqslant i \leqslant m,$$

Using the basic theory for second order homogeneous ODEs, we get

Theorem 5.3. Let $\gamma : [a, b] \to M$ be a geodesic, then for any $X_{\gamma(a)}, Y_{\gamma(a)} \in T_{\gamma(a)}M$, there exists a unique Jacobi field X along γ so that

$$X(a) = X_{\gamma(a)} \text{ and } \nabla_{\dot\gamma(a)} X = Y_{\gamma(a)}.$$

Moreover, the set of Jacobi fields along γ is a linear space of dimension $2m$ (which is canonically isomorphic to $T_{\gamma(a)}M \oplus T_{\gamma(a)}M$).

The following consequence is fundamental:

Corollary 5.4. If $X(t)$ is a Jacobi field along γ, and X is not identically zero, then the zeroes of X are discrete.

Proof. If X has a sequence of zeroes that converges to t_0, then $X^1(t) = \cdots = x^m(t) = 0$ for a sequence of points t_k converging to $\gamma(t_0)$. It follows that $X^i(t_0) = 0$ and $\dot X^i(t_0) = 0$ for all i, i.e. $X(t_0) = 0$, $\nabla_{\dot\gamma(t_0)} X = 0$. By uniqueness, $X \equiv 0$. □

Jacobi Fields as Variational Fields of Geodesic Variation

Now we prove that each Jacobi field X along a geodesic γ can be realized as the variation field of a geodesic variation of γ (So the space of all the Jacobi fields along γ describes all possible ways that γ can vary in "the space of all geodesics" infinitesimally):

Theorem 5.5. A vector field X along a geodesic γ is a Jacobi field if and only if X is the variation field of some geodesic variation of γ.

Proof. We have seen that the variation field of any geodesic variation of γ is a Jacobi field. Now we suppose X is a Jacobi field along γ and construct the desired geodesic variation. For simplicity we parameterize γ as $\gamma : [0,1] \to M$, so $\gamma(t) = \exp_{\gamma(0)}(t\dot\gamma(0))$ is defined for $0 \leqslant t \leqslant 1$. It follows that for any (p, Y_p) in a small neighborhood of $(\gamma(0), \dot\gamma(0))$, the exponential map $\exp_p(tY_p)$ is defined for $0 \leqslant t \leqslant 1$.

Let $\xi : (-\varepsilon, \varepsilon) \to M$ be the geodesic with initial conditions

$$\xi(0) = \gamma(0),\ \dot\xi(0) = X_{\gamma(0)}.$$

Let $T(s)$, $W(s)$ be parallel vector fields along ξ with

$$T(0) = \dot\gamma(0) \text{ and } W(0) = \nabla_{\dot\gamma(0)}X.$$

Define $f : [0,1] \to (-\varepsilon, \varepsilon) \to M$ by

$$f(t,s) = \exp_{\xi(s)}(t(T(s) + sW(s))).$$

As we mentioned above, for ε small enough, f is well-defined. Moreover, $f(t,0) = \gamma(t)$, so f is a geodesic variation of γ. Let V be the variation field of f. Since both V and X are Jacobi fields along γ, to show $V = X$, it is enough to show

$$V(0) = X_{\gamma(0)} \text{ and } \nabla_{\dot\gamma(0)}V = \nabla_{\dot\gamma(0)}X.$$

The first one follows from

$$V(0) = f_s(0,0) = \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} f(0,s) = \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} \xi(s) = X_{\gamma(0)}.$$

For the second one, we start with the fact $\widetilde\nabla_{\partial/\partial t} f_s = \widetilde\nabla_{\partial/\partial s} f_t$. Evaluate the left hand side at $(0,0)$ we get

$$(\widetilde\nabla_{\partial/\partial t} f_s)_{0,0} = \left.(\widetilde\nabla_{\partial/\partial t} f_s(t,0))\right|_{t=0} = \nabla_{\dot\gamma(0)}V,$$

and evaluate the right hand side at $(0,0)$ and use the fact

$$f_t(0,s) = (\mathrm{d}\exp_{\xi(s)})_0 \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} (t(T(s) + sW(s))) = T(s) + sW(s)$$

we get

$$(\widetilde\nabla_{\partial/\partial s} f_t)_{0,0} = \left.(\widetilde\nabla_{\partial/\partial s} f_t(0,s))\right|_{s=0} = \nabla_{X_{\gamma(0)}}(T(s) + sW(s)) = W(0) = \nabla_{\dot\gamma(0)}X.$$

So we get $\nabla_{\dot\gamma(0)}V = \nabla_{\dot\gamma(0)}X$ and thus completes the proof. □

Note that given any Jacobi field V along a geodesic γ, there exist many geodesic variations of γ whose variation fields are V [analogue: given any vector v at a point p, there exist many curves whose tangent vector at p is v]. In the proof above we give an explicit formula for one such geodesic variations, namely,

$$f(t,s) = \exp_{\xi(s)}(t(T(s) + sW(s))), \tag{5.2}$$

where ξ is a geodesic with $\xi(0) = \gamma(0)$ and $\dot\gamma(0) = V(0)$, and T, W are parallel vector fields along ξ with $T(0) = \dot\gamma(0)$ and $W(0) = \nabla_{\dot\gamma(0)}V$.

5.1.2 Jacobi Fields with Special Conditions

5.1.3 Normal Jacobi Fields

The obviously Jacobi fields $\dot\gamma$, $t\dot\gamma$ [and their linear combinations] along γ are both tangent to γ and are not so interesting in applications. Very often we need to rule out them and mainly consider normal Jacobi fields.

Definition 5.6. A Jacobi field along γ is called a **normal Jacobi field** if it is perpendicular to $\dot\gamma$ along γ.

It turns out that for any Jacobi field, the tangential components must be a linear combination of $\dot\gamma$ and $t\dot\gamma$:

Proposition 5.7. For any Jacobi field X along γ, there exists c^1, $d^1 \in \mathbb{R}$ so that

$$X^\perp = X - c^1 t\dot\gamma - d^1\dot\gamma$$

is a normal Jacobi field along γ.

Proof. By Lemma 5.2, $\langle X, \dot\gamma\rangle$ is a linear function along γ, i.e.

$$\langle X, \dot\gamma\rangle = c_1 t + d_1$$

for some constant c_1, $d_1 \in \mathbb{R}$. Now we let

$$X^\perp = X - c^1 t\dot\gamma - d^1\dot\gamma$$

with $c^1 = \dfrac{c_1}{|\dot\gamma|^2}$, $d^1 = \dfrac{d_1}{|\dot\gamma|^2}$. Then it is a Jacobi field along γ since it is a linear combination of Jacobi fields along γ, and it is normal since

$$\langle X^\perp, \dot\gamma\rangle = c_1 t + d_1 - c^1 t|\dot\gamma^2| - d^2|\dot\gamma|^2 = 0.$$

□

Note that if $X^\perp$ is a normal Jacobi field along γ, then

$$\langle \nabla_{\dot\gamma} X^\perp, \dot\gamma\rangle = \frac{\mathrm{d}}{\mathrm{d}t}\langle X^\perp, \dot\gamma\rangle - \langle X^\perp, \nabla_{\dot\gamma}\dot\gamma\rangle = 0$$

and thus $\nabla_{\dot\gamma} X^\perp \perp \dot\gamma$. It follows

Corollary 5.8. A Jacobi field X along γ is normal if and only if

$$\langle X(a), \dot\gamma(a)\rangle = \langle \nabla_{\dot\gamma(a)} X, \dot\gamma(a)\rangle = 0.$$

In particular, the set of normal Jacobi fields form a linear space of dimension $2m-2$.

Proof. With $X = X^\perp + c^1 t\dot\gamma + d^1\dot\gamma$, we have

$$\begin{aligned}\langle X(a), \dot\gamma(a)\rangle &= (c^1 a + d^1)|\dot\gamma|^2,\\ \langle \nabla_{\dot\gamma(a)} X, \dot\gamma(a)\rangle &= \langle \nabla_{\dot\gamma(a)}(c^1 t\dot\gamma + d^1\dot\gamma), \dot\gamma(a)\rangle = c^1|\dot\gamma|^2.\end{aligned}$$

The conclusion follows. □

Corollary 5.9. Let X be a Jacobi field so that

$$\langle X(t_1), \dot\gamma(t_1)\rangle = \langle X(t_2), \dot\gamma(t_2)\rangle = 0$$

for two distinct numbers t_1, t_2. Then X is a normal Jacobi field.

Proof. This follows from Lemma 5.2, i.e. $\langle X, \dot\gamma\rangle$ is a linear function along γ, and the fact that a linear function has no more than one zero unless it is identically zero. □

Normal Jacobi Fields on Spaces with Constant Sectional Curvature

Let (M, g) be a Riemannian manifold with constant sectional curvature k, i.e.

$$R(X, Y)Z = -k(\langle X, Z\rangle Y - \langle Y, Z\rangle X).$$

Let γ be a normal geodesic in M, and X a normal Jacobi field along γ. Then

$$R(\dot\gamma, X)\dot\gamma = -k(\langle\dot\gamma, \dot\gamma\rangle X - \langle X, \dot\gamma\rangle\dot\gamma) = -kX.$$

So the equation for a normal Jacobi field X along γ becomes

$$\nabla_{\dot\gamma}\nabla_{\dot\gamma}X + kX = 0.$$

Again we take an orthonormal frame $\{e_i(t)\}$ along γ so that

- $e_1(t) = \dot\gamma(t)$,
- each $e_i(t)$ is parallel along γ,

as we did in the proof of Theorem 5.3, and write

$$X = \sum_{i=2}^{m} X^i(t)e_i(t),$$

then the equation for the coefficient $X^i(t)$ becomes

$$\ddot{X}^i(t) + kX^i(t) = 0, \quad 2 \leqslant i \leqslant m.$$

The solution to this equation is

$$X^i(t) = \begin{cases} c^i\dfrac{\sin(\sqrt{k}t)}{\sqrt{k}} + d^i\cos(\sqrt{k}t), & \text{if } k > 0,\\ c^it + d^i, & \text{if } k = 0,\\ c^i\dfrac{\sinh(-\sqrt{k}t)}{\sqrt{-k}} + d^i\cosh(\sqrt{-k}t), & \text{if } k < 0, \end{cases}$$

where c^i, d^i are constants, and

$$\cosh(t) = \frac{\mathrm{e}^t + \mathrm{e}^{-t}}{2}, \quad \sinh(t) = \frac{\mathrm{e}^t - \mathrm{e}^{-t}}{2}$$

are the hyperbolic cosine and hyperbolic sin functions. Sometimes people denote

$$sn_k(t) = \begin{cases} \dfrac{\sin(\sqrt{k}t)}{\sqrt{k}}, & k > 0\\ t, & k = 0\\ \dfrac{\sinh(\sqrt{-k}t)}{\sqrt{-k}}, & k < 0 \end{cases} \quad \text{and} \quad cn_k(t) = sn_k'(t) = \begin{cases} \cos(\sqrt{k}t), & k > 0\\ 1, & k = 0\\ \cosh(\sqrt{-k}t), & k < 0 \end{cases}$$

so that we can write $X^i(t) = c^i sn_k(t) + d^i cn_k(t)$.

Jacobi Fields with $V(0) = 0$

For simplicity, let $a = 0$ for the defining interval $[a, b]$ of γ. In many applications we need geodesic variations that fix one end, i.e. with $\gamma_s(0) = \gamma(0)$ for all s. Of course the Jacobi field for such geodesic variations satisfies $V(0) = 0$. Conversely, if V is a Jacobi field along γ with $V(0) = 0$, then in (5.2) we may take

$$\xi(s) \equiv \gamma(0), \quad T(s) \equiv \dot\gamma(0), \quad W(s) \equiv \nabla_{\dot\gamma(0)}V$$

and get an explicit geodesic variation with one end fixed, whose variation field is V:

Proposition 5.10. If V is a Jacobi field along geodesic γ with $V(0) = 0$, then

$$f(t,s) = \exp_{\gamma(0)}(t(\dot\gamma(0) + s\nabla_{\dot\gamma(0)}V)).$$

is a geodesic variation of γ with $\gamma_s(0) = \gamma(0)$ and whose variation field is V.

In particular, by calculating the variation field of the above variation via its formula, we get

Corollary 5.11. If V is a Jacobi field along geodesic γ with $V(0) = 0$, then

$$V(t) = f_s(t,0) = (\mathrm{d}\exp_{\gamma(0)})_{t\dot\gamma(0)}(t\nabla_{\dot\gamma}V).$$

Taylor's Expansion of the Jacobi Field with $V(0) = 0$

Now let V, W be Jacobi fields along a geodesic γ with

$$V(0) = 0,\ \nabla_{\dot\gamma(0)}V = X_p \in T_pM \text{ and } W(0) = 0,\ \nabla_{\dot\gamma(0)}V = Y_p \in T_pM.$$

According to Corollary 5.11, we have

$$V(t) = (\mathrm{d}\exp_p)_{t\dot\gamma(0)}(tX_p) \text{ and } W(t) = (\mathrm{d}\exp_p)_{t\dot\gamma(0)}(tY_p).$$

Let $f(t) = \langle V(t), W(t)\rangle$. Then we have

$$\begin{aligned} f(0) &= \langle V(0), W(0)\rangle = 0,\\ f'(0) &= \langle \nabla_{\dot\gamma(0)}V, W(0)\rangle + \langle V(0), \nabla_{\dot\gamma(0)}W\rangle = 0,\\ f''(0) &= \langle \nabla_{\dot\gamma(0)}\nabla_{\dot\gamma}V, W(0)\rangle + 2\langle \nabla_{\dot\gamma(0)}V, \nabla_{\dot\gamma(0)}W\rangle + \langle V(0), \nabla_{\dot\gamma(0)}\nabla_{\dot\gamma}W\rangle = 2\langle X_p, Y_p\rangle. \end{aligned}$$

To compute more derivatives, we note that in view of $V(0) = 0$,

$$\nabla_{\dot\gamma(0)}\nabla_{\dot\gamma}V = R(\dot\gamma(0), V(0))\dot\gamma(0) = 0,$$

and similarly $\nabla_{\dot\gamma(0)}\nabla_{\dot\gamma}W = 0$. So [We abbreviate the k^{th} composition $\nabla_{\dot\gamma}\cdots\nabla_{\dot\gamma}$ to $\nabla_{\dot\gamma}^{(k)}$]

$$\begin{aligned} f'''(0) &= \sum_{l=0}^{3}\binom{3}{l}\langle \nabla_{\dot\gamma}^{(3-l)}V, \nabla_{\dot\gamma}^{(l)}W\rangle(0) = 0,\\ f''''(0) &= \sum_{l=0}^{4}\binom{4}{l}\langle \nabla_{\dot\gamma}^{(4-l)}V, \nabla_{\dot\gamma}W\rangle(0) = 4\langle \nabla_{\dot\gamma}^{(3)}V, \nabla_{\dot\gamma}W\rangle(0) + 4\langle \nabla_{\dot\gamma}V, \nabla_{\dot\gamma}^{(3)}W\rangle(0). \end{aligned}$$

To calculate the third order derivative, we note that if we take the $(k-2)^{\text{th}}$ covariant derivative of the Jacobi field equation for V, then

$$\nabla_{\dot\gamma}^{(k)}V - \sum_{l=0}^{k-2}\binom{k-2}{l}(\nabla_{\dot\gamma}^{(k-2-l)}R)(\dot\gamma, \nabla_{\dot\gamma}^{(l)}V)\dot\gamma = 0,$$

where we used the facts

$$\nabla_W(R(X,Y)Z) = (\nabla_W R)(X,Y)Z + R(\nabla_W X, Y)Z + R(X, \nabla_W Y)Z + R(X,Y)\nabla_W Z$$

and $\nabla_{\dot\gamma}\dot\gamma = 0$. Taking $k = 3$, we get

$$\langle V(t), W(t)\rangle = \langle X_p, Y_p\rangle t^2 - \frac{1}{3}Rm(\dot\gamma(0), X_p, \dot\gamma(0), Y_p)t^4 + \mathcal{O}(t^5).$$

In particular, if we take $W = V$ and assume $|X_p| = 1$, then

$$|V(t)|^2 = t^2 - \frac{1}{3}Rm(\dot\gamma(0), X_p, \dot\gamma(0), X_p)t^4 + \mathcal{O}(t^5).$$

5.2 Immediate Applications of Jacobi Field to Curvature

We have studied Jacobi fields along a geodesic γ, which control all possible geodesic variations of γ. Now we give some immediate applications.

5.2.1 Geometric Interpretations of Various Curvatures

Taylor's Expansion of Metric Tensor in Riemannian Normal Coordinates

Recall that if V, W are Jacobi fields along a geodesic γ with

$$V(0) = 0,\ \nabla_{\dot\gamma(0)} V = X_p \in T_pM \text{ and } W(0) = 0,\ \nabla_{\dot\gamma(0)} V = Y_p \in T_pM,$$

then the function $f(t) = \langle V(t), W(t)\rangle$ has the Taylor's expansion

$$f(t) = \langle X_p, Y_p\rangle t^2 - \frac{1}{3} Rm(\dot\gamma(0), X_p, \dot\gamma(0), Y_p)t^4 + \mathcal{O}(t^5).$$

For the first application, we calculate the next term in the Taylor's expansion of any Riemannian metric tensor in any Riemannian normal coordinate system. Recall that with any Riemannian normal coordinate system centered at p,

$$g_{ij}(p) = \delta_{ij} \text{ and } \partial_k g_{ij}(p) = 0.$$

We now prove that the next term encodes the curvature information:

Theorem 5.12. With respect to Riemannian normal coordinates near p, the functions g_{ij}'s admit the following Taylor expansion at $x = 0$,

$$g_{ij}(x) = \delta_{ij} - \frac{1}{3} R_{ikjl}(p) x^k x^l + \mathcal{O}(|x^3|). \tag{5.3}$$

Proof. Let $(U; x^1, \dots, x^m)$ be a Riemannian normal coordinate system near p. Fix x^i's and let γ be the geodesic starting at p in the direction $X_p = x^i \partial_i$,

$$\gamma(t) = (tx^1, \dots, tx^m), \quad 0 \leqslant t \leqslant \varepsilon.$$

For each $1 \leqslant i \leqslant m$, consider a geodesic variation

$$f_i(t, s) = (tx^1, \dots, t(x^i + s), \dots, tx^m).$$

Its variation field $V_i = t\partial_i$ is thus a Jacobi field along γ, which satisfies

$$V_i(0) = 0, \quad \nabla_{\dot\gamma(0)} V_i = \partial_i.$$

So if we let

$$h(t) = t^2 g_{ij}(tx^1, \dots, tx^m) = \langle V_i(t), V_j(t)\rangle,$$

then

$$\begin{aligned} g_{ij}(tx^1, \dots, tx^m) &= \frac{1}{t^2}\langle V_i(t), V_j(t)\rangle \\ &= \frac{1}{t^2}\left(\delta_{ij} t^2 - \frac{1}{3} Rm(X_p, \partial_i, X_p, \partial_j) t^4 + \mathcal{O}(t^5)\right) \\ &= \delta_{ij} - \frac{1}{3} Rm(\partial_i, X_p, \partial_j, X_p) t^2 + \mathcal{O}(t^3) \\ &= \delta_{ij} - \frac{1}{3} Rm(\partial_i, \partial_k, \partial_j, \partial_l)(tx^k)(tx^l) + \mathcal{O}(t^3). \end{aligned}$$

This proves the theorem. ∎

Remark. One can continue to calculate $\nabla^{(k)}_{\dot\gamma(0)} V_i$'s and get a full expansion of g_{ij} in Riemannian normal coordinates. For example, the next two terms are

$$\frac{1}{6} R_{iklj;r} x^k x^l x^r + \left(\frac{1}{20} R_{iklj;rs} + \frac{2}{45} R^m_{kil} R_{rjsm}\right) x^k x^l x^r x^s.$$

Taking derivative of (5.3), we get

$$\partial_r g_{ij} = -\frac{1}{3}R_{irjl}x^l - \frac{1}{3}R_{ikjr}x^k + \mathcal{O}(|x|^2).$$

Taking derivative again and evaluate at p, we get

$$\partial_s\partial_r g_{ij}(0) = -\frac{1}{3}R_{irjs}(p) - \frac{1}{3}R_{isjr}(p).$$

As a consequence, we get Riemann's original definition of the curvature tensor:

Corollary 5.13. With respect to Riemannian normal coordinates, one has

$$R_{ijkl}(p) = \frac{1}{2}(\partial_i\partial_l g_{jk} + \partial_j\partial_k g_{il} - \partial_i\partial_k g_{jl} - \partial_j\partial_l g_{ik})(0).$$

Proof. The right hand side equals

$$\frac{1}{6}(R_{jlik} + R_{jilk} + R_{ikjl} + R_{ijkl} - R_{jkil} - R_{jikl} - R_{iljk} - R_{ijlk})(p),$$

which equals $R_{ijkl}(p)$ by using symmetries of the Riemann curvature tensor. □

Geometric Meaning of Sectional Curvature

Now we are ready to give geometric interpretations of curvatures. We start with sectional curvature:

Theorem 5.14. Let $\Pi_p \subset T_pM$ be a 2-dimensional plane. Denote by C_r^0 the circle of radius r in Π_p centered at p, and $C_r = \exp_p(C_r^0)$. Let L_r be the length of C_r. Then

$$\lim_{r\to 0}\frac{2\pi r - L_r}{r^3} = \frac{\pi}{3}K(\Pi_p).$$

Proof. Take an orthonormal basis $\{e_1, \dots, e_m\}$ of T_pM so that Π_p is spanned by e_1, e_2, and consider the normal coordinate system with respect to $\{e_i\}$. Then for r small, the circle C_r has equation

$$C_r : x^1(t) = r\cos t,\ x^2(t) = r\sin t,\ x^k(t) = 0\ (k \geqslant 3).$$

It follows that

$$\begin{aligned}
|\dot{C}_r(t)|^2 &= g_{ij}(C_r(t))\dot{x}^i(t)\dot{x}^j(t)\\
&= \left(1 - \frac{1}{3}R_{1212}x^2x^2\right)\dot{x}^1\dot{x}^1 + \left(1 - \frac{1}{3}R_{2121}x^1x^1\right)\dot{x}^2\dot{x}^2 - 2\frac{1}{3}R_{1221}x^1x^2\dot{x}^1\dot{x}^2 + \mathcal{O}(r^5)\\
&= r^2 - \frac{r^4}{3}K(\Pi_p) + \mathcal{O}(r^5).
\end{aligned}$$

So

$$\begin{aligned}
L_r = \mathrm{Length}(C_r) = \int_0^{2\pi}|\dot{C}_r|\mathrm{d}t &= r\int_0^{2\pi}\sqrt{1 - \frac{r^2}{3}K(\Pi_p) + \mathcal{O}(r^3)}\mathrm{d}t\\
&= 2\pi r - \frac{\pi}{3}K(\Pi_3)r^3 + \mathcal{O}(r^4).
\end{aligned}$$

This implies the theorem. □

So the sectional curvature measures the derivation of the length of small geodesic circles centered at p to the standard circles of the same radius in Euclidean plane.

Geometric Meaning of Ricci Curvature

With the Taylor's expansion of g_{ij} at hand, it is easy to get the Taylor's expansion of $\det(g_{ij})$: According to Theorem 5.12 we get

$$(g_{ij}) = I + \left(-\frac{1}{3}R_{ikjl}(p)x^k x^l + \mathcal{O}(|x^3|)\right)$$

which implies[1]

$$\log(g_{ij}) = \left(-\frac{1}{3}R_{ikjl}(p)x^k x^l + \mathcal{O}(|x^3|)\right)$$

and thus

$$\begin{aligned}\det(g_{ij}) = \det(\mathrm{e}^{\log(g_{ij})}) = \mathrm{e}^{\mathrm{tr}(\log(g_{ij}))} &= \mathrm{e}^{-\frac{1}{3}Rc_{kl}(p)x^k x^l + \mathcal{O}(|x|^3)} \\ &= 1 - \frac{1}{3}Rc_{kl}(p)x^k x^l + \mathcal{O}(|x|^3).\end{aligned}$$

As an immediate consequence, we get the Taylor's expansion for the volume element:

Corollary 5.15. In Riemannian normal coordinates centered at p, one has

$$\sqrt{\det(g_{ij})} = 1 - \frac{1}{6}Rc_{kl}(p)x^k x^l + \mathcal{O}(|x|^3).$$

In particular, we can prove that the Ricci curvature measures the change of the volume element in the given direction:

Corollary 5.16. Let $u_p \in S_pM$ be a unit tangent vector at p, and let $\gamma(t)$ be the geodesic starting at p with $\dot\gamma(0) = u_p$. Then

$$\sqrt{\det(g_{ij}(\gamma(t)))} = 1 - \frac{Ric(u_p)}{6}t^2 + \mathcal{O}(t^3)$$

Proof. Take an orthonormal basis of T_pM with $e_1 = u_p$. With respect to the associated Riemannian normal coordinates, the geodesic $\gamma(t) = \exp_p(tu_p)$ is given by

$$\gamma : x^1(t) = t, \quad x^2 = \cdots = x^m = 0.$$

It follows

$$\sqrt{\det(g_{ij}(\gamma(t)))} = 1 - \frac{1}{6}Rc_{11}(p)t^2 + \mathcal{O}(t^3) = 1 - \frac{1}{6}Ric(u_p)t^2 + \mathcal{O}(t^3).$$

□

Geometric Meaning of Scalar Curvature

Finally we study the scalar curvature S. We have

Proposition 5.17. For r small enough,

$$\mathrm{Vol}(B(p,r)) = \omega_m r^m \left(1 - \frac{S(p)}{6(m+2)}r^2 + \mathcal{O}(|r^3|)\right),$$

where ω_m is the volume of the unit ball in $\mathbb{R}^m$.

Proof. By definition

$$\begin{aligned}\mathrm{Vol}(B(p,r)) &= \int_{B_r(0)} \sqrt{\det(g_{ij})}\mathrm{d}x^1\cdots\mathrm{d}x^m \\ &= \int_{B_r(0)} \left(1 - \frac{1}{6}Rc_{kl}(p)x^k x^l\right)\mathrm{d}x^1\cdots\mathrm{d}x^m + \mathcal{O}(r^3) \\ &= \omega_m r^m - \frac{Rc_{kl}(p)}{6}\int_{B_r(0)} x^k x^l \mathrm{d}x^1\cdots\mathrm{d}x^m + \mathcal{O}(r^3).\end{aligned}$$

[1]Here we use $\log(I+A) = A + \mathcal{O}(|A|^2)$ and $\mathrm{e}^A = 1 + A + \mathcal{O}(|A|^2)$ for matrix A.

An elementary computation shows

$$\int_{B_r(0)} x^k x^l \mathrm{d}x^1 \cdots \mathrm{d}x^m = \frac{\omega_m}{m+2} r^{m+2} \delta^{kl}$$

and the conclusion follows. □

So the scalar curvature measures the derivation of the volume of a geodesic ball to the standard Euclidean ball with the same radius.

Corollary 5.18. The surface area of geodesic sphere $S(p,r)$ is

$$\mathrm{Area}(S(p,r)) = m\omega_m r^{m-1} - \frac{S(p)}{6}\omega_m r^{m+1} + \mathcal{O}(r^{m+2}).$$

5.2.2 Cartan's Local Isometry Theorem

Cartan's Local Isometry Theorem

In view of Theorem 5.12, one may anticipate that "curvature determines the Riemannian metric". Although this is not true in the most general sense, there are many theorems supporting this philosophy. In what follows we prove a theorem of Cartan in this direction. More precisely, let (M,g) and $(\widetilde{M},\widetilde{g})$ be two Riemannian manifolds, $p \in M$ and $\widetilde{p} \in \widetilde{M}$. Let $B(p,r)$ and $B(\widetilde{p},r)$ be normal neighborhoods of p and $\widetilde{p}$ respectively. Given a smooth map $\phi : B(p,r) \to B(\widetilde{p},r)$, one may ask: under what condition will ϕ be an isometry? Of course if ϕ is an isometry, then

- $L = \mathrm{d}\phi_p : (T_pM, g_p) \to (T_{\widetilde{p}}\widetilde{M}, \widetilde{g}_{\widetilde{p}})$ is a linear isometry,
- "the curvature tensor at corresponding points are the same".

Cartan's theorem claims that the converse is true.

We need more explanation for the second condition above. How to compare the curvature tensor at q and $\varphi(q)$? We need to identify the tangent spaces T_qM and $T_{\phi(q)}\widetilde{M}$ first. How? We already have the map L which identifies T_pM with $T_{\widetilde{p}}\widetilde{M}$. For $q \neq p$ we simply parallel transport vectors in T_qM and in $T_{\phi(q)}\widetilde{M}$ along geodesics to get vectors in T_pM and in $T_{\phi(q)}\widetilde{M}$ respectively, and then apply the map L.

Now we start to state Cartan's theorem. Suppose

$$L : (T_pM, g_p) \to (T_{\widetilde{p}}\widetilde{M}, \widetilde{g}_{\widetilde{p}})$$

is a linear isometry. Then one may define a map

$$\phi = \exp_{\widetilde{p}} \circ L \circ (\exp_p)^{-1} : B(p,r) \to B(\widetilde{p},r).$$

For any $q \in B(p,r)$, we let $\gamma = \gamma_q : [0,1] \to M$ be the unique geodesic in $B(p,r)$ with $\gamma(0) = p$, $\gamma(1) = q$, and let $\widetilde{\gamma} = \phi \circ \gamma$. Note that $\widetilde{\gamma}$ is the unique geodesic in $B(\widetilde{p},r) \subset \widetilde{M}$ with $\widetilde{\gamma}(0) = \widetilde{p} = \phi(p)$, $\widetilde{\gamma}(1) = \widetilde{q} = \phi(q)$ and $\dot{\widetilde{\gamma}} = L(\dot{\gamma}(0))$. Define

$$L_q = P^{\widetilde{\gamma}} \circ L \circ (P^{\gamma})^{-1} : T_qM \to T_{\phi(q)}\widetilde{M},$$

Theorem 5.19 (Cartan's Local Isometry Theorem)**.** If for any $q \in B(p,r)$ and any $u,v,w \in T_qM$, one has

$$L_q(R(u,v)w) = \widetilde{R}(L_q(u), L_q(v))L_q(w),$$

then ϕ is an isometry, and $\mathrm{d}\phi_q = L_q$ for all $q \in B(p,r)$.

Note that for constant curvature spaces, the condition holds trivially. So we get

Corollary 5.20. If both (M,g) and $(\widetilde{M},\widetilde{g})$ has constant sectional curvature k, then for any $p \in M$ and $\widetilde{p} \in \widetilde{M}$, one can find a neighborhood $U \ni p$ and $\widetilde{U} \ni \widetilde{p}$ so that (U,g) and $(\widetilde{U},\widetilde{g})$ are isometric.

which implies Riemann's theorem for constant sectional curvature spaces, namely, Theorem 3.42.

Proof of Cartan's Local Isometry Theorem

We first prove $|\mathrm{d}\phi_q(v)| = |v|$ for any $v \in T_qM$. By using polarization this implies that $\mathrm{d}\phi_q$ preserves inner products. Since ϕ is already a diffeomorphism, we conclude that ϕ is an isometry.

The idea to prove $|\mathrm{d}\phi_q(v)| = |v|$ is: realize both v and $\mathrm{d}\phi_q(v)$ as Jacobi field vector at endpoints, and compare the length of two Jacobi fields at each point. We first construct a Jacobi field V along γ with $V(0) = 0$, $V(1) = v$. By Corollary 5.11, if V is such a Jacobi field, then

$$V(t) = (\mathrm{d}\exp_p)_{t\dot\gamma(0)}(t\nabla_{\dot\gamma(0)}V),$$

which implies $v = (\mathrm{d}\exp_p)_{\dot\gamma(0)}(\nabla_{\dot\gamma(0)}V)$. As a result, V is the unique Jacobi field with $V(0) = 0$ and $\nabla_{\dot\gamma(0)}V = (\mathrm{d}\exp_p)^{-1}_{t\dot\gamma(0)}(v)$.

Next we construct the Jacobi field $\widetilde{V}$ along $\widetilde{\gamma}$ with $\widetilde{V}(0) = 0$ and $\widetilde{V}(1) = \mathrm{d}\phi_q(v)$. In fact, we may simply take $\widetilde{V}$ to be the Jacobi field along $\widetilde{\gamma}$ with $\widetilde{V}(0) = 0$ and $\widetilde{V}_{\dot{\widetilde{\gamma}}(0)}V = L(\nabla_{\dot\gamma(0)}V)$. Since $\exp_p(\dot\gamma(0)) = \gamma(1) = q$, it follows

$$\begin{aligned}
\mathrm{d}\phi_q(v) &= (\mathrm{d}\exp_{\widetilde{p}})_{L(\dot\gamma(0))} \circ L \circ (\mathrm{d}\exp_p^{-1})_q(v) \\
&= (\mathrm{d}\exp_{\widetilde{p}})_{L(\dot\gamma(0))} \circ L \circ (\mathrm{d}\exp_p)^{-1}_{\dot\gamma(0)}(v) \\
&= (\mathrm{d}\exp_{\widetilde{p}})_{\dot{\widetilde{\gamma}}(0)}(L(\nabla_{\dot\gamma(0)}V)) = \widetilde{V}(1),
\end{aligned}$$

where the last equality follows from Corollary 5.11.

To prove $|V(1)| = |\widetilde{V}(1)|$, we apply a standard trick: Let $e_1(t) = \dot\gamma(t)$, $e_2(t)$, …, $e_m(t)$ be an orthonormal frame that is parallel along γ. Let $V(t) = V^i(t)e_i(t)$, then

$$|V(1)|^2 = \sum(V^i(1))^2.$$

Moreover, $V^i(t)$ is the solution to the Jacobi equation

$$\ddot{V}^i(t) - \langle R(\dot\gamma, e_k)\dot\gamma, e_i\rangle V^k = 0,$$

with initial conditions $V^i(0) = 0$ and $\nabla_{\dot\gamma(0)}V = \dot{V}^i(0)e_i(0)$.

Similarly, we let $\widetilde{e}_1(t) = \dot{\widetilde{\gamma}}(t)$, $\widetilde{e}_2(t)$, …, $\widetilde{e}_m(t)$ be the parallel orthonormal frame along $\widetilde{\gamma}$ with $\widetilde{e}_i(0) = L(e_i(0))$. Then $L_{\gamma(t)}(e_j(t)) = \widetilde{e}_j(t)$ for all j, and thus

$$\langle\widetilde{R}(\dot{\widetilde{\gamma}}, \widetilde{e}_k)\dot{\widetilde{\gamma}}, \widetilde{e}_i\rangle = \langle R(\dot\gamma, e_k)\dot\gamma, e_i\rangle.$$

As a result, if $\widetilde{V} = \widetilde{V}^i\widetilde{e}_j(t)$, then $\widetilde{V}^i$'s satisfy exactly the same equations and the same initial conditions as V^i's, and thus $\widetilde{V}^i = V^i$ for all $1 \leqslant i \leqslant m$. It follows

$$|\widetilde{V}(1)|^2 = \sum(\widetilde{V}^i(1))^2 = \sum(V^i(1))^2 = |V(1)|^2.$$

Finally, the fact $\widetilde{V} = V^i\widetilde{e}_i$ that we just proved also implies

$$(\mathrm{d}\phi_q)(v) = \widetilde{V}(1) = V^i(1)\widetilde{e}_i(1) = V^iL_q(e_i(1)) = L_q(V(1)) = L_q(v)$$

and thus the proof is completed.

5.3 Conjugate Point and Applications

5.3.1 Conjugate Points

The Index Form

We want to understand the mechanism for geodesics to fail to be length minimizing among nearby curves with the same endpoints. So we go back to the second variation formula for a proper variation $f(t,s) = \gamma_s(t)$ of a geodesic $\gamma : [a,b] \to M$ with variation field X,

$$\left.\frac{\mathrm{d}^2}{\mathrm{d}s^2}\right|_{s=0} E(\gamma_s) = \int_a^b (\langle R(\dot\gamma, X)\dot\gamma, X\rangle + \langle\nabla_{\dot\gamma}X, \nabla_{\dot\gamma}X\rangle)\mathrm{d}t.$$

If the second variation is positive, then γ is length minimizing among these γ_s's for s small. Since any vector field can be realized as a variation field [see Exercise Sheet 3], we are led to study the quadratic form

$$I(X,X) := \int_a^b (\langle R(\dot\gamma, X)\dot\gamma, X\rangle + \langle \nabla_{\dot\gamma} X, \nabla_{\dot\gamma} X\rangle)\mathrm{d}t.$$

for any vector field X along γ, and thus are led to study its polarization

$$\begin{aligned} I(X,X) &:= \int_a^b (\langle R(\dot\gamma, X)\dot\gamma, X\rangle + \langle \nabla_{\dot\gamma} X, \nabla_{\dot\gamma} X\rangle)\mathrm{d}t \\ &= \int_a^b \langle R(\dot\gamma, X)\dot\gamma - \nabla_{\dot\gamma}\nabla_{\dot\gamma} X, Y\rangle \mathrm{d}t + \langle \nabla_{\dot\gamma} X, Y\rangle \Big|_a^b. \end{aligned}$$

In many applications one need to consider a variation whose variation field is only continuous and piecewise smooth. So in this most general case, the **index form** $I = I^\gamma$ of the geodesic γ is a symmetric bilinear form defined on

$$\mathcal{V} = \mathcal{V}_\gamma = \{X \text{ is a continuous piecewise smooth vector field along } \gamma\}$$

by the formula

$$\begin{aligned} I(X,Y) &= \int_a^b (\langle R(\dot\gamma, X)\dot\gamma, Y\rangle + \langle \nabla_{\dot\gamma} X, \nabla_{\dot\gamma} Y\rangle)\mathrm{d}t \\ &= \int_a^b \langle R(\dot\gamma, X)\dot\gamma - \nabla_{\dot\gamma}\nabla_{\dot\gamma} X, Y\rangle \mathrm{d}t + \langle \nabla_{\dot\gamma} X, Y\rangle \Big|_a^b - \sum_{j=1}^k \langle \nabla_{\dot\gamma(t_j^+)} X - \nabla_{\dot\gamma(t_j^-)} X, Y\rangle, \end{aligned} \tag{5.4}$$

where $a < t_1 < \cdots < t_k < b$ are those points where X is not smooth, and $\nabla_{\dot\gamma(t_j^+)} X$ means $\lim_{t\to t_j^+} \nabla_{\dot\gamma(t)} X$.

