C3.11 Riemannian Geometry

Prof Jason D. Lotay

jason.lotay@maths.ox.ac.uk

Hilary Term 2023–2024

Course overview

Riemannian Geometry is the study of curved spaces and provides an important tool with diverse applications from group theory to general relativity. The surprising power of Riemannian Geometry is that we can use local information to derive global results.

This course will study the key notions in Riemannian Geometry: geodesics and curvature. Building on the theory of surfaces in $\mathbb{R}^3$, we will describe the notion of Riemannian submanifolds, and study Jacobi fields, which exhibit the interaction between geodesics and curvature. We will prove the Hopf–Rinow theorem, which shows that various notions of completeness are equivalent on Riemannian manifolds, and classify the spaces with constant curvature. The highlight of the course will be to see how curvature influences topology. We will see this by proving the Cartan–Hadamard theorem, Bonnet–Myers theorem and Synge's theorem.

Prerequisities. We will assume familiarity with material from the Differentiable Manifolds, so we recommend you read through the lecture notes of that course. An understanding of the theory of surfaces in $\mathbb{R}^3$, and topological notions such as covering spaces and the fundamental group would also be very helpful.

Disclaimer. These lecture notes are intended to cover the essential course material, but there are no pictures and possibly a few typos. The lectures will contain additional motivation and intuition which will greatly help you to understand the ideas in the course. Moreover, I would suggest combining these lecture notes with material from the recommended reading below.

Recommended texts

- W. Boothby, An Introduction to Differentiable Manifolds and Riemannian Geometry, 2nd edition, (Academic Press, 1986).
- M.P. do Carmo, Riemannian Geometry, (Birkhause, 1992).
- S. Gallot, D. Hulin and J. Lafontaine, Riemannian Geometry, (Springer, 1987).
- J.M. Lee, Riemmanian Manifolds: An Introduction to Curvature, (Springer, 1997).

The most relevant material is M.P. do Carmo, Riemannian Geometry, §0-7, §8.1-8.4 and §9.

Contents

1 Riemannian manifolds: definitions and examples

Riemannian geometry is the study of smooth curved objects, which play a role in analysis, engineering (like imaging), group theory, number theory, physics (especially gravity) and topology. The smooth curved objects in question are called Riemannian manifolds and the basic examples come from surfaces. There are three key examples:

- the flat plane $\mathbb{R}^2$ (which is flat or zero curvature);
- the sphere $\mathcal{S}^2$ (which is positively curved);
- the hyperbolic plane $\mathcal{H}^2$ (which is negatively curved).

These three examples give the basic models for what objects with zero, positive and negative curvature look like even in higher dimensions. (Another way to think about areas of negative curvature is a saddle, like regions near points on the inner circle of a torus in $\mathbb{R}^3$.)

In this course there are two key notions in Riemannian geometry that we will study:

- geodesics – "shortest paths between points"
- curvature – "area of small geodesic triangles: fat (bigger than in flat space/sum of angles$>\pi$) means positive curvature, thin (smaller than flat space/sum of angles$<\pi$) means negative curvature"

Both of these ideas are primarily "local": they are things you can work out near a given point on your object. (This is clear on the torus in $\mathbb{R}^3$ where points on the inner circle are where the torus is negatively curved but on the outer circle it looks more like a piece of a sphere, so is positively curved.)

One of the most striking things about Riemannian geometry is that we can take local information (particularly curvature) and deduce "global" results (particularly concerning topology). Let M be an n-dimensional Riemannian manifold and let K denote its curvature. Some highlights of Riemannian geometry include (stated roughly):

- If $K \leq 0$ then M is essentially $\mathbb{R}^n$ topologically (Cartan–Hadamard Theorem).
- If $K \geq \delta > 0$ then M has finite diameter (and is therefore compact) and there are only finitely many distinct closed loops (Bonnet–Myers Theorem).
- If $\frac{1}{4} < K \leq 1$ then M is essentially the n-dimensional sphere $\mathcal{S}^n$ topologically (Sphere Theorem).

One of the main aims of this course will be the precise statement and understanding of these and similar results. Along the way we will develop the language and tools necessary to formulate and tackle problems in many areas of mathematics.

1.1 Definition

Riemannian geometry was invented by Riemann in his habilitation thesis, which he first announced through his inaugural lecture in 1854, in what must rank as one of the greatest maths job talks ever given! The key idea is to have a notion of a way of measuring distance which varies from point to point, known as a Riemannian metric. We give a fake definition which will give the necessary intuition.