Index Form v.s. Jacobi Fields

Now we relate the index form I with Jacobi fields along $\gamma : [a,b] \to M$. Denote

$$\mathcal{V}^0 = \mathcal{V}^0_\gamma := \{X \in \mathcal{V} \mid X(a) = 0, X(b) = 0\}.$$

We have

Proposition 5.21. Let $V \in \mathcal{V}$. Then V is a Jacobi field along γ if and only if for any $X \in \mathcal{V}^0$, $I(V,X) = 0$.

Proof. ($\Rightarrow$) According to (5.4), if $X \in \mathcal{V}^0$ and V is a Jacobi field (which has to be smooth) along γ, then $I(V,X) = 0$.
($\Leftarrow$) Conversely assume $V \in \mathcal{V}$ satisfies $I(V,X) = 0$ for all $X \in \mathcal{V}^0$, and

$$a = t_0 < t_1 < \cdots < t_k < t_{k+1} = b$$

is a subdivision of $[a,b]$ so that V is smooth on each $[t_j, t_{j+1}]$.

- First take a smooth function $f : [a,b] \to \mathbb{R}$ with $f(t_i) = 0$ for all i and $f(t) > 0$ for all $t \notin \{t_0, t_1, \ldots, t_{k+1}\}$, and define

  $$X = f(t)(R(\dot\gamma, V)\dot\gamma - \nabla_{\dot\gamma}\nabla_{\dot\gamma} V).$$

  Then $X \in \mathcal{V}^0$ and so

  $$0 = I(V,X) = \int_a^b f(t)|R(\dot\gamma, V)\dot\gamma - \nabla_{\dot\gamma}\nabla_{\dot\gamma} V|^2 \mathrm{d}t.$$

  It follows that V is a Jacobi field on each (t_j, t_{j+1}).

- Next let's choose any $X' \in \mathcal{V}^0$ with

$$X'(t_i) = \nabla_{\dot{\gamma}(t_i^+)} V - \nabla_{\dot{\gamma}(t_i^-)} V.$$

Then

$$0 = I(V, X') = -\sum_{i=1}^{k} |\nabla_{\dot{\gamma}(t_i^+)} V - \nabla_{\dot{\gamma}(t_i^-)} V|^2.$$

It follows that V is of class $\mathcal{C}^1$ at each t_i.

By uniqueness, V is smooth. So V is a Jacobi field. □

Conjugate Points

In particular, if there is nonzero Jacobi field V along γ with $V(t_1) = V(t_2) = 0$, then $I^{\overline{\gamma}}(V, V) = 0$, where $\overline{\gamma} = \gamma|_{[t_1, t_2]}$, and thus $I^{\overline{\gamma}}$ is not positive definite. As a result, $\overline{\gamma}$ may fail to be length minimizing for the variation with variation field V.

Definition 5.22. Let (M, g) be a Riemannian manifold, $\gamma : [a, b] \to M$ a geodesic, and $t_1 \neq t_2 \in [a, b]$. If there exists a Jacobi field V along γ which is not identically zero, such that $V(t_1) = V(t_2) = 0$, then we say $\gamma(t_2)$ is **conjugate** to $\gamma(t_1)$ along γ.

Note that according to Corollary 5.9, any Jacobi field V along γ satisfying $V(t_1) = 0$ and $V(t_2) = 0$ (where $t_1 \neq t_2$) must be a normal Jacobi field. So if $q = \gamma(t_2)$ is a conjugate point of $p = \gamma(t_1)$ along γ, then

$$\mathcal{J}_{\gamma, t_1, t_2} = \{V \mid V \text{ is a Jacobi field along } \gamma \text{ with } V(t_1) = V(t_2) = 0\}$$

is a vector subspace of the space $\mathcal{J}_\gamma^\perp$ of normal Jacobi fields along γ.

Definition 5.23. If $q = \gamma(t_2)$ is a conjugate point of $p = \gamma(t_1)$ along γ, we call

$$n_{\gamma, t_1}(t_2) := \dim \mathcal{J}_{\gamma, t_1, t_2}$$

the **multiplicity** of the conjugate point q to p along γ.

By definition, if q is conjugate to p along a geodesic γ, then p is conjugate to q along the geodesic $-\gamma$, with the same multiplicity. We have

Lemma 5.24. Suppose $\dim M = m$, then $n_{\gamma, t_1}(t_2) \leqslant m - 1$.

Proof. As we have seen in Section 5.1, a Jacobi field V is uniquely determined by $V(t_1)$ and $\nabla_{\dot{\gamma}(t_1)} V$. Moreover, V is normal implies $\nabla_{\dot{\gamma}(t_1)} V \in (\dot{\gamma}(t_1))^\perp$. So $\mathcal{J}_{\gamma, t_1, t_2}$ is isomorphic to a subspace of

$$\{(0, v) \mid v \in (\dot{\gamma}(t_1))^\perp\} \subset \{(u, v) \mid u, v \in T_{\gamma(t_1)} M\}$$

and the conclusion follows. □

Example. Consider the round sphere (S^m, g_{round}) whose sectional curvature is 1. Let $\gamma : [0, l] \to M$ be a normal geodesic starting from any p. Then in Section 5.1 we have seen that any normal Jacobi field along γ with $V(0) = 0$ must be of the form

$$V(t) = \sum_{i=2}^{m} c^i \sin(t) e_i(t),$$

where $\{e_i(t)\}$ is a parallel orthonormal frame along γ, with $e_1(t) = \dot{\gamma}(t)$. It follows

- if γ has length less than π, then there is no conjugate point of p,
- if the length of γ is between π and 2π, then the antipodal point $\gamma(\pi) = -p$ is the only conjugate point to the north pole along any geodesic starting at p, and its multiplicity equals $m - 1$.

We may also repeat the same argument for $(\mathbb{R}^m, g_0)$ and $(\mathbb{H}^m, g_{\text{hyperbolic}})$, and arrive at the conclusion that there is no conjugate point at all. In fact the same result holds for any Riemannian manifold whose sectional curvatures are non-positive:

Proposition 5.25. Let (M, g) be a Riemannian manifold whose sectional curvature is non-positive. Then any $p \in M$ has no conjugate point along any geodesic.

Proof. Let γ be any geodesic from $\gamma(0) = p$ and V any nonzero normal Jacobi field along γ with $V(0) = 0$. Let $f(t) = \langle V(t), V(t)\rangle$. Then

$$f'(t) = 2\langle \nabla_{\dot\gamma(t)} V, V\rangle$$

and thus, in view of $R(\dot\gamma, V, \dot\gamma, V) = -K(\dot\gamma, V)|\dot\gamma|^2|V|^2 \geqslant 0$,

$$f''(t) = 2\langle \nabla_{\dot\gamma}\nabla_{\dot\gamma} V, V\rangle + 2|\nabla_{\dot\gamma} V|^2 = 2R(\dot\gamma, V, \dot\gamma, V) + 2|\nabla_{\dot\gamma} V|^2 \geqslant 0.$$

Since $f(0) = 0$, $f'(0) = 0$ and $f''(0) > 0$, we conclude $f(t) > 0$ for $t > 0$. In other words, V has no other zeroes along γ. So p has no conjugate point along γ. □

Remark. In the proof we also get $f''(t) \geqslant -K(t)|\dot\gamma|^2 f(t)$, where $K(t) = K(\dot\gamma(t), V(t))$, from which one may derive better lower bounds of f.

5.3.2 Conjugate Points via Critical Points of the Exponential Map

Conjugate Points v.s. the Exponential Map

It turns out that conjugate points are exactly those points where the exponential map fails to be diffeomorphism. In fact, suppose $\gamma : [0, l] \to M$ is a geodesic, and V is a Jacobi field along γ with $V(0) = 0$, then by Corollary 5.11,

$$V(t) = (\mathrm{d}\exp_p)_{t\dot\gamma(0)}(t\nabla_{\dot\gamma(0)} V).$$

It follows

$q = \gamma(t_0)$ is conjugate to $p = \gamma(0)$

$\iff$ there is a nonzero Jacobi field V along γ so that $V(0) = 0$ and $V(t_0) = 0$

$\overset{(*)}{\iff} \nabla_{\dot\gamma(0)} V \neq 0$ and $0 = V(t_0) = (\mathrm{d}\exp_p)_{t_0\dot\gamma(0)}(t_0\nabla_{\dot\gamma(0)} V)$

$\iff \ker(\mathrm{d}\exp_p)_{t_0\dot\gamma(0)} \neq 0$

Moreover, $(*)$ also implies that $\mathcal{J}_{\gamma,p,q}$ is isomorphic to $\ker(\mathrm{d}\exp_p)_{t_0\dot\gamma(0)}$. So we get the following useful characterization of conjugate point:

Theorem 5.26. Let $\gamma : [0, l] \to M$ be a geodesic. Then $q = \gamma(t_0)$ is a conjugate point of $p = \gamma(0)$ if and only if $\exp_p$ is singular at $t_0\dot\gamma(0)$. Moreover, the multiplicity

$$n_{\gamma,p}(q) = \dim\ker(\mathrm{d}\exp_p)_{t_0\dot\gamma(0)}.$$

The Cartan-Hadamard Theorem

As an immediate application, we prove

Theorem 5.27 (Cartan-Hadamard)**.** Let (M, g) be a complete Riemannian manifold with non-positive sectional curvature, then

(1) for any $p \in M$, the exponential map $\exp_p : T_pM \to M$ is a covering map.

(2) if M is also simply connected, then $\exp_p$ is a diffeomorphism.

Proof. According to Proposition 5.25 and Theorem 5.26,

$$\exp_p : T_pM \to M$$

is a local diffeomorphism everywhere. So (1) follows from Corollary 4.38. If M is simply connected, then any covering map to M must be a homeomorphism. Since $\exp_p$ is also a local diffeomorphism, it must be diffeomorphism. □

Definition 5.28. A complete simply-connected Riemannian manifold with non-positive curvature is called a **Cartan-Hadamard manifold**, or an **Hadamard manifold**.

Remark. Let (M, g) be a complete Riemannian manifold. We say $p \in M$ is a **pole** of (M, g) if $\exp_p : T_pM \to M$ is non-singular everywhere, i.e. p has no conjugate point along any geodesic. Repeating the proof of Cartan-Hadamard theorem word by word, we can prove

Theorem 5.29. If (M, g) has a pole p, then $\exp_p : T_pM \to M$ is a smooth covering.

The Killing-Hopf Theorem

As another application we prove

Theorem 5.30 (Killing-Hopf)**.** Let (M, g) be a complete Riemannian manifold of constant sectional curvature k, then the Riemannian universal cover of (M, g) is

(a) $\left(S^m, \frac{1}{k} g_{\text{round}}\right)$ if $k > 0$,

(b) $(\mathbb{R}^m, g_0)$ if $k = 0$,

(c) $\left(H^m, -\frac{1}{k} g_{\text{hyperbolic}}\right)$ if $k < 0$.

Proof. It is enough to work on the Riemannian universal covering of (M, g) directly, i.e. prove that if (M, g) is also simply connected, then (M, g) is isometric to one of these model spaces above. It is also enough to prove the theorem for $k = 0, \pm 1$.

$\boxed{\text{Case 1: } k = -1 \text{ or } k = 0}$ Write $(\mathbb{S}^m, \overline{g}) = (H^m, g_{\text{hyperbolic}})$ for $k = -1$, and $(\mathbb{S}^m, \overline{g}) = (\mathbb{R}^m, g_0)$ for $k = 0$. Choose any point $\widetilde{p} \in \mathbb{S}^m$ and fix any linear isometry

$$L : (T_{\widetilde{p}}\mathbb{S}^m, g_{\widetilde{p}}) \to (T_pM, g_p)$$

and consider

$$F = \exp_p \circ L \circ (\exp_{\widetilde{p}})^{-1} : (\mathbb{S}^m, \overline{g}) \to (M, g).$$

By Cartan's local isometry theorem, F is a local isometry. By Cartan-Hadamard theorem, $\exp_p : T_pM \to M$ is a diffeomorphism. So F is a diffeomorphism, and thus an isometry.

$\boxed{\text{Case 2: } k = 1}$ Again we start with a point $\widetilde{p} \in S^m$ and fix any linear isometry

$$L : (T_{\widetilde{p}}S^m, g_{\widetilde{p}}) \to (T_pM, g_p).$$

Since

$$\exp_{\widetilde{p}} : B_\pi(0) \subset T_{\widetilde{p}}S^m \to S^m \setminus \{-\widetilde{p}\}$$

is a diffeomorphism, by Cartan's local isometry theorem the map

$$F_1 = \exp_p \circ L \circ (\exp_{\widetilde{p}})^{-1} : (S^m \setminus \{-\widetilde{p}\}, g_{\text{round}}) \to (M, g)$$

is a local isometry. Similarly we start with $\widetilde{q} \neq \pm\widetilde{p}$ and get a local isometry

$$F_2 = \exp_{F_1(\widetilde{q})} \circ (dF_1)_{\widetilde{q}} \circ (\exp_{\widetilde{q}})^{-1} : (S^m \setminus \{-\widetilde{q}\}, g_{\text{round}}) \to (M, g).$$

Note that by construction,

$$F_2(\widetilde{q}) = \exp_{F_1(\widetilde{q})} \circ (dF_1)_{\widetilde{q}} \circ (\exp_{\widetilde{q}})^{-1}(q) = \exp_{F_1(\widetilde{q})} \circ (dF_1)_{\widetilde{q}}(0) = \exp_{F_1(\widetilde{q})}(0) = F_1(\widetilde{q})$$

and

$$(dF_2)_{\widetilde{q}} = (d\exp_{F_1(\widetilde{q})})_0 \circ (dF_1)_{\widetilde{q}} \circ (d\exp_{\widetilde{q}})_0^{-1} = (dF_1)_{\widetilde{q}}.$$

So by Lemma 5.31 below, we have $F_1 = F_2$ on $S^m \setminus \{-\widetilde{p}, -\widetilde{q}\}$. So we may glue F_1 and F_2 to get a local isometry

$$F : (S^m, g_{\text{round}}) \to (M, g).$$

Finally by Ambrose theorem, F is a covering map and thus a diffeomorphism. So F is the desired isometry. □

It remains to prove

Lemma 5.31. Let M be connected. If $f_i : (M, g) \to (\widetilde{M}, \widetilde{g})$ $(i = 1, 2)$ are two local isometries, and if there exists $p \in M$ with

$$f_1(p_0) = f_2(p_0) \quad \text{and} \quad (\mathrm{d}f_1)_{p_0} = (\mathrm{d}f_2)_{p_0},$$

then $f_1 = f_2$.

Proof. Consider the subset

$$A = \{p \in M \mid f_1(p) = f_2(p) \text{ and } (\mathrm{d}f_1)_p = (\mathrm{d}f_2)_p\}$$

of M, and apply the standard connectedness argument.

- By assumption $p_0 \in A$, so A is non-empty.
- Obviously A is closed.
- Take any $p \in A$ and consider normal ball $B(p, r)$ so that both f_1 and f_2 are isometries on $B(p, r)$. The fact $p \in A$ implies that both f_1 and f_2 map the "radial geodesics" in $B(p, r)$ to the "radial geodesics" in $B(f(p), r)$, which implies $f_1 = f_2$ on $B(p, r)$, and as a result, $\mathrm{d}f_1 = \mathrm{d}f_2$ on $B(p, r)$. So $B(p, r) \subset A$, i.e. A is also open.

Since M is connected, we conclude $A = M$. □

Note that in the proof of Theorem 5.30, the first step is "start with any $p \in M$ and $\widetilde{p}$ in the model space" and "fix any linear isometry L". So we may fix on orthonormal basis of T_pM and an orthonormal basis at $T_{\widetilde{p}}\mathbb{S}_k$ and take L to be the linear isometry that maps the first basis to the second basis. As a result, the isometry we get has such L as its differential at p. On the other hand, according to Lemma 5.31, such an isometry is unique. So we get

Proposition 5.32. Let (M, g), $(\widetilde{M}, \widetilde{g})$ be two simply connected Riemannian manifolds of constant sectional curvature k. Then for any $p \in M$, $\widetilde{p} \in \widetilde{M}$,, any orthonormal basis $\{e_1, \ldots, e_m\}$ of T_pM, and any orthonormal basis $\{e'_1, \ldots, e'_m\}$ of $T_{\widetilde{p}}\widetilde{M}$, there is a unique isometry $\varphi : (M, g) \to (\widetilde{M}, \widetilde{g})$ such that

$$\varphi(p) = \widetilde{p} \quad \text{and} \quad (\mathrm{d}\varphi)_p(e_i) = \widetilde{e}_i \quad (1 \leqslant i \leqslant m).$$

5.4 The Index Form

5.4.1 Length Minimizing Through Index Form

Index Form as Hessian

Last times we defined the index form $I = I^\gamma$ of the geodesic $\gamma : [0, l] \to M$,

$$I(X, Y) = \int_a^b (\langle R(\dot\gamma, X)\dot\gamma, Y\rangle + \langle \nabla_{\dot\gamma} X, \nabla_{\dot\gamma} Y\rangle)\mathrm{d}t$$

defined on

$$\mathcal{V} = \mathcal{V}_\gamma = \{V \text{ is a continuous piecewise smooth vector field along } \gamma\}.$$

If $0 < t_1 < \cdots < t_k < l$ are those points where V is not smooth, then

$$I(X, Y) = \int_a^b \langle R(\dot\gamma, X)\dot\gamma - \nabla_{\dot\gamma}\nabla_{\dot\gamma} X, Y\rangle \mathrm{d}t + \langle \nabla_{\dot\gamma} X, Y\rangle\Big|_a^b - \sum_{j=1}^k \langle \nabla_{\dot\gamma}(t_j^+)X - \nabla_{\dot\gamma(t_j^-)}X, Y\rangle \tag{5.5}$$

Recall that geodesics γ with $\gamma(0) = p$ and $\gamma^l = q$ are precisely the critical points of the energy functional defined on the space $\mathcal{C}^{0,l}_{p,q}$ of "piecewise smooth curves with fixed endpoints parameterized on $[0,l]$". The index form $I = I^\gamma$ originates from the second variation, namely for a proper variation $\gamma_s(t)$ of γ,

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2}\bigg|_{s=0} E(\gamma_s) = \int_0^l (\langle R(\dot\gamma, X)\dot\gamma, X\rangle + \langle \nabla_{\dot\gamma}X, \nabla_{\dot\gamma}X\rangle)\mathrm{d}t = I(X,X),$$

where $X \in \mathcal{V}^0$ is the variation field of γ_s. So what about $I(X,Y)$? It is also second variation of the energy functional E or the length functional L, but with respect to "mixed directions": Let $\gamma_{r,s}(t)$ be a two-parameter variation of $\gamma = \gamma_{0,0}$ with fixed endpoints. Denote the variation fields corresponding to the two parameter directions by X and Y. Obviously $X,\ Y \in \mathcal{V}^0$. Then as in Exercise Sheet 3, one can prove

$$\frac{\partial^2}{\partial r \partial s}\bigg|_{r=s=0} E(\gamma_{r,s}) = I(X,Y).$$

So the index form I of γ, restricted the subspace $\mathcal{V}^0$, can be regarded as the Hessian of the energy/length functional defined on $\mathcal{C}^{0,l}_{p,q}$ at the critical point γ.

Recall from calculus: at a critical point p of a multi-variable smooth function f,

- if $\mathrm{Hess}_p(f)$ is positive definite, then p is an isolated local minimum,
- if $\mathrm{Hess}_p(f)(v,v) < 0$ for some v, then p cannot be a local minimum,
- if $\mathrm{Hess}_p(f)$ is positive semi-definite but not positive definite, we can't draw a conclusion on the behavior of f near p.

It turns out that the same phenomena happens for the energy/length functional:

Theorem 5.33. Let $\gamma : [0,l] \to M$ be a geodesic from $p = \gamma(0)$ to $q = \gamma(l)$. Then

(1) The index form I is positive definite on $\mathcal{V}^0 \iff p$ has no conjugate point along γ.

- Moreover in this case γ is an "isolated" length minimizing among nearby curves: there exists $\varepsilon > 0$ so that for any piecewise smooth curve $\overline{\gamma} : [0,l] \to M$ from p to q satisfying $\mathrm{dist}(\gamma(t), \overline{\gamma}(t)) < \varepsilon$, we have $L(\overline{\gamma}) \geqslant L(\gamma)$, with equality hold if and only if $\overline{\gamma}$ is a re-parametrization of γ.

(2) There exists $X \in \mathcal{V}^0$ with $I(X,X) < 0 \iff$ there exists $\overline{t} < l$ such that $\overline{q} = \gamma(\overline{t})$ is conjugate to p along γ.

- Moreover in this case γ is not length minimizing among nearby curves: there is a proper variation of γ so that $L(\gamma_s) < L(\gamma)$ for $0 < |s| < \varepsilon$.

(3) The index form I is positive semi-definite but not positive definite on $\mathcal{V}^0 \iff q$ is the first conjugate point of p along γ.

Remark. Note that we only claim that γ is minimizing among nearby curves. It is possible that there exists other shorter geodesics from p to q. For example, for the cylinders there is no conjugate point (since the sectional curvature is 0), but there are infinitely many geodesics between any given two points: each is minimizing among "nearby curves", but only one of them is minimizing among all curves.

As a corollary of part (1), we get the following important property:

Corollary 5.34. Suppose $p = \gamma(0)$ has no conjugate point along $\gamma : [0,l] \to M$. If V is a Jacobi field along γ, and $X \in \mathcal{V}$ satisfies $X(0) = V(0)$, $X(l) = V(l)$, then

$$I(V,V) \leqslant I(X,X),$$

with equality holds if and only if $X = V$.

Proof. Since X is a Jacobi field and $X(0) = V(0)$, $X(l) = V(l)$, by the equation (5.5),

$$I(V,V) = I(V,X).$$

It follows from part (1) of Theorem 5.33 that

$$0 \leqslant I(V-X, V-X) = I(V,V) - 2I(V,X) + I(X,X) = -I(V,V) + I(X,X),$$

with equality holds if and only if $V - X = 0$. ☐

Proof of Theorem 5.33, Part (1)

We need the following lemma which is essentially Theorem 4.19:

Lemma 5.35. Suppose $\mathcal{E}_p = \mathcal{E} \cap T_pM$ contains a line segment $[0, l]X_p$. Let $\varphi : [0, l] \to \mathcal{E}_p$ be a piecewise smooth curve with $\varphi(0) = 0$, $\varphi(l) = lX_p$. Then for

$$\gamma(t) := \exp_p(tX_p) \ (0 \leqslant t \leqslant l) \quad \text{and} \quad \overline{\gamma}(t) := \exp_p(\varphi(t)) \ (0 \leqslant t \leqslant l),$$

we have $L(\overline{\gamma}) \geqslant L(\gamma)$. Moreover, if $\exp_p$ is non-singular along $[0, l]X_p$, then the equality $L(\overline{\gamma}) = L(\gamma)$ holds if and only if $\overline{\gamma}$ is a re-parametrization of γ.

Proof. Without loss of generality, we may assume $\varphi(t) \neq 0$ for all $t \in (0, l)$. Write $\varphi(t) = r(t)e(t)$, where $r(t) = |\varphi(t)|$ and $e(t) \in S_pM$ has unit length. Then

$$\dot{\varphi}(t) = \dot{r}(t)e(t) + r(t)\dot{e}(t),$$

and $e(t) \perp \dot{e}(t)$. According to Gauss lemma,

$$|\dot{\overline{\gamma}}(t)| = |(\mathrm{d}\exp_p)_{\varphi(t)}\dot{\varphi}(t)| \geqslant |(\mathrm{d}\exp_p)_{\varphi(t)}(\dot{r}(t)e(t))| = |\dot{r}(t)|.$$

Therefore,

$$L(\overline{\gamma}) \geqslant \int_0^l |\dot{r}(t)|\mathrm{d}t \geqslant |r(l) - r(0)| = ||\varphi(l)| - |\varphi(0)|| = l|X_p| = L(\gamma).$$

If $\exp_p$ is non-singular along $[0, l]X_p$, then for ε small enough $\exp_p$ is non-singular in the ε-neighborhood of $[0, l]X_p$. Now suppose the equality holds, then we have

$$(\mathrm{d}\exp_p)_{\varphi(t)}\dot{e}(t) = 0,$$

thus $\dot{e}(t) = 0$ for all t. It follows that $e(t)$ is a constant unit vector, i.e. $\varphi(t) = r(t)e$ for some $e \in S_pM$. Obviously $e = X_p/|X_p|$ is the direction vector of X_p. Moreover, $\dot{r}$ cannot change sign. So $\overline{\gamma} = \exp_p\left(\frac{r(t)}{|X_p|}X_p\right)$ is a re-parametrization of γ. □

Remark. There is no conflict with Theorem 5.33 (2). Suppose γ contains a conjugate point of p. Since the exponential map $\exp_p$ is not even a local diffeomorphism at a conjugate point of p, it is possible that a curve that is close to γ cannot be realized as the image of a curve in T_pM with the same endpoints under the exponential map.

Now we prove the "moreover" and $\Leftarrow$ part in (1) of Theorem 5.33.

[(1) Moreover] Suppose p has no conjugate point along γ. Find a subdivision

$$0 = t_0 < t_1 < \cdots < t_k < t_{k+1} = l$$

and open neighborhoods V_i, $1 \leqslant i \leqslant k$, of the line segment $[t_i, t_{i+1}]\dot{\gamma}(0)$ in T_pM so that $\exp_p$ is a diffeomorphism on each V_i. Denote $U_i = \exp_p(V_i)$. According to our assumption on $\overline{\gamma}$, for ε small enough, $\overline{\gamma}([t_i, t_{i+1}]) \subset U_i$. Now define

$$\varphi(t) = (\exp_p|_{V_i})^{-1}(\overline{\gamma}(t)), \quad t_{i-1} \leqslant t \leqslant t_i.$$

Then $\varphi(t)$ is a piecewise smooth curve in T_pM connecting 0 to $l\dot{\gamma}(0)$ with $\exp_p(\varphi(t)) = \overline{\gamma}(t)$. So the conclusion follows from Lemma 5.35.

[(1) $\Longleftarrow$] According to the part [(1) Moreover] that we just proved, if $p = \gamma(0)$ has no conjugate point along γ, then for any $X \in \mathcal{V}^0$, $I(X, X) \geqslant 0$. (Otherwise one can construct a variation with $L(\gamma_s) < L(\gamma)$.) If I is not positive definite on $\mathcal{V}^0$, then $I(Y, Y) = 0$ for some $Y \in \mathcal{V}^0$. It follows that for any $Z \in \mathcal{V}^0$ and any $\lambda \in \mathbb{R}$,

$$0 \leqslant I(Y - \lambda Z, Y - \lambda Z) = -2\lambda I(Y, Z) + \lambda^2 I(Z, Z).$$

As a consequence, $I(Y, Z) = 0$ for any $Z \in \mathcal{V}^0$. In other words, Y is a Jacobi field. Since $q = \gamma(l)$ is not a conjugate point of p, and $Y(0) = 0$, $Y(l) = 0$, we must have $Y \equiv 0$. So I is positive definition on $\mathcal{V}^0$.

Proof of Theorem 5.33, Part (2) and (3)

Lemma 5.36. Suppose $q = \gamma(t_0)$ is NOT conjugate to $p = \gamma(0)$ along a geodesic $\gamma : [0, l] \to M$. Then for any $X_p \in T_pM$ and $X_q \in T_qM$, there exists a unique Jacobi field V along γ so that $V(0) = X_p$ and $V(t_0) = X_p$.

Proof. Let $\mathcal{J}_\gamma$ be the space of all Jacobi fields along γ. Define a mapping

$$\begin{aligned} \Theta : \mathcal{J}_\gamma &\to T_pM \times T_qM, \\ V &\mapsto \Theta(V) = (V(0), V(t_0)). \end{aligned}$$

Since q is not a conjugate point of p, Θ is injective. But Θ is linear, and $\dim \mathcal{J}_\gamma = \dim(T_pM \times T_qM) = 2m$ are of same dimension, so Θ is an linear isomorphism. ☐

Now we prove part (2) and (3) of Theorem 5.33.

$\boxed{\text{(2) Moreover}}$ and $\boxed{(2) \Longleftarrow}$ Let X be a nonzero Jacobi field along γ with $X(0) = 0$, $X(\bar{t}) = 0$. Note that $\nabla_{\dot\gamma(\bar t)}X \neq 0$, otherwise X will be identically zero. Let Z be a smooth vector field along γ with

$$Z(0) = 0,\ Z(l) = 0,\ Z(\bar t) = -\nabla_{\dot\gamma(\bar t)}X.$$

Denote $\gamma_1 = \gamma|_{[0,\bar t]}$, $\gamma_2 = \gamma|_{[\bar t, l]}$, $Z_1 = Z|_{[0,\bar t]}$ and $Z_2 = Z|_{[\bar t, l]}$. For any $\eta > 0$ we put

$$Y_\eta(t) := \begin{cases} Y_\eta^1 = X(t) + \eta Z_1(t), & \text{for } 0 \leqslant t \leqslant \bar t, \\ Y_\eta^2 = \eta Z_2(t), & \text{for } \bar t \leqslant t \leqslant l. \end{cases}$$

Then

$$I^{\gamma_1}(Y_\eta^1, Y_\eta^1) = -\eta |\nabla_{\dot\gamma(\bar t)}X|^2 + \eta^2 I^{\gamma_1}(Z_1, Z_1)$$

and thus

$$I^{\gamma}(Y_\eta, Y_\eta) = -\eta |\nabla_{\dot\gamma(\bar t)}X|^2 + \eta^2 I^{\gamma_1}(Z_1, Z_1) + \eta^2 I^{\gamma_2}(Z_2, Z_2) < 0$$

for η small enough. This proves the theorem.

$\boxed{(3) \Longleftarrow}$ Fix any $c \in (0, l)$. For any $X \in \mathcal{V}^0$, we may write $X = X^i(t)e_i(t)$, where $\{e_i\}$ are orthonormal and parallel along γ with $e_1 = \dot\gamma$. Let

$$X^c(t) = X^i\left(\frac{lt}{c}\right) e_i(t), \quad 0 \leqslant t \leqslant c.$$

Then X^c is a vector field along $\gamma^c := \gamma|_{[0,c]}$ with $X^c(0) = 0$, $X^c(c) = 0$. Since p has no conjugate point along $\gamma|_{[0,c]}$, we get from (1) that $I(X^c, X^c) \geqslant 0$. It follows that $I(X, X) = \lim_{c \to a} I(X^c, X^c) \geqslant 0$. So I is positively semi-definite on $\mathcal{V}^0$.

Obviously I is not positively definite on $\mathcal{V}^0$, since if X is a nonzero Jacobi field along γ with $X(a) = 0$, $X(b) = 0$, then we have $I(X, X) = 0$.

$\boxed{(1) \Longrightarrow}$ This follows from (2) ($\Longleftarrow$) and (3) ($\Longleftarrow$).

$\boxed{(2) \Longrightarrow}$ This follows from (1) ($\Longleftarrow$) and (3) ($\Longleftarrow$).

$\boxed{(3) \Longrightarrow}$ This follows from (1) ($\Longleftarrow$) and (2) ($\Longleftarrow$).

5.4.2 Morse Index Theorem

Morse Index of a Geodesic

Let's continue our "finite dimension manifold" v.s. "infinite dimension space $\mathcal{C}_{p,q}^{0,l}$" analogue a bit further. For a finite dimensional manifold M, there is a remarkable theory, known as Morse theory, that relates the topology of M to the behavior of critical points of a Morse function [the existence is guaranteed by Sard's theorem]:

Suppose $f \in \mathcal{C}^\infty(M)$ is a Morse function, i.e. all critical points of f are non-degenerate [i.e. the Hessian $\mathrm{Hess}_p(f)$ at any critical point p of f is non-singular], then M has the homotopy type of a CW-complex whose λ-cells are in one-to-one correspondence with critical points of index λ of

f [the index of a critical point p is the number of negative eigenvalues of $\mathrm{Hess}_p(f)$]. For example, if a compact manifold M admits a Morse function with only two critical points, then they have to be maximum and minimum, so their indexes have to be m and 0. In this case one can prove that M is homeomorphic to S^m.
What is the analogue in our setting? We already have

- critical points $\leftrightsquigarrow$ geodesics $\gamma : [0, l] \to M$.
- Hessian at the critical point $\leftrightsquigarrow$ the index form I of γ defined on $\mathcal{V}^0$.

So "γ is a non-degenerate critical point" should be translated to "the index form I of γ defined on $\mathcal{V}^0$ has trivial null space". On the other hand, we have seen last time $I(V, W) = 0$ for all $W \in \mathcal{V}^0$ if and only if V is Jacobi field. So

$$\begin{aligned} &\text{the index form } I \text{ of } \gamma \text{ defined on } \mathcal{V}^0 \text{ has trivial null space} \\ \Longleftrightarrow\ &\text{there is no Jacobi field } V \text{ along } \gamma \text{ with } V(0) = V(l) = 0 \\ \Longleftrightarrow\ &q = \gamma(l) \text{ is not a conjugate point of } p = \gamma(0) \text{ along } \gamma. \end{aligned}$$

So we get

- non-degenerate critical point $\leftrightsquigarrow$ geodesics $\gamma : [0, l] \to M$ so that $q = \gamma(l)$ is not a conjugate point of $p = \gamma(0)$ along γ,
- the index of a non-degenerate critical point $\leftrightsquigarrow$ the dimension of the maximal subspace of $\mathcal{V}^0$ on which I is negative definite.

By the explanations above, we are naturally led to study

Definition 5.37. Let (M, g) be a Riemannian manifold, and $\gamma : [0, l] \to M$ a geodesic. We will call

$$\mathrm{ind}(\gamma) = \max \dim\{\mathcal{A} \subset \mathcal{V}^0 \mid I|_{\mathcal{A}} \text{ is negatively definite}\}$$

the **index** of γ, and call

$$N(\gamma) = \dim\{X \in \mathcal{V}^0 \mid I(X, X) = 0 \text{ for all } Y \in \mathcal{V}^0\}$$

the **nullity** of γ.

Remark. The geometric meanings of $N(\gamma)$ and $\mathrm{ind}(\gamma)$ are clear:

(1) The nullity $N(\gamma)$ equals the multiplicity of $q = \gamma(l)$ as a conjugate point of $p = \gamma(0)$. In particular, $N(\gamma) = 0$ if $q = \gamma(l)$ is not conjugate to p along γ.

(2) The index $\mathrm{ind}(\gamma)$ is "the number of independent directions" towards which γ can be deformed to shorter curves with the same endpoints.

Morse Index Theorem

It turns out that there is a well-developed infinite dimensional Morse theory:
As in the finite dimensional case one can show that there exists p, q so that all geodesics connecting p and q are non-degenerate in the above sense. Again $\mathcal{C}_{pq} = \mathcal{C}_{p,q}^{0,l}$ has the homotopy type of a CW-complex whose λ-cells are in one-to-one correspondence with geodesics of index λ. This also gives the topological information of M, since $\mathcal{C}_{pq}$ is homotopy equivalent to $\mathcal{C}_{pp}$, and $\pi_{k+1}(M) \cong \pi_k(\mathcal{C}_{pp})$. As an application we can explain Serre's result we mentioned in the remark on Page 90: By using algebraic topology, Serre proved $H_i(\mathcal{C}_{pq}, \mathbb{R}) \neq 0$ for infinitely many i, which implies that for any k, there exists a geodesic whose index is great than k. By Morse index theorem below, there must be infinitely many geodesics connecting p and q.
So it is important to study the index of a geodesic. In what follows we will denote by γ^t the geodesic $\gamma|_{[0,t]}$, so that the corresponding index and nullity are $\mathrm{ind}(\gamma^t)$ and $N(\gamma^t)$. The following theorem is fundamental in the direction:

Theorem 5.38 (Morse Index Theorem)**.** For any geodesic $\gamma : [0,l] \to M$, we have

$$\mathrm{ind}(\gamma) = \sum_{0<t<l} N(\gamma^t) < \infty.$$

As a consequence, we get a quantitative version of Theorem 5.33 (2) "moreover",

Corollary 5.39. For any geodesic $\gamma : [0,l] \to M$, $\gamma(0)$ has only finitely many conjugate points along γ. If we denote these conjugate points (except possibly $\gamma(b)$) by $\gamma(t_1), \dots, \gamma(t_k)$ $(a < t_1 < \cdots < t_k < b)$, then

$$\mathrm{ind}(\gamma) = \sum_{j=1}^{k} n_{\gamma,0}(t_j).$$

Proof of Morse Index Theorem

To prove the theorem, the first step is to reduce the infinite dimension space $\mathcal{V}^0$ to a finite dimensional subspace $\mathcal{T}_1$, so that the index and nullity of $I|_{\mathcal{V}^0}$ is the same as $I|_{\mathcal{T}_1}$. As a consequence, $\mathrm{ind}(\gamma)$ is finite.