Fake definition: A Riemannian metric g on a manifold M is a smooth choice of positive definite inner product on each tangent space, i.e. for each $p \in M$ we have a symmetric bilinear map $g_p : T_pM \times T_pM \to \mathbb{R}$ which is positive definite.

Remark. We recall that for an n-dimensional manifold M and $p \in M$ we denote by T_pM the *tangent space* to M at p, which is an n-dimensional vector space. We also recall that the tangent bundle

$$TM = \cup_{p\in M} T_pM$$

is a vector bundle of rank n over M whose sections, which we denote by $\Gamma(TM)$, are the *vector fields* on M.

To formally define a Riemannian metric we recall another important vector bundle on a manifold. If M is an n-dimensional manifold we let

$$S^2T_p^*M = \{\text{symmetric bilinear maps } g_p : T_pM \times T_pM \to \mathbb{R}\}.$$

Then

$$S^2T^*M = \cup_{p\in M} S^2T_p^*M$$

is a vector bundle of rank $\frac{1}{2}n(n+1)$ over M. The true definition of Riemannian metric becomes clear.

Definition 1.1. Let M be a manifold. A *Riemannian metric* g on M is a section of S^2T^*M, i.e. $g \in \Gamma(S^2T^*M)$, which is positive definite (meaning that g_p is positive definite for all $p \in M$). We will often simply say that g is a metric on M for brevity. (We will see that Riemannian metrics are, in fact, related to metrics in the sense of metric spaces later.)

A *Riemannian manifold* (M, g) is a manifold M with a Riemannian metric g on M.

Remark. Every manifold admits a Riemannian metric – see the Differentiable Manifolds lecture notes for a proof. Of course the Riemannian metric is not unique (in fact, there are always infinitely many on any given manifold), and the geometry of the Riemannian manifold can vary wildly even though one has the same underlying manifold. For a simple example, consider the sphere, ellipsoid and dumbbell in $\mathbb{R}^3$, which are all diffeomorphic to the 2-sphere $\mathcal{S}^2$, but clearly have very different geometries as curved objects (i.e. Riemannian manifolds).

Therefore, the interesting questions become: how does the choice of manifold restrict the possible Riemannian metric and, conversely, what does the existence of a certain type of Riemannian metric encode about the ambient manifold? These will be central questions that will guide us throughout this course.

1.2 Examples

Let us try to understand what Riemannian metrics are more concretely. Any inner product can be viewed as a symmetric matrix. For example, if $\langle ., . \rangle$ is an inner product on $\mathbb{R}^n$ then there is a symmetric matrix A such that if $x, y \in \mathbb{R}^n$ are vectors then

$$\langle x, y\rangle = x^{\mathrm{T}} Ay.$$

The inner product is positive definite if and only if all of the eigenvalues of A are positive.

Therefore, at each point $p \in M$, we can view g_p as a symmetric matrix, and so (locally) we can think of g as a symmetric matrix of functions. Let us see this in practice on $\mathbb{R}^n$. This is actually all we will need to understand since the picture is local.

Remark. Recall that, if we are given coordinates $(x_1, \ldots, x_n)$ on $\mathbb{R}^n$ then we have the standard vector fields

$$\partial_i = \frac{\partial}{\partial x_i}$$

for $i = 1, \ldots, n$ on $\mathbb{R}^n$, which are everywhere linearly independent. Therefore any vector field X on $\mathbb{R}^n$ can be written as

$$X = \sum_{i=1}^n a_i\partial_i$$

for smooth functions $a_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \ldots, n$. We shall use this notation throughout the course.

Example. On $\mathbb{R}^n$, we have the standard Riemannian metric g_0 which is given by

$$g_0(\sum_{i=1}^n a_i\partial_i, \sum_{j=1}^n b_j\partial_j) = \sum_{i=1}^n a_ib_i,$$

i.e. thinking of tangent vectors as vectors in $\mathbb{R}^n$, g_0 is just the usual dot product on $\mathbb{R}^n$.

We see that

$$g_0(\partial_i, \partial_i) = 1 \quad \text{and} \quad g_0(\partial_i, \partial_j) = 0 \quad \text{if } i \neq j.$$

Hence, with respect to the basis $\{\partial_1, \ldots, \partial_n\}$, the matrix of g_0 is the identity matrix, as we would expect.

As we well know, the matrix of a map depends on the choice of basis. We see this concretely in the next example, which is based on polar coordinates.