To construct $\mathcal{T}_1$, let's recall that near any point one can find a strongly convex neighborhood U, so that any two points in U can be connected by a unique minimal geodesic which is contained in U. Now we use finitely many such strongly convex neighborhoods $U_0, U_1, \dots, U_k$ to cover γ, and take $0 = t_0 < t_1 < \cdots < t_k < t_{k+1} = l$ so that $\gamma([t_j, t_{j+1}]) \subset U_j$. Define

$$\mathcal{T}_1 = \{X \in \mathcal{V}^0 \mid X \text{ is a Jacobi field along each } \gamma|_{[t_j,t_{j+1}]}\}.$$

Note that $\gamma|_{[t_j,t_{j+1}]}$ contains no pair of conjugate points since $\exp_p$ is non-singular on U_j for all $q \in U_j$. By Lemma 5.36, if $X \in \mathcal{T}_1$, then $X(t_j)$ and $X(t_{j+1})$ determine $X|_{[t_j,t_{j+1}]}$. As a consequence, the map

$$\Phi : \mathcal{T}_1 \to T_{\gamma(t_1)}M \oplus \cdots \oplus T_{\gamma(t_k)}M,$$
$$\Phi(X) = (X(t_1), \dots, X(t_k))$$

is a linear isomorphism. In particular,

$$\dim \mathcal{T}_1 = mk > \infty.$$

Next define

$$\mathcal{T}_2 = \{X \in \mathcal{V}^0 \mid X(t_1) = 0, \dots, X(t_k) = 0\}.$$

Then

- Obviously $\mathcal{T}_1 \cap \mathcal{T}_2 = \{0\}$. On the other hand, for any $X \in \mathcal{V}^0$, if we let $V = \Phi^{-1}(X_1(t_1), \dots, X(t_k)) \in \mathcal{T}_1$, then $X - V \in \mathcal{T}_2$. So $\mathcal{V}^0 = \mathcal{T}_1 \oplus \mathcal{T}_2$.
- By formula (5.5), for any $X \in \mathcal{T}_1$, $Y \in \mathcal{T}_2$ we have $I(X,Y) = 0$.
- By Corollary 5.34, for any $Y \in \mathcal{T}_2$ we have $I(Y,Y) = \sum I^{\gamma_i}(Y|_{\gamma_i}, Y|_{\gamma_i}) > 0$. Thus $I|_{\mathcal{T}_2}$ is positive definite.

As a result, we get

$$\mathrm{ind}(\gamma) = \max \dim\{\mathcal{A} \subset \mathcal{T}_1 \mid I|_{\mathcal{A}} \text{ is negatively definite}\} \tag{5.6}$$

and thus $\mathrm{ind}(\gamma) \leqslant \dim \mathcal{T}_1 < +\infty$.

To get the precise formula of $\mathrm{ind}(\gamma)$ in the theorem, we need the following lemmas which reveals the continuity property of the index $\mathrm{ind}(\gamma^t)$:

Lemma 5.40. $\mathrm{ind}(\gamma^t)$ is non-decreasing in t.

Proof. Let $0 < t < s$. For any $X \in \mathcal{V}^0_{\gamma^t}$, we can extend X to $X' \in \mathcal{V}^0_{\gamma^s}$ by zero extension. Since $I^{\gamma^t}(X,X) = I^{\gamma^s}(X',X')$, the conclusion follows. □

Lemma 5.41. $\operatorname{ind}(\gamma^t) = \operatorname{ind}(\gamma^{t-\varepsilon})$ for ε small enough.

Proof. Suppose $t \in (t_j, t_{j+1})$, where t_j is a subdivision of $(0, l)$ as above. Then by (5.6),

$$\operatorname{ind}(\gamma^t) = \{\mathcal{A} \subset \mathcal{T}_1^{\gamma^t} \mid I|_{\mathcal{A}} \text{ is negatively definite}\},$$

where $\mathcal{T}_1^{\gamma^t}$ is the set of piecewise smooth vector fields in $\mathcal{V}^0_{\gamma^t}$ which are Jacobi fields on each (t_i, t_{i+1}). Under the isomorphism $\Phi^t : \mathcal{T}_1^{\gamma^t} \to T_{\gamma(t_1)}M \oplus \cdots \oplus T_{\gamma(t_k)}M$ the restriction of I on $\mathcal{T}_1^{\gamma^t}$ becomes a bilinear form I^t defined on $T_{\gamma(t_1)}M \oplus \cdots \oplus T_{\gamma(t_k)}M$. It depends smoothly on t since for $\vec{X}$, $\vec{Y} \in T_{\gamma(t_1)}M \oplus \cdots \oplus T_{\gamma(t_k)}M$, by (5.5),

$$I^t(\vec{X}, \vec{Y}) = \sum_{i<j} I^{\gamma_i}(X^i, Y^j) + I^{\widetilde{\gamma}_t}(X^t, Y^t) = \langle \nabla_{\dot{\gamma}(t_j^+)} X^t, Y \rangle + \text{terms independent of } t,$$

where $\gamma^i = \gamma|_{[t_i, t_{i+1}]}$ and $X^i = (\Phi^t)^{-1}(\vec{X})|_{[t_i, t_{i+1}]}$ for $i < j$, while $\widetilde{\gamma}^t = \gamma|_{[t_j, t]}$ and $X^t = (\Phi^t)^{-1}(\vec{X})|_{[t_j, t]}$. So if I^t is negative definite on a subspace, so is $I^{t-\varepsilon}$ for ε small, i.e. $\operatorname{ind}(\gamma^{t-\varepsilon}) \geqslant \operatorname{ind}(\gamma^t)$. The conclusion follows from Lemma 5.40. □

Lemma 5.42. $\operatorname{ind}(\gamma^{t+\varepsilon}) = \operatorname{ind}(\gamma^t) + N(\gamma^t)$ for ε small enough.

Proof. As in the proof of Lemma 5.41, we identify the restriction of I on $\mathcal{T}_1^{\gamma^t}$ with the bilinear form I^t defined on $T_{\gamma(t_1)}M \oplus \cdots \oplus T_{\gamma(t_k)}M$ which depends smoothly on t. Then I^t is positive definite on a subspace of dimension $mk - \operatorname{ind}(\gamma^t) - N(\gamma^t)$. It follows that $I^{t+\varepsilon}$ is positive definite on this subspace for ε small enough. This implies

$$\operatorname{ind}(\gamma^{t+\varepsilon}) \leqslant mk - (mk - \operatorname{ind}(\gamma^t) - N(\gamma^t)) = \operatorname{ind}(\gamma^t) + N(\gamma^t).$$

On the other hand, we have

- By the same continuity arugument, $I^{t+\varepsilon}$ is negative definite on the space where I^t is negative definite.
- $I^{t+\varepsilon}$ is negative definite on the null space of I^t: If $0 \neq \vec{X}$ is in the null space of I^t, then $(\Phi^t)^{-1}(\vec{X})$ is a Jacobi field along γ^t and $X_{t_j} \neq 0$ (since there is no Jacobi field that vanishes at both t_j and t). So if we denote by $\widetilde{X}^t$ the zero extension of X^t to $[t_j, t+\varepsilon]$, then by Corollary 5.34,
$$I^{\widetilde{\gamma}_{t+\varepsilon}}(X^{t+\varepsilon}, X^{t+\varepsilon}) < I^{\widetilde{\gamma}_{t+\varepsilon}}(\widetilde{X}^t, \widetilde{X}^t) = I^{\widetilde{\gamma}_t}(X^t, X^t)$$
and thus $I^{t+\varepsilon}(\vec{X}, \vec{X}) < I^t(\vec{X}, \vec{X}) = 0$.

So we get $\operatorname{ind}(\gamma^{t+\varepsilon}) \geqslant \operatorname{ind}(\gamma^t) + N(\gamma^t)$ which finishes the proof. □

Proof of Morse Index Theorem. According to the previous lemmas, we have

- $\operatorname{ind}(\gamma^t)$ is a non-decreasing and left-continuous **step** function.
- For t small, $\operatorname{ind}(\gamma^t) = 0$ (because γ^t is a minimal geodesic for t small).
- If $\gamma(t)$ is not a conjugate point of $\gamma(0)$, then $\operatorname{ind}(\gamma^\tau)$ is constant for $\tau \in (t-\varepsilon, t+\varepsilon)$.
- If $\gamma(t)$ is a conjugate point of $\gamma(0)$, then $\operatorname{ind}(\gamma^\tau)$ is a step function in $(t-\varepsilon, t+\varepsilon)$ with a "jump from right" of size $N(\gamma^t)$ at $\tau = t$.

The theorem follows from these facts. □

5.5 Cut Locus

5.5.1 The Conjugate and Cut Locus

The Conjugate Locus

Let (M, g) be a complete Riemannian manifold.

Definition 5.43. For any $p \in M$, we call

$$\widetilde{\mathrm{Con}}(p) = \{v \in T_pM \mid (\mathrm{d}\exp_p)_v \text{ is singular}\}$$

the **conjugate locus** of p in T_pM, and call

$$\mathrm{Con}(p) = \{q \in M \mid q \text{ is a conjugate point of } p \text{ along some geodesic}\}$$

the **conjugate locus** of p in M.

Obviously $\widetilde{\mathrm{Con}}(p)$ is a closed subset in T_xM. We first prove

Lemma 5.44. The conjugate locus $\widetilde{\mathrm{Con}}(p)$ and $\mathrm{Con}(p)$ are measure zero sets in T_pM and M respectively.

Proof. By Morse index theorem, for each $v \in S_pM$, the set $\{tv \mid t > 0\} \cap \widetilde{\mathrm{Con}}(p)$ is discrete. So the conclusion follows. □

As we have seen last time, the first conjugate point in each direction is important: a geodesic γ starting from p is locally length minimizing if the endpoint is before the first conjugate point, and is not locally length minimizing if the endpoint is after the first conjugate point. We define $k : S_pM \to \mathbb{R} \cup \{+\infty\}$ to be the function

$$k(v) = \inf\{t > 0 \mid tv \in \widetilde{\mathrm{Con}}(p)\},$$

and we set $k(v) = +\infty$ if there is no conjugate point in the direction v.

Proposition 5.45. The function k is continuous.

Proof. Suppose $v_i \in S_pM$ and $v_i \to v$. We want to prove $k(v_i) \to k(v)$.

We need an observation:

Continuity Observation: Suppose $v_i \in S_pM$ and $v_i \to v$. Fix l and denote

$$\gamma_i = \exp(tv_i) \quad \text{and} \quad \gamma(t) = \exp(tv) \quad (0 \leqslant t \leqslant l),$$

By the smooth dependence of geodesics on initial values, γ_i converges uniformly to γ. By using parallel transport [first from $\gamma_i(t)$ to $\gamma_i(0) = p$ along γ_i, then from $p = \gamma(0)$ to $\gamma(t)$ along γ] one can identify each $T_{\gamma(t)}M$. As a result, we get an identification $\Psi_i : \mathcal{V}^0_{\gamma_i} \to \mathcal{V}^0_\gamma$. Furthermore, for any $X \in \mathcal{V}^0_\gamma$, again by smooth dependence, $I^{\gamma_i}(\Psi_i^{-1}(X), \Psi_i^{-1}(X)) \to I^\gamma(X, X)$.

Case 1: $k(v) = c < +\infty$ We want to prove $k(v_i) \to k(v) = c$.

- Take $l = c + \varepsilon$. Then $\mathrm{ind}(\gamma) \geqslant 1$, i.e. $I^\gamma(X, X) > 0$ for some $X \in \mathcal{V}^0_\gamma$. By continuity observation above, $I^{\gamma_i}(\Psi_i^{-1}(X), \Psi_i^{-1}(X)) < 0$ for i large enough. So $\mathrm{ind}(\gamma_i) \geqslant 1$ and thus $k(v_i) \leqslant l + \varepsilon$ for i large enough.
- Take $l = c - \varepsilon$. Since $\mathcal{V}^0_\gamma$ is infinite dimensional, we can't conclude that I^{γ_i} is positive definite directly from the fact I^γ is positive definite. However, from the proof of Morse index theorem, the maximal negative definite space of I^γ can be taken to be a subspace in

  $$\mathcal{T}_1^\gamma = \{X \in \mathcal{V}^0_\gamma \mid X \text{ is Jacobi on each } [t_i, t_{i+1}]\},$$

  or any other direct sum complement of

  $$\mathcal{T}_2^\gamma = \{V \in \mathcal{V}^0_\gamma \mid X(t_1) = \cdots = X(t_k) = 0\},$$

where $0 = t_0 < t_1 < \cdots < t_k < t_{k+1} = l$ are chosen so that $\gamma([t_j, t_{j+1}]) \subset U_j$, and $U_0, U_1, \ldots, U_k$ are strongly convex open subsets that cover γ. Since γ_i converges to γ uniformly, we have $\gamma_i([t_j, t_{j+1}]) \subset U_j$ for i large enough. We may use the same partition for all γ_i. Now we have $\Psi_i(\mathcal{T}_2^{\gamma_i}) = \mathcal{T}_2^{\gamma}$, and thus

$$\mathcal{V}_{\gamma_i}^0 = \Psi_i^{-1}(\mathcal{V}_\gamma^0) = \Psi^{-1}(\mathcal{T}_1^\gamma \oplus \mathcal{T}_2^\gamma) = (\Psi_i)^{-1}(\mathcal{T}_1^\gamma) \oplus \mathcal{T}_2^{\gamma_i},$$

so $\mathrm{ind}(\gamma_i)$ also equals the maximal dimension of subspace in $(\Psi_i)^{-1}(\mathcal{T}_1^\gamma)$ on which I^{γ_i} is negative definite. Now suppose $I^{\gamma_i}(X_i, X_i) < 0$ for $\Psi_i(X_i) \in \mathcal{T}_1^\gamma$ which can be taken so that $\sup |X_i| = 1$, then we may take a convergent subsequence of $\Psi_i(X_i) \in \mathcal{T}_1^\gamma$ and conclude the existence of $X \neq 0$ with $I^\gamma(X, X) \leqslant 0$, which contradicts with the fact I^γ is positive definite. So we get $k(v_i) \geqslant l - \varepsilon$ for i large enough.

Case 2: $k(v) = +\infty$ Suppose to the contrary that $k(v_i)$ has a bounded subsequence. Without less of generality, suppose $k(v_i) \leqslant c$ for all i. Take $l = c$. By the same argument above we get a contradiction. □

As a consequence, we get

Corollary 5.46. The set of first conjugate points of p in T_pM is closed.

Proof. If t_iv_i are first conjugate points of p and $t_iv_i \to v$, then

$$k(v/|v|) = \lim_{i\to\infty} k(v_i) = \lim_{i\to\infty} t_i = |v|.$$

So v is the first conjugate point of p in the direction $v/|v|$. □

The Cut Locus

Let γ be the normal geodesic in (M, g) with $\gamma(0) = p$ and $\dot\gamma(0) = v$. Suppose (M, g) is complete so that γ can be defined on $\mathbb{R}$. Let's concentrate for $t > 0$, which corresponds to the part of the geodesic in the direction v. For t small $\gamma|_{[0,t]}$ is length minimizing between $\gamma(0)$ and $\gamma(t)$. For general t, it may happen that either $\gamma|_{[0,t]}$ is length minimizing between $\gamma(0)$ and $\gamma(t)$ for all $t > 0$, or there exists t_0 such that $\gamma|_{[0,t]}$ is no longer length minimizing between $\gamma(0)$ and $\gamma(t)$ for all $t > t_0$.

Definition 5.47. Let (M, g) be a complete Riemannian manifold, $p \in M$ a point, and $\gamma : [0, \infty) \to M$ a normal geodesic with $\gamma(0) = p$. If

$$t_0 := \sup\{t \mid \gamma([0, t]) \text{ is a minimizing geodesic}\} < +\infty$$

then we will call $\gamma(t_0)$ the **cut point** of p along γ.

- The **cut locus** of p in M is defined to be the set $\mathrm{Cut}(p)$ of all cut points of p along all geodesics that start from p.
- The **cut locus** of p in T_pM is defined to be the set $\widetilde{\mathrm{Cut}(p)}$ of all vectors $v \in T_pM$ so that $\exp_p(v)$ is a cut point.

Remark. If M is compact, then $\mathrm{Cut}(p) \neq \varnothing$ for all p.

Example. On $\mathbb{R}^m$ and $\mathbb{H}^m$ (endowed with the canonical metrics), there exists only one normal minimizing geodesic joining any two given points. So $\mathrm{Cut}(p) = \varnothing$ for all p.

Example. For S^m with the round metric, $\mathrm{Cut}(p) = \{\overline{p}\}$ for any $p \in M$, where $\overline{p} = -p$ is the antipodal point of p. Note that $\overline{p}$ is also the first conjugate point of p.

Example. For the cylinder $S^1 \times \mathbb{R}$ endowed with the canonical metric, if $p = (\mathrm{e}^{\mathrm{i}\theta_0}, z_0)$, then $\mathrm{Cut}(p) = \{(\mathrm{e}^{\mathrm{i}(\theta_0+\pi)}, z) \mid z \in \mathbb{R}\}$ is the vertical line "opposite to p". Note that p has no conjugate points at all.

By definition we have

Lemma 5.48. For any $q \notin \mathrm{Cut}(p)$, there exists a unique minimizing geodesic joining p to q.

Proof. If there exist two minimizing geodesics γ, σ of length l joining p to q, then γ is minimizing between p and q, and is no longer minimizing after q:

The curve $\overline{\gamma}$ defined by connecting σ with $\gamma|_{[l,l+\varepsilon]}$ is a piecewise smooth but not smooth curve connecting p to $\gamma(l+\varepsilon)$ whose length is $l+\varepsilon$. But according to the first variation formula, any piecewise smooth but not smooth curve is not a minimizing curve [c.f. Corollary 4.48]. We conclude that $\gamma|_{[0,l+\varepsilon]}$ is also not a minimizing curve, since it has the same length as $\overline{\gamma}$.
So $q \in \mathrm{Cut}(p)$. □

Cut Points v.s. First Conjugate Points

The following theorem relates cut points with first conjugate points:

Theorem 5.49. Suppose $\gamma(t_0)$ is the cut point of $p = \gamma(0)$ along a normal geodesic γ, then at least one of the following assertion holds:

(1) $\gamma(t_0)$ is the first conjugate point of p along γ.

(2) $\gamma(t_0)$ is the first point along γ so that there exists another normal geodesic $\sigma \neq \gamma$ from p to $\gamma(t_0)$ with length $L(\sigma) = t_0 = L(\gamma|_{[0,t_0]})$.

Proof. Take a decreasing sequence $t_i \to t_0^+$. Let σ_i be a normal minimizing geodesic connecting p to $\gamma(t_i)$. Then by definition of cut point, $s_i := L(\sigma_i) < t_i$. Note that $\{\dot{\sigma}_i(0)\}$ is sequence in the unit sphere S_pM. By passing to a subsequence, we may assume $\dot{\sigma}_i(0) \to X_p \in S_pM$. Let σ be the normal geodesic with $\sigma(0) = p$, $\dot{\sigma}(0) = X_p$. Then by continuity, σ is a minimizing geodesic connecting p to $\gamma(t_0)$, thus $L(\sigma) = t_0$.
Case 1: $X_p = \dot{\gamma}(0)$. Since $s_i < t_i$, we have $s_i\dot{\sigma}_i(0) \neq t_i\dot{\gamma}(0)$. But

$$\exp_p(s_i\dot{\sigma}(0)) = \sigma_i(s_i) = \gamma(t_i) = \exp_p(t_i\dot{\gamma}(0)),$$

so $\exp_p$ is not a local diffeomorphism near $t_0\dot{\gamma}(0)$. So $\gamma(t_0)$ is a conjugate point of p. Obviously it is the first conjugate point, otherwise $\gamma([0,t_0])$ is not minimizing.
Case 2: $X_p \neq \dot{\gamma}(0)$. Then σ is a geodesic that is different from γ. We have

$$t_0 = L(\gamma|_{[0,t_0]}) \leqslant L(\sigma) = \lim_{i\to\infty} L(\sigma_i) \leqslant \lim_{i\to\infty} t_i = t_0.$$

So $L(\sigma) = t_0$. To show that $\gamma(t_0)$ is the first point along γ with this property, we argue by contradiction. If there exists a $\overline{t} < t_0$ and a normal geodesic $\overline{\sigma}$ connecting p to $\gamma(\overline{t})$ so that $L(\overline{\sigma}) = \overline{t}$, then by the argument in the proof of Lemma 5.48, $\gamma|_{[0,t_0]}$ is not a minimizing curve. This contradicts with the definition of cut point. □

Corollary 5.50. If $q \in \mathrm{Cut}(p)$, then $p \in \mathrm{Cut}(q)$.

Proof. If q is the cut point of p along γ, then γ is minimizing between p and q. It follows that the "opposite geodesic" $-\gamma$ is also minimizing between q and p. Moreover, by the theorem above, either q is the first conjugate point of p along γ, or there exists a different normal geodesic σ joint p to q which has length $L(\sigma) = \mathrm{dist}(p,q)$. In both cases $-\gamma$ is no longer minimizing after p. So $p \in \mathrm{Cut}(q)$. □

Remark. One can show that the function $f : SM \to \mathbb{R}\cup\{\infty\}$ defined by

$$f(p,X_p) = \begin{cases} t_0, & \text{if } \gamma_{p,X_p}(t_0) \text{ is the cut point of } p \text{ along } \gamma, \\ +\infty, & \text{if } p \text{ has no cut point along } \gamma_{p,X_p}. \end{cases}$$

is a continuous function. It follows that $\mathrm{Cut}(p)$ is a closed subset in M. It has measure zero since there is at most one cut point in each direction in M. Note that by definition $f \leqslant k$.

5.5.2 The Distance Function

Smoothness of Distance Function

Now let's fix $p \in M$ and consider the distance function

$$d_p : M \to \mathbb{R},$$
$$d_p(q) = \mathrm{dist}(p, q).$$

As we have already seen, d_p is a continuous function. However, it is not hard to see that $d_p \notin \mathcal{C}^\infty(M)$. In fact, d_p is never smooth at p.

Example. Consider (S^2, g_{S^2}). Let $\overline{p} = -p$ be the antipodal point of p. Then for q near $\overline{p}$, $d_p(q) = \pi - d_{\overline{p}}(q)$. It follows that d_p is also not smooth at $\overline{p}$.

Theorem 5.51. The function d_p is smooth on $M \setminus \mathrm{Cut}(p) \cup \{p\}$. Moreover, for each $q \in M \setminus \mathrm{Cut}(p) \cup \{p\}$, if we let γ^q be the unique normal minimizing geodesic from p to q, then the gradient of d_p at q is

$$(\nabla d_p)(q) = \dot{\gamma}^q(d_p(q)).$$

Proof. For each $q \in M \setminus \mathrm{Cut}(p) \cup \{p\}$, let γ^q be the unique normal minimizing geodesic from p to q and denote $X^q = \dot{\gamma}^q(0) \in S_pM$. Let

$$A = \{L(\gamma^q)X^q \mid q \in M \setminus \mathrm{Cut}(p) \cup \{p\}\}.$$

Then $A \subset T_pM \setminus \{0\}$ is an open set and

$$\exp_p : A \to M \setminus \mathrm{Cut}(p) \cup \{p\}$$

is smooth. Moreover, at each vector in A, $\exp_p$ is non-singular and thus a local diffeomorphism. Since $\exp_p$ is globally one-to-one on A, it is a diffeomorphism from A to $M \setminus \mathrm{Cut}(p) \cup \{p\}$. It follows that

$$\exp_p^{-1} : M \setminus \mathrm{Cut}(p) \cup \{p\} \to A \subset T_pM \setminus \{0\}$$

is smooth. Thus $d_p(q) = |\exp_p^{-1}(q)|$ is smooth on $M \setminus \mathrm{Cut}(p) \cup \{p\}$.

To calculate its gradient at q, we choose any $X_q \in T_qM$ and let $\sigma(s)$ be a smooth curve in $M \setminus \mathrm{Cut}(p) \cup \{p\}$ tangent to X_q at $q = \sigma(0)$. Now we consider the variation of γ^q so that γ^q_s be the unique minimizing geodesic from p to $\sigma(s)$. Observe that the variation field vector [which is a Jacobi field] of this variation at the point q is exactly X_q. So according to the first variation formula,

$$X_q(d_p) = \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} d_p(\sigma(s)) = \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} L(\gamma^q_s) = \langle X_q, \dot{\gamma}^q(d_p(q))\rangle.$$

It follows that $(\nabla d_p)(q) = \dot{\gamma}^q(d_p(q))$. □

Remark. One can show that if there exists two minimizing geodesic from p to q, then d_p is not differentiable at q.

Hessian of the Distance Function

By using the second variation formula one can calculate the Hessian of d_p on $M \setminus \mathrm{Cut}(p) \cup \{p\}$. Recall that the Hessian of a smooth function f is

$$\begin{aligned}(\nabla^2 f)_q(X_q, Y_q) = (X_q Y f - \nabla_{X_q} Y)f &= \nabla_{X_q}(\langle \nabla f, Y\rangle) - \langle \nabla f, \nabla_{X_q} Y\rangle \\ &= \langle \nabla_{X_q} \nabla f, Y\rangle.\end{aligned}$$

Now let $\gamma_s : [0, l] \to M$ be geodesic variation of γ by minimizing geodesics with $\gamma_s(0) = p$ [so its variation field X is a normal Jacobi field along γ with $X(0) = 0$]. Then

$$(\nabla^2 d_p)_q(X_q, Y_q) = \langle (\nabla_X \nabla d_p)_q, Y_q\rangle = \langle \nabla_{X(q)} \dot{\gamma}^q, Y_q\rangle = \langle \nabla_{\dot{\gamma}^q(l)} X, Y_q\rangle, \quad \forall Y_q \in T_qM.$$

So we proved

Proposition 5.52. Suppose $q \notin \mathrm{Cut}(p) \cup \{p\}$. Let $\gamma : [0, l] \to M$ be the unique length minimizing normal geodesic connecting p to q, and let X be a normal Jacobi field along γ with $X(0) = 0$. Denote $X_q = X(l)$. Then for any $Y_q \in T_qM$,

$$(\nabla^2 d_p)_q(X_q, Y_q) = \langle \nabla_{\dot\gamma(l)} X, Y_q \rangle.$$

Here are two special cases that will be quite useful later:

Corollary 5.53. Suppose $q \notin \mathrm{Cut}(p) \cup \{p\}$, and $\gamma : [0, l] \to M$ the unique length minimizing normal geodesic connecting p to q.

(1) For any $Y_q \in T_qM$, $(\nabla^2 d_p)_q(\dot\gamma(l), Y_q) = \langle \nabla_{\dot\gamma(l)} \dot\gamma, Y_q \rangle = 0$.

(2) For any normal Jacobi field X along γ with $X(0) = 0$,

$$(\nabla^2 d_p)_q(X_q, X_q) = \langle \nabla_{\dot\gamma(q)} X, X_q \rangle = I(X, X).$$

Note that the "singularity" of d_p at the point p is not too bad: one can always remove the singularity at p by considering the function d_p^2 instead. So it is reasonable to study $(\nabla^2 d_p^2)_p$. In general, for any smooth function $f \in \mathcal{C}^\infty(M)$ and $X_p \in T_pM$, if we let γ be the geodesic $\gamma(t) = \exp_p(tX_p)$, then "the second order derivative of f along γ" is

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} f \circ \gamma(t) = \frac{\mathrm{d}}{\mathrm{d}t}\frac{\mathrm{d}}{\mathrm{d}t}(f \circ \gamma) = \frac{\mathrm{d}}{\mathrm{d}t}\langle \nabla f, \dot\gamma \rangle = \langle \nabla_{\dot\gamma} \nabla f, \dot\gamma \rangle = (\nabla^2 f)_{\gamma(t)}(\dot\gamma, \dot\gamma).$$

So we get

Lemma 5.54. $(f \circ \gamma)''(t) = (\nabla^2 f)_{\gamma(t)}(\dot\gamma, \dot\gamma)$.

On the other hand, if X_p is a normal vector, i.e. γ is a normal geodesic, then for t small enough we have $d_p^2(\gamma(t)) = t^2$. So we get, for $X_p \in S_pM$,

$$(\nabla^2 d_p^2)_p(X_p, X_p) = 2g(X_p, X_p)$$

and thus by polarization, we get

Proposition 5.55. The Hessian of d_p^2 at p is

$$(\nabla^2 d_p^2)_p = 2g.$$

Chapter 6

Comparison Theorems

6.1 Theorems on Curvature v.s. Topology

6.1.1 Complete Riemannian Manifolds with Non-positive Curvature

A Hessian Comparison for d_p^2

Let (M, g) be a Cartan-Hadamard manifold, i.e. a complete simply-connected Riemannian manifold with non-positive curvature. We have seen that there is no conjugate point for such manifolds, and by Cartan-Hadamard theorem, $\exp_p : T_pM \to M$ is diffeomorphism. In particular, between any pair of points there is a unique geodesic [which has to be minimizing], and there is no cut point for Cartan-Hadamard manifolds.

We first prove that d_p^2 is strictly convex on Cartan-Hadamard manifolds:

Proposition 6.1. For any p in a Cartan-Hadamard manifold (M, g), $\nabla^2 d_p^2 \geqslant 2g$. Moreover, if the sectional curvature is negative, then $\nabla^2 d_p^2 > 2g$ at any $q \neq p$.

Proof. Let γ be a normal geodesic with $\gamma(0) = p$. For any $v \in T_qM$, where $q = \gamma(l)$,

$$(\nabla^2 d_p^2)_q(\dot\gamma(l), v) = \langle \nabla_{\dot\gamma(l)} \nabla d_p^2, v\rangle = \langle \nabla_{\dot\gamma(l)}(2t\dot\gamma), v\rangle = 2\langle \dot\gamma(l), v\rangle.$$

It remains to prove that for $v \perp \dot\gamma(l)$, one has $(\nabla^2 d_p^2)_q(v, v) \geqslant 2\langle v, v\rangle$.

Let V be the normal Jacobi field along γ with $V(0) = 0$, $V(l) = v$. Then

$$(\nabla^2 d_p^2)_{\gamma(t)}(V, V) = \langle \nabla_{V(t)} \nabla d_p^2, V\rangle = \langle \nabla_{V(t)}(2t\dot\gamma), V\rangle = 2t\langle \nabla_{V(t)}(\dot\gamma), V\rangle = 2t\langle \nabla_{\dot\gamma(t)} V, V\rangle.$$

Denote $f(t) = \dfrac{\langle \nabla_{\dot\gamma(t)} V, V(t)\rangle}{\langle V(t), V(t)\rangle}$. Then $\lim\limits_{t\to 0^+} f(t) = +\infty$, since $V(t) = t\nabla_{\dot\gamma(0)} V + \mathcal{O}(t^2)$. By definition we have

$$f'(t) = \frac{\langle \nabla_{\dot\gamma(t)} \nabla_{\dot\gamma(t)} V, V\rangle + \langle \nabla_{\dot\gamma(t)} V, \nabla_{\dot\gamma(t)} V\rangle}{\langle V, V\rangle} - 2\frac{\langle \nabla_{\dot\gamma(t)} V, V\rangle^2}{\langle V, V\rangle^2} \geqslant -f^2(t),$$

where we used the fact $\langle \nabla_{\dot\gamma(t)} \nabla_{\dot\gamma(t)} V, V\rangle = \langle R(\dot\gamma, V)\dot\gamma, V\rangle = -K(\dot\gamma, V)|\dot\gamma \wedge V|^2 \geqslant 0$ and Cauchy-Schwartz inequality $\langle \nabla_{\dot\gamma(t)} V, \nabla_{\dot\gamma(t)} V\rangle\langle V, V\rangle \geqslant \langle \nabla_{\dot\gamma(t)} V, V\rangle^2$. It follows

$$t \geqslant \int_0^t \frac{-f'(\tau)}{f^2(\tau)} \mathrm{d}\tau = \int_0^t \left(\frac{1}{f(\tau)}\right)' \mathrm{d}\tau = \frac{1}{f(t)} - \lim_{\tau\to 0^+} \frac{1}{f(\tau)} = \frac{1}{f(t)},$$

So we get $tf(t) \geqslant 1$ for all t, and the first conclusion follows.

For the second conclusion, one just notice $f'(t) > -f^2(t)$ for $t > 0$. $\square$

The Fundamental Group of Riemannian Manifolds with $K \leqslant 0$

As a first application, we prove

Theorem 6.2 (Cartan)**.** Let (M, g) be a Cartan-Hadamard manifold, and $\varphi : M \to M$ an isometry with $\varphi^n = \mathrm{Id}$ for some n. Then φ admits a fixed point.

Proof. Fix $p \in M$ and consider the function

$$\begin{aligned} g : M &\to \mathbb{R}, \\ q &\mapsto g(q) = d^2(q, p) + d^2(q, \varphi(p)) + \cdots + d^2(q, \varphi^{n-1}(p)). \end{aligned}$$

Then g is strictly convex and $g(q) \to +\infty$ as $d(q, p) \to +\infty$. So g admits a unique minimum at some point $\widetilde{p}$. Since $g(\varphi(q)) = g(q)$ we conclude $\varphi(\widetilde{p}) = \widetilde{p}$. □

As a corollary we get

Corollary 6.3. Let (M, g) be a complete Riemannian manifold with non-positive sectional curvature. Then $\pi_1(M)$ is torsion free [i.e. no nontrivial finite order element].

Proof. If $\pi_1(M)$ admits a finite order element τ, then the corresponding Deck transformation $f_\tau : \widetilde{M} \to \widetilde{M}$ is of finite order, and thus by Cartan's theorem above, f_τ admits a fixed point. This implies $f_\tau = \mathrm{Id}$ and $\tau = e$. □

As an immediate consequence, we see

Corollary 6.4. For any compact manifold M, $\mathbb{RP}^m \times M$ admits no metric of non-positive sectional curvature.

A Weak Cosine Law for Cartan-Hadamard Manifolds

As a second application of the convexity of d_p^2, we prove the following weak cosine law for Cartan-Hadamard manifolds.

Proposition 6.5. Let (M, g) be Cartan-Hadamard manifold. Consider the geodesic triangle with vertices $p_1, p_2, p_3 \in M$. Let a, b, c be the lengths of sides and A, B, C be the corresponding opposite angles. Then

(1) $a^2 + b^2 - 2ab\cos C \leqslant c^2$.

(2) $A + B + C \leqslant \pi$.

Further more, if the sectional curvature is negative, then these inequalities are strict.

Proof. Let γ be a normal geodesic from p_3 to p_1, and let $f(t) = d^2(p_2, \gamma(t))$. Then

$$f(0) = d^2(p_2, p_3) = a^2$$

and

$$f'(0) = 2d(p_2, p_3)\langle \nabla d_{p_2}, \dot{\gamma}(0)\rangle = -2a\cos C.$$

By Lemma 5.54, $f''(\tau) = (\nabla^2 d_p^2)_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))$. By Proposition 6.1, $f''(\tau) \geqslant 2$ for all τ. Thus we get

$$c^2 = f(b) \geqslant f(0) + f'(0)b + b^2 = a^2 + b^2 - 2ab\cos C.$$

To prove (2), one may compare the triangle in the plane with sides a, b, c [which satisfies the triangle inequality since they are distances of three points in a Riemannian manifold]. Denote the angles by A', B', C'. Then by the cosine law in $\mathbb{R}^2$ we get

$$A \leqslant A', \; B \leqslant B', \; C \leqslant C',$$

which implies $A + B + C \leqslant \pi$.

Finally if the sectional curvature is negative, then by Proposition 6.1, $f''(\tau) > 2$ for $\tau \neq 0$ and the conclusion follows. □

Preissman's Theorem

What if M is not simply connected? We have just seen that if M admits a non-positive sectional curvature metric, then $\pi_1(M)$ is torsion free. It turns out that if M admits a negative sectional curvature metric, then any nontrivial abelian subgroup of $\pi_1(M)$ is an infinite cyclic group generated by one element:

Theorem 6.6 (Preissman)**.** Let (M, g) be a compact Riemannian manifold with negative sectional curvature, and let $\{1\} \neq H \subset \pi_1(M)$ be a nontrivial abelian subgroup of the fundamental group. Then $H \cong \mathbb{Z}$.

Remark. The theorem was strengthened by Byers to: under the same assumption, any nontrivial solvable subgroup of $\pi_1(M)$ is infinite cyclic.

Example. For any closed surface M_g of genus $g \leqslant 2$, there is Riemannian metric of constant negative sectional curvature. The fundamental group of M_g is

$$\langle a_1, b_1, \ldots, a_g, b_g \mid a_1 b_1 a_1^{-1} b_1^{-1} \cdots a_1 b_1 a_1^{-1} b_1^{-1} = e \rangle,$$

Which is not abelian, but all its abelian subgroups are isomorphic to $\mathbb{Z}$.

As an immediate consequence, we see

Corollary 6.7. Suppose $m \geqslant 2$. For any compact manifold M, $\mathbb{T}^m \times M$ admits no metric of negative sectional curvature.

Remark. It was first proved by Gao and Yau in 1986 that any compact manifold of dimension 3 admits a metric with negative Ricci curvature. Then in 1994, Lohkamp proved that any manifold of dimension at least 3 admits a complete Riemannian metric of negative Ricci curvature. So there is no topological constraint for a manifold of dimension $\geqslant 3$ to admit Riemannian metrics with negative Ricci curvature.

The idea of proof is as follows: Realize Deck transformations associated with all $\alpha \in H$ as "a discrete family of translations along a fixed geodesic". As a result, a nontrivial discrete subgroups of H corresponds to a nontrivial subgroup of $\mathbb{R}$, which is isomorphic to $\mathbb{Z}$.

Translations in Cartan-Hadamard Manifolds

So we need to introduce the concept of translation.

Definition 6.8. Let (M, g) be a complete simply-connected Riemannian manifold, and $\gamma : \mathbb{R} \to M$ a geodesic. An isometry $f : (M, g) \to (M, g)$ is called a **translation** along γ if f has no fixed point, and $f(\mathrm{im}(\gamma)) = \mathrm{im}(\gamma)$.

Let (M, g) be any complete Riemannian manifold and $\pi : \widetilde{M} \to M$ be the universal covering. We endow with $\widetilde{M}$ the pull back metric $\widetilde{g} = \pi^* g$. Recall that for each element $\alpha \in \pi_1(M)$, one can define a deck transformation $f_\alpha : \widetilde{M} \to \widetilde{M}$: for each $\widetilde{p} \in \widetilde{M}$, there is a loop γ based at $p = \pi(\widetilde{p})$ whose homotopy class is α. Let $\widetilde{\gamma}$ be the lift of γ with starting point $\widetilde{p}$. Define $f_\alpha(\widetilde{p})$ be the endpoint of $\widetilde{\gamma}$.
One can prove that

- f_α is well-defined,
- f_α is an isometry,
- $f_\beta \circ f_\alpha = f_{\beta\alpha}$ for all α, $\beta \in \pi_1 M$,
- f_α has no fixed point if $\alpha \neq e$.

Now suppose $e \neq \alpha \in \pi_1(M)$, and let γ be a minimizing closed geodesic in the homotopy class α. Let $\widetilde{\gamma}$ be a lift of γ to $\widetilde{M}$. Then by definition $\widetilde{\gamma}$ is a geodesic in $(\widetilde{M}, \widetilde{g})$, and $f_\alpha(\mathrm{im}(\widetilde{\gamma})) = \mathrm{im}(\widetilde{\gamma})$. So we get

Lemma 6.9. $f_\alpha : (\widetilde{M}, \widetilde{g}) \to (\widetilde{M}, \widetilde{g})$ is a translation along $\widetilde{\gamma}$ for any $e \neq \alpha \in \pi_1(M)$.

As a consequence of this lemma, we prove

Corollary 6.10. Suppose (M, g) has negative sectional curvature, then any translation $f : \widetilde{M} \to \widetilde{M}$ fixed only one geodesic[1].

Proof. Suppose there are two geodesics $\widetilde{\gamma}_1$ and $\widetilde{\gamma}_2$ in $\widetilde{M}$ such that $f(\widetilde{\gamma}_i) = \widetilde{\gamma}_i$. First we claim that $\widetilde{\gamma}_1 \cap \widetilde{\gamma}_2 = \varnothing$. Otherwise either $\widetilde{\gamma}_1 \cap \widetilde{\gamma}_2 = \{\widetilde{p}\}$ for some $\widetilde{p}$, which implies $f(\widetilde{p}) = \widetilde{p}$ which is a contradiction (since f has no fixed point), or there are at least two points in $\widetilde{\gamma}_1 \cap \widetilde{\gamma}_2$, which contradicts with Cartan-Hadamard theorem.

Now choose $\widetilde{p}_i \in \widetilde{\gamma}_i$, and let $\widetilde{\gamma}_3$ and $\widetilde{\gamma}_5$ be the minimizing geodesic connecting $\widetilde{p}_1$, $\widetilde{p}_2$ and connecting $f(\widetilde{p}_1)$, $\widetilde{p}_2$ respectively. Let $\widetilde{\gamma}_4 = f(\widetilde{\gamma}_3)$ be the minimizing geodesic connecting $f(\widetilde{p}_1)$, $\widetilde{p}_2$ respectively. Let $\widetilde{\gamma}_4 = f(\widetilde{\gamma}_3)$ be the minimizing geodesic connecting $f(\widetilde{p}_1)$, $f(\widetilde{p}_2)$. Since f is an isometry, the "corresponding angles" at $\widetilde{p}_1$ and at $\widetilde{p}_2$ are the same, similarly for $\widetilde{p}_2$. As a result, the angles of the two geodesic triangle $\widetilde{p}_1\widetilde{p}_2 f(\widetilde{p}_1)$ and $\widetilde{p}_2 f(\widetilde{p}_1) f(\widetilde{p}_2)$ add up to at least 2π, which contradicts with Proposition 6.5. □

Proof of Preissman's Theorem

As above we denote by $\widetilde{M}$ the universal covering of M, and f_α the deck transformation described above associated to $\alpha \in \pi_1(M)$.

First fix $\alpha \in H$ and let $\widetilde{\gamma}$ be the geodesic that is invariant under f_α. Then for any $\beta \in H$, one has $f_{\beta\alpha} = f_{\alpha\beta}$ since H is abelian. So

$$f_\beta(\mathrm{im}(\widetilde{\gamma})) = f_\beta(f_\alpha(\mathrm{im}(\widetilde{\gamma}))) = f_\alpha(f_\beta(\mathrm{im}(\widetilde{\gamma}))).$$

By the Corollary 6.10, one must have

$$f_\beta(\mathrm{im}(\widetilde{\gamma})) = \mathrm{im}(\widetilde{\gamma}), \quad \forall \beta \in H.$$

As a consequence, $\widetilde{\gamma}$ is invariant under all f_α's for $\alpha \in H$.

Now we denote $\widetilde{p}_0 = \widetilde{\gamma}(0)$. Since $\widetilde{\gamma}$ is invariant under f_β, for each $\beta \in H$, there is a unique $t_\beta \in \mathbb{R}$ so that

$$\widetilde{\gamma}(t_\beta) = f_\beta(\widetilde{p}_0).$$

Note that this implies

$$\widetilde{\gamma}(t_\beta + t) = f_\beta(\widetilde{\gamma}(t))$$

for any t, since as t varies, both sides are geodesics with the same initial condition.

Now we define a map

$$\varphi : H \to \mathbb{R},$$
$$\varphi(\beta) = t_\beta.$$

Claim 6.11. φ is a group homomorphism: For any β_1, $\beta_2 \in H$,

$$\widetilde{\gamma}(t_{\beta_1} + t_{\beta_2}) = f_{\beta_1} \circ f_{\beta_2}(\widetilde{p}_0) = f_{\beta_1\beta_2}(\widetilde{p}_0) = \widetilde{\gamma}_{\beta_1\beta_2}.$$

So we have $\varphi(\beta_1\beta_2) = t_{\beta_1\beta_2} = t_{\beta_1} + t_{\beta_2}$.

Claim 6.12. φ is injective: Suppose $\varphi(\beta) = 0$, then $\widetilde{p}_0 = \widetilde{\gamma}(0) = f_\beta(\widetilde{p}_0)$. So $\beta = e \in \pi_1(M)$.

Claim 6.13. The image of φ is not dense in $\mathbb{R}$. Pick a strongly convex geodesic ball $U = B_r(p)$ centered at $p = \pi(\widetilde{p}_0)$ so that $\pi^{-1}(U) = \bigcup_\delta U_\delta$, where each U_δ is diffeomorphic to U under the covering map $\pi : \widetilde{M} \to M$ and are disjoint. Denote U_0 be the one so that $\widetilde{p}_0 \in U_0$. Then for each $\beta \neq e$, $f_\beta(\widetilde{p}_0) \notin U_0$. So

$$|t_\beta| = d(\widetilde{p}_0, f_\beta(\widetilde{p}_0)) \geqslant r$$

for any $\beta \neq e$.

As a consequence of the first two claims, H is an additive subgroup of $\mathbb{R}$. But we know that any additive subgroup of $\mathbb{R}$ is either dense or infinite cyclic. So the theorem is proved.

[1]Here different parametrizations will be viewed as the same

6.1.2 Complete Riemannian Manifolds with Positive Curvature

Now let's turn to Riemannian manifolds with positive curvature.

Synge's Theorem

Another application of the second variation formula to Riemannian manifolds with positive curvature is

Theorem 6.14 (Synge)**.** Let (M, g) be a compact Riemannian manifold with positive sectional curvature.

(1) If M is even dimensional and orientable, then M is simply connected.

(2) If M is odd dimensional, then M is orientable.

Since $\mathbb{RP}^m$ admits a positive sectional curvature metric, given the fact "$\pi_1(\mathbb{RP}^m) \cong \mathbb{Z}_2$ for $m \geqslant 2$", we conclude that "$\mathbb{RP}^m$ is orientable if and only if m is odd".

Corollary 6.15. If (M, g) is a compact even dimensional Riemannian manifold of positive sectional curvature, and M is not orientable, then $\pi_1(M) \cong \mathbb{Z}_2$.

Proof. Let $\overline{M}$ be the orientable double covering of M, endowed with the induced pull-back metric. Then $\overline{M}$ is orientable and satisfies all the conditions in Synge theorem. It follows that $\overline{M}$ is simply connected and thus $\pi_1(M) \cong \mathbb{Z}_2$. □

As a consequence, $\mathbb{RP}^2 \times \mathbb{RP}^2$ admits no metric of postive sectional curvature since $\pi_1(\mathbb{RP}^2 \times \mathbb{RP}^2) \cong (\mathbb{Z}_2)^2$. We remark that it is still unknown whether $S^2 \times S^2$ admits a positive sectional curvature metric: this is the well-known Hopf conjecture.

Remark. In the odd dimensional case we cannot say too much of its fundamental group. For example, for each k there is a lens space $S^3/\mathbb{Z}_k$ that has constant sectional curvature 1 and fundamental group $\mathbb{Z}_k$.

Proof of Synge's Theorem

We first prove

Lemma 6.16. Let (M, g) be an orientable Riemannian manifold, and $\gamma : [a, b] \to M$ be a smooth map loop, i.e. $\gamma(a) = \gamma(b) := p$. Then the parallel transport $P^\gamma_{a,b} : T_pM \to T_pM$ has determinant 1.

Proof. In Section 2.2 we have seen $P^\gamma_{a,b} \in O(T_pM)$. It remains to show $\det P^\gamma_{a,b} > 0$. For this purpose we take a positive m-form ω on M, and let $\{e_i\}$ be a positive basis of T_pM, i.e. $\omega(e_1, \ldots, e_m) > 0$. Let $e_j(t) = P^\gamma_{a,t}(e_j)$ be the parallel transport of $\{e_i\}$ along γ. Then

$$\omega(e_1(t), \ldots, e_m(t)) \neq 0$$

for all t. It follows that $\omega(e_1(t), \ldots, e_m(t)) \neq 0$. But

$$\omega(e_1(b), \ldots, e_m(b)) = (\det P^\gamma_{a,b})\omega(e_1, \ldots, e_m),$$

so we must have $\det P^\gamma_{a,b} > 0$. □

Proof of Synge's Theorem. (1) Suppose M is not simply connected. Then there exists a nontrivial closed geodesic $\gamma : [0, 1] \to M$ which is length minimizing in its free homotopy class. Since the parallel transport $P^\gamma_{0,1} \in SO(T_pM)$ and satisfies

$$P^\gamma_{0,1}(\dot\gamma(0)) = \dot\gamma(0),$$

we can find $X_p \in E_p = (\dot\gamma(0))^\perp$ such that

$$P^\gamma_{0,1}(X_p) = X_p.$$

(Here, we used the condition that $\dim M$ is even, so that $\dim E$ is odd!)

Now let $X(t)$ be the parallel vector field along γ with $X(0) = X_p$. Then

$$X(1) = P^\gamma_{0,1}(X_p) = X_p.$$

Thus for the variation γ_s of γ whose variation field is X, we have

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2}\bigg|_{s=0} E(\gamma_s) = \int_0^1 \langle R(\dot\gamma, X)\dot\gamma - \nabla_{\dot\gamma}\nabla_{\dot\gamma}X, X\rangle \mathrm{d}t = \int_0^1 R(\dot\gamma, X, \dot\gamma, X)\mathrm{d}t < 0.$$

This contradicts with the fact γ is minimizing in its homotopy class.

(2) Suppose M is not orientable, then there is a smooth loop $\gamma : [0,1] \to M$ and a frame $\{e_1(t), \dots, e_m(t)\}$ so that $e_1(1) \wedge \cdots \wedge e_m(1) = -e_1(0) \wedge \cdots \wedge e_m(0)$. On the other hand, we write $\widetilde{e}_i(t) = P^\gamma_{0,t}(e_i(0))$, then there is a nonzero function $f(t)$ so that $\widetilde{e}_1(t) \wedge \cdots \wedge \widetilde{e}_m(t) = f(t)e_1(t) \wedge \cdots \wedge e_m(t)$. So we must have $\det(P^\gamma_{0,1}) = -1$ since m is odd. By continuity, any smooth loop in the same homotopy class satisfies $\det(P^\gamma_{0,1}) = -1$. Let γ be the minimizing geodesic in this class. Since $P^\gamma_{0,1}(\dot\gamma(0)) = \dot\gamma(0)$, we see

$$\det P^\gamma_{0,1}\bigg|_E = -1,$$

where $E = (\dot\gamma)^\perp$ is the orthogonal complement of $\dot\gamma(0)$ in T_pM. Since E is even dimensional, complex eigenvalues appear in conjugate pairs and -1 appear by odd times. So again we conclude that there exists $X_p \in E$ so that $P^\gamma_{0,1}(X_p) = X_p$. Now repeat the variation argument above we conclude that γ is not minimizing in its homotopy class, a contradiction. □

Bonnet-Myers Theorem

What about Ricci curvature?

Theorem 6.17 (Bonnet-Myers)**.** Let (M, g) be a complete connected Riemannian manifold. Suppose there is a constant $\kappa > 0$ such that

$$Ric(X_p) \geqslant (m-1)\kappa, \quad \forall X_p \in SM.$$

Then M is compact, and its diameter is bounded by

$$\mathrm{diam}(M) := \sup_{p,q\in M} \mathrm{dist}(p,q) \leqslant \frac{\pi}{\sqrt{\kappa}}.$$

Remark. (1) One cannot weaken the condition on Ricci curvature to $Ric > 0$ or even $K > 0$. For example, consider the paraboloid

$$\{(x,y,z) \in \mathbb{R}^3 \mid z = x^2 + y^2\}.$$

It is a surface of revolution with $K > 0$, which is not compact.

(2) The estimate is optimal in the following sense: Let M be the standard sphere of radius $\frac{1}{\sqrt{\kappa}}$, then it has Ricci curvature $(m-1)\kappa$ and diameter $\frac{\pi}{\sqrt{\kappa}}$. (Note: the diameter here is not the standard diameter as a subset in $\mathbb{R}^m$.)

(3) We will prove the following result of S. Y. Cheng later: If (M, g) satisfies the conditions of the Bonnet-Myers theorem and $\mathrm{diam}(M) = \frac{\pi}{\sqrt{\kappa}}$, then (M, g) is isometric to the standard sphere of radius $\frac{1}{\sqrt{\kappa}}$.

Corollary 6.18. Let (M, g) be a complete Riemannian manifold whose Ricci curvature is bounded below by a positive number. Then $\pi_1(M)$ is finite.

Proof. Let $\widetilde{M}$ be the universal covering of M, endowed with the pullback metric $\widetilde{g} = \pi^* g$. Then $(\widetilde{M}, \widetilde{g})$ is also a complete Riemannian manifold whose Ricci curvature is bounded below by a positive number. By Bonnet-Myers theorem, $\widetilde{M}$ is compact. As a consequence, $\pi : \widetilde{M} \to M$ has to be a finite covering. So $\pi_1(M)$ is finite. □

In particular, we see that if M, N are compact, $\pi_1(M)$ is infinite, then $M \times N$ admits no Riemannian metric of positive Ricci curvature.

Proof of Bonnet-Myers Theorem

Proof. For any $p, q \in M$, let $\gamma : [0,1] \to M$ be a minimizing geodesic joining p to q. It's enough to show $L(\gamma) \leqslant \frac{\pi}{\sqrt{\kappa}}$ (which implies compactness of M by Hopf-Rinow theorem).

By contradiction, suppose that $L(\gamma) = l > \frac{\pi}{\sqrt{\kappa}}$. Let $\{e_i(t)\}$ be parallel vector fields along γ which form an orthonormal basis at each point $\gamma(t)$ and so that $e_1 = \frac{\dot{\gamma}(t)}{l}$. For $j = 2, \ldots, m$, we define

$$V_j(t) = \sin(\pi t) e_j(t).$$

Then $V_j(0) = V_j(1) = 0$, and $\nabla_{\dot{\gamma}} \nabla_{\dot{\gamma}} V_j = -\pi^2 \sin(\pi t) e_j(t)$. Thus

$$I(V_j, V_j) = \int_0^1 \langle R(\dot{\gamma}, V_j)\dot{\gamma} - \nabla_{\dot{\gamma}} \nabla_{\dot{\gamma}} V_j, V_j \rangle \mathrm{d}t = \int_0^1 \sin^2(\pi t)(\pi^2 + l^2 R(e_1, e_j, e_1, e_j))\mathrm{d}t.$$

Summing over $2 \leqslant j \leqslant m-1$, we get

$$\sum_{j=2}^m I(V_j, V_j) = \int_0^1 \sin^2(\pi t)((m-1)\pi^2 - l^2 Ric(e_1))\mathrm{d}t < 0.$$

So there exists some $j \geqslant 2$ so that $I(V_j, V_j) < 0$. It follows that there exists $\overline{q} = \gamma(t_0)$ with $0 < t_0 < 1$ which is conjugate to p along γ. In particular, γ is not length minimizing. A contradiction. □

6.2 Rauch Comparison Theorem

Now we begin to study the so-called comparison theorems. As we have seen last time, a comparison on curvature tensor will induce a comparison on Jacobi fields, which would further give a comparison of geometry (triangles for the Cartan-Hadamard manifolds) or analysis (Hessian of d_p^2 for the Cartan-Hadamard manifolds), or restrict the possible behavior of geodesics (as in the proof of Synge's theorem and Bonnet-Myers theorem). In all these cases we finally arrive at some restrictions on global geometry/topology of the manifold. In the next couple sections we develop such ideas more systematically.

6.2.1 The Index Comparison

Basic Index Comparison Lemma

In the proof of Synge's Theorem and Bonnet-Myers Theorem, we used parallel vector fields to construct variations of a given geodesic. The advantage of parallel vector fields is that inner products (and thus lengths, angles) are preserved along geodesic. Another class of vector fields that are widely used in constructing variations are Jacobi fields, which are variation fields of geodesic variations. Usually one start with two geodesics on two manifolds whose curvatures are pointwise comparable, then compare two Jacobi fields with same initial value on these geodesics.

So our basic setting for comparison is the following:

- Let (M, g) and $(\widetilde{M}, \widetilde{g})$ be Riemannian manifolds of dimension m.
- Let $\gamma : [0, a] \to M$ and $\widetilde{\gamma} : [0, a] \to \widetilde{M}$ be normal geodesics with $\gamma(0) = p$ and $\widetilde{\gamma}(0) = \widetilde{p}$.
- For each $t \in [0, a]$, let

$$K^-(t) = \min\{K(\Pi_{\gamma(t)}) \mid \Pi_{\gamma(t)} \subset T_{\gamma(t)}M,\ \dim \Pi_{\gamma(t)} = 2 \text{ and } \dot{\gamma}(t) \in \Pi_{\gamma(t)}\},$$
$$\widetilde{K}^+(t) = \max\{\widetilde{K}(\widetilde{\Pi}_{\widetilde{\gamma}(t)}) \mid \widetilde{\Pi}_{\gamma(t)} \subset T_{\widetilde{\gamma}(t)}\widetilde{M},\ \dim \widetilde{\Pi}_{\widetilde{\gamma}} = 2 \text{ and } \dot{\widetilde{\gamma}}(t) \in \widetilde{\Pi}_{\widetilde{\gamma}(t)}\}.$$

- We say two vectors $X_c \in T_{\gamma(c)}M$ and $\widetilde{X}_c \in T_{\widetilde{\gamma}(c)}\widetilde{M}$ are roughly the same if

$$|X_c| = |\widetilde{X}_c| \quad \text{and} \quad \langle X_c, \dot{\gamma}(c) \rangle = \langle \widetilde{X}_c, \dot{\widetilde{\gamma}}(c) \rangle.$$

Now let X, $\widetilde{X}$ be Jacobi fields along γ and $\widetilde{\gamma}$ respectively, with $X(0) = 0$ and $\widetilde{X}(0) = 0$. To compare X and $\widetilde{X}$, we usually assume either

$$\text{“}X(a) \text{ and } \widetilde{X}(a) \text{ are roughly the same”},$$

or

$$\text{“}\nabla_{\dot{\gamma}(0)}X \text{ and } \widetilde{\nabla}_{\dot{\widetilde{\gamma}}(0)}\widetilde{X} \text{ are roughly the same”}.$$

The following lemma is quite obvious whose proof is left as an exercise:

Lemma 6.19. Let X, $\widetilde{X}$ be Jacobi fields along γ and $\widetilde{\gamma}$ with $X(0) = \widetilde{X}(0) = 0$, and write

$$X = c\dot{\gamma} + dt\dot{\gamma} + X^{\perp}, \quad \widetilde{X} = \widetilde{c}\dot{\widetilde{\gamma}} + \widetilde{d}t\dot{\widetilde{\gamma}} + \widetilde{X}^{\perp}.$$

Suppose either "$X(a)$ and $\widetilde{X}(a)$ are roughly the same", or "$\nabla_{\dot{\gamma}(0)}X$ and $\widetilde{\nabla}_{\dot{\widetilde{\gamma}}}\widetilde{X}$ are roughly the same", then $c = \widetilde{c}$ and $d = \widetilde{d}$.

As a result, to compare two Jacobi fields whose initial or boundary values are roughly the same, it is enough to compare their "normal components".

Now we prove the basic index comparison theorem:

Theorem 6.20. Let X, $\widetilde{X}$ be Jacobi fields along γ, $\widetilde{\gamma}$ such that $X(0) = 0$, $\widetilde{X}(0) = 0$, and suppose $X(a)$ and $\widetilde{X}(a)$ are roughly the same. Assume further that

(1) γ has no conjugate points on $[0, a]$.

(2) $\widetilde{K}^{+}(t) \leqslant K^{-}(t)$ holds for all $t \in [0, a]$.

then

$$I(X, X) \leqslant I(\widetilde{X}, \widetilde{X}).$$

Moreover, if $K^{+}(t) < K^{-}(t)$ for some $t < a$, then $I(X, X) < I(\widetilde{X}, \widetilde{X})$.

Proof. By Lemma 6.19 we may assume X, $\widetilde{X}$ are normal. Let $\{e_1(t), \dots, e_m(t)\}$ and $\{\widetilde{e}_1(t), \dots, \widetilde{e}_m(t)\}$ be orthonormal frames that are parallel along γ and $\widetilde{\gamma}$, such that

$$e_1(t) = \dot{\gamma}(t), \ \widetilde{e}_1(t) = \dot{\widetilde{\gamma}}(t), \text{ and } e_2(a) = X(a)/\alpha, \ \widetilde{e}_2(a) = \widetilde{X}(a)/\alpha,$$

where $\alpha = |X(a)| = |\widetilde{X}(a)| \neq 0$ since γ has no conjugate point. If we denote

$$X(t) = X^i(t)e_i(t), \quad \widetilde{X}(t) = \widetilde{X}^i(t)\widetilde{e}_i(t)$$

respectively, then obviously we have

- $X^i(0) = \widetilde{X}^i(0) = 0$ for all i,
- $X^2(a) = \widetilde{X}^2(a) = \alpha$ and $X^i(a) = \widetilde{X}^i(a) = 0$ for all $i \neq 2$,
- $X^1(t) = \widetilde{X}^1(t) = 0$ for all $t \in [0, a]$ (since both X and $\widetilde{X}$ are normal).

As in the proof of Bonnet-Myers theorem we transplant $\widetilde{M}$ to γ by defining

$$Y(t) = \widetilde{X}^i(t)e_i(t).$$

Then $Y(0) = 0$, $Y(a) = X(a)$. Since X is a Jacobi field,

$$I(X, X) \leqslant I(Y, Y).$$

On the other hand,

$$\begin{aligned}
I(Y,Y) &= \int_0^a (|\nabla_{\dot\gamma} Y|^2 + \langle R(\dot\gamma, Y)\dot\gamma, Y\rangle)\mathrm{d}t \\
&= \int_0^a \left(\sum(\dot{\widetilde{X}}^i(t))^2 - \sum(\widetilde{X}^i(t))^2 K(\dot\gamma, Y)\right)\mathrm{d}t \\
&\leqslant \int_0^a \left(\sum(\dot{\widetilde{X}}^i(t))^2 - \sum(\dot{\widetilde{X}}^i(t))^2 K^-(t)\right)\mathrm{d}t \\
&\leqslant \int_0^a \left(\sum(\dot{\widetilde{X}}^i(t))^2 - \sum(\dot{\widetilde{X}}^i(t))^2 K^+(t)\right)\mathrm{d}t \\
&\leqslant \int_0^a \left(|\widetilde\nabla_{\dot{\widetilde\gamma}} \widetilde X|^2 + \langle \widetilde R(\dot{\widetilde\gamma}, \widetilde X)\dot{\widetilde\gamma}, \widetilde X\rangle\right)\mathrm{d}t \\
&= I(\widetilde X, \widetilde X).
\end{aligned}$$

It follows that $I(X,X) \leqslant I(\widetilde X, \widetilde X)$.

Finally if $K^+(t) < K^-(t)$ for some $t < a$, then the second inequality above is strict, and thus $I(X,X) < I(\widetilde X, \widetilde X)$. □

Local Hessian Comparison

As a consequence of the basic index comparison theorem, we prove

Theorem 6.21 (Local Hessian Comparison)**.** Let (M,g), $(\widetilde M, \widetilde g)$ be complete Riemannian manifolds, $\gamma : [0,a] \to M$ and $\widetilde\gamma : [0,a] \to \widetilde M$ be minimizing normal geodesics in M and $\widetilde M$ respectively, so that

$$\widetilde K^+(t) \leqslant K^-(t) \text{ holds for all } t \in [0,a].$$

Fix $0 < b < a$ and write $q = \gamma(b)$, $\widetilde q = \widetilde\gamma(b)$. Suppose $X_q \in T_qM$ and $\widetilde X_{\widetilde q} \in T_{\widetilde q}\widetilde M$ are roughly the same. Then

$$\nabla^2 d_p(X_q, X_q) \leqslant \widetilde\nabla^2 \widetilde d_{\widetilde p}(\widetilde X_{\widetilde q}, \widetilde X_{\widetilde q}).$$

Moreover, the equality holds if and only if $\widetilde K^+(t) = K^-(t)$ for all $t \in [0,b]$.

Proof. Since γ and $\widetilde\gamma$ are length minimizing, and $b < a$, we see $q \notin \mathrm{Cut}(p)$ and $\widetilde q \notin \mathrm{Cut}(\widetilde p)$. Let X be the Jacobi field with $X(0) = 0$, $X(b) = X_q$, then as we have seen in Section 5.5,

$$(\nabla^2 d_p)_q(X_q, X_q) = \langle \nabla_{\dot\gamma(q)} X, X_q\rangle = I(X,X).$$

Now the conclusion follows from the index comparison theorem above. □

Since $\Delta = \mathrm{tr}\nabla^2$, by taking trace we get, under the same assumptions,

$$\Delta d_p(q) \leqslant \widetilde\Delta \widetilde d_{\widetilde p}(\widetilde q).$$

6.2.2 Rauch's Jacobi Field Comparison Theorem

Rauch Comparison Theorem

Now we state and prove Rauch's comparison theorem, which relates the sectional curvature of a Riemannian manifold to the length of Jacobi fields (and thus the rates at which geodesics spread apart).

Theorem 6.22 (Rauch Comparison Theorem)**.** Let X, $\widetilde X$ be Jacobi fields along γ, $\widetilde\gamma$ with $X(0) = \widetilde X(0) = 0$, such that $\nabla_{\dot\gamma(0)} X$ and $\widetilde\nabla_{\dot{\widetilde\gamma}(0)} \widetilde X$ are roughly the same. Assume

(1) γ has no conjugate points on $[0,a]$.

(2) $\widetilde K^+(t) \leqslant K^-(t)$ holds for all $t \in [0,a]$.

Then $\widetilde{\gamma}$ has no conjugate points on $[0, a]$, and for all $t \in [0, a]$,

$$|X(t)| \leqslant |\widetilde{X}(t)|.$$

Moreover, if there is $0 < t_0 < t$ such that $K^+(t_0) < K^-(t_0)$, then $|X(t)| < |\widetilde{X}(t)|$.

Proof. Again by Lemma 6.19 we may assume X, $\widetilde{X}$ are normal. We denote

$$u(t) = |X(t)|^2, \quad \widetilde{u}(t) = |\widetilde{X}(t)|^2.$$

Then $\widetilde{u}(t)/u(t)$ is well-defined. Moreover, since

$$X(t) = t\nabla_{\dot{\gamma}(0)}X + \mathcal{O}(t^2),$$

we have

$$\lim_{t\to 0} \frac{\widetilde{u}(t)}{u(t)} = \lim_{t\to 0} \frac{t^2|\nabla_{\dot{\widetilde{\gamma}}(0)}\widetilde{X}|^2 + \mathcal{O}(t^3)}{t^2|\nabla_{\dot{\gamma}(0)}X|^2 + \mathcal{O}(t^3)} = \frac{|\nabla_{\dot{\widetilde{\gamma}}(0)}\widetilde{X}|^2}{|\nabla_{\dot{\gamma}(0)}X|^2} = 1.$$

Therefore, to prove $|X| \leqslant |\widetilde{X}|$, it is enough to prove $\dfrac{\mathrm{d}}{\mathrm{d}t}\dfrac{\widetilde{u}(t)}{u(t)} \geqslant 0$, or equivalently,

$$\dot{\widetilde{u}}(t)u(t) - \widetilde{u}(t)\dot{u}(t) \geqslant 0.$$

Since γ has no conjugate point, $u(t) > 0$ for all $t \in (0, a]$. Let $c \leqslant a$ be the greatest number so that $\widetilde{u}(t) > 0$ on $(0, c)$. For any $b \in (0, c)$, we define

$$X_b(t) = \frac{X(t)}{|X(b)|}, \quad \widetilde{X}_b(t) = \frac{\widetilde{X}(t)}{|\widetilde{X}(b)|}.$$

Applying Theorem 6.20 to X_b, $\widetilde{X}_b$ on $[0, b]$ we get

$$\langle \nabla_{\dot{\gamma}(b)}X_b, X_b(b)\rangle = I(X_b, X_b) \leqslant I(\widetilde{X}_b, \widetilde{X}_b) = \langle \widetilde{\nabla}_{\dot{\widetilde{\gamma}}(b)}\widetilde{X}_b, \widetilde{X}_b(b)\rangle,$$

i.e.

$$\frac{1}{2}\frac{\dot{u}(b)}{u(b)} = \frac{\langle \nabla_{\dot{\gamma}(b)}X, X(b)\rangle}{\langle X(b), X(b)\rangle} \leqslant \frac{\langle \widetilde{\nabla}_{\dot{\widetilde{\gamma}}(b)}\widetilde{X}, \widetilde{X}(b)\rangle}{\langle \widetilde{X}(b), \widetilde{X}(b)\rangle} = \frac{1}{2}\frac{\dot{\widetilde{u}}(b)}{\widetilde{u}(b)}.$$

So for any $t \in (0, c)$, we have $\dfrac{\dot{u}(t)}{u(t)} \leqslant \dfrac{\dot{\widetilde{u}}(t)}{\widetilde{u}(t)}$. This is exactly what we need.

To summary, we proved that $|X(t)| \leqslant |\widetilde{X}(t)|$ for $t \in (0, c)$. If $c < a$, then

$$|\widetilde{X}(c)| \geqslant |X(c)| > 0,$$

contradicting with the choice of c. So we must have $c = a$. In particular, $\widetilde{\gamma}$ has no conjugate points on $[0, a]$.

Finally if there is $0 < t_0 < t$ such that $K^+(t_0) < K^-(t_0)$, then

$$I(X_{t_0}, X_{t_0}) < I(\widetilde{X}_{t_0}, \widetilde{X}_{t_0}),$$

and thus the inequality for t is strict. This completes the proof. □

Although Rauch comparison theorem is stated for two general manifolds whose sectional curvatures are comparable, in most applications one of the two manifolds is a model space that has constant sectional curvature, and the second one is the manifold under study whose curvature is bounded either from below or from above.

For example, according to the explicit formula for Jacobi fields on spheres, the distance between any two consecutive conjugate points of the sphere with sectional curvature κ is $\pi/\sqrt{\kappa}$. As a result we get

Corollary 6.23. Suppose the sectional curvature of (M, g) satisfies

$$0 \leqslant C_1 \leqslant K \leqslant C_2,$$

where C_1, C_2 are constants. Let γ be any geodesic in M. Then the distance D between any two consecutive conjugate points of γ satisfies

$$\frac{\pi}{\sqrt{C_2}} \leqslant D \leqslant \frac{\pi}{\sqrt{C_1}}.$$

One should compare this with Sturm comparison theorem in ODE.

Rauch Comparison Theorem: Second Form

Since any Jacobi field X along γ with $X(0) = 0$ can be written explicitly as

$$X(t) = t(\mathrm{d}\exp_p)_{t\dot{\gamma}(0)}(\nabla_{\dot{\gamma}(0)}X),$$

one can rewrite Rauch comparison theorem above as

Theorem 6.24 (Rauch Comparison Theorem: Second Form)**.** Let γ, $\widetilde{\gamma}$ be geodesics with $p = \gamma(0)$, $\widetilde{p} = \widetilde{\gamma}(0)$, and suppose $X_p \in T_pM$, $\widetilde{X}_{\widetilde{p}} \in T_{\widetilde{p}}\widetilde{M}$ are roughly the same. Assume also

(1) γ has no conjugate points on $[0, a]$,

(2) $\widetilde{K}^+(t) \leqslant K^-(t)$ holds for all $t \in [0, a]$.

Then

$$|(\mathrm{d}\exp_p)_{t\dot{\gamma}(0)}X_p| \leqslant |(\mathrm{d}\exp_{\widetilde{p}})_{t\dot{\widetilde{\gamma}}(0)}\widetilde{X}_p|.$$

Moreover, the equality is strict for t if there exists $0 < t_0 < t$ with $K^+(t_0) < K^-(t_0)$.

Note that for any Riemannian manifold (M, g) and any point $p \in M$, (T_pM, g_p) is a Riemannian manifold that has constant sectional curvature 0. Applying Theorem 6.24 to (M, g) and (T_pM, g_p) we get

Corollary 6.25. Let (M, g) be a complete Riemannian manifold with sectional curvature $K \leqslant 0$. Then for any $p \in M$, any $X_p \in T_pM$ and $Y_p \in T_pM = T_{X_p}(T_pM)$,

$$|(\mathrm{d}\exp_p)_{X_p}(Y_p)| \geqslant |Y_p|.$$

In particular, for any curve γ in T_pM, one has

$$L(\gamma) \leqslant L(\exp_p \circ \gamma).$$

Moreover, the equality is strict if $K < 0$.

Note that by using this corollary one can give another proof of Proposition 6.5 (the weak cosine law for Cartan-Hadamard manifolds) in Section 6.1.

6.3 The Global Hessian and Toponogov Comparison

Last section we proved various local comparison theorems that holds away from cut locus. In this section we turn to global comparison that holds on M.