Example. Suppose we are on $\mathbb{R}^2$ and we define polar coordinates on $\mathbb{R}^2 \setminus \{0\}$ as usual:

$$x_1 = r\cos\theta \quad \text{and} \quad x_2 = r\sin\theta.$$

Formally, if we define $f : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}^2 \setminus \{0\}$ by

$$f(r, \theta) = (r\cos\theta, r\sin\theta),$$

then f is a local diffeomorphism and if we denote the standard basis vector fields on $\mathbb{R}^+ \times \mathbb{R}$ in the coordinates (r, θ) by ∂_r and ∂_θ then, if we denote the pushforward of f by f_*,

$$X = f_*(\partial_r) = \cos\theta\partial_1 + \sin\theta\partial_2 = \frac{x_1}{r}\partial_1 + \frac{x_2}{r}\partial_2,$$
$$Y = f_*(\partial_\theta) = -r\sin\theta\partial_1 + r\cos\theta\partial_2 = -x_2\partial_1 + x_1\partial_2$$

are everywhere linearly independent on $\mathbb{R}^2 \setminus \{0\}$. It is usual to simply write $X = \partial_r$ and $Y = \partial_\theta$.

We see that, with respect to the standard Euclidean metric, we have

$$g_0(X, X) = \frac{x_1^2 + x_2^2}{r^2} = 1, \quad g_0(X, Y) = 0, \quad g_0(Y, Y) = x_1^2 + x_2^2 = r^2.$$

Hence, g_0 with respect to the basis X, Y on $\mathbb{R}^2 \setminus \{0\}$ is

$$\begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}.$$

So, we see that even though g_0 is independent of the choice of basis, the matrix of g_0 changes, and does not even have to be constant. Moreover, we see that this matrix we have written down has eigenvalues 1 and r^2 so is positive definite as long as $r \neq 0$, which we have assumed.

Remark. The polar coordinates example also shows the following key issue. Since the matrix of g_0 has eigenvalues 1 and r^2 in that example, it is tempting to say that g_0 degenerates (i.e. is no longer positive definite) when $r = 0$. Of course, this is not the case, and instead what is happening is that our polar coordinates degenerate at $r = 0$. Therefore, to check that one has a well-defined Riemannian metric using local coordinates, one needs to be wary that degeneracy in the matrix could be a result of bad choices of coordinates, rather than a genuine failure for the metric to be well-defined.

Example. Let $M \subseteq \mathbb{R}^n$ be a manifold. We can define a Riemannian metric on M by $g_p(X, Y) = g_0(X, Y)$, since if $X, Y \in T_pM$ then $X, Y \in T_p\mathbb{R}^n$. We call this the *induced (Riemannian) metric* on M.

Remark. In the case where M is a surface in $\mathbb{R}^3$ then the induced metric is nothing other than the *first fundamental form* of M.

Example. In particular, by our previous example, we get that the n-sphere $\mathcal{S}^n$ has a Riemannian metric g induced from the Euclidean metric on $\mathbb{R}^{n+1}$, where we let

$$\mathcal{S}^n = \left\{ (x_0, \ldots, x_n) \in \mathbb{R}^{n+1} \,:\, \sum_{i=0}^{n} x_i^2 = 1 \right\}.$$

We call g the *standard round metric* on $\mathcal{S}^n$.

Example. Suppose we are on $\mathcal{S}^2$ and we take standard local coordinates

$$f(\theta,\phi) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$$

where $\theta \in (0,\pi)$ and $\phi \in \mathbb{R}$ (typically, we restrict ϕ to $[0,2\pi)$). We see that if we take the vector fields

$$\begin{aligned} X_1 &= f_*(\partial_\theta) = \cos\theta\cos\phi\partial_1 + \cos\theta\sin\phi\partial_2 - \sin\theta\partial_3 \\ X_2 &= f_*(\partial_\phi) = -\sin\theta\sin\phi\partial_1 + \sin\theta\cos\phi\partial_2 \end{aligned}$$

on $\mathcal{S}^2 \setminus \{N,S\}$, where $N=(0,0,1)$ and $S=(0,0,-1)$ are the North and South poles respectively, they are every linearly independent (and usually just called ∂_θ and ∂_ϕ). Moreover, with respect to the induced metric g on $\mathcal{S}^2$ we have

$$g(X_1,X_1)=1, \quad g(X_1,X_2)=0, \quad g(X_2,X_2)=\sin^2\theta.$$

So we can identify the induced metric on $\mathcal{S}^2$ with the matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & \sin^2\theta \end{pmatrix}$$

away from the poles. We see, in fact, that the eigenvalues of the matrix are 1 and $\sin^2\theta$ and so the matrix is positive definite if and only if $\sin\theta \neq 0$, i.e. we are not at the poles.