6.3.1 The Hessian Comparison Theorem: Global Form

Hessian of Distance for M_κ^m

Let M_κ^m be a space form, i.e. a complete connected Riemannian manifold with constant sectional curvature κ. Let $\gamma : [0, l] \to M_\kappa^m$ be a normal geodesic in M_κ^m from p to $q \notin \mathrm{Cut}(p) \cup \{p\}$, then for any $X_q \in (\dot\gamma)^\perp$, the Jacobi field V along γ with $V(0) = 0$ and $V(l) = X_q$ is

$$V(t) = \frac{sn_\kappa(t)}{sn_\kappa(l)} X_q(t),$$

where $X_q(t)$ is the parallel vector field along γ with $X_q(l) = X_q$, and we used

$$sn_\kappa(t) = \begin{cases} \dfrac{\sin(\sqrt{\kappa}t)}{\sqrt{\kappa}}, & \kappa > 0 \\ t, & \kappa = 0 \\ \dfrac{\sinh(\sqrt{-\kappa}t)}{\sqrt{-\kappa}}, & \kappa < 0 \end{cases} \quad \text{and} \quad cn_\kappa(t) = sn_\kappa'(t) = \begin{cases} \cos(\sqrt{\kappa}t), & \kappa > 0, \\ 1, & \kappa = 0, \\ \cosh(\sqrt{-\kappa}t), & \kappa < 0. \end{cases}$$

As a result, for any $Y_q \in T_qM$,

$$(\nabla^2 d_p)_q(X_q, Y_q) = \langle \nabla_{\dot\gamma(l)} V, Y_q \rangle = \frac{cn_\kappa(l)}{sn_\kappa(l)} \langle X_q, Y_q \rangle.$$

So the Hessian of d_p at the point q, with respect to an orthonormal basis $e_1(l) = \dot\gamma(l), e_2(l), \ldots, e_m(l)$, is

$$(\nabla^2 d_p)_q = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & \dfrac{cn_\kappa(l)}{sn_\kappa(l)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \dfrac{cn_\kappa(l)}{sn_\kappa(l)} \end{pmatrix}.$$

The Hessian matrix is almost a constant matrix, with the only exception that there is a zero for the top-left entry. We will carefully choose a function f so that $\nabla^2(f \circ d_p)$ is a constant matrix. For this purpose we calculate, for any $X_q, Y_q \in T_qM$,

$$\begin{aligned} \nabla^2(f \circ d_p)(X_q, Y_q) = \langle \nabla_{X_q} \nabla(f \circ d_p), Y_q \rangle &= \langle \nabla_{X_q}(f'(d_p) \nabla d_p), Y_q \rangle \\ &= f'(d_p) \langle \nabla_{X_q} \nabla d_p, Y_q \rangle + f''(d_p) \langle \nabla d_p, X_q \rangle \langle \nabla d_p, Y_q \rangle \\ &= f'(d_p) \nabla^2 d_p(X_q, Y_q) + f''(d_p) \langle \dot\gamma(l), X_q \rangle \langle \dot\gamma(l), Y_q \rangle, \end{aligned}$$

where γ is the minimizing geodesic from p to q. As a result, for $\nabla^2(f \circ f_p)$ to be a constant matrix under the given basis, we should choose f so that

$$f'(t) \frac{cn_\kappa(t)}{sn_\kappa(t)} = f''(t).$$

So the simplest solution is to take f to satisfy $f' = sn_\kappa(t)$, i.e. take f to be

$$md_\kappa(r) = \int_0^r sn_\kappa(t) \mathrm{d}t = \begin{cases} \dfrac{1 - \cos(\sqrt{\kappa}r)}{\kappa}, & \text{if } \kappa > 0, \\ \dfrac{r^2}{2}, & \text{if } \kappa = 0, \\ \dfrac{1 - \cosh(\sqrt{-\kappa}r)}{\kappa}, & \text{if } \kappa < 0. \end{cases}$$

It follows

$$\nabla^2(md_\kappa \circ d_p)_q = cn_\kappa(l) \mathrm{Id}.$$

The Cosine Law in M^m_κ

It is easy to check $cn_\kappa = 1-\kappa md_\kappa(r)$. So if we let $\varphi(t) = md_\kappa \circ d_p \circ \gamma(t)$, where γ is a geodesic in M^m_κ away from $\mathrm{Cut}(p)\cup\{p\}$, then

$$\varphi''(t) = cn_\kappa \circ d_p \circ \gamma(t) = 1 - \kappa md_\kappa(d_p \circ \gamma(t)) = 1 - \kappa\varphi(t).$$

As an application, we may derive the cosine law in M^m_κ. Consider a geodesic triangle ΔABC in M^m_κ with side lengths a, b, c and angles A, B, C, where for $\kappa > 0$ we assume $a, b, c < \pi/\sqrt{\kappa}$. Let $\gamma_1 : [0,a] \to M^m_\kappa$ be the normal geodesic with $\gamma_1(0) = B$, $\gamma_1(a) = C$ and let $\gamma_2 : [0,b] \to M^m_\kappa$ be the normal geodesic with $\gamma_2(0) = C$, $\gamma_2(b) = A$.

Now take $p = B$ and $\gamma = \gamma_2$, i.e. consider $\varphi(t) = md_\kappa \circ d_B \circ \gamma_2(t)$. Then

$$\varphi(0) = md_\kappa(a), \quad \varphi(b) = md_\kappa(c)$$

and

$$\varphi' = sn_\kappa(a)\langle\dot\gamma_1(a), \dot\gamma_2(0)\rangle = -sn_\kappa(a)\cos C.$$

So if $\kappa = 0$, we get

$$\varphi(t) = \frac{a^2}{2} - a\cos Ct + \frac{1}{2}t^2$$

and thus $\varphi(b) = md_\kappa(c)$ becomes $c^2 = a^2 + b^2 - 2ab\cos C$.

For $\kappa \neq 0$, we get

$$\varphi(t) = \frac{1}{\kappa} + c_1 sn_\kappa(t) + c_2 cn_\kappa(t).$$

The initial conditions $\varphi(0) = md_\kappa(a)$ and $\varphi'(0) = -sn_\kappa(a)\cos C$ implies $c_1 = -sn_\kappa(a)\cos C$, $c_2 = -\frac{1}{\kappa}cn_\kappa(a)$. So the equation $\varphi(b) = md_\kappa(c)$ becomes the following cosine law in M^m_κ,

$$cn_\kappa(c) = cn_\kappa(a)cn_\kappa(b) + \kappa sn_\kappa(a)sn_\kappa(b)\cos C.$$

As a direct corollary, we get

Corollary 6.26. In M^m_κ, take an angle α with side lengths l_1 and l_2 fixed. Let $f(\alpha)$ be the distance between the end points. Then $f(\alpha)$ is increasing in α.

Compare in the Barrier Sense

It turns out that the Hessian and Laplacian comparison theorems holds globally on the whole of M in several weak sense: in the barrier sense, in the viscosity sense and in the distribution sense[2]. Here we only discuss the first one, since the condition is often easier to check. The notion of barrier sense was first introduced by Calabi in 1958:

Definition 6.27. Let f be a continuous function defined on (M,g).

(1) If $g \in \mathcal{C}^2(U)$ is defined in a neighborhood U of p, and

$$f(p) = g(p), \quad \text{and} \quad f(q) \leqslant g(q),\ \forall q \in U,$$

then we call g an **upper barrier function** of f at p.

(2) If for any $\varepsilon > 0$, there is an upper barrier function g_ε of f at p, such that $\Delta g_\varepsilon(p) \leqslant c+\varepsilon$, then we say

$$\Delta f(p) \leqslant c \text{ in the barrier sense.}$$

(3) If for any normal geodesic σ with $\sigma(0) = q$, one has $(f\circ\sigma)''(0) \leqslant c$ in the barrier sense, then we say

$$(\nabla^2 f)(q) \leqslant c\cdot \mathrm{Id} \text{ in the barrier sense.}$$

[2]It can be proven that if $\Delta f \leqslant g$ holds in the barrier sense, then it also holds in the viscosity sense and in the distribution sense.

Example. Note that by taking $M = (a,b)$, we get a definition of "$f''(t_0) \leqslant c$ in the barrier sense" for continuous function $f : (a,b) \to \mathbb{R}$. For example, consider $f(x) = -|x|$. Then $g = 0$ is a upper barrier function of f at $x = 0$. As a result, $f''(0) \leqslant 0$ in the barrier sense.

As observed by Calabi, one can easily construct upper barrier functions for the distance function:

Example. If γ is a minimizing geodesic from p to q, then for $0 < \eta < d(p,q)$, the function $r_\eta(x) = \eta + d(x, \gamma(\eta))$ is an upper barrier function for d_p at q.

The proof of the following lemma will be left as an exercise.

Lemma 6.28. Let $f : (a,b) \to \mathbb{R}$ be continuous.

(1) If f is $\mathcal{C}^2$, then $f''(t_0) \leqslant c$ in the barrier sense if and only if $f''(t_0) \leqslant c$ in the usual sense.

(2) If f takes its minimum at t_0, and $f''(t_0) \leqslant c$ in the barrier sense, then $c \geqslant 0$.

(3) If $f'' \leqslant 0$ in the barrier sense, then f is concave.

We also mention the following Hopf strong maximal principle without proof.

Theorem 6.29 (Hopf-Calabi Strong Maximum Principle)**.** Let $\Omega \subset M$ be a connected open set. Suppose $\Delta f \leqslant 0$ in M in the barrier sense, and f has an interior minimum, then f is constant on Ω.

The Global Hessian Comparison Theorem

Now we are ready to state and prove

Theorem 6.30 (The Global Hessian Comparison Theorem)**.** Let (M,g) be a Riemannian manifold with sectional curvature $K \geqslant \kappa$. Then for any $p \in M$,

$$\nabla^2(md_\kappa \circ d_p) \leqslant cn_k \circ d_p \cdot \mathrm{Id} \text{ in the barrier sense.}$$

Proof. According to the (local) Hessian comparison theorem that we proved last time, the theorem holds at smooth points q of d_p. We first prove the conclusion at the point p: For any normal geodesic σ with $\sigma(0) = p$, we have

$$md_\kappa \circ d_p \circ \sigma(t) = md_\kappa(|t|) = md_\kappa(t).$$

Thus $(md_\kappa \circ d_p \circ \sigma)''(0) = md_\kappa''(0) = cn_\kappa(0)$.

It remains to prove the conclusion for a non-smooth point $q \neq p$ of d_p. So we let $\gamma : [0,l] \to M$ be a minimizing normal geodesic from p to q, and let σ be a normal geodesic with $\sigma(0) = q$. Note that by Bonnet-Myers theorem, $l \leqslant \dfrac{\pi}{\sqrt{\kappa}}$ if $\kappa > 0$.

For $0 < \eta < d(p,q)$ small, the function

$$r_\eta(x) = \eta + d(x, \gamma(\eta))$$

is an upper barrier function of d_p at q. Apply Hessian comparison to r_η we get

$$\begin{aligned}
(md_\kappa \circ r_\eta \circ \sigma)''(0) &= md_\kappa'(l)(r_\eta \circ \sigma)''(0) + md_\kappa''(l)\langle \dot\gamma(l), \dot\sigma(0)\rangle^2 \\
&= sn_\kappa(l)(\nabla^2 d_{\gamma(\eta)})_q(\dot\sigma(0), \dot\sigma(0)) + cn_\kappa(l)\langle \dot\gamma(l), \dot\sigma(0)\rangle^2 \\
&\leqslant sn_\kappa(l)\frac{cn_\kappa(l-\eta)}{sn_\kappa(l-\eta)}|\dot\sigma^\perp(0)|^2 + cn_\kappa(l)\langle \dot\gamma(l), \dot\sigma(0)\rangle^2 \\
&= \frac{sn_\kappa(l)cn_\kappa(l-\eta) - cn_\kappa(l)sn_\kappa(l-\eta)}{sn_\kappa(l-\eta)}|\dot\sigma^\perp(0)|^2 + cn_\kappa(l)(|\dot\sigma^\perp(0)|^2 + \langle \dot\gamma(l), \dot\sigma(0)\rangle^2) \\
&= \frac{sn_\kappa(\eta)}{sn_\kappa(l-\eta)}|\dot\sigma^\perp(0)|^2 + cn_\kappa(l).
\end{aligned}$$

So if $\kappa \leqslant 0$ or if $l < \dfrac{\pi}{\sqrt{\kappa}}$ when $\kappa > 0$, given any $\varepsilon > 0$, for η is small enough, we have

$$(md_\kappa \circ r_\eta \circ \sigma)''(0) \leqslant cn_\kappa(l) + \varepsilon,$$

which implies

$$\nabla^2(md_\kappa \circ d_p) \leqslant cn_\kappa \circ d_p \text{ in the barrier sense.}$$

For $\kappa > 0$ and $l = \dfrac{\pi}{\sqrt{\kappa}}$, we will prove (M, g) is isomorphic to S^m_κ, in which case we may take σ such that $\dot\sigma(0) = \dot\gamma(l)$, and the desired conclusion follows. □

6.3.2 The Toponogov Comparison Theorem

The purpose of this section is to prove a very useful global comparison theorem, due to Toponogov in 1959. It quantifies the assertion (c.f. Problem Sheet 4) that a pair of geodesics emanating from a point p spread apart more slowly in a region of high curvature than they would in a region of low curvature.

Geodesic Triangles and Hinges

Definition 6.31. Let (M, g) be complete.

(1) A **geodesic triangle** ΔABC consists of three points A, B, C in M (which are called the **vertices**) and three minimizing normal geodesics (which are called the **sides**) γ_{AB}, γ_{BC}, γ_{CA} joining each two of them.

If only two sides, say γ_{AB} and γ_{AC}, are minimizing, while the third is a normal geodesic [which need not be minimizing] that satisfies the triangle inequality

$$L(\gamma_{BC}) \leqslant L(\gamma_{AB}) + L(\gamma_{AC}),$$

then we will call ΔABC a **generalized geodesic triangle**.

(2) A **geodesic hinge** $\angle BAC$ consists of a point A in M (which is again called the **vertex**) and two minimizing normal geodesics γ_{AB}, γ_{AC} (called the **sides**) emanating from A, with end points B and C in M.

If one side is minimizing, while the other side is normal geodesic [which need not be minimizing], we call $\angle BAC$ a **generalized geodesic hinge**.

In what follows when we say hinge or triangle, we always mean generalized geodesic hinge or generalized geodesic triangle.

Remark. In the definition of generalized geodesic hinge, we required that at least one curve is minimizing. Otherwise the Toponogov comparison theorem below may fail: If one take two geodesics of length $\dfrac{\pi}{\sqrt{\kappa}}$ in $M_{\kappa+\varepsilon}$, then the other endpoints of the comparing hinge in M_κ will meet.

Lemma 6.32. Let (M, g) be a complete Riemannian manifold of dimension m whose sectional curvature $K \geqslant \kappa$. Then

(1) Let $\angle BAC$ be a generalized geodesic hinge in M. If $\kappa > 0$ we further assume that all the sides of $\angle ABC$ have lengths no more than $\dfrac{\pi}{\sqrt{\kappa}}$. Then there is a generalized geodesic hinge $\angle\widetilde{B}\widetilde{A}\widetilde{C}$ in M^m_κ with same angle and the same corresponding side lengths. [We will call it a comparing hinge.]

(2) Let ΔABC be a generalized geodesic triangle in M. If $\kappa > 0$ we further assume that all the sides of ΔABC have lengths no more than $\dfrac{\pi}{\sqrt{\kappa}}$. Then there is a triangle $\Delta\widetilde{A}\widetilde{B}\widetilde{C}$ in M^m_κ whose corresponding sides have the same length as ΔABC. [We will call it a comparing triangle.]

Toponogov Comparison Theorem

Now we state and prove Toponogov Comparison Theorem, in which we can actually compare the distance functions instead of only comparing their Hessian.

Theorem 6.33 (Toponogov Comparison Theorem)**.** Let (M, g) be a complete Riemannian manifold with sectional curvature $K \geqslant \kappa$. Then

(1) (**Hinge Version**) Let $\angle BAC$ be a generalized geodesic hinge in M and $\angle\widetilde{B}\widetilde{A}\widetilde{C}$ a comparing hinge in M^m_κ.[3] If $\kappa > 0$ we further assume that the sides of $\angle BAC$ have lengths no more than $\dfrac{\pi}{\sqrt{\kappa}}$. Then $\operatorname{dist}(B, C) \leqslant \operatorname{dist}(\widetilde{B}, \widetilde{C})$.

(2) (**Triangle Version**) Let ΔABC be a generalized geodesic triangle in M and $\Delta\widetilde{A}\widetilde{B}\widetilde{C}$ a comparing triangle in M^m_κ. If $\kappa > 0$ we further assume that all the sides of ΔABC have lengths no more than $\dfrac{\pi}{\sqrt{\kappa}}$. Then the angles in ΔABC opposites to the minimizing geodesics are greater than the corresponding angles in $\Delta\widetilde{A}\widetilde{B}\widetilde{C}$.

Proof. We first observe that according to the cosine law in M^2_κ, for any hinge with sides γ_1, γ_2 and angle α, the function

$$f(\alpha) = d(\gamma_0(l_0), \gamma_1(l_1))$$

is increasing for $\alpha \in (0, \pi)$. So the Hinge version implies the triangle version.

To prove the Hinge version, we need

Lemma 6.34. Let $f : [0, l] \to \mathbb{R}$ be a continuous function that is differentiable at $t = 0$, with $f(0) = 0$ and $f'(0) \leqslant 0$, where $l \leqslant \dfrac{\pi}{\sqrt{\kappa}}$ if $\kappa > 0$. Moreover, assume

$$f''(t) + \kappa f(t) \leqslant \text{ in the barrier sense,}$$

then $f(t) \leqslant 0$ for all $t \in [0, l]$.

Proof of Lemma 6.34. Let $f_\varepsilon(t) = f(t) - \varepsilon sn_\kappa(t)$, then $f_\varepsilon(0) = 0$ and $f'_\varepsilon(0) \leqslant -\varepsilon < 0$. So $f_\varepsilon < 0$ for $t > 0$ small enough. In what follows we prove $f_\varepsilon \leqslant 0$ for all $t \in [0, l]$. Letting $\varepsilon \to 0$ we get the desired conclusion.

By contradiction we let t_0 be the smallest positive root of f_ε.

Case 1: $\kappa \leqslant 0$. Suppose $f_\varepsilon\Big|_{[0,t_0]}$ takes its minimum at t_1. Then we get,

$$f''_\varepsilon(t_1) + \kappa f_\varepsilon(t_1) = f''(t_1) + \kappa f(t_1) - \varepsilon(sn''_\kappa(t_1) + \kappa sn_\kappa(t_1)) \leqslant 0$$

in the barrier sense. By Lemma 6.28 (2), we get $-\kappa f_\varepsilon(t_1) \geqslant 0$, i.e. $f_\varepsilon(t_1) \geqslant 0$, a contradiction.

Case 2: $\kappa > 0$. We may assume $t_0 < l$, otherwise we are done. Take $\delta > 0$ small so that $\left[-\delta, \dfrac{\pi}{\sqrt{\kappa+\delta}} - \delta\right] \supset [0, t_0]$. Let $\phi(t) = -\sin(\sqrt{\kappa+\delta}(t+\delta))$, so that $\phi''(t) + (K+\delta)\phi = 0$. Suppose $\dfrac{f_\varepsilon}{\phi}\Big|_{[0,t_0]}$ takes its maximum at t_1. Let $g_{\varepsilon,\varepsilon'}$ be an upper barrier function of f_ε at t_1, i.e.

$$g_{\varepsilon,\varepsilon'}(t_1) = f_\varepsilon(t_1), \qquad g_{\varepsilon,\varepsilon'}(t) \geqslant f_\varepsilon(t) \text{ near } t_1,$$

and such that

$$g''_{\varepsilon,\varepsilon'}(t_1) \leqslant (-\kappa) f_\varepsilon(t_1) + \varepsilon'.$$

Then t_1 is a maximum for $\dfrac{g_{\varepsilon,\varepsilon'}}{\phi}$ since $\phi < 0$. It follows

$$\left(\frac{g_{\varepsilon,\varepsilon'}}{\phi}\right)'(t) = \frac{g'_{\varepsilon,\varepsilon'}(t)\phi(t) - g_{\varepsilon,\varepsilon'}(t)\phi'(t)}{\phi^2(t)}$$

[3] Obviously can replace M^m_κ by M^2_κ.

equals 0 at t_1, and thus

$$0 \geqslant \left(\frac{g_{\varepsilon,\varepsilon'}}{\phi}\right)''(t_1) = \frac{g''_{\varepsilon,\varepsilon'}(t_1)\phi(t_1) - g_{\varepsilon,\varepsilon'}(t_1)\phi''(t_1)}{\phi^2(t_1)}$$
$$= \frac{g''_{\varepsilon,\varepsilon'}(t_1) + (K+\delta)g_{\varepsilon,\varepsilon'}(t_1)}{\phi(t_1)} \geqslant \frac{\varepsilon' + \delta g_{\varepsilon,\varepsilon'}(t_1)}{\phi(t_1)}.$$

Letting $\varepsilon' \to 0$ we get $f_\varepsilon(t_1) = g_{\varepsilon,\varepsilon'}(t_1) \geqslant 0$, a contradiction. □

For simplicity we denote $\gamma_0 = \gamma_{AB}$, $\gamma_1 = \gamma_{AC}$ and denote $l_0 = L(\gamma_0)$, $l_1 = L(\gamma_1)$, $\alpha = \angle BAC$. Assume γ_0 is minimizing. For $\varepsilon > 0$ small, let

$$\rho_\varepsilon(t) = d(\gamma_0(l_0 - \varepsilon), \gamma_1(t)), \qquad t \in [0, l_1].$$

Then ρ_ε is smooth for $t > 0$ small enough, $\rho_\varepsilon(0) = l_0 - \varepsilon$ and

$$\rho'_\varepsilon(0) = \langle -\dot\gamma_0(0), \dot\gamma_1(0)\rangle = -\cos\alpha.$$

By the global Hessian comparison theorem,

$$(md_\kappa \circ \rho_\varepsilon)''(t) \leqslant cn_\kappa \circ \rho(t) = 1 - \kappa md_\kappa \circ \rho(t), \text{ in the barrier sense.}$$

We may perform the same computation in M^2_κ to conclude that for $\widetilde\rho_\varepsilon(t) = d_{\widetilde\gamma_0(l_0-\varepsilon)}(\widetilde\gamma_1(t))$, one has $\widetilde\rho_\varepsilon(0) = l_0 - \varepsilon$, $\widetilde\rho'_\varepsilon(0) = -\cos\alpha$ and

$$(md_\kappa \circ \widetilde\rho_\varepsilon)''(t) = 1 - \kappa md_\kappa \circ \widetilde\rho(t).$$

So if we let $f(t) = md_\kappa \circ \rho_\varepsilon - md_\kappa \circ \widetilde\rho_\varepsilon$, then

$$f''(t) + \kappa f(t) \leqslant 0 \text{ in the barrier sense,}$$

and $f(0) = 0$, $f'(0) = sn_\kappa(\rho_\varepsilon(0))\rho'_\varepsilon(0) - sn_\kappa(\widetilde\rho'_\varepsilon(0))\widetilde\rho'_\varepsilon(0) = 0$. By Lemma 6.34, we have $f(t) \leqslant 0$ for all $t \in [0, l_1]$. It follows that $\rho_\varepsilon(t) \leqslant \widetilde\rho_\varepsilon(t)$ for all $t \in [0, l_1]$. Letting $\varepsilon \to 0$ we get the desired conclusion. □

Application to Fundamental Group

As an application, we prove

Theorem 6.35 (Gromov)**.** Let (M, g) be a complete Riemannian manifold with sectional curvature $K \geqslant 0$. Then $\pi_1(M)$ is generated by no more than

$$C(m) = \frac{\text{Vol}(S^{m-1})}{\text{Vol}(S^{m-1}(\pi/6))}$$

generators, where $S^{m-1}(\pi/6)$ is a geodesic ball of radius $\pi/6$ in S^{m-1}.

Proof. We will consider $\pi_1(M)$ as the group of Deck transformations on the universal covering $\widetilde M$. Fix $\widetilde p \in \widetilde M$ and choose inductively a generating set of $\pi_1(M)$ as follows:

- We first choose $e \neq g_1 \in \pi_1(M)$ so that
$$d(\widetilde p, g_1 \cdot \widetilde p) \leqslant d(\widetilde p, g \cdot \widetilde p), \qquad \forall g \in \pi_1(M) \setminus \{e\}.$$
- Suppose $g_1, \dots, g_{k-1}$ are chosen. We then choose $g_k \notin \langle g_1, \dots, g_{k-1}\rangle$ so that
$$d(\widetilde p, g_k \cdot \widetilde p) \leqslant d(\widetilde p, g \cdot \widetilde p), \qquad \forall g \in \pi_1(M) \setminus \langle g_1, \dots, g_{k-1}\rangle.$$

Let $\widetilde{\gamma}_k$ be a minimizing geodesic in $(\widetilde{M}, \widetilde{g})$ from $\widetilde{p}$ to $g_k \cdot \widetilde{p}$. We claim that the angle between any two such geodesics is at least $\dfrac{\pi}{3}$. Then the conclusion follows.

We prove that claim by contradiction. Suppose the angle between $\widetilde{\gamma}_k$ and $\widetilde{\gamma}_{k+l}$ is less than $\dfrac{\pi}{3}$. For simplicity we denote $\widetilde{l}_k = d(\widetilde{p}, g_k \cdot \widetilde{p})$. Then according to the Toponogov comparison theorem,

$$d(g_{k+l} \cdot \widetilde{p}, g_k \cdot \widetilde{p})^2 < l_k^2 + l_{k+1}^2 - l_k l_{k+1} \leqslant l_{k+1}^2.$$

This implies

$$d(\widetilde{p}, g_{k+l}^{-1} g_k \cdot \widetilde{p}) = d(g_{k+1} \cdot \widetilde{p}, g_k \cdot \widetilde{p}) < l_{k+l} = d(\widetilde{p}, g_{k+l} \cdot \widetilde{p}),$$

which contradicts with the choice of g_{k+l}. □

Remark. By the same way one can prove the following theorem of Gromov:

Theorem 6.36 (Gromov)**.** For k negative, there is a constant $C = C(m, k, D)$ so that for any complete Riemannian manifold (M, g) with $K \geqslant k$ and $\operatorname{diag}(M, g) \leqslant D$, the fundamental group $\pi_1(M)$ is generated by no more than $C(m, k, D)$ generators.

Note that a bound on diameter is needed. To see this one can look at the example of surface of genus g.

Chapter 7

Volume Comparison and Applications

7.1 The Laplacian and Volume Comparison

In this section we discuss comparison under Ricci curvature condition. We first prove the Laplacian comparison theorem, which can be viewed as an averaged version of the Hessian comparison (under slightly stronger condition). Then we prove the very useful Bishop-Gromov volume comparison theorem, which can be viewed as an "integrated" version of the Laplacian comparison theorem.

7.1.1 The Laplacian Comparison Theorem

Recall that if (M, g), $(\widetilde{M}, \widetilde{g})$ are complete Riemannian manifolds, $\gamma : [0, a] \to M$ and $\widetilde{\gamma} : [0, a] \to \widetilde{M}$ are minimizing normal geodesics with

$$\widetilde{K}^+(t) \leqslant K^-(t) \text{ holds for all } t \in [0, a].$$

and if $X_q \in T_qM$ and $\widetilde{X}_{\widetilde{q}} \in T_{\widetilde{q}}\widetilde{M}$ are roughly the same, where $q = \gamma(b)$, $\widetilde{q} = \widetilde{\gamma}(b)$ and $0 < b < a$, then we have

$$\nabla^2 d_p(X_q, X_q) \leqslant \widetilde{\nabla}^2 \widetilde{d}_{\widetilde{p}}(\widetilde{X}_{\widetilde{q}}, \widetilde{X}_{\widetilde{q}}).$$

Moreover, the equality holds if and only if $\widetilde{K}^+(t) = K^-(t)$ for all $t \in [0, b]$.

Since $\Delta = \mathrm{tr}\nabla^2$, the Hessian comparison theorem will imply a Laplacian comparison $\Delta d_p(q) \leqslant \widetilde{\Delta}\widetilde{d}_{\widetilde{p}}(\widetilde{q})$. Note that by taking the trace of the Hessian, what we get is (up to a constant) "the average of the Hessian". As a result, one can anticipate to weaken the comparison condition from sectional curvature to a weaker "averaged version", namely, a comparison condition on Ricci curvature.

Theorem 7.1 (The Laplacian Comparison Theorem)**.** Let (M, g) be a Riemannian manifold, and $\gamma : [0, l] \to M$ a minimizing normal geodesic with $\gamma(0) = p$. Suppose

$$Ric(\dot{\gamma}(t)) \geqslant (m-1)\kappa$$

We denote $\widetilde{\Delta}_\kappa$, $\widetilde{d}$, $\widetilde{\gamma}$ etc the corresponding objects in M^m_κ. Then

$$\Delta_p(\gamma(t)) \leqslant \widetilde{\Delta}\widetilde{d}_{\widetilde{p}}(\widetilde{\gamma}(t)), \quad \forall 0 < t < l.$$

Moreover, $\Delta d_p(\gamma(b)) \leqslant \widetilde{\Delta}\widetilde{d}_{\widetilde{p}}(\widetilde{\gamma}(b))$ for some $b < l$ if and only if for any $0 \leqslant t \leqslant b$, $Ric(\dot{\gamma}(t)) = (m-1)\kappa$, and any normal Jacobi field X along $\gamma\Big|_{[0,b]}$ with $X(0) = 0$ is almost parallel, i.e. is of the form $X = \dfrac{sn_\kappa(t)}{sn_\kappa(b)}e(t)$, where e is a parallel vector field along γ.

Proof. Fix $b < l$. As usual we let $\{e_1(t), \dots, e_m(t)\}$ be a parallel orthonormal frame along γ with $e_1(t) = \dot\gamma(t)$, and let $\{\widetilde{e}_1(t), \dots, \widetilde{e}_m(t)\}$ be a parallel orthonormal frame along $\widetilde\gamma$ with $\widetilde{e}_1(t) = \dot{\widetilde\gamma}(t)$. For any $i \geqslant 2$, let $X_i(\tau)$ be the normal Jacobi field along $\gamma\Big|_{[0,b]}$ with $X_i(0) = 0$ and $X_i(b) = e_i(b)$, and let $\widetilde{X}_i(\tau)$ be the normal Jacobi field along $\widetilde\gamma\Big|_{[0,b]}$ with $\widetilde{X}_i(0) = 0$ and $\widetilde{X}_i(b) = \widetilde{e}_i(b)$. Then for $q = \gamma(b)$ we have

$$\Delta d_p(q) = \sum_{i=2}^m (\nabla^2 d_p)_q(e_i(b), e_i(b)) = \sum_{i=2}^m I(X_i, X_i)$$

and similarly for $\widetilde{q} = \widetilde\gamma(b)$, $\widetilde\Delta \widetilde{d}_{\widetilde{p}}(\widetilde{q}) = \sum_{i=2}^m I(\widetilde{X}_i, \widetilde{X}_i)$. It remains to prove

$$\sum_{i=2}^m I(X_i, X_i) \leqslant \sum_{i=2}^m I(\widetilde{X}_i, \widetilde{X}_i).$$

We shall apply the same trick that we played in the proof of the basic index comparison lemma, namely we transplant $\widetilde{X}_i$ to γ. For this purpose we first recall that the condition "$\widetilde{M}$ has constant sectional curvature κ along $\widetilde\gamma$" (c.f. Problem Sheet 4), the normal Jacobi field $\widetilde{X}_i$ is given by $\widetilde{X}^i(t) = \frac{sn_\kappa(t)}{sn_\kappa(b)} \widetilde{e}_i(t)$. So for each $2 \leqslant i \leqslant m$ we define on $\gamma\Big|_{[0,b]}$ a vector field

$$X_i'(t) = \frac{sn_\kappa(t)}{sn_\kappa(b)} e_i(t).$$

Obviously X_i' has the same boundary condition as the Jacobi field X_i. So we get $I(X_i, X_i) \leqslant I(X_i', X_i')$. Now the conclusion follows from

$$\begin{aligned}
\sum I(X_i', X_i') &= \sum \int_0^b (|\nabla_{\dot\gamma} X_i'|^2 + Rm(\dot\gamma, X_i', \dot\gamma, X_i'))\mathrm{d}t \\
&= \sum_i \int_0^b \left(\left(\frac{sn_\kappa'(t)}{sn_\kappa(b)}\right)^2 - \left(\frac{sn_\kappa(t)}{sn_\kappa(b)}\right)^2 K(\dot\gamma, e_i) \right) \mathrm{d}t \\
&= \int_0^b \left((m-1)\left(\frac{sn_\kappa'(t)}{sn_\kappa(b)}\right)^2 - \left(\frac{sn_\kappa(t)}{sn_k(b)}\right)^2 Ric(\dot\gamma) \right) \mathrm{d}t \\
&\leqslant \int_0^b \left((m-1)\left(\frac{sn_\kappa'(t)}{sn_k(b)}\right)^2 - \left(\frac{sn_\kappa(t)}{sn_\kappa(b)}\right)^2 (m-1)\kappa \right) \mathrm{d}t \\
&= \sum \int_0^b (|\nabla_{\dot{\widetilde\gamma}} \widetilde{X}_i|^2 + \widetilde{Rm}(\dot{\widetilde\gamma}, \widetilde{X}_i, \dot{\widetilde\gamma}, \widetilde{X}_i))\mathrm{d}t = \sum I(\widetilde{X}_i, \widetilde{X}_i).
\end{aligned}$$

From the proof we see that equality holds if and only if "$Ric(\dot\gamma) = (m-1)\kappa$, and $X_i = X_i'$ for all i". Since any normal Jacobi field X along γ with $X(0) = 0$ is a linear combination of these X_i, the conclusion follows. □

Remark. As we have seen, $(\nabla^2 d_p)_q(X_q, Y_q) = \langle \nabla_{X_1} N, Y_q \rangle$, where $N = \partial_r$ is the outward unit normal vector of the geodesic sphere $S(p, d(p,q))$ at q. In other words, we may replace $\nabla_{X_1} N$ by the shape vector $(\nabla_{X_1} N)^\perp$ (c.f. Problem Sheet Exercise 2.9) and conclude that $\nabla^2 d_p$ is the second fundamental form of the geodesic sphere $S(p, d(p,q))$. It follows that $\Delta d_p(q) = \mathrm{tr}(\nabla^2 d_p)$ is the trace of the second fundamental form. i.e. the mean curvature of the geodesic sphere $S(p, d(p,q))$ at q.

7.1.2 The Bishop-Gromov Volume Comparison Theorem

The Volume Measure in Coordinates

Recall that the Riemannian volume density is defined in a chart (φ, U, V) as

$$\mathrm{d}V_g = \sqrt{G} \circ \varphi^{-1} \mathrm{d}x^1 \cdots \mathrm{d}x^m,$$

where $G = \det(g_{ij})$ and $dx^1 \cdots dx^m$ is the Lebesgue measure on $\mathbb{R}^m$.

Now let (M, g) be a complete Riemannian manifold. Although in general there is no global chart, there is a large chart that covers almost the whole of M, namely $\{\exp_p^{-1}, M \setminus \mathrm{Cut}(p), \Sigma(p)\}$, which is defined except for the measure zero closed subset $\mathrm{Cut}(p)$, where

$$\Sigma(p) = \exp^{-1}(M \setminus \mathrm{Cut}(p))$$

is an open star-shaped domain in T_pM. In particular, on T_pM we may use polar coordinates and write

$$dx^1 \cdots dx^m = r^{m-1} dr d\Theta,$$

where $d\Theta$ is the usual surface measure on S^{m-1}. Combine this with the chart $\{\exp_p^{-1}, M \setminus \mathrm{Cut}(p), \Sigma(p)\}$ we get

$$dV_g = \sqrt{G}(\exp_p(r\Theta)) r^{m-1} dr d\Theta, \quad r\Theta \in \Sigma(p).$$

We denote

$$\mu_p(r, \Theta) = \begin{cases} \sqrt{G}(\exp_p(r\Theta)) r^{m-1}, & r\Theta \in \Sigma(p), \\ 0, & r\Theta \notin \Sigma(p). \end{cases}$$

Note that by definition

$$\overline{B_r(p)} = \exp_p(\overline{B_r(0)}) = \exp_p(\overline{B_r(0)} \cap \overline{\Sigma(p)}).$$

Since $\mathrm{Cut}(p)$ is of measure zero in M, we get

$$\mathrm{Vol}(B_r(p)) = \int_{B_r(0) \cap \Sigma(p)} \mu(t, \Theta) dt d\Theta = \int_{B_r(0)} \mu(t, \Theta) dt d\Theta.$$

Example. We may calculate the function $\mu_p(r, \Theta)$ for the three model spaces,

- $\mathbb{R}^m$: $\mu(r, \Theta) = r^{m-1}$.
- S^m: $\mu(r, \Theta) = \sin^{m-1}(r)$.
- $\mathbb{H}^m$: $\mu(r, \Theta) = \sinh^{m-1}(r)$.

The Volume Measure via Jacobi Fields

Observe that the three functions $\mu(r, \Theta)$ in the previous example are closely related to the Jacobi fields on the three model spaces. This is not a coincidence:

Proposition 7.2. Given any $\Theta \in S_pM$ and write $\gamma(t) = \exp_p(t\Theta)$. Then for any basis $v_2, \ldots, v_m$ of $\dot\gamma(0)^\perp$, if we let $V_j(t)$ $(i \geqslant 2)$ be the normal Jacobi fields along γ with $V_j(0) = 0$ and $\nabla_{\dot\gamma(0)} V_j = v_j$, then for any point $r\Theta \in \Sigma(p)$,

$$\mu_p(r, \Theta) = \left| \frac{\det(V_2(r), \ldots, V_m(r))}{\det(\nabla_{\dot\gamma(0)} V_2, \ldots, \nabla_{\dot\gamma(0)} V_m)} \right|.$$

Proof. Using $v_1 = \Theta, v_2, \ldots, v_m$ as a basis one can define a set of global linear coordinates $u_1, \ldots, u_m$ on T_pM,

$$u \in T_pM \ \rightsquigarrow \ u = u_1 v_1 + \cdots + u_m v_m.$$

Then we have

$$du_1 \cdots du_m = \frac{r^{m-1}}{|\det(v_1, \ldots, v_m)|} dr d\Theta.$$

Since V_j is a Jacobi field along γ with $V_j(0) = 0$, we have

$$V_j = t(d\exp_p)_{t\Theta}(v_j).$$

Note that under $\exp_p^{-1}$, $(u_1, \ldots, u_m)$ also gives a coordinate system near $\exp_p(r\Theta)$ for $r\Theta \in \Sigma(p)$. With this coordinate system, we have (at $\exp_p(r\Theta)$)

$$\partial_1 = \dot\gamma(r), \quad \partial_j = \frac{1}{r} V_j(r) \quad (j \geqslant 2).$$

It follows

$$g_{ij}(\exp_p(r\Theta)) = \langle \partial_i, \partial_j \rangle \Big|_{\exp_p(r\Theta)} = \frac{1}{r^2}\langle V_i(r), V_j(r)\rangle$$

for $i, j \geqslant 2$. Since γ is a normal geodesic and since each V_j is a normal Jacobi field, we have $\langle \partial_1, \partial_1 \rangle = 1$ and $\langle \partial_1, \partial_i \rangle = 0$ for $i \geqslant 2$. So we get, at $\exp_p(r\Theta)$,

$$G = \det(g_{ij}) = r^{-2m+2}\det(\langle V_i, V_j\rangle)_{i,j\geqslant 2} = r^{-2m+2}\det(V_2(r), \ldots, V_m(r))^2.$$

It follows

$$\mathrm{d}V_g = \sqrt{G}(\exp_p(r\Theta))\mathrm{d}u_1 \cdots \mathrm{d}u_m = \left| \frac{\det(V_2(r), \ldots, V_m(r))}{\det(\nabla_{\dot\gamma(0)}V_2, \ldots, \nabla_{\dot\gamma(0)}V_m)} \right| \mathrm{d}r\mathrm{d}\Theta.$$

This completes the proof. □

The Volume Measure v.s. the Laplacian of Distance

The crucial observation is

Lemma 7.3. Suppose $\Theta \in S_pM$ and $r\Theta \in \Sigma(p)$, then

$$\frac{\mu_p'(r,\Theta)}{\mu_p(r,\Theta)} = (\Delta d_p)(\exp_p(r\Theta)),$$

where the derivative is taken with respect to r.

Proof. Let $\gamma(t) = \exp_p(t\Theta)$ $(0 \leqslant t \leqslant r)$ be the normal geodesic starting at p in the direction Θ. Consider a parallel orthonormal frame $\{e_i(t)\}$ along γ with $e_1(t) = \dot\gamma(t)$. Let $V_j(t)$ be a Jacobi field along γ such that

$$V_j(0) = 0 \text{ and } V_j(r) = e_j(r).$$

Then at $q = \gamma(r)$ we have

$$\Delta d_p(\exp_p(r\Theta)) = \sum_{i=2}^m (\nabla^2 d_p)_q(e_i(t), e_i(t)) = \sum_{i=2}^m I(V_i, V_i).$$

On the other hand, if we denote $A(t) = (\langle V_i(t), V_j(t)\rangle)_{i,j\geqslant 2}$, then $A(r) = \mathrm{Id}$, and the derivative $d(t) = \det A(t) = (\det(V_2(t), \ldots, V_m(t)))^2$ is

$$d'(t) = d(t)\mathrm{tr}(A^{-1}(t)A'(t)).$$

Thus we get

$$\frac{\mu_p'(r,\Theta)}{\mu_p(r,\Theta)} = \frac{1}{2}\frac{d'(r)}{d(r)} = \frac{1}{2}\mathrm{tr}(A'(r)) = \sum_{j=2}^m \langle V_j(r), \nabla_{\dot\gamma(t)}V_j(r)\rangle = \sum_{j=2}^m I(V_j, V_j),$$

so the conclusion follows. □

Comparison of Volume Elements

In particular if we denote by $\mu_\kappa(r)$ the function $\mu(r,\Theta)$ for the space M_κ^m, i.e.

$$\mu_\kappa(r) = \begin{cases} \sin^{m-1}(\sqrt{\kappa}r), & \kappa > 0, \\ r^{m-1}, & \kappa = 0, \\ \sinh^{m-1}(\sqrt{-\kappa}r), & \kappa < 0, \end{cases}$$

then by Laplace comparison theorem,

$$\frac{\mu_p'(r,\Theta)}{\mu_p(r,\Theta)} \leqslant \frac{\mu_\kappa'(r)}{\mu_\kappa(r)}$$

for any complete Riemannian manifold with $Ric \geqslant (m-1)\kappa$, as long as $r\Theta \in \Sigma(p)$.

Proposition 7.4. If (M, g) is a complete Riemannian manifold with $Ric \geqslant (m-1)\kappa$, then for any fixed $\Theta \in S_pM$,

(1) the function $\dfrac{\mu_p(r,\Theta)}{\mu_\kappa(r)}$ is non-increasing in r.

(2) $\lim\limits_{r\to 0^+} \dfrac{\mu_p(r,\Theta)}{\mu_\kappa(r)} = 1$, and thus $\mu_p(r,\Theta) \leqslant \mu_\kappa(r)$ for all $r > 0$.

(3) if $\mu_p(t,\Theta) = \mu_\kappa(t)$ for $t \in [a, r]$ and any Θ, then $B(p, r)$ is isometric to $B_\kappa(r)$.

Proof. (1) The monotonicity of $\dfrac{\mu_p(r,\Theta)}{\mu_\kappa(r)}$ follows from

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\log \frac{\mu_p(t,\Theta)}{\mu_\kappa(t)}\right) = \frac{\mu_p'(t,\Theta)}{\mu_p(t,\Theta)} - \frac{\mu_\kappa'(t)}{\mu_\kappa(t)} \leqslant 0.$$

(2) By $V_j(r) = r\nabla_{\dot\gamma(0)}V_j + \mathcal{O}(r^2)$ we get $\mu_p(r,\Theta) = r^{m-1} + \mathcal{O}(r^m)$. The result follows.
(3) If $\mu_p(t,\Theta) = \mu_\kappa(t)$ for $t \leqslant r$ and any Θ, then

$$(\Delta d_p)(\exp_p(t\Theta)) = \frac{\mu_p'(t,\Theta)}{\mu_p(t,\Theta)} = \frac{\mu_\kappa'(t)}{\mu_\kappa(t)} = (\Delta_\kappa d_\kappa)(t).$$

It follows that $Ric(\dot\gamma(t)) = (m-1)\kappa$, and (since Θ and thus γ are arbitrary) any normal Jacobi field along any geodesic starting at p is almost parallel. By Problem Sheet 4, (M, g) has constant sectional curvature, and thus the constant has to be κ. Finally by Cartan's local isometry theorem, $B(p, r)$ is isometric to $B_\kappa(r)$ in M_κ^m. □

The Bishop-Gromov Volume Comparison Theorem

Note that for t large, it may happen that $t\Theta \notin \Sigma(p)$. However, in this case $\mu_p(t,\Theta) = 0$. So the monotonicity holds for all t. It is this simple observation that leads to a global comparison instead of a local comparison inside the interior radius. In fact, by integrating the volume density we get

Theorem 7.5 (Bishop-Gromov)**.** If (M, g) is a complete Riemannian manifold with $Ric \geqslant (m-1)\kappa$, and $p \in M$ is an arbitrary point. Let $S_\kappa(r)$ and $B_\kappa(r)$ be the metric sphere and the metric ball of radius r in M_κ^m. Then the functions

$$\frac{\mathrm{Area}(S(p,r))}{\mathrm{Area}(S_\kappa(r))} \text{ and } \frac{\mathrm{Vol}(B(p,r))}{\mathrm{Vol}(B_\kappa(r))}$$

are non-increasing in r, and both tends to 1 as $r \to 0^+$. Moreover, the quotient is a constant for $r \in [r_1, r_2]$ if and only if $B(p, r_2)$ is isometric to $B_\kappa(r_2)$.

Proof. By definition

$$\mathrm{Area}(S(p,r)) = \int_{S^{m-1}} \mu_p(r,\Theta)\mathrm{d}\Theta,$$
$$\mathrm{Vol}(B(p,r)) = \int_0^r \int_{S^{m-1}} \mu_p(r,\Theta)\mathrm{d}\Theta\mathrm{d}r.$$

Thus if we denote the surface area of the sphere $S^{m-1} \subset \mathbb{R}^m$ by ω_{m-1}, then

$$\frac{\mathrm{Area}(S(p,r))}{\mathrm{Area}(S_\kappa(r))} = \frac{\int_{S^{m-1}} \mu_p(r,\Theta)\mathrm{d}\Theta}{\int_{S^{m-1}} \mu_\kappa(r)\mathrm{d}\Theta} = \frac{1}{\omega_{m-1}} \int_{S^{m-1}} \frac{\mu_p(r,\Theta)}{\mu_\kappa(r)}\mathrm{d}\Theta$$

is non-increasing and tends to 1 as $r \to 0$. As a consequence,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}r}\left(\log \frac{\mathrm{Vol}(B(p,r))}{\mathrm{Vol}(B_\kappa(r))}\right) &= \frac{\mathrm{Area}(S(p,r))}{\mathrm{Vol}(B(p,r))} - \frac{\mathrm{Area}(S_\kappa(r))}{\mathrm{Vol}(B_\kappa(r))} \\ &= \frac{\displaystyle\int_0^r (\mathrm{Area}(S(p,r))\mathrm{Area}(S_\kappa(t)) - \mathrm{Area}(S_\kappa(r))\mathrm{Area}(S(p,r)))\mathrm{d}t}{\mathrm{Vol}(B(p,r))\mathrm{Vol}(B_\kappa(r))} \\ &\leqslant 0\end{aligned}$$

and thus $\dfrac{\mathrm{Vol}(B(p,r))}{\mathrm{Vol}(B_\kappa(r))}$ is also non-increasing in r, and tends to 1 as $t \to 0$. ☐

Corollary 7.6. We have

$$\mathrm{Area}(S(p,r)) \leqslant \mathrm{Area}(S_\kappa(r)), \text{ and } \mathrm{Vol}(B(p,r)) \leqslant \mathrm{Vol}(B(p,r))$$

for all $r \geqslant 0$. Moreover, equality holds if and only if $B(p,r)$ is isometric to $B_\kappa(r)$.

7.2 Applications of the Volume Comparison Theorem

In last section we proved the Bishop-Gromov comparison theorem, namely for any complete Riemannian manifold satisfying $Ric \geqslant (m-1)\kappa$, the functions

$$\frac{\mathrm{Area}(S(p,r))}{\mathrm{Area}(S_\kappa(r))} \text{ and } \frac{\mathrm{Vol}(B(p,r))}{\mathrm{Vol}(B_\kappa(r))}$$

are non-increasing in r, and both tends to 1 as $r \to 0^+$. Today we shall give some applications.

7.2.1 Applications to Geometric Quantities

As an application of the volume comparison theorem, we get a one-sentence proof of Bonnet-Myers theorem: If $d(p,q) > \dfrac{\pi}{\sqrt{\kappa}}$, then $0 < \mathrm{Area}\left(S\left(p, \dfrac{\pi}{\sqrt{\kappa}}\right)\right) \leqslant \mathrm{Area}\left(S_\kappa\left(\dfrac{\pi}{\sqrt{\kappa}}\right)\right) = 0$, contradiction.

As an immediate consequence,

Corollary 7.7. Let (M,g) be a complete Riemannian manifold with $Ric \geqslant (m-1)\kappa > 0$, then $\mathrm{Vol}(M,g) \leqslant \mathrm{Vol}(S^m_\kappa)$, and the equality holds if and only if (M,g) is isometric to S^m_κ.

In fact, with a bit more work, one can prove

Theorem 7.8 (S. Y. Cheng)**.** Let (M,g) be a complete Riemannian manifold with $Ric \geqslant (m-1)\kappa > 0$, and $\mathrm{diam}(M,g) = \dfrac{\pi}{\sqrt{\kappa}}$, then (M,g) is isometric to S^m_κ.

Proof. By Bishop-Gromov volume comparison theorem, for any $p \in M$,

$$\frac{\mathrm{Vol}\left(B\left(p, \frac{\pi}{2\sqrt{\kappa}}\right)\right)}{\mathrm{Vol}(M)} = \frac{\mathrm{Vol}\left(B\left(p, \frac{\pi}{2\sqrt{\kappa}}\right)\right)}{\mathrm{Vol}(B_{\pi/\sqrt{\kappa}}(p))} \geqslant \frac{\mathrm{Vol}\left(B_\kappa\left(\frac{\pi}{2\sqrt{\kappa}}\right)\right)}{\mathrm{Vol}\left(B_\kappa\left(\frac{\pi}{\sqrt{\kappa}}\right)\right)} = \frac{1}{2}.$$

Now let p, $q \in M$ so that $\mathrm{dist}(p,q) = \dfrac{\pi}{\sqrt{\kappa}}$. Then the above inequality implies

$$\mathrm{Vol}\left(B\left(p, \frac{\pi}{2\sqrt{\kappa}}\right)\right) \geqslant \frac{1}{2}\mathrm{Vol}(M), \qquad \mathrm{Vol}\left(B\left(q, \frac{\pi}{2\sqrt{\kappa}}\right)\right) \geqslant \frac{1}{2}\mathrm{Vol}(M).$$

Since $B\left(p, \frac{\pi}{2\sqrt{\kappa}}\right) \cap B\left(q, \frac{\pi}{2\sqrt{\kappa}}\right) = \varnothing$, we must have

$$\frac{\mathrm{Vol}\left(B\left(p, \frac{\pi}{2\sqrt{\kappa}}\right)\right)}{\mathrm{Vol}(M)} = \frac{\mathrm{Vol}\left(B_\kappa\left(\frac{\pi}{2\sqrt{\kappa}}\right)\right)}{\mathrm{Vol}\left(B_\kappa\left(\frac{\pi}{\sqrt{\kappa}}\right)\right)} = \frac{1}{2}, \quad \frac{\mathrm{Vol}\left(B\left(q, \frac{\pi}{2\sqrt{\kappa}}\right)\right)}{\mathrm{Vol}(M)} = \frac{\mathrm{Vol}\left(B_\kappa\left(\frac{\pi}{2\sqrt{\kappa}}\right)\right)}{\mathrm{Vol}\left(B_\kappa\left(\frac{\pi}{\sqrt{\kappa}}\right)\right)} = \frac{1}{2}.$$

So $\overline{B\left(p, \frac{\pi}{2\sqrt{\kappa}}\right)} \cup \overline{B\left(q, \frac{\pi}{2\sqrt{\kappa}}\right)} = M$. According to Bishop-Gromov comparison theorem, $B\left(p, \frac{\pi}{2\sqrt{\kappa}}\right)$ and $B\left(q, \frac{\pi}{2\sqrt{\kappa}}\right)$ are both isometric to hemisphere in S^m_κ. It follows that $\mathrm{Vol}(M, g) = \mathrm{Vol}(S^m_\kappa)$ and thus M is isometric to M^m_κ. □

Remark. The maximal diameter theorem was first proved by Toponogov under the stronger assumption $K \geqslant \kappa$.

Volume Growth Rate

Another immediate consequence of volume comparison theorem is

Corollary 7.9. Let (M, g) be a complete Riemannian manifold with $Ric \geqslant 0$, then

$$\mathrm{Vol}(B(p, r)) \leqslant \mathrm{Vol}(B_0(r)) = \omega_m r^m,$$

where ω_m is the volume of unit ball in $\mathbb{R}^m$.

Note that for any p, q, with $l = d(p, q)$, one has

$$B(p, r) \subset B(q, r + l) \subset B(p, r + 2l).$$

It follows that the **asymptotic volume ratio**

$$\alpha_M := \lim_{r \to \infty} \frac{\mathrm{Vol}(B(p, r))}{\omega_m r^m}$$

is independent of p, and $\alpha_M \leqslant 1$, with equality if and only if (M, g) is isometric with $(\mathbb{R}^m, g_0)$.

Remark. We say (M, g) has **large volume growth** [or **Euclidean volume growth**] if $\alpha_M > 0$. It has been proved by Li (1986) and Anderson (1990) that $|\pi_1(M)|$ is bounded by $\frac{1}{\alpha_M}$, and in particular, M is simply connected if $\alpha_M > \frac{1}{2}$. In 1994 Perelman proved that there exists $\delta_m > 0$ such that if $\alpha_M \geqslant 1 - \delta_m$, then M is contractible.

Remark. The asymptotic volume ratio α_M appears in many problems. For example, S. Brendle recently proved the following isoperimetric inequality: Let (M, g) be a complete noncompact Riemannian manifold with nonnegative Ricci curvature, then for any compact domain Ω in M with boundary $\partial\Omega$,

$$\frac{(\mathrm{Area}(\partial\Omega))^{1/(m-1)}}{(\mathrm{Vol}(\Omega))^{1/m}} \geqslant \alpha_M^{1/m(m-1)} \frac{(\mathrm{Area}(\partial B_0(1)))^{1/(m-1)}}{(\mathrm{Vol}(B_0(1)))^{1/m}}.$$

It is easy to construct manifolds with $Ric \geqslant 0$ and $\alpha_M = 0$. For example, the Riemannian manifold $M = S^\kappa \times \mathbb{R}^l$ (with the product metric) has nonnegative Ricci curvature, and the volume $\mathrm{Vol}(B(p, r))$ is of order r^l. In particular, $S^{m-1} \times \mathbb{R}$ has "linear" volume growth. It turns out that for any complete non-compact Riemannian manifold (M, g) with $Ric \geqslant 0$, this is the least possible volume growth:

Theorem 7.10 (Calabi-Yau)**.** Let (M, g) be a complete non-compact Riemannian manifold with $Ric \geqslant 0$. Then there exists a positive constant c depending only on p and m so that

$$\mathrm{Vol}(B(p, r)) \geqslant cr, \quad \forall r > 2.$$

Proof. (Following Gromov.) Since M is complete and non-compact, for any $p \in M$ there exists a ray, i.e. a geodesic $\gamma : [0, \infty) \to M$ with $\gamma(0) = p$ such that $\mathrm{dist}(p, \gamma(t)) = t$ for all $t > 0$. (Exercise: Prove the existence of a ray.)

For any $t > \frac{3}{2}$, using the Bishop-Gromov volume comparison theorem, we get

$$\frac{\mathrm{Vol}(B(\gamma(t), t+1))}{\mathrm{Vol}(B(\gamma(t), t-1))} \leqslant \frac{\omega_m (t+1)^m}{\omega_m (t-1)^m} = \frac{(t+1)^m}{(t-1)^m}.$$

On the other hand, by triangle inequality, $B(p, 1) \subset B(\gamma(t), t+1) \setminus B(\gamma(t), t-1)$. So

$$\frac{\mathrm{Vol}(B(p,1))}{\mathrm{Vol}(B(\gamma(t), t-1))} \leqslant \frac{\mathrm{Vol}(B(\gamma(t), t+1) \setminus B(\gamma(t), t-1))}{\mathrm{Vol}(B(\gamma(t), t-1))} \leqslant \frac{(t+1)^m - (t-1)^m}{(t-1)^m},$$

i.e.

$$\mathrm{Vol}(B(\gamma(t), t-1)) \geqslant \mathrm{Vol}(B(p,1)) \frac{(t-1)^m}{(t+1)^m - (t-1)^m} \geqslant C(m) \mathrm{Vol}(B(p,1)) t,$$

where $C(m)$ is the infimum of the function $\frac{1}{t} \frac{(t-1)^m}{(t+1)^m - (t-1)^m}$ on $\left[\frac{3}{2}, \infty\right)$, which is positive. Now the theorem follows the fact

$$B(p, r) \supset B\left(\gamma\left(\frac{r+1}{2}\right), \frac{r+1}{2} - 1\right).$$

□

Remark. A complete non-compact Riemannian manifolds with

$$\liminf_{r \to \infty} \frac{\mathrm{Vol}(B(p,r))}{r} = C > 0$$

are called manifold with linear volume growth. Unlike the maximal volume growth case, the constant $c = c(p, m)$ depends on p and may tends to 0 as $p \to \infty$.

7.2.2 Applications to the Fundamental Group

The Fundamental Group of Compact Manifolds

We start with a theorem concerning the topology of manifolds:

Theorem 7.11. For any compact manifold M, $\pi_1(M)$ is finitely generated.

For a topological proof, we refer to Hatcher's book Algebraic Topology (Corollary A.8 and A.9). In what follows we will give a couple geometric proofs. As we can see, by introducing a Riemannian metric on M, we are able to find a nice generator set that satisfies additional requirements and thus is better in applications.

To state the geometric versions of Theorem 7.11, we first endow a Riemannian metric g on M. As usual we let $\pi : \widetilde{M} \to M$ be the universal covering, endowed with the pull-back metric $\widetilde{g} = \pi^* g$. Fix any $\widetilde{p} \in \widetilde{M}$ and consider the fundamental domain centered at $\widetilde{p}$,

$$K = \{\widetilde{q} \in \widetilde{M} \mid d(\widetilde{q}, \widetilde{p}) \leqslant d(\widetilde{q}, g \cdot \widetilde{p}),\ \forall e \neq g \in \pi_1(M)\}.$$

For simplicity we denote $d_0 = \mathrm{diam}(M, g)$. It is not hard to check

(1) K is compact and $\pi(K) = M$.

(2) For any $e \neq g \in \pi_1(M)$, $g \cdot K \cap K \subset \partial K$.

(3) $\widetilde{M} = \bigcup_{g \in \pi_1(M)} g \cdot K$.

(4) $K \subset B(\widetilde{p}, d_0)$.

Our first geometric strengthened version of Theorem 7.11 is a set of generators with an estimate of the "length" of any element via such generators:

Theorem 7.12. The set

$$\Gamma = \{g \in \pi_1(M) : \exists x, y \in K \text{ such that } d(x, g \cdot y) \leqslant 1\}$$

is a finite generator set of $\pi_1(M)$. Moreover, for any $g \in \pi_1(M)$, if there exists x, $y \in K$ such that $d(x, g \cdot y) \leqslant s$, then there exists $g_1, \ldots, g_s \in \Gamma$ such that $g = g_1 \cdots g_s$.

Proof. For any $g \in \Gamma$, if we pick x, $y \in K$ so that $d(x, g \cdot y) \leqslant 1$, then for any $\widetilde{q} \in K$,

$$d(g \cdot \widetilde{q}, \widetilde{p}) \leqslant d(g \cdot \widetilde{q}, g \cdot \widetilde{p}) + d(g \cdot \widetilde{p}, g \cdot y) + d(g \cdot \widetilde{p}, g \cdot y) + d(g \cdot y, x) + d(x, \widetilde{p}) \leqslant 3d_0 + 1.$$

So we get $g \cdot K \subset B(\widetilde{p}, 3d_0 + 1)$ and as a result, Γ is a finite set.

To prove the second conclusion, we proceed by induction. The conclusion holds trivially for $s = 1$. Suppose it holds for $s = k$, and there exists x, $y \in K$ such that $d(x, g \cdot y) \leqslant k + 1$. On the minimizing geodesic connecting x and $g \cdot y$, one can find a point, which, in view of (3) above, can be written as $g' \cdot z$ for some $g' \in \pi_1(M)$ and $z \in K$, such that $d(x, g' \cdot z) \leqslant 1$ and $d(g' \cdot z, g \cdot y) \leqslant k$. By induction hypothesis, $g' \in \Gamma$ and $(g')^{-1} g = g_1 \cdots g_k$ for some $g_1, \ldots, g_k \in \Gamma$. This completes the proof. □

In a second geometric strengthened version of Theorem 7.11, we give another set of generators with relations of given form:

Theorem 7.13 (Gromov)**.** Let (M, g) be a compact Riemannian manifold. Then one can find a finite generator set $\Gamma = \{g_1, \ldots, g_n\}$ so that

- $d(\widetilde{p}, g_i \cdot \widetilde{p}) \leqslant 2d_0$ for $1 \leqslant i \leqslant n$.
- all relations among these generators are of the form $g_i g_j g_k^{-1} = e$.

Proof. Since M is compact, one can find a triangulation of M so that the distance between any pair of adjacent vertices is less than $\varepsilon < \mathrm{inj}(M)$ and such that any loop is homotopic to a loop in the 1-skeleton of the triangulation. Let $\{v_1, \ldots, v_q\}$ be the set of vertices, and e_{ij} the set of edges (so e_{ij} is only defined for some i, j). Fix $p = \pi(\widetilde{p})$ and realize $\pi_1(M)$ with basepoint p. Let σ_i be the minimal geodesic from p to v_i. Then for any edge e_{ij}, $\sigma_{ij} := \sigma_i e_{ij} \sigma_j^{-1}$ is a loop based at p, with

$$L(\sigma_{ij}) \leqslant 2d_0 + \varepsilon.$$

Since any loops in M based at p is homotopic to a loop in the 1-skeleton of the triangulation, and since $\sigma_i e_{ij} e_{jk} \sigma_k^{-1} \sim \sigma_i \sigma_{ij} \sigma_{jk} \sigma_k^{-1}$, the loops σ_{ij}'s generates $\pi_1(M)$.

Observe that if three vertices are adjacent to each other, then they span a 2-simplex Δ_{ijk}, and the loop $\sigma_{ij} \sigma_{jk} \sigma_{ki}$ is null-homotopic, i.e. $\sigma_{ij} \sigma_{jk} \sigma_{ik}^{-1} = e$. So each 2-simplex gives rise to a relation among σ_{ij}'s of the given form. Conversely, if σ is null-homotopic loop in the 1-skeleton based at p, then σ is contractible in the 2-skeleton, and thus there is set of 2-simplex Δ_{ijk} so that the relation $\sigma = e$ is a product of the elementary relations of the form $\sigma_{ij} \sigma_{jk} \sigma_{ik}^{-1} = e$.

Finally by discreteness of the action of $\pi_1(M)$ on $\widetilde{M}$ one has, for ε small enough,

$$\{g \in \pi_1(M) \mid d(\widetilde{p}, g \cdot \widetilde{p}) < 2d_0 + \varepsilon\} = \{g \in \pi_1(M) \mid d(\widetilde{p}, g \cdot \widetilde{p}) \leqslant 2d_0\}.$$

This finishes the proof. □

Detour: Growth of a Finitely Generated Group

We give some abstract definitions in algebra. Let G be a finitely generated group, and $\Gamma = \{g_1, \ldots, g_N\}$ a generator set. The **growth function** of G with respect to Γ is defined to be the number of group elements that can be represented as a product of at most k generators, i.e.

$$N_G^\Gamma(k) = \#\{g \in G \mid \exists l \leqslant k \text{ and } g_{i_1}, \ldots, g_{i_l} \in \Gamma, \text{ s.t.} g = g_{i_1}^{\pm 1} \cdots g_{i_l}^{\pm 1}\}.$$

We will use $|g|$ to represent the smallest l such that $g = g_{i_1}^{\pm 1} \cdots g_{i_l}^{\pm 1}$, and call it the **length** of g with respect to the given generator set.

Definition 7.14. Let G be finitely generated, and Γ is a finite set that generates G.

(1) We say that G is of **polynomial growth** (of order n) if

$$N_G^\Gamma(k) \leqslant ck^n$$

for some constant c depending only on G, Γ.

(2) We say G is of **exponential growth** if there is $c > 0$ and $a > 1$ such that

$$N_G^\Gamma(k) \geqslant ca^k.$$

Remark. Note that if Γ' is another finite set of generators, then there exists integers c_1, c_2 so that any element of Γ can be represented via at most c_1 elements of Γ', and any element of Γ' can be represented via at most c_2 elements of Γ. It follows that

$$N_G^\Gamma(k) \geqslant N_G^{\Gamma'}(c_1 k), \qquad N_G^{\Gamma'} \geqslant N_G^\Gamma(c_2 k).$$

So the concept of polynomial/exponential growth is independent of the choice of the generator set.

Example. Here are two simple examples:

- For $G = \mathbb{Z} \oplus \mathbb{Z}$, we may take $\Gamma = \{(1,0),(0,1)\}$ and it is quite obvious that $N_G^\Gamma(k) \approx 2k^2$. It follows that $\mathbb{Z} \oplus \mathbb{Z}$ has quadratic growth.
- For $G = \mathbb{Z} * \mathbb{Z}$, we may take $\Gamma = \{(1,0),(0,1)\}$ and it is not hard to get $N_G^\Gamma(k) \approx 4\sum_{l=0}^{k} 3^l = 2 \cdot 3^{k+1}$. It follows that $\mathbb{Z} * \mathbb{Z}$ has exponential growth.

Growth of Fundamental Group: Negative Curvature Case

Back to Riemannian manifolds. Let (M,g) be a compact Riemannian manifold. According to Theorem 7.11, $\pi_1(M)$ is finitely generated. It is natural to study its growth rate. As one can imagine, the growth rate of $\pi_1(M)$ is related to the volume growth rate of the Riemannian universal covering $(\widetilde{M}, \widetilde{g})$ of (M,g), since the fundamental group $\pi_1(M)$ acts isometrically on $\widetilde{M}$ as the group of deck transformations.

In view of Lohkamp's result that any smooth manifold M of dimension $m \geqslant 3$ admits a Riemannian metric with $Ric < 0$, there is no control to $\pi_1(M)$ for manifolds with negative Ricci curvature. It turns out that for a compact manifold to admit negative sectional curvature, then its fundamental group must be of exponential growth. To prove this we need the following volume comparison theorem (in reverse direction inside injectivity radius, with sectional curvature upper bound condition) whose proof is left as an exercise:

Theorem 7.15. If (M,g) is a complete Riemannian manifold with sectional curvature $K \leqslant \kappa$, then for $r < \min\left(\mathrm{inj}(p), \dfrac{\pi}{\sqrt{\kappa}}\right)$, one has

(1) The functions $\dfrac{\mathrm{Area}(S(p,r))}{\mathrm{Area}(S_\kappa(r))}$ and $\dfrac{\mathrm{Vol}(B(p,r))}{\mathrm{Vol}(B_\kappa(r))}$ are non-decreasing in r,

(2) we have

$$\mathrm{Area}(S(p,r)) \geqslant (S_\kappa(r)), \text{ and } \mathrm{Vol}(B(p,r)) \geqslant \mathrm{Vol}(B(p,r)).$$

Moreover, equality holds if and only if $B(p,r)$ is isometric to $B_\kappa(r)$.

Using this we can prove

Theorem 7.16 (Milnor)**.** Let (M,g) be a compact Riemannian manifold with $K < 0$, then $\pi_1(M)$ has exponential growth.

Proof. Consider the generator set Γ as in Theorem 7.12. For any $y \in B(\widetilde{p}, d_0 + s)$, there exists $g \in \pi_1(M)$ so that $y \in g.K$. By Theorem 7.12, $|g| \leqslant s$ with respect to Γ. In other words, $B(\widetilde{p}, d_0 + s)$ can be covered by $N_G^\Gamma(s)$ sets of the form $g \cdot K$. So

$$\mathrm{Vol}(B(\widetilde{p}, d_0 + s)) \leqslant N_G^\Gamma(s)\mathrm{Vol}(K).$$

On the other hand, since M is compact, there exists $\kappa > 0$ so that $K \leqslant -\kappa$. Note that in this case the injectivity radius is $+\infty$. So by Theorem 7.15,

$$\mathrm{Vol}(B(\widetilde{p}, d_0 + s)) \geqslant \mathrm{Vol}(B_{-\kappa}(d_0 + s)).$$

Finally use the fact

$$\mathrm{Vol}(B_{-\kappa}(r)) = \int_0^r \int_{S^{m-1}} \left(\frac{\sinh(\sqrt{\kappa} t)}{\sqrt{\kappa}} \right)^{m-1} \mathrm{d}\Theta \mathrm{d}t \geqslant ce^{\sqrt{\kappa} r}$$

we get the desired estimate

$$N_G^\Gamma(s) \geqslant \frac{\mathrm{Vol}(B(\widetilde{p}, d_0 + s))}{\mathrm{Vol}(K)} \geqslant \frac{\mathrm{Vol}(B_{-\kappa}(d_0 + s))}{\mathrm{Vol}(K)} \geqslant ce^{\sqrt{\kappa}(d_0+s)}.$$

□

Growth of Fundamental Group: Non-compact

From the proof above we see that reason for "the fundamental group of 'a compact Riemannian manifold with negative sectional curvature' to be of exponential grown" is that the geodesic balls in the negative curvature space M_κ^m have exponential volume growth. On the other hand, the geodesic balls in the Euclidean space M_0^m has volume growth of order m. So one may naturally anticipate

Proposition 7.17. Let (M, g) be a compact Riemannian manifold with $Ric \geqslant 0$, then $\pi_1(M)$ has polynomial growth of order $\leqslant m$.

This is true, and it is a direct consequence of the following more general theorem:

Theorem 7.18 (Milnor)**.** Let M be a complete Riemannian manifold with $Ric \geqslant 0$ and let $G \subset \pi_1(M)$ be any finitely generated subgroup. Then G has polynomial growth of order no more than m.

Proof. Let Γ be a finite set of generators of G. Fix a point $\widetilde{p} \in \widetilde{M}$ and let

$$l = \max\{\mathrm{dist}(\widetilde{p}, g_i\widetilde{p}) \mid g_i \in \Gamma\}.$$

Then by triangle inequality, for any $g = g_{i_1} \cdots g_{i_k} \in \Gamma^k \subset G$, $\mathrm{dist}(\widetilde{p}, g\widetilde{p}) \leqslant kl$. On the other hand side, we can pick

$$\delta = \frac{1}{3} \min\{\mathrm{dist}(\widetilde{p}, g\widetilde{p}) \mid e \neq g \in G\} > 0$$

so that the balls $B(g\widetilde{p}, \delta)$ are all disjoint for $g \in G$. So $B(\widetilde{p}, kl + \delta) \supset \bigcup_{g \in \Gamma^k} B(g\widetilde{p}, \delta)$ and thus

$$\mathrm{Vol}(B(\widetilde{p}, kl + \delta)) \geqslant N_G^\Gamma \mathrm{Vol}(B(p, \delta)).$$

Since the Riemannian universal covering $(\widetilde{M}, \widetilde{g})$ has non-negative Ricci curvature, applying the Bishop-Gromov's volume comparison theorem we get

$$N_G^\Gamma \leqslant \frac{\mathrm{Vol}(B(\widetilde{p}, kl + \delta))}{\mathrm{Vol}(B(p, \delta))} \leqslant \frac{(kl + \delta)^m}{\delta^m} \leqslant ck^m.$$

□

As a consequence one gets

Corollary 7.19. For $g \geqslant 2$, $\Sigma_g \times \mathbb{R}^k$ admits no complete metric of $Ric \geqslant 0$.

Example. Let H be the Heisenberg group

$$H = \left\{ \begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} : x, y, z \in \mathbb{R} \right\}$$

and let $H_{\mathbb{Z}}$ be the subgroup that contains elements with $x, y, z \in \mathbb{Z}$. Then one can show that the growth rate of $H_{\mathbb{Z}}$ is of order 4. So $H/H_{\mathbb{Z}}$ admits no metric with $Ric \geqslant 0$.

Remark. According to a celebrated theorem of Gromov, if G is finitely generated and has polynomial growth, then G is virtually nilpotent [i.e. G contains a nilpotent subgroup of finite index]. As we have seen in Section 6.2, if (M, g) has sectional curvature $K \geqslant 0$, then $\pi_1(M)$ is finitely generated. It was conjectured by Milnor in 1968 that

Conjecture. For any complete Riemannian manifold (M, g) with $Ric \geqslant 0$, the fundamental group $\pi_1(M, g)$ is finitely generated.

The conjecture was proved for $m = 3$ by Liu in 2013, and was disproved for $m \geqslant 7$ by Brue-Naber-Semola in 2023. It is still open for $m = 4, 5, 6$.

Finiteness of Fundamental Group

Finally we apply Theorem 7.13 and the volume comparison to prove

Theorem 7.20 (Anderson)**.** In the set of all Riemannian manifolds M with

$$Ric \geqslant (m-1)\kappa, \quad \mathrm{Vol}(M, g) \geqslant V, \quad \mathrm{diam}(M) \leqslant D,$$

there are only finitely many isomorphism types of the fundamental groups $\pi_1(M)$.

Proof. Let $\Gamma = \{g_1, \ldots, g_n\}$ be a set of generators as in Theorem 7.13. Then there are at most 2^{n^3} isomorphism types of abstract groups G generated by Γ. It remains to show that n is bounded, which is a consequence of volume comparison. In fact, consider the fundamental domain K again. Then $g \cdot K$'s are disjoint. Moreover, for each $g \in \Gamma$, $g \cdot K$ is contained in $B(\widetilde{p}, 4d_0)$. So by volume comparison.

$$nV \leqslant n\mathrm{Vol}(M, g) = m\mathrm{Vol}(K) \leqslant \mathrm{Vol}(B(\widetilde{p}, 4d_0)) \leqslant \mathrm{Vol}(B_\kappa(4d_0)) \leqslant \mathrm{Vol}(B_\kappa(4D))$$

and thus the conclusion follows. □

Remark. This is one of the many finiteness theorems that people have proved for Riemannian manifolds satisfying suitable bounds on curvature/volume/diameter etc. Note that this theorem fails if we drop the volume lower bound or diameter upper bound condition:

- Consider the set of all lens space $L(p, q) = S^3/\mathbb{Z}_p$. Then they have constant sectional curvature 1 (and thus constant Ricci curvature $m - 1$), and constant diameter[1] $\pi/2$, but volume tends to 0 as p tends to infinity. Obviously this set has infinitely many different fundamental groups.
- Consider the set of surfaces Σ_g endowed with hyperbolic metrics, i.e. with constant sectional curvature -1. Then they have constant Ricci curvature $-(m-1)$, volume bounded below (tends to infinity by Gauss-Bonnet), but the diameters are not bounded above.

[1] To see this one first note that for the lens space $L(p, q)$, viewed as $S^3 \subset \mathbb{C}^2$ quotient by the action $k \cdot (z_1, z_2) = (\mathrm{e}^{2\pi ik/p} z_1, \mathrm{e}^{2\pi ikq/p} z_2)$, any smooth curve connecting $\pi(1, 0)$ and $\pi(0, 1)$ is the quotient of a curve connecting $(1, 0)$ and a point of the form $(0, \mathrm{e}^{2\pi il/p})$. So $\mathrm{diam}(L(p, q)) \geqslant \pi/2$. On the other hand, by the Grove-Shiohama theorem that we will prove next time, we can't have $\mathrm{diam}(L(p, q)) > \pi/2$.

7.3 The Sphere Theorem

7.3.1 Critical Point Theory of Distance Functions

A Glimpse into Morse Theory

As we mentioned in Section 5.4, Morse theory is a basic tool in differential topology relates the topology of M to the critical points of a Morse function on M, and the theory has many applications in Riemannian geometry.

On major theme in Morse theory is to study the change of topology of sub-level sets $M_a = \{x \mid f(x) \leqslant a\}$ as a varies. Two crucial facts in Morse theory are

Theorem 7.21 (Isotopy Lemma)**.** *Suppose $f \in \mathcal{C}^\infty(M)$, $f^{-1}([a,b])$ is compact, and $f^{-1}([a,b]) \cap \mathrm{Crit}(f) = \varnothing$. Then M_a is diffeomorphic [and is a deformation retract] to M_b.*

Idea of Proof. "Push" M_b down to M_a along trajectories of $\frac{\nabla f}{|\nabla f|}$ (which are of constant speed and are perpendicular to each level set $f = c$). The topology is not changed during this procedure. [See my notes on smooth manifolds for detail.] □

One can show that on any smooth manifold, there are lots of "good Morse functions" [i.e. the critical points are disjoint, non-degenerate and take different values].

Theorem 7.22. *Suppose $f \in \mathcal{C}^\infty(M)$, p is a non-degenerate critical point of f, $f^{-1}([c-\varepsilon, c+\varepsilon])$ is compact, and $f^{-1}([c-\varepsilon, c+\varepsilon]) \cap \mathrm{Crit}(f) = \{p\}$. Then $M_{c+\varepsilon}$ is homotopy equivalent to "$M_{c-\varepsilon}$ with a λ-cell attached", where λ is the index of p.*

As a result, one can detect the homotopy type of M from a good Morse function. A useful theorem in differential topology that can be used to produce a sphere is

Theorem 7.23 (Brown)**.** *If M is a compact manifold, $M = U_1 \cup U_2$, and U_1, U_2 are both homeomorphic to $\mathbb{R}^m$, then M is homeomorphic to S^m.*

As a consequence, one has

Theorem 7.24 (Reeb)**.** *If M is compact, $f \in \mathcal{C}^\infty(M)$ is a Morse function that has only two critical points, then M is homeomorphic to S^m.*

Proof. The two critical points have to be the maximum/minimum of f. Take a close to the minimal value of f and b closed to the maximal value of f, so that both $f^{-1}((-\infty, a))$ and $f^{-1}((b, +\infty))$ are homeomorphic to $\mathbb{R}^m$. Take b' between b and the maximal value of f. By Theorem 7.21, $f^{-1}((-\infty, a))$ is homeomorphic to $f^{-1}((-\infty, b'))$. Since $M = f^{-1}((-\infty, b')) \cap f^{-1}((b, +\infty))$, by Brown's theorem, M is homeomorphic to S^m. □

Note that in this proof we avoided the use of Theorem 7.22.

Critical Points of the Distance Function

Now let (M, g) be a Riemannian manifold, and $p \in M$ be a point. In some sense the distance functions d_p's are the most natural functions that are defined on M. Although $d_p \notin \mathcal{C}^\infty(M)$, Grove and Shiohama succeeded in developing a Morse theory for d_p in 1977 which played an important role in studying the topology of Riemannian manifolds. To get an idea let's examine the behavior of d_p on (M, g):

- As we have seen, the distance function d_p is smooth at any $q \notin \mathrm{Cut}(p) \cup \{p\}$, with $(\nabla d_p)_q = \dot\gamma(d(p,q))$, where γ is the unique minimizing normal geodesic from p to q. In particular, $|\nabla d_p| = 1$ at any $q \notin \mathrm{Cut}(p) \cup \{p\}$. As a result, these points are not critical points of d_p.
- The singularity of d_p at the point p is not too bad, since it is the only minimum of d_p, and the change of topology near p is well-understood. This point can be regarded as a "trivial critical point" of d_p.

- So we are more interested in those points $q \in \mathrm{Cut}(p)$. They are candidates of critical points for d_p. To get a better idea, let's take a closer look at the example $S^1 \times \mathbb{R}$: given any $p = (\mathrm{e}^{\mathrm{i}\theta}, z_0) \in S^1 \times \mathbb{R}$, $\mathrm{Cut}(p) = \{(\mathrm{e}^{-\mathrm{i}\theta}, z) \mid z \in \mathbb{R}\}$. When will the topology of $M_a = \{q \mid d(p,q) < a\}$ change as a varies? Obviously
  - the topology of M_a will not change for $a < \pi$ [no critical points there],
  - the topology of M_a will change when a pass the value π, i.e. pass the cut point $\widetilde{p} = (\mathrm{e}^{-\mathrm{i}\theta}, z_0)$,
  - the topology of M_a will not change for $a > \pi$, although there are two non-smooth points of d_p for each such a.

  Why the topology for M_a will not change for $a > \pi$? Because although there are two minimizing geodesics meeting at one point $q \in \mathrm{Cut}(p)$ with $d_p(q) = a$, their directions at q lies in the same open half space. As a result, there is one direction that one can "flow-out" M_a to M_b ($b > a$), and that is a direction whose angles with both geodesics are obtuse. Why such "flow-out" argument fails for $a = \pi$, i.e. at $\widetilde{p} = (\mathrm{e}^{-\mathrm{i}\theta}, z_0)$? Because for the two geodesics meeting at $\widetilde{p}$, one can't find such a direction whose angles with both geodesics are obtuse!

We are thus led to the following definition:

Definition 7.25. A point $q \neq p$ is called a **critical point** of d_p [or a critical point of p] if for all $X_q \in T_qM$, there exists a minimizing geodesic γ from $q = \gamma(0)$ to p so that

$$\langle \dot{\gamma}(0), X_q \rangle \geqslant 0,$$

i.e. the angle α between $\dot{\gamma}(0)$ and X_q is no more than $\frac{\pi}{2}$.

The set of all such critical points of d_p will be denoted as $\mathrm{CP}(p)$. Note that if q is not a critical point of d_p, then the tangent vector of all minimizing geodesic from q to p lie in an open half space of T_qM.

Examples of Critical Points of the Distance Function

Example. Here are some immediate examples:

- $M = S^2$ the standard sphere: the only critical point of p is its antipodal $\overline{p}$.
- $M = S^1 \times \mathbb{R}^1$ the cylinder: the only critical point of $({}^{\mathrm{i}\theta}, z)$ is $(\mathrm{e}^{-\mathrm{i}\theta}, z)$.
- $M = S^1 \times S^1$ the flat torus with fundamental domain a square centered at p: the critical points are the two midpoints of the sides and the corner point.
- If γ is closed geodesic of length $2l$ so that both $\gamma\Big|_{[0,l]}$ and $\gamma\Big|_{[l,2l]}$ are minimal, then $\gamma(l)$ is a critical point of $\gamma(0)$.

Recall that for any point $q \notin \mathrm{Cut}(p)$, there is a unique minimizing geodesic joining p to q. So $\exp_p$ is injective on an open ball $B^p(0,r) \subset T_pM$ if $B(p,r) \subset M \setminus \mathrm{Cut}(p)$. Moreover, for most points in $\mathrm{Cut}(p)$, there exists at least two minimizing normal geodesic to p (c.f. Problem Sheet 3). So we conclude

$$\mathrm{inj}_p(M,g) = \mathrm{dist}(p, \mathrm{Cut}(p)) \text{ and } \mathrm{inj}(M,g) = \inf p \in M \mathrm{dist}(p, \mathrm{Cut}(p)).$$

Proposition 7.26. If $q \in \mathrm{Cut}(p)$ is not conjugate to p and

$$d(p,q) = d(p, \mathrm{Cut}(p)),$$

then there are exactly two minimizing normal geodesic γ and σ from p to q, and $\dot{\sigma}(l) = -\dot{\gamma}(l)$. In particular, q is a critical point of p.

Proof. We have seen in Theorem 5.49 that there are at least two minimizing normal geodesic γ, σ from p to q. We shall prove $\dot\gamma(l) = -\dot\gamma(l)$. Suppose not, then there exists $X_q \in T_qM$ with $|X_q| = 1$, such that

$$\langle X_q, \dot\gamma(l)\rangle < 0 \text{ and } \langle X_q, \dot\sigma(l)\rangle < 0.$$

Since q is not conjugate to p along γ, there exists $U \ni l\dot\gamma(0)$ such that $\exp_p\Big|_U$ is a diffeomorphism. Let

$$\xi(s) = (\exp_p|_U)^{-1}\exp_q(sX_q).$$

Then $\gamma_s(t) = \exp_p\left(\frac{t}{l}\xi(s)\right)$ is a geodesic variation of γ. By the first variation formula,

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0} L(\gamma_s) = \langle X_q, \dot\gamma(l)\rangle < 0.$$

So for $s > 0$ small enough, $L(\gamma_s) < L(\gamma) = l$.

Similarly one can construct a geodesic variation

$$\sigma_s = \exp_p\left(\frac{s}{l}\eta(s)\right),$$

where $\eta(s) = (\exp_p|_V)^{-1}\exp_q(sX_q)$ of the minimizing geodesic σ so that $L(\sigma_s) < L(\sigma) = l$ for $s > 0$ small enough. Note that for each s, both γ_s and σ_s are geodesics from p to $\exp_q(sX_q)$. Moreover, for $s > 0$ small enough,

$$l_s := d(p, \exp_q(sX_q)) \leqslant L(\gamma_s) \leqslant l.$$

So $\exp_p$ is NOT injective on $B^p\left(0, \frac{l_s + l}{2}\right)$, which contradicts with the fact that $\exp_p$ is a diffeomorphism on $B^p(0, l)$, since $l = d(p, \mathrm{Cut}(p))$. □

The Isotopy Lemma for d_p

As in the usual Morse theory, the following fact is crucial in all applications.

Theorem 7.27 (The Isometry Lemma)**.** Suppose (M, g) is complete, $b > a > 0$, and $d_p^{-1}([a, b]) \cap \mathrm{CP}(d) = \varnothing$. Then M_a is diffeomorphic [and is a deformation retract] to M_b.

Proof. For any point $q \notin \mathrm{CP}(p)$, then there exists $X_q \in T_qM$ so that for any minimizing geodesic γ from q to p, the angle

$$\angle(X_q, \dot\gamma(0)) < \frac{\pi}{2}.$$

Take a locally finite covering $\{U_{q_i}\}$ of $\overline{B(p, b)} \setminus B(p, a)$ using such neighborhoods, and a smooth partition of unity $\{\rho_i\}$ subordinate to this covering. Let $X = \sum \rho_i X^{q_i}$. Clearly X is a smooth non-vanishing vector field on $\overline{B}(p, b) \setminus B(p, a)$, since

$$\langle X(\overline{q}, \dot{\overline{\gamma}}(0))\rangle = \sum \rho_i \langle X^{q_i}(\overline{X}^{q_i}(\overline{q}), \dot\gamma(0))\rangle > 0, \quad \forall \overline{q} \in \overline{B}(p, b) \setminus B(p, a).$$

We normalize X so that $|X(\overline{q})| = 1$ at each $\overline{q}$, and then repeat the proof of Theorem 7.21. More precisely, for any $\sigma^{\overline{q}}(t) \in \overline{B(p, b)} \setminus B(p, a)$ we let $\overline{\gamma}_t$ be the minimizing geodesic from $\sigma^{\overline{q}}(t)$ to p. Then by the first variation formula,

$$\frac{\mathrm{d}}{\mathrm{d}t}(d_p(\sigma^{\overline{q}}(t))) = \frac{\mathrm{d}}{\mathrm{d}t}L(\overline{\gamma}_t) = -\langle X(\sigma^{\overline{q}}(t)), \dot{\overline{\gamma}}_t(0)\rangle.$$

Fix $t_1 < t_2$ so that $\sigma^{\overline{q}}([t_1, t_2]) \subset \overline{B(p, t_2)} \setminus B(p, t_1)$. By compactness, $\exists \varepsilon > 0$ so that

$$-\langle X(\sigma^{\overline{q}}(t)), \dot{\overline{\gamma}}_t(0)\rangle \leqslant -\cos\left(\frac{\pi}{2} - \varepsilon\right) < 0$$

for all $t \in [t_1, t_2]$. It follows

$$d_p(\sigma^{\overline{q}}(t_2)) - d_p(\sigma^{\overline{q}}(t_1)) = \int_{t_1}^{t_2} \frac{\mathrm{d}}{\mathrm{d}t}(d_p(\sigma^{\overline{q}}(t)))\mathrm{d}t \leqslant -(t_2 - t_1)\cos\left(\frac{\pi}{2} - \varepsilon\right) < 0.$$

So as t increases, d_p is strictly decreasing along the integral curves $\sigma^{\overline{q}}(t)$ of X inside $\overline{B(p,t_2)}\backslash B(p,t_1)$ as. So the flow of X gives the desired diffeomorphism. ☐

Since the topology changes after the "farthest point", we get

Corollary 7.28. Let (M,g) be a compact Riemannian manifold, $p \in M$, and q is a farthest point from p, then q is a critical point of p.

In particular, if $d(p,q) = \mathrm{diam}(M,g)$, then for any $X_p \in T_pM$, there is a minimal geodesic γ from $p = \gamma(0)$ to q so that $\langle \dot\gamma(0), X_p\rangle \geqslant 0$.

The Reeb Theorem for d_p

Although there is no Morse lemma and there is no index for the critical points of a distance function, near the trivial critical point p of d_p the sub-level set is still an m-ball. Similar phenomena happens near a "non-degenerate (=discrete) farthest point". So it is not amazing that we still have the following analogue to the Reeb theorem for d_p:

Corollary 7.29. Let (M,g) be a compact Riemannian manifold and $p \in M$. If d_p has only one non-trivial critical point $q \neq p$, then M is homeomorphic to S^m.

Proof. According to Corollary 7.28, q has too be the only farthest point of p. Take r_1 small so that both $B(p,r_1)$ and $B(q,r_1)$ are homeomorphic to $\mathbb{R}^m$. Take $r_2 \in (r_1, d(p,q))$ large so that $B(p,r_2) \cup B(q,r_1) = M$. Then d_p has no critical point in $\overline{B(p,r_2)} \setminus B(p,r_1)$. By the isometry lemma, $B(p,r_2)$ is homeomorphic to $B(p,r_1)$, and thus is homeomorphic to $\mathbb{R}^m$. By Brown's theorem, M is homeomorphic to S^m. ☐

7.3.2 Some Sphere Theorems

The Diameter Sphere Theorem of Grove-Shiohama

As a first application of the critical point theory of distance function, we shall prove the following diameter sphere theorem:

Theorem 7.30 (Grove-Shiohama)**.** Let (M,g) be a complete connected Riemannian manifold with

$$K > \frac{1}{4} \text{ and } \mathrm{diam}(M,g) \geqslant \pi,$$

then M is homeomorphic to S^m.

Proof. Since M is compact, there exists $\kappa > \frac{1}{4}$ so that $K \geqslant \kappa$. By Bonnet-Myers, $\mathrm{diam}(M,g) \leqslant \frac{\pi}{\sqrt{\kappa}}$. In view of Cheng's maximal diameter theorem, we may assume

$$\mathrm{diam}(M,g) = l < \frac{\pi}{\sqrt{\kappa}}$$

Let p, $q \in M$ so that $d(p,q) = l = \mathrm{diam}(M,g)$. By Corollary 7.28, q is a critical point of p. By Corollary 7.29, it is enough to prove that p has no other critical points.

Suppose to the contrary, $\overline{p} \neq q$ is a critical point of p. Denote $l' = d(p,\overline{q})$ and $l'' = d(q,\overline{q})$. Let γ be a minimizing normal geodesic from $q = \gamma(0)$ to $\overline{q} = \gamma(l'')$. By definition of critical points, there exists a minimizing normal geodesic σ from $\overline{q} = \sigma(0)$ to $p = \sigma(l')$ so that

$$\alpha = \angle(-\dot\gamma(l), \dot\sigma(0)) \leqslant \frac{\pi}{2}.$$

Apply the Toponogov comparison theorem (triangle version), we conclude that there is a geodesic triangle in $S^m\left(\frac{1}{\sqrt{\kappa}}\right)$ whose sides have lengths l, l', l'', so that the opposite angle of l is $\widetilde{\alpha} \leqslant \frac{\pi}{2}$. Since $\pi \leqslant l < \frac{\pi}{\sqrt{\kappa}}$, we get

$$\frac{\pi}{2} \leqslant \frac{l}{2}\sqrt{\kappa}l < \pi \text{ and } 0 < \sqrt{\kappa}l',\ \sqrt{\kappa}l'' < \pi.$$

Then by the cosine law in $S^m\left(\frac{1}{\sqrt{\kappa}}\right)$,

$$\begin{aligned} 0 > \cos(\sqrt{\kappa}l) &= \cos(\sqrt{\kappa}l')\cos(\sqrt{\kappa}l'') + \sin(\sqrt{\kappa}l')\sin(\sqrt{\kappa}l'')\cos(\widetilde{\alpha}) \\ &\geqslant \cos(\sqrt{\kappa}l')\cos(\sqrt{k}l''), \end{aligned}$$

which implies that exactly one of l' and l'' is strictly greater than $\frac{\pi}{2\sqrt{\kappa}}$, and the other is strictly smaller than $\frac{\pi}{2\sqrt{\kappa}}$. Without loss of generality, assume $0 < l'' < \frac{\pi}{2\sqrt{\kappa}}$. Then

$$\cos(\sqrt{\kappa}l) \geqslant \cos(\sqrt{\kappa}l')\cos(\sqrt{\kappa}l'') > \cos(\sqrt{\kappa}l').$$

In other words, $l < l'$. This contradicts with the fact $l = \operatorname{diam}(M, g)$. □

The Topological Sphere Theorem

In 1940s, Hopf asked the following question: If a simply connected complete Riemannian manifold (M, g) has sectional curvature close to 1, is it homeomorphic to or even diffeomorphic to the sphere S^m?

In 1951 Rauch gave a positive answer: he proved that (M, g) is homeomorphic to S^m if it has sectional curvature $\delta \leqslant K \leqslant 1$, where $\delta \approx 0.75$ is the solution to $\sin(\sqrt{\delta}\pi) = \sqrt{\delta}/2$. The pinching constant δ was improved to $1/4$ by Berger (1960, for m even) and Klingenberg (1961, for m odd).

Theorem 7.31 (Topological Sphere Theorem, Rauch-Berger-Klingenberg)**.** Let (M, g) be a complete simply connected Riemannian manifold with $\frac{1}{4} < K \leqslant 1$. Then M is homeomorphic to S^m.

Remark. The constant $\frac{1}{4}$ is sharp, since $\mathbb{CP}^m$ has sectional curvature $\frac{1}{4} \leqslant K \leqslant 1$. In fact, suppose (M, g) is a complete and simply connected, then

(1) (Berger 1983) If m is even, then there exists $\varepsilon(m) > 0$ so that if $\frac{1}{4} - \varepsilon(m) \leqslant K \leqslant 1$, then M is either homeomorphic to S^m or diffeomorphic to one of the CROSSes: $\mathbb{CP}^{m/2}$, $\mathbb{HP}^{m/4}$, $\mathrm{Ca}\mathbb{P}^2$.

(2) (Abresch-Myers, 1994) If m is odd, then there exists $\varepsilon > 0$ so that if $\frac{1}{4} - \varepsilon \leqslant K \leqslant 1$, then M is homeomorphic to S^m.

Remark. The other half of Hopf's question was proved by Brendle and Schoen:

Theorem 7.32 (Differential Sphere Theorem, Brendle-Schoen 2009)**.** Let (M, g) be complete and simply connected, such that any $p \in M$, $0 < \sup_{\Pi_p} K(\Pi_p) < 4 \inf_{\Pi_p} K(\Pi_p)$, then M is diffeomorphic to S^m.

Remark (Sphere Theorem in Lower Dimensions)**.** (1) For $m = 2$: let M be an oriented compact surface with $K > 0$, then by the Gauss-Bonnet formula M is diffeomorphic to S^2.

(2) For $m = 3$, by introducing the method of Ricci flow, R. Hamilton proved in 1982 that if (M, g) is a 3 dimensional compact Riemannian manifold with $Ric > 0$, then (M, g) is diffeomorphic to S^3.

(3) For $m = 4$ there is a very interesting conformal sphere theorem:

Theorem 7.33 (Chang-Gursky-Yang 2003)**.** Let (M, g) be a compact 4-manifold whose Yamabe invariant is positive. Suppose $\int_M |W|^2 \mathrm{d}v < 16\pi^2\chi(M)$, then M is diffeomorphic to S^4 or $\mathbb{RP}^4$.

In view of Grove-Shiohama's diameter sphere theorem, to prove topological sphere theorem it is enough to prove

Theorem 7.34 (Klingenberg Injectivity Radius Estimate)**.** Let (M, g) be a complete simply connected Riemannian manifold with $\frac{1}{4} < K \leqslant 1$, then $\mathrm{inj}(M, g) \geqslant \pi$.

In what follows we will prove Klingenberg's injectivity radius estimate for m even. The odd case is more involved.

Klingenberg Lemma

We need

Lemma 7.35 (Klingenberg Lemma)**.** Let (M, g) be a compact Riemannian manifold whose sectional curvature satisfies $K \leqslant C$ for some constant C. Then either

$$\mathrm{inj}(M, g) \geqslant \frac{\pi}{\sqrt{C}}$$

or there exists a closed geodesic γ in M whose length is minimum among all closed geodesics, such that

$$\mathrm{inj}(M, g) = \frac{1}{2}L(\gamma).$$

Proof. Take $p \in M$ and $q \in \mathrm{Cut}(p)$ so that $\mathrm{dist}(p, q) = \mathrm{inj}(M, g)$. If q is conjugate to p along some minimizing geodesic, then by Corollary 6.23,

$$\mathrm{inj}(M, g) = \mathrm{dist}(p, q) \geqslant \frac{\pi}{\sqrt{C}}.$$

If q is not conjugate to p, then by Proposition 7.26, there exists two minimizing normal geodesics σ, τ joining p to q so that $\dot\sigma(l) = -\dot\tau(l)$, where $l = \mathrm{dist}(p, q)$. Since p is also a cut point of q, and by definition p realized the distance from q to $\mathrm{Cut}(q)$. It follows that $\dot\sigma(0) = -\dot\tau(0)$. So σ and τ together form a closed geodesic. If we denote this closed geodesic by γ, then

$$\mathrm{inj}(M, g) = \frac{1}{2}L(\gamma).$$

Finally we prove γ has minimal length among all closed geodesics: Otherwise if there is another closed geodesic γ' with length $L(\gamma') < L(\gamma)$, and let p', q' be two "antipodal" points on γ', i.e. $\mathrm{dist}(p', q') = \frac{1}{2}L(\gamma')$, then by definition there is a point q'' on γ' which lies in $\mathrm{Cut}(p')$, and $\mathrm{dist}(p', q'') \leqslant \frac{1}{2}L(\gamma') < \mathrm{inj}(M, g)$. Contradiction. □

Proof of Klingenberg Injectivity Radius Estimate, m Even

In what follows we only need to assume M is orientable (which implies M is simply connected by Synge's theorem).

By Bonnet-Myers's theorem, M is compact. So there exists $p \in M$ and $q \in \mathrm{Cut}(p)$ so that $\mathrm{dist}(p, q) = \mathrm{inj}(M, g) =: l$. Suppose the theorem fails, i.e. $l < \pi$. Then by Corollary 6.23, q is not conjugate to p. So according to Klingenberg lemma, there exists a closed normal geodesic γ in M passing $p = \gamma(0)$ and $q = \gamma(l)$ whose length is $L(\gamma) = 2l < 2\pi$.

Since M is of even-dimension and is oriented, by repeating the proof of Synge's theorem, we can find a vector field $X(t)$ parallel along γ with

$$X(2l) = X(0) = X_p \in \dot\gamma(0)^\perp,$$

so that the variation of γ with variation field X satisfies

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2}\bigg|_{s=0} E(\gamma_s) = -\int R(\dot{\gamma}, X, \dot{\gamma}, X)\mathrm{d}t < 0.$$

In other words, $L(\gamma_s) < L(\gamma)$ for all small $s \neq 0$.

Denote $p_s = \gamma_s(0)$ and let $q_s = \gamma_s(l_s)$ be the point on γ_s which is farthest to p_s. Then

$$\mathrm{dist}(p_s, q_s) < l = \mathrm{inj}(M, g),$$

so there exists a unique normal minimizing geodesic σ_s joint $q_s = \sigma_s(0)$ to p_s. Since $\lim_{s\to 0} q_s = q$, there exists a sequence $s_i \to 0$ so that $\dot{\sigma}_{s_i}(0)$ converges to a unit vector $Y_q \in T_qM$. By continuity, $\sigma(t) = \exp_q(tY_q)$ is a minimizing normal geodesic connecting q to p. In what follows we will show $\dot{\sigma}(0) \perp \dot{\gamma}(l)$, so that σ is not one of the two parts of γ. As a consequence, we get three minimizing geodesic from q to p. This contradicts with Proposition 7.26.

It remains to prove $\dot{\sigma}(0) \perp \dot{\gamma}(l)$. We let $\sigma_{s,t}$ be the minimizing normal geodesic from $p_s = \gamma_s(0)$ to $\gamma_s(t)$ for $\gamma_s(t)$ close to $q_s = \gamma_s(l_s)$. Then $\sigma_{s,t}$ is a variation of $\sigma_s = \sigma_{s,l_s}$. By the choice of q_s, $L(\sigma_{s,t}) \leqslant L(\sigma_s)$. So according to the first variation formula,

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=l_s} E(\sigma_{s,t}) = -\langle \dot{\sigma}_s(0), \dot{\gamma}_s(l_s)\rangle.$$

It follows that $\dot{\gamma}(0) \perp \dot{\gamma}_s(l_s)$. Passing to the subsequence s_i and taking limit, we get $\dot{\sigma}(0) = Y_q \perp \dot{\gamma}(l)$.

Remark. By checking the prove above one can see that for the case $m = \dim M$ even, it's enough to assume that (M, g) is oriented and satisfies the weaker curvature condition that there exists ε such that $0 < \varepsilon < K \leqslant 1$.

Chapter 8

Bochner and Spectral Theory

8.1 Bochner's Technique and Applications

In studying the relation between the curvatures of a Riemannian manifold and its geometry/topology, another very useful method is the so-called Bochner technique.

8.1.1 Bochner's Formula

Bochner's Formula

We start with

Theorem 8.1. Let (M, g) be Riemannian manifold, and $X \in \Gamma^\infty(TM)$.

(1) If ∇X is symmetric, i.e. $\langle \nabla_u X, v\rangle = \langle \nabla_v X, u\rangle$ for all u, $v \in T_xX$, then

$$\frac{1}{2}\Delta(|X|^2) = |\nabla X|^2 + \langle X, \nabla(\mathrm{div}X)\rangle + Rc(X,X).$$

(2) If ∇X is anti-symmetric, i.e. $\langle \nabla_u X, v\rangle = -\langle \nabla_v X, u\rangle$ for all u, $v \in T_xX$, then

$$\frac{1}{2}\Delta(|X|^2) = |\nabla X|^2 - Rc(X,X).$$

Proof. (1) With Riemannian normal coordinates centered at x, we have

$$\nabla_{\partial_i}\partial_j(x) = 0, \quad \forall i, j.$$

Recall at x one can write

$$(\nabla^2 f)_x(\partial_i, \partial_j) = (\partial_i\partial_j f)(x) \text{ and } (\Delta f)(x) = \sum(\partial_i\partial_i f)(x).$$

It follows that at x,

$$\begin{aligned}
\frac{1}{2}\Delta(|X|^2) = \frac{1}{2}\sum_i \partial_i\partial_i\langle X, X\rangle &= \sum_i \partial_i\langle \nabla_{\partial_i}X, X\rangle \\
&\overset{\bigstar}{=} \sum_i \partial_i\langle \nabla_X X, \partial_i\rangle \\
&= \sum_i \langle \nabla_{\partial_i}\nabla_X X, \partial_i\rangle \\
&= \sum_i (\langle \nabla_X\nabla_{\partial_i}X, \partial_i\rangle - \langle \nabla_{[X,\partial_i]}X, \partial\rangle - \langle R(X,\partial_i)X, \partial_i\rangle) \\
&= \sum_i (\langle \nabla_X\nabla_{\partial_i}X, \partial_i\rangle - \langle \nabla_{[X,\partial_i]}X, \partial_i\rangle + Rm(X,\partial_i,X,\partial_i)) \\
&= \sum_i (\langle \nabla_X\nabla_{\partial_i}X, \partial_i\rangle - \langle \nabla_{[X,\partial_i]}X, \partial_i\rangle) + Rc(X,X).
\end{aligned}$$

Note that if we write $X = X^i\partial_i$, then

$$\mathrm{tr}(\nabla X) = \sum_i \langle \nabla_{\partial_i} X, \partial_i\rangle \overset{\clubsuit}{=} \sum_i \partial_i \langle X, \partial_i\rangle = \sum_i \partial_i X^i = \mathrm{div}X.$$

It follows

$$\sum_i \langle \nabla_X \nabla_{\partial_i} X, \partial_i\rangle = \sum_i X\langle \nabla_{\partial_i} X, \partial_i\rangle = X(\mathrm{div}X) = \langle X, \nabla \mathrm{div}X\rangle.$$

On the other hand, since $[\partial_i, X] = \nabla_{\partial_i}\partial - \ \nabla_X \partial_i = \nabla_{\partial_i} X$,

$$\begin{aligned} -\sum_i \langle \nabla_{[X,\partial_i]} X, \partial_i\rangle &= \sum_i \langle \nabla_{\nabla_{\partial_i} X} X, \partial i\rangle \\ &\overset{\bigstar}{=} \sum_i \langle \nabla_{\partial_i} X, \nabla_{\partial_i} X\rangle \\ &= |\nabla X|^2. \end{aligned}$$

So the conclusion follows.

(2) If ∇X is anti-symmetric, then there will be a negative sign at the right hand side of the two $\overset{\bigstar}{=}$, and we will get 0 after the $\overset{\clubsuit}{=}$. So the conclusion follows. □

Bochner's Formula for Smooth Functions

In particular, if $u \in \mathcal{C}^\infty(M)$, then $X = \nabla u$ is smooth and $\nabla X = \nabla^2 u$ is symmetric. Moreover, $\mathrm{div}X = \mathrm{div}\nabla u = \Delta u$. It follows

Theorem 8.2. For any $u \in \mathcal{C}^\infty(M)$,

$$\frac{1}{2}\Delta(|\nabla u|^2) = |\nabla^2 u|^2 + \langle \nabla u, \nabla(\Delta u)\rangle + Rc(\nabla u, \nabla u).$$

Sometimes one need to replace the Hessian term $|\nabla^2 u|^2$ by a simpler one. Note that by Cauchy-Schwartz inequality, for any $A = (a_{ij})$,

$$|A|^2 = \sum_{ij} |a_{ij}|^2 \geqslant \sum_I a_{ii}^2 \geqslant \frac{1}{m}\left(\sum_i a_{ii}\right)^2 = \frac{1}{m}(\mathrm{tr}A)^2.$$

As a result, we get $|\nabla^2 u|^2 \geqslant \frac{1}{m}(\Delta u)^2$ and thus

Corollary 8.3. For any $u \in \mathcal{C}^\infty(M)$,

$$\frac{1}{2}\Delta(|\nabla u|^2) \geqslant \frac{1}{m}(\Delta u)^2 + \langle \nabla u, \nabla(\Delta u)\rangle + Rc(\nabla u, \nabla u).$$

Remark. In particular, if $Ric \geqslant (m-1)\kappa$, then for any $u \in \mathcal{C}^\infty(M)$,

$$\frac{1}{2}\Delta(|\nabla u|^2) \geqslant \frac{1}{m}(\Delta u)^2 + \langle \nabla u, \nabla(\Delta u)\rangle + (m-1)\kappa|\nabla u|^2. \tag{8.1}$$

Conversely, if this inequality holds for any $u \in \mathcal{C}^\infty(M)$, then we must have $Ric \geqslant (m-1)\kappa$. To see this, given any $x_0 \in M$ and any $X_0 \in T_{x_0}M$, we take $u \in \mathcal{C}^\infty(M)$ so that $\nabla u(x_0) = X_0$ and $(\nabla^2 u)_{x_0} = c\mathrm{Id}$. Then by Bochner formula and (8.1),

$$|\nabla^2 u|^2 + Rc(X_0, X_0) = \frac{1}{2}\Delta(|\nabla u|^2) - \langle \nabla u, \nabla(\Delta u)\rangle \geqslant \frac{1}{m}(\Delta u)^2 + (m-1)\kappa|X_0|^2,$$

On the other hand, our choice of u implies $|\nabla^2 u|^2 = \frac{1}{m}(\Delta u)^2$, so we get

$$Ric \geqslant (m-1)\kappa.$$

The condition (8.1) is used in discrete geometric analysis (on graphs one can define ∇ and Δ but not curvature tensor) as a definition of "$Ric \geqslant (m-1)\kappa$".

Bochner Formula for Closed 1-forms

Recall from Section 1.2 that given any smooth 1-form $\omega \in \Omega^1(M)$, the musical isomorphism produce a smooth vector field $X = \sharp\omega$. It is not hard to check

$$|X| = |\omega|, \ |\nabla X|^2 = |\nabla\omega|^2.$$

Now suppose $\omega \in \Omega^1(M)$ is a closed 1-form. Then locally ω is exact, i.e. locally of the form $\omega = \mathrm{d}u$. As a result, $X = \sharp\omega = \nabla u$, and ∇X is symmetric[1]. So we may apply part (1) of Theorem 8.1 to get

$$\frac{1}{2}\Delta(|\omega|^2) = |\nabla\omega|^2 + \langle \sharp\omega, \nabla \mathrm{div}\sharp\omega\rangle + Rc(\sharp\omega, \sharp\omega).$$

To proceed let's do some local computation. Suppose $\omega = \omega_i \mathrm{d}x^i$. Then

$$\sharp\omega = \sum_i \omega_i\partial_i, \quad \mathrm{div}\sharp\omega = \sum_i \partial_i\omega_i, \quad \nabla\mathrm{div}\sharp\omega = \sum_{i,j}\partial_j\partial_i\omega_i\partial_j.$$

and thus

$$\langle \sharp\omega, \nabla\mathrm{div}\sharp\omega\rangle = \sum_{i,j}\omega_j\partial_j\partial_i\omega_i.$$

On the other hand, as we have seen in Section 1.4, for any smooth function f,

$$\mathrm{div}(f\sharp\omega) = f\mathrm{div}\sharp\omega + \langle\nabla f, \sharp\omega\rangle.$$

Now we assume M is compact. After integration we get

$$0 = \int_M \mathrm{div}(f\sharp\omega)\mathrm{d}V_g = \int_M (f\mathrm{div}\sharp\omega + \langle\nabla f, \sharp\omega\rangle)\mathrm{d}V_g = \int_M (f\mathrm{div}\sharp\omega + \langle \mathrm{d}f, \omega\rangle)\mathrm{d}V_g.$$

It follows that

$$\langle \mathrm{d}f, \omega\rangle_{L^2} = \int_M \langle\mathrm{d}f,\omega\rangle\mathrm{d}V_g = \int_M f(-\mathrm{div}\sharp\omega)\mathrm{d}V_g = \langle f, -\mathrm{div}\sharp\omega\rangle_{L^2}.$$

So if we define $\delta\omega = -\mathrm{div}\sharp\omega$. Then $\delta : \Omega^1(M) \to \mathcal{C}^\infty(M)$ is the L^2-dual of d,

$$\langle\mathrm{d}f,\omega\rangle_{L^2} = \langle f, \delta\omega\rangle_{L^2}, \quad \forall f \in \mathcal{C}^\infty(M), \ \omega\in\Omega^1(M).$$

For any closed 1-form ω we define $\Delta\omega = \mathrm{d}\delta\omega$. Then locally

$$\Delta\omega = \mathrm{d}\left(-\sum_i \partial_i\omega_i\right) = -\sum_{i,j}\partial_j\partial_i\omega_i$$

and thus

$$\langle\omega, \Delta\omega\rangle = -\sum_{i,j}\omega_j\partial_j\partial_i\omega_i = -\langle\sharp\omega, \nabla\mathrm{div}\sharp\omega\rangle.$$

So we end with

Theorem 8.4 (Bochner's Formula for Closed 1-form)**.** Let (M, g) be a compact Riemannian manifold, then for any closed 1-form $\omega \in \Omega^1(M)$,

$$\frac{1}{2}\Delta(|\omega|^2) = |\nabla\omega|^2 - \langle\omega,\Delta\omega\rangle + Rc(\sharp\omega,\sharp\omega).$$

[1] In fact one can show that ∇X is symmetric if and only if $\omega = \flat X$ is closed.

Harmonic k-form

More generally, one can define $\delta : \Omega^k(M) \to \Omega^{k-1}(M)$ so that

$$\langle \omega, \mathrm{d}\eta \rangle_{L^2} = \langle \delta\omega, \eta \rangle_{L^2}, \quad \forall \omega \in \Omega^k(M),\ \eta \in \Omega^{k-1}(M),$$

and define the **Hodge Laplacian** on all smooth k-forms to be

$$\Delta := \mathrm{d}\delta + \delta\mathrm{d} : \Omega^k(M) \to \Omega^k(M).$$

One can check that when $k = 0$ this definition coincides with the Laplace-Beltrami operator Δ on smooth functions (and thus differed with $\Delta = \mathrm{tr}\nabla^2$ by a negative sign). A differential form $\omega \in \Omega^k(M)$ is called a **harmonic k-form** if $\Delta\omega = 0$. In view of the fact

$$\langle \omega, \Delta\omega \rangle_{L^2} = \langle \omega, \mathrm{d}\delta\omega \rangle_{L^2} + \langle \omega, \delta\mathrm{d}\omega \rangle_{L^2} = \langle \mathrm{d}\omega, \mathrm{d}\omega \rangle_{L^2} + \langle \delta\omega, \delta\omega \rangle_{L^2}$$

and the definition of Δ, we have

Proposition 8.5. $\omega \in \Omega^k(M)$ is harmonic if and only if $\mathrm{d}\omega = 0$ and $\delta\omega = 0$.

Now suppose ω is a harmonic 1-form on compact Riemannian manifold (M, g). Then ω is closed, and thus by Theorem 8.4,

$$0 = \int_M \frac{1}{2}\Delta(|\omega|^2) = \int_M |\nabla\omega|^2 + \int_M Rc(\sharp\omega, \sharp\omega).$$

So we get

Theorem 8.6 (Bochner)**.** Let (M, g) be a compact Riemannian manifold, then

(1) Suppose $Ric \geqslant 0$. If $\Delta\omega = 0$, then $\nabla\omega = 0$.

(2) Suppose $Ric \geqslant 0$, and $Ric > 0$ at one point. If $\Delta\omega = 0$, then $\omega = 0$.

According to the famous Hodge theory, the space of harmonic k-forms is isomorphic to the de Rham cohomology group $H^k_{dR}(M)$. So we conclude

Corollary 8.7. Let (M, g) be a closed oriented Riemannian manifold, $Ric \geqslant 0$, and $Ric > 0$ at one point. Then $b_1(M) = 0$.

Remark. There is a Bochner-Weitzenböck formula that generalize the Bochner formula above to k-forms using which one can prove: If (M, g) is a closed Riemannian manifold with non-negative curvature operator, then all harmonic forms of order $1 \leqslant k \leqslant m - 1$ on M are parallel.

Bochner Formula for Killing Forms

Now let's turn to part (2) in Theorem 8.1. As we have seen in Problem Sheet 1, ∇X is anti-symmetric if and only if X is a Killing field on (M, g). So we get

Corollary 8.8. For any Killing vector field on M,

$$\frac{1}{2}\Delta(|X|^2) = |\nabla X|^2 - Rc(X, X).$$

As a result,

Theorem 8.9 (Bochner, 1946)**.** Any Killing vector field of a compact Riemannian manifold with negative Ricci curvature must be zero.

Since the space of all Killing vector fields on (M, g) is the Lie algebra of the isometry group $\mathrm{Iso}(M, g)$ [which is a Lie group], and the isometry group of any compact Riemannian manifold is compact, we conclude that the isometry group of any compact Riemannian manifold with negative Ricci curvature must be a finite group.

8.1.2 Cheeger-Gromoll Splitting Theorem

Cheeger-Gromoll Splitting Theorem

Let (M, g) be a complete non-compact connected Riemannian manifold. Recall that a line in M is a normal geodesic $\gamma : \mathbb{R} \to M$ so that

$$d(\gamma(a), \gamma(b)) = |a - b|, \quad \forall a, b \in \mathbb{R}.$$

Unlike the case of rays, it is possible that there is no ray in a complete non-compact Riemannian manifold. For example, a cylinder $\mathbb{R} \times S^1$ admits many lines, while the paraboloid $z = x^2 + y^2$ admits no line at all.

As another application of Bochner formula, we prove the following structure theorem for Riemannian manifolds with positive Ricci curvature that admit lines:

Theorem 8.10 (Cheeger-Gromoll, 1971)**.** Let (M, g) be a complete non-compact Riemannian manifold with $Ric \geqslant 0$. Suppose there exists a line in M. Then (M, g) is isometric to $\mathbb{R} \times N$, where N is an $(m-1)$-dimensional complete Riemannian manifold with $Ric \geqslant 0$.

The idea is to construct a function on M which behaves like the function $f(x, r) = r$ on $N \times \mathbb{R}$, so that the level sets of f gives the desired component N. So what is the specialty of the function $f(x, r) = r$? It is smooth, with gradient $\nabla f = \partial_r$ which has length 1, and has Hessian $\nabla^2 f = 0$ (so that ∇f is parallel). It is in the proof of "$\nabla^2 f = 0$" that we need Bochner's formula.

To apply Bochner's formula one need the function to be smooth. However, the construction below uses the distance, and thus the function is only Lipschitz. To solve this problem we will apply theories on PDEs in the barrier sense. More precisely, we need the Hopf-Calabi strong maximum principle,

Theorem 8.11 (Hopf-Calabi Strong Maximum Principle)**.** Let $\Omega \subset M$ be a connected open set. Suppose $\Delta f \leqslant 0$ in M in the barrier sense, and f attains an interior minimum, then f is constant on Ω.

We also need the well-known Weyl lemma to increase regularity:

Theorem 8.12 (Weyl Lemma)**.** If $\Delta f = 0$ in the barrier sense, then f is smooth.

Busemann Function

The function that we need is the so-called Busemann function (introduced by Busemann in 1955), defined as follows.

Since (M, g) is complete and non-compact, for any ray $\gamma : [0, +\infty) \to M$, let

$$b_\gamma^t : M \to \mathbb{R},$$
$$b_\gamma^t(x) = t - d(x, \gamma(t)).$$

By triangle inequality, it is easy to see

- $b_\gamma^t(x) \leqslant d(\gamma(0), x)$,
- for any $t < s$, one has $b_\gamma^s(x) - b_\gamma^t(x) = (s - t) + d(x, \gamma(t)) - d(x, \gamma(s)) \geqslant 0$,
- $|b_\gamma^t(x) - b_\gamma^t(y)| \leqslant d(x, y)$.

As a result, the limit

$$b_\gamma(x) := \lim_{t \to +\infty} (t - d(x, \gamma(t)))$$

is well-defined and is Lipschitz with Lipschitz constant 1. We call the function $b_\gamma : M \to \mathbb{R}$ the **Busemann function** associated with γ.

By Laplacian comparison theorem, formally we have

$$\Delta b_\gamma(x) \geqslant - \lim_{t \to +\infty} \frac{m-1}{d(x, \gamma(t))} = 0.$$

This can be proved rigorously by constructing a lower barrier. Let's admit this:

Proposition 8.13. Let b_γ be the Busemann function associated with a ray γ, then $\Delta b_\gamma \geqslant 0$ in the barrier sense.

Now let $l : (-\infty, +\infty) \to M$ be a line in M, and let $\gamma_+, \gamma_- : [0, +\infty) \to M$ be the two rays in l defined by $\gamma_+(t) = l(t)$ and $\gamma_-(t) = l(-t)$. Then by Proposition 8.13,

$$\Delta(b_{\gamma_+}(x) + b_{\gamma_-}(x)) \geqslant 0.$$

On the other hand, by definition we have

$$2t = d(\gamma_-(t), \gamma_+(t)) \leqslant d(\gamma_-(t), x) + d(\gamma_+(t), x),$$

which implies

$$b_{\gamma_+}(x) + b_{\gamma_-}(x) \leqslant 0.$$

Note that on the line l, if we denote $x = l(s)$, then

$$b_{\gamma_+}(x) + b_{\gamma_-}(x) = \lim_{t\to\infty}(2t - d(x, l(t)) - d(x, l(-t))) = \lim_{t\to\infty}(2t - t - s - t + s) = 0.$$

So we conclude that $b_{\gamma_+}(x) + b_{\gamma_-}(x)$ is a subharmonic function that achieves its maximum at an interior point. By Theorem 8.11,

$$b_{\gamma_+}(x) + b_{\gamma_-}(x) = 0.$$

In other words, $b_{\gamma_+}(x) = -b_{\gamma_-}(x)$, and thus

$$\Delta b_{\gamma_+}(x) = -\Delta b_{\gamma_-}(x) \leqslant 0.$$

So we arrive at

$$\Delta b_{\gamma_+}(x) = -\Delta b_{\gamma_-}(x) = 0.$$

By Theorem 8.12, $b_{\gamma_+}, b_{\gamma_-} \in \mathcal{C}^\infty(M)$. Moreover, since b_{γ_+} has Lipschitz constant 1,

$$|\nabla b_{\gamma_+}| \leqslant 1.$$

Also note that by definition, on the line l we have $b_{\gamma_+}(l(s)) = s$ and thus

$$|\nabla b_{\gamma_+}| = 1.$$

The Proof of Cheeger-Gromoll Splitting Theorem

Now we apply Bochner formula to the function b_{γ_+}, to get

$$\frac{1}{2}\Delta(|\nabla b_{\gamma_+}|^2) = |\nabla^2 b_{\gamma_+}|^2 + Ric(\nabla b_{\gamma_+}, \nabla b_{\gamma_+}) \geqslant 0.$$

So again, $|\nabla b_{\gamma_+}|^2$ is a sub-harmonic function that achieves its interior maximum. So by the Hopf strong maximum principle again,

$$|\nabla b_{\gamma_+}|^2 = 1.$$

It follows that $|\nabla b_{\gamma_+}| = 1$ and in particular, ∇b_{γ_+} is a complete vector field. Moreover, it follows that

$$\Delta(|\nabla b_{\gamma_+}|) = 0,$$

and thus $|\nabla^2 b_{\gamma_+}|^2 = 0$, i.e.

$$\nabla^2 b_{\gamma_+} = 0.$$

Finally we construct the splitting. Let $M_t = b_{\gamma_+}^{-1}(t)$. Since $|\nabla b_{\gamma_+}| = 1$, any $t \in \mathbb{R}$ is a regular value of b_{γ_+}. So M_t is a smooth submanifold of M of dimension $m-1$.

Denote $N = M_0$. Let $\varphi_s : M \to M$ be the flow of the vector field ∇b_{γ_+}. Then φ_s is diffeomorphism. Moreover, for any $s \in \mathbb{R}$ and any $x \in N = M_0$, we have $\varphi_s(x) \in M_s$. So we get a smooth map

$$\Phi : \mathbb{R} \times N \to M,$$

$$\Phi(s, p) := \varphi_s(p)$$

which is bijective, and whose inverse

$$\begin{aligned}\Phi^{-1} :& M \to \mathbb{R} \to \mathbb{N}, \\ & x \mapsto (b_{\gamma_+(x)}, \varphi_{-b_{\gamma_+(x)}}(x))\end{aligned}$$

is smooth. So Φ is a diffeomorphism.

It remains to prove that Φ is an isometry. Note that if we let $\gamma_p(s) = \varphi_s(p)$ be the integral curve passing p, then

$$\dot{\gamma}_p = \nabla b_{\gamma_+} \;\Rightarrow\; \nabla \dot{\gamma}_p = \nabla^2 b_{\gamma_+} = 0$$

which implies that γ_p is the geodesic $\gamma_p(s) = \exp_p(sX_p)$, where $X_p = \nabla b_{\gamma_+}(p)$. As a result, we have

- Φ is a radial isometry: We have $|\partial_s| = 1$ and

$$|\mathrm{d}\Phi_{(s,p)}(\partial_s)| = |\dot{\gamma}_p| = |\nabla b_{\gamma_+}| = 1.$$

- Φ maps "vectors orthogonal to radial direction ∂_s" to "vectors orthogonal to radial direction $\mathrm{d}\Phi_{(s,p)}(\partial_s)$": For any $X_0 \in T_pN = T_pM_0$, we have

$$(\mathrm{d}\Phi)_{(s,p)}(0, X_0) = (\mathrm{d}\varphi_s)_p(X_0) \in T_{\varphi_s(p)}M_s \perp \nabla b_{\gamma_+}(\varphi_s(p)) = \mathrm{d}\Phi_{(s,p)}(\partial_s).$$

- Φ preserves the length (and thus the inner product by polarization) of all vectors orthogonal to ∂_s: For any $X_0 \in T_p$, we may extend X_0 to a local coordinate vector field $\widetilde{X}_0$ on TN such that $[\partial_s, \widetilde{X}_0] = 0$. Then

$$\nabla_{\dot{\gamma}_p(s)}((\mathrm{d}\varphi_s)_p(X_0)) = \nabla_{\mathrm{d}\varphi_s(\widetilde{X}_0)}(\nabla b_{\gamma_+}) - [(\nabla b_{\gamma_+})(\varphi_s(p)), (\mathrm{d}\varphi_s)_p\widetilde{X}_0].$$

  The first term vanishes since $\nabla(\nabla b_{\gamma_+}) = 0$, while the second term vanishes since it equals $\mathrm{d}\varphi_s([\partial_s, \widetilde{X}_0]) = 0$. So we conclude that $(\mathrm{d}\varphi_s)_p(X_0)$ is parallel along $\gamma_p(s)$, and thus

$$|\mathrm{d}\Phi_{(s,p)}(0, X_p)| = |\mathrm{d}\varphi_s(X_0)| = |X_0|.$$

So we conclude that (M, g) is isometric to $\mathbb{R} \times N$. Finally since (M, g) has non-negative Ricci curvature, and N is a Riemannian submanifold, and $K(\partial_s, X_0) = 0$, we conclude that N has non-negative Ricci curvature.

8.2 Spectral Geometry

Spectral geometry is the branch of differential geometry that studies the relations between the spectrum of the Laplace-type operator and the underline geometry. There are many beautiful results that have been proved, and at the meantime there are also many open problems to be studied in the future. In this last section, we apply Bochner formula to spectral geometry.

8.2.1 Spectral Geometry

Eigenvalues and Eigenfunctions

In spectral geometry, there are three typical spectral problems:

(1) **Closed Setting** Let (M, g) be a closed connected Riemannian manifold. We call λ an **eigenvalue** of Δ if there exists smooth function $u \neq 0$ so that[2]

$$\Delta u + \lambda u = 0.$$

[2]Here we use $\Delta = \mathrm{div}\nabla = \mathrm{tr}(\nabla^2)$. If one uses $\Delta = -\mathrm{div} = -\mathrm{tr}(\nabla^2) = \mathrm{d}\delta + \delta\mathrm{d}$, then the equation should be $\Delta u = \lambda u$.

(2) Let (Ω, g) be a compact connected Riemannian manifold with boundary $\partial\Omega$.

(a) **Dirichlet Setting** We call a number λ a **Dirichlet eigenvalue** of Δ if there exists a smooth function $u \neq 0$ so that

$$\begin{cases} \Delta u + \lambda u = 0, & \text{in } \Omega, \\ u = 0, & \text{on } \partial\Omega. \end{cases}$$

(b) **Neumann Setting** We call a number λ a **Neumann eigenvalue** of Δ if there exists a smooth function $u \neq 0$ so that

$$\begin{cases} \Delta u + \lambda u = 0, & \text{in } \Omega, \\ \partial_\nu u = 0, & \text{on } \partial\Omega. \end{cases}$$

where ∂_ν represents the outer normal derivative.

We have seen from Problem Sheet 1 that

- All eigenvalues of Δ are non-negative real numbers.
- $\lambda = 0$ is always an eigenvalue for the closed problem and the Neumann eigenvalue problem (with eigenfunctions the constant functions), and $\lambda = 0$ is not an eigenvalue of the Dirichlet problem.
- If u and v are eigenfunctions of different eigenvalues, then $\langle u, v\rangle_{L^2} = 0$.

According to the standard spectral theory in functional analysis, one can prove

Theorem 8.14. In all three settings above, each eigenvalue has finite multiplicity and the eigenvalues of Δ form an increasing sequence that tends to ∞, namely

$$0 = \lambda_0 < \lambda_1 \leqslant \lambda_2 \leqslant \lambda_3 \leqslant \cdots \leqslant \lambda_k \leqslant \cdots \to \infty$$

for the closed eigenvalues and the Neumann eigenvalues, and

$$0 < \lambda_1 < \lambda_2 \leqslant \lambda_3 \leqslant \cdots \leqslant \lambda_k \leqslant \cdots \to \infty$$

for the Dirichlet eigenvalues. Moreover, one can choose an eigenbasis so that they form a complete orthonormal basis of $L^2(M)$ or $L^2(\Omega)$.

The simplest example being

Example. For S^1, the Laplacian eigenvalues are the squares $0, 1, 1, 4, 4, 9, 9, \ldots$, with eigenfunctions $\cos(kx)$ and $\sin(kx)$. Since these functions already form an orthonormal basis, there are no other eigenvalues/eigenfunctions.

Similarly for $\mathbb{T}^m = S^1 \times \cdots \times S^1$, equipped with the standard flat metric, the eigenvalues are numbers of the form $k_1^2 + \cdots + k_m^2$, with eigenfunctions $\cos(k \cdot x)$ and $\sin(k \cdot x)$, and again they form an orthonormal basis. [Note that in this case, the multiplicity is complicated since there may be many different ways to represent a given positive integer as the sum of m squares.]

Example. One can show that the eigenvalues of the standard sphere S^m are $k(k + m - 1)$ ($k = 0, 1, 2, \ldots$), with multiplicity $n_k = \binom{m+k}{m} - \binom{m+k-2}{m}$.

Unfortunately, other than very few examples like the sphere/the torus/the projective spaces etc (or rectangles/balls/annulus etc in the case of manifold with boundary), for most Riemannian manifolds there is no way to calculate its eigenvalues explicitly. There are two major problem in spectral geometry:

- **The Direct Problem** Given information of (M, g) or (Ω, g), what can we say about these eigenvalues/eigenfunctions?
- **The Inverse Problem** Given the sequence of eigenvalues, what can we say about the geometry of (M, g) or (Ω, g)?

The First Eigenvalue λ_1

The first non-zero eigenvalue λ_1 is very important and has received much attention. Although in general one can't calculate it explicitly, we do have a variational characterization as follows. Given any smooth function $\varphi \neq 0$, we call

$$R(\varphi) = \frac{\displaystyle\int_M |\nabla\varphi|^2 \mathrm{d}V_g}{\displaystyle\int_M \varphi^2 \mathrm{d}V_g}$$

the **Rayleigh quotient** of φ. Then

Theorem 8.15 (Variational Characterization of λ_1)**.** For the closed or the Neumann eigenvalue problem,

$$\lambda_1 = \inf\{R(\varphi) \mid 0 \neq \varphi \in H^1(M), \int_M \varphi = 0\},$$

while for the Dirichlet eigenvalue problem,

$$\lambda_1 = \inf\{R(\varphi) \mid 0 \neq \varphi \in H_0^1(M)\}.$$

Proof. For any φ in the given space, we may expand $\varphi = \sum_{k=1}^{\infty} c_k u_k$. We may assume $\|\varphi\|_{L^2} = 1$, i.e. $\sum c_k^2 = 1$. Then

$$R(\varphi) = \sum_{k \geqslant 1} \lambda_k c_k^2 \geqslant \sum_{k \geqslant 1} \lambda_1 c_k^2 = \lambda_1.$$

On the other hand, if we take $\varphi = u_1$, then $R(\varphi) = R(u_1) = \lambda_1$. □

Remark. For example, given bounded domain Ω, the Poincaré inequality states that there exists constant C so that

$$\int_\Omega |u|^2 \mathrm{d}V_g \leqslant C \int_\Omega |\nabla u|^2 \mathrm{d}V_g, \quad \forall u \in H_0^1(\Omega).$$

Now in view of the above theorem, the smallest (=the best) constant C for the Poincaré inequality to be true is the reciprocal of the first Dirichlet eigenvalue of Ω.

Remark. One also has variational characterization of higher eigenvalues λ_k for all k.

8.2.2 Some Results on the First Eigenvalue λ_1

Now suppose (M, g) is closed and we focus on the first non-zero eigenvalue.

Lichnerowitz Estimate for λ_1

Now we apply Bochner formula to prove a lower bound estimate for λ_1.

Theorem 8.16 (Lichnerowitz)**.** Let (M, g) be a closed Riemannian manifold with $Ric \geqslant (m-1)\kappa$ for some $\kappa > 0$. Then the first eigenvalue

$$\lambda_1 \geqslant m\kappa.$$

Proof. According to Corollary 8.3, for any $u \in \mathcal{C}^\infty(M)$,

$$\frac{1}{2}\Delta(|\nabla u|^2) \geqslant \frac{1}{m}(\Delta u)^2 + \langle \nabla u, \nabla(\Delta u)\rangle + Rc(\nabla u, \nabla u).$$

So if we take u be an eigenfunction, i.e. $\Delta u + \lambda u = 0$, then we get

$$\frac{1}{2}\Delta|\nabla u|^2 \geqslant -\frac{\lambda}{m} u\Delta u - \lambda\langle \nabla u, \nabla u\rangle + Rc(\nabla u, \nabla u). \tag{8.2}$$

Integrate over M and apply the Green's formula

$$-\int_M u\Delta u \mathrm{d}x = \int_M |\nabla u|^2 \mathrm{d}x$$

we get

$$0 \geqslant \int_M \left(\frac{\lambda}{m} - \lambda + (m-1)\kappa\right) |\nabla u|^2 \mathrm{d}x.$$

This implies

$$\frac{\lambda}{m} - \lambda + (m-1)\kappa \leqslant 0,$$

i.e. $\lambda \geqslant m\kappa$. □

Obata's λ_1 Rigidity Theorem

One can prove that the first eigenvalue of the standard sphere S^m is m. In fact, this is the only case where $\lambda_1 = m$ if (M, g) satisfies the conditions in Theorem 8.16.

Theorem 8.17 (Obata)**.** Let (M, g) be a closed Riemannian manifold with $Ric \geqslant (m-1)\kappa$ for some $\kappa > 0$. If $\lambda_1 = m\kappa$, then (M, g) is isometric to the sphere S^m_κ.

Proof. Without loss of generality we may assume $\kappa = 1$. If $\lambda_1 = m$, then from the proof above we see

$$Rc(\nabla u, \nabla u) = (m-1)|\nabla u|^2.$$

Since $\Delta(u^2) = 2u\Delta u + 2|\nabla u|^2$ (see Problem Sheet 1 for $-\Delta$), from (8.2) we get

$$\frac{1}{2}\Delta(|\nabla u|^2 + u^2) \geqslant -u\Delta u - m|\nabla u|^2 + (m-1)|\nabla u|^2 + u\Delta u + |\nabla u|^2 = 0.$$

It follows $\Delta(|\nabla u|^2 + u^2) \equiv 0$ since its integral over M is 0. In other words,

$$|\nabla u|^2 + u^2 = \text{constant}.$$

We normalize u so that $\max_M u^2 = 1$. Since $\nabla u = 0$ at the maximum/minimum points of u, we get

$$|\nabla u|^2 + u^2 = 1 \text{ and } \max_M u = -\min_M u = 1.$$

Now let $p, q \in M$ be points such that $u(p) = -1$, $u(q) = 1$. Let $l = d(p, q)$ and let $\gamma : [0, l] \to M$ be a normal geodesic from p to q. Let $f(t) = u(\gamma(t))$. Then

$$\frac{f'(t)}{\sqrt{1 - f^2(t)}} \leqslant \frac{|\nabla u(\gamma(t))|}{\sqrt{1 - u(\gamma(t))^2}} = 1.$$

Integrating both sides we get

$$\pi = \int_0^l \frac{f'(t)}{\sqrt{1 - f^2(t)}} \mathrm{d}t \leqslant \int_0^l \mathrm{d}t = l = d(p, q).$$

So $\mathrm{diam}(M, g) \geqslant \pi$. But by Bonnet-Myers, $\mathrm{diam}(M, g) \leqslant \pi$. So $\mathrm{diam}(M, g) = \pi$. Finally by Cheng's maximal diameter theorem, (M, g) is isomorphic to S^m. □

Reilly's Formula

Let Ω be a compact smooth manifold with smooth boundary $M = \partial\Omega$. Then one can define the second fundamental form of M (as a Riemannian submanifold of Ω) as follows: For any $p \in M$, the vector-valued second fundamental form II at p is a symmetric bilinear map

$$\begin{aligned} \mathrm{II} : T_pM \times T_pM &\to N_pM, \\ (X, Y) &\mapsto (\nabla^\Omega_X \overline{Y})^\perp, \end{aligned}$$

where $\overline{X}$, $\overline{Y}$ are smooth vector fields whose value at p are X and Y respectively [According to Problem Sheet 2, $\mathrm{II}(X,Y)$ is well-defined and is symmetric]. Since in the hypersurface case there is only one normal dimension, we may study the (scalar-valued) second fundamental form

$$h : T_pM \times T_pM \to \mathbb{R},$$
$$(X,Y) \mapsto h(X,Y) := -\langle \mathrm{II}(X,Y), \nu\rangle.$$

If we pick a local orthonormal coordinate system $\{e_i\}$ near $p \in M$, where e_{m+1} is the out normal direction, then for any $X = X^i e_i$, $Y = Y^j e_j \in T_pM$, one has

$$h(X,Y) = \sum_{i,j=1}^{m} h_{ij} X^i Y^j,$$

where $h_{ij} = -\langle \nabla_{e_i} e_j, e_{m+1}\rangle = \langle \nabla_{e_i} e_{m+1}, e_j\rangle$. The trace of h,

$$H := \mathrm{tr}(h) = \sum_i h_{ii},$$

is known as the **mean curvature** of M at p.

By integrating Bochner formula, one can prove the following useful formula obtained by R. Reilly in 1977:

Theorem 8.18 (Reilly's Formula)**.** Let Ω be a compact Riemannian manifold of dimension $m+1$, with smooth boundary $M = \partial\Omega$. Then for any $f \in \mathcal{C}^\infty(\Omega)$,

$$\frac{m}{m+1}\int_\Omega (\Delta^\Omega f)^2 \geqslant \int_M H f_\nu^2 + 2\int_M f_\nu \Delta^M f + \int_M h(\nabla^M f, \nabla^M f) + \int_\Omega Rc^\Omega(\nabla f, \nabla f).$$

Moreover, the equality holds if and only if $f_{ij} = \dfrac{\Delta^\Omega f}{m+1}\delta_{ij}$, i.e. $\nabla^2 f = \dfrac{\Delta^\Omega f}{m+1}\mathrm{Id}$.

Proof. For simplicity we write $\Delta^\Omega f = g$, and write $f\Big|_{\partial\Omega} = u$. So in what follows we may abbreviate $\Delta^\Omega f = \Delta f$, $\nabla^\Omega f = \nabla f$ and $\Delta^M u = \Delta u$, $\nabla^M u = \nabla u$. By Bochner formula, we have

$$\frac{1}{2}\Delta(|\nabla f|^2) \geqslant \frac{1}{m+1} g^2 + \langle \nabla f, \nabla g\rangle + Rc^\Omega(\nabla f, \nabla f),$$

with equality if and only if $\nabla^2 f = \dfrac{\Delta f}{m+1}\mathrm{Id}$. Integrate and in view of Green's formula

$$\int_\Omega \langle \nabla f, \nabla g\rangle = -\int_\Omega g\Delta f + \int_{\partial\Omega} g f_\nu$$

we get

$$\begin{aligned}\frac{1}{2}\int_\Omega \Delta(|\nabla f|^2) &\geqslant \frac{1}{m+1}\int_\Omega g^2 + \int_\Omega \langle \nabla f, \nabla g\rangle + \int_\Omega Rc^\Omega(\nabla f, \nabla f)\\ &= \frac{-m}{m+1}\int_\Omega g^2 + \int_M g f_\nu + \int_\Omega Rc^\Omega(\nabla f, \nabla f).\end{aligned}$$

In what follows we will prove

$$\frac{1}{2}\int_\Omega \Delta(|\nabla f|^2) = -\int_M H f_\nu^2 + \int_M g f_\nu - 2\int_M f_\nu \Delta u - \int_M h(\nabla u, \nabla u) \tag{8.3}$$

from which the theorem follows.

We choose orthonormal frame near p so that $e_{m+1}(p) = \nu(p)$ and $\nabla_{e_{m+1}} e_{m+1} = 0$. Cover M by such coordinate neighborhoods and let ρ_α be a partition of unity subordinate to this covering (together with the open set $\Omega \setminus M$). As we have seen in Section 1.4, $\sum\limits_\alpha \rho_\alpha \mathrm{div}(X) = \sum\limits_\alpha \mathrm{div}(\rho_\alpha X)$. So

$$\frac{1}{2}\int_\Omega \Delta(|\nabla f|^2) = \frac{1}{2}\sum_\alpha \int_\Omega \rho_\alpha \Delta(|\nabla f|^2) = \frac{1}{2}\sum_\alpha \int_\Omega \mathrm{div}(\rho_\alpha \nabla(|\nabla f|^2)).$$

So by divergence theorem,

$$\frac{1}{2}\int_\Omega \Delta(|\nabla f|^2) = \frac{1}{2}\sum_\alpha \int_M \langle \rho_\alpha \nabla(|\nabla f|^2), \nu\rangle = \frac{1}{2}\sum_\alpha \int_M \rho_\alpha \partial_\nu(|\nabla f|^2).$$

Thus we may compute in the above local coordinates. Since $|\nabla f|^2 = \sum_i (e_i f)^2$,

$$\begin{aligned}\frac{1}{2}\int_M \rho_\alpha \partial_\nu(|\nabla f|^2) &= \int_M \rho_\alpha \sum_{i=1}^{m+1}(e_i f)(e_{m+1}e_i f)\\ &= \int_M \rho_\alpha \left[(e_{m+1}f)(e_{m+1}e_{m+1}f) + \sum_{i=1}^{m}(e_i f)(e_{m+1}e_i f)\right].\end{aligned}$$

Note that $\nabla_{e_{m+1}} e_{m+1} = 0$ in the neighborhood. So we get, at all $x \in M$,

$$\begin{aligned}e_{m+1}e_{m+1}f &= \sum_{i=1}^{m+1}[e_ie_if - (\nabla_{e_i}e_i)f] - \sum_{i=1}^{m}[e_ie_if - (\nabla_{e_i}e_i)f]\\ &= \sum_{i=1}^{m+1}[e_ie_if - (\nabla_{e_i}e_i)f] - \sum_{i=1}^{m}[e_ie_if - (\nabla_{e_i}e_i)^T f] + \sum_{i=1}^{m}(\nabla_{e_i}e_i)^\perp f\\ &= \Delta f - \Delta u + \sum_{i=1}^{m}\langle \nabla_{e_i}e_i, e_{m+1}\rangle f_\nu\\ &= g - \Delta u - Hf_\nu.\end{aligned}$$

For the second term we use

$$\begin{aligned}e_{m+1}e_i f &= e_ie_{m+1}f + (\nabla_{e_{m+1}}e_i)f - (\nabla_{e_i}e_{m+1})f\\ &= e_ie_{m+1}f + \sum_{j=1}^{m+1}\langle \nabla_{e_{m+1}}e_i, e_j\rangle f_j - \sum_{j=1}^{m+1}\langle \nabla_{e_i}e_{m+1}, e_j\rangle f_j\\ &= e_ie_{m+1}f + \sum_{j=1}^{m}\langle \nabla_{e_{m+1}}e_i, e_j\rangle f_j - \sum_{j=1}^{m} h_{ij}f_j,\end{aligned}$$

where in the last step we used the facts $\langle \nabla_{e_{m+1}}e_i, e_{m+1}\rangle = -\langle e_i, \nabla_{e_{m+1}}e_{m+1}\rangle = 0$, $\langle \nabla_{e_i}e_{m+1}, e_{m+1}\rangle = \frac{1}{2}e_i|e_{m+1}|^2 = 0$ and $h_{ij} = \langle \nabla_{e_i}e_{m+1}, e_j\rangle$. So we get three terms. For the first one we have

$$\sum_\alpha \int_M \rho_\alpha \sum_{i=1}^{m}(e_if)(e_ie_{m+1}f) = \sum_\alpha \int_M \rho_\alpha \langle \nabla^M f, \nabla^M f_\nu\rangle = -\int_M (\Delta u)f_\nu.$$

For the second term, we have $\sum_{i,j=1}^{m}(e_if)\langle \nabla_{e_{m+1}}e_i, e_j\rangle f_j = 0$ since

$$S := \sum_{i,j=1}^{m}(e_if)\langle \nabla_{e_{m+1}}e_i, e_j\rangle f_j = \sum_{i,j}\langle \nabla_{e_{m+1}}e_i, e_j\rangle f_if_j = -\sum_{i,j=1}^{m}\langle e_i, \nabla_{e_{m+1}}e_j\rangle f_if_j = -S.$$

Finally for the last term,

$$\sum_{i,j=1}^{m}(e_if)\langle \nabla_{e_i}e_{m+1}, e_j\rangle f_j = \sum_{i,j=1}^{m} h_{ij}f_if_j = h(\nabla u, \nabla u).$$

So we get the desired equality (8.3). □

Remark. If we don't apply Cauchy-Schwartz inequality at the first step, then

$$\int_\Omega ((\Delta f)^2 - |\nabla^2 f|^2) = \int_M (Hf_\nu^2 + 2f_\nu \Delta^M f + h(\nabla^M f, \nabla^M f)) + \int_\Omega Rc(\nabla f, \nabla f).$$

Yau's Conjecture

A Riemannian submanifold M^m of N is called minimal if it has mean curvature $H = 0$. Minimal submanifolds are very important objects in Riemannian geometry, especially the branch "submanifold geometry". As an application we prove

Theorem 8.19 (Choi-Wang, 1984)**.** Let M^m be a compact connected embedded oriented minimal hypersurface in a compact oriented Riemannian manifold N^{m+1}. Suppose N has Ricci curvature $Ric^N \geqslant m\kappa > 0$, then $\lambda_1(M) \geqslant \dfrac{m\kappa}{2}$.

Proof. Since $Ric^N > 0$, by Bochner theorem (c.f. Corollary 8.7), $b_1(N) = 0$. Let Ω be a tubular neighborhood of M. Then the Mayer–Vietoris sequence of de Rham cohomologies,

$$0 \to H^0(N) \to H^0(N \setminus M) \oplus H^0(\Omega) \to H^0((N \setminus M) \cap \Omega) \to H^1(N)$$

becomes

$$0 \to \mathbb{R} \to H^0(N \setminus M) \oplus \mathbb{R} \to \mathbb{R} \oplus \mathbb{R} \to 0.$$

It follows that $H^0(N \setminus M) \cong \mathbb{R} \oplus \mathbb{R}$, i.e. $N \setminus M$ contains exactly two connected components. We denote

$$N \setminus M = \Omega_1 \cup \Omega_2, \quad \partial\Omega_1 = \partial\Omega_2 = M.$$

Now let $u \in \mathcal{C}^\infty(M)$ be an eigenfunction associated to $\lambda_1 = \lambda_1(M)$, i.e. $\Delta_M u + \lambda_1 u = 0$. Without loss of generality, we assume

$$\int_M h(\nabla u, \nabla u) \mathrm{d}V_g \geqslant 0,$$

where we regard M as $\partial\Omega_1$. [If this inequality is not true, then the analogue inequality for Ω_2 holds and we proceed with Ω_2 instead of Ω_1.] Let f be a solution to

$$\begin{cases} \Delta^N f = 0, & \text{in } \Omega_1 \\ f = u, & \text{on } M = \partial\Omega_1. \end{cases}$$

By Reilly's formula,

$$0 \geqslant -2\lambda \int_M u f_\nu + \int_M h(\nabla u, \nabla u) + m\kappa \int_{\Omega_1} |\nabla f|^2 \geqslant -2\lambda \int_M u f_\nu + m\kappa \int_{\Omega_1} |\nabla f|^2.$$

Since $\Delta f = 0$, by Green's formula we get

$$\int_M u f_\nu = \int_{\partial\Omega_1} f f_\nu = \int_{\Omega_1} |\nabla f|^2.$$

thus

$$0 \geqslant (-2\lambda_1 + m\kappa) \int_{\Omega_1} |\nabla f|^2.$$

It follows $\lambda_1 \geqslant \dfrac{m\kappa}{2}$. □

In particular, if we take $N^{m+1} = S^{m+1}$ we get

$$\lambda_1(M) \geqslant \frac{m}{2}.$$

This lower bound is half of the conjectured bound by Yau in 1982:

Conjecture (Yau)**.** For any embedded minimal hypersurface M of S^{m+1}, one has

$$\lambda_1(M) \geqslant m.$$

References

[1] F. Flaherty and M.P. do Carmo. *Riemannian Geometry*. Mathematics: Theory & Applications. Birkhäuser Boston, 2013.

[2] S. Gallot, D. Hulin, and J. Lafontaine. *Riemannian Geometry*. Universitext. Springer Berlin Heidelberg, 2004.

[3] P. Petersen. *Riemannian Geometry*. Graduate Texts in Mathematics. Springer New York, 2006.

[4] T. Sakai. *Riemannian Geometry*. Fields Institute Communications. American Mathematical Soc., 1996.