Example. In contrast, if we are on $\mathcal{S}^3$ and we take the vector fields

$$\begin{aligned} E_1 &= -x_1\partial_0 + x_0\partial_1 - x_3\partial_2 + x_2\partial_3, \\ E_2 &= -x_2\partial_0 + x_3\partial_1 + x_0\partial_2 - x_1\partial_3, \\ E_3 &= -x_3\partial_0 - x_2\partial_1 + x_1\partial_2 + x_0\partial_3, \end{aligned}$$

then these are everywhere linearly independent on $\mathcal{S}^3$, and with respect to the induced metric g we have

$$g(E_i,E_j) = \delta_{ij}$$

so globally g can be viewed as the identity matrix.

Remark. The previous two examples indicate how there is sometimes trade-off between making a natural choice of basis of vector fields to write the metric, but then the metric coefficients become functions, or choosing a special basis so that the metric becomes constant (usually the identity matrix). We will see that both types of bases have their uses.

Remark. On any parallelizable manifold M we can find a basis for the vector fields on (M,g) so that g can be globally viewed as the identity matrix. The same argument shows that for any Riemannian manifold, on any chart we can choose a basis for the vector fields so that the metric is given by the identity matrix on that chart.

Example. Let $f:\mathbb{R}^2\to\mathbb{R}^3$ be given by

$$f(\theta,\phi) = ((2+\cos\theta)\cos\phi, (2+\cos\theta)\sin\phi, \sin\theta)$$

so that $f(\mathbb{R}^2)$ is the 2-torus $T^2 \subseteq \mathbb{R}^3$.

Then

$$\begin{aligned} X_1 &= f_*(\partial_\theta) = -\sin\theta\cos\phi\partial_1 - \sin\theta\sin\phi\partial_2 + \cos\theta\partial_3, \\ X_2 &= f_*(\partial_\phi) = -(2+\cos\theta)\sin\phi\partial_1 + (2+\cos\theta)\cos\phi\partial_2 \end{aligned}$$

are vector fields on T^2 which are everywhere linearly independent (and again usually just called ∂_θ and ∂_ϕ). We see that, with respect to the induced metric g, we have

$$g(X_1, X_1) = 1, \quad g(X_1, X_2) = 0, \quad g(X_2, X_2) = (2 + \cos\theta)^2.$$

So, we can identify g with the matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & (2+\cos\theta)^2 \end{pmatrix}.$$

We see that this matrix is positive definite everywhere, and so gives a global formula for the metric.

1.3 Pullback and local metrics

The examples show that when we have a diffeomorphism $f : M \to N$ between manifolds M and N, vector fields $X, Y \in \Gamma(TM)$ on M and a metric h on N, then we can see how the metric acts on the pushforward vector fields $f_*(X)$ and $f_*(Y)$. This then seems to give us a metric g on M defined by $g(X, Y) = h(f_*(X), f_*(Y))$. This leads us to the recall the following definition.

Definition 1.2. Let $f : M \to N$ be a smooth map between manifolds M and N and let h be a Riemannian metric on N. We define the pullback f^*h of h by f as:

$$(f^*h)_p(X, Y) = h_{f(p)}(df_p(X), df_p(Y))$$

for $p \in M$ and $X, Y \in T_pM$. If X, Y are vector fields on M then

$$(f^*h)(X, Y) = h(f_*(X), f_*(Y)).$$

We saw in the Differentiable Manifolds course that the pullback can take a Riemannian metric on N to a Riemannian metric on M as follows.

Proposition 1.3. *Let M be a manifold and let (N, h) be a Riemannian manifold. Let $f : M \to N$ be an immersion (so $\mathrm{d}f_p$ is injective for all $p \in M$). Then $g = f^*h$ is a Riemannian metric on M.*

Remark. In particular, if f is a diffeomorphism then f^*h is a Riemannian metric, since the differential $\mathrm{d}f_p$ is an isomorphism.

Proof. Let $p \in M$ and let $X, Y \in T_pM$. Since h is symmetric and bilinear and smooth and f is smooth, we see that g is symmetric and bilinear and smooth, so we only need to check that it is positive definite.

We see that

$$g_p(X, X) = h_{f(p)}(\mathrm{d}f_p(X), \mathrm{d}f_p(X)) \geq 0$$

and $g_p(X, X) = 0$ if and only if $\mathrm{d}f_p(X) = 0$. But $\mathrm{d}f_p$ is injective so $\mathrm{d}f_p(X) = 0$ if and only if $X = 0$. Hence g is positive definite and thus g is a Riemannian metric. □

Example. Let $M \subseteq \mathbb{R}^n$ be a manifold and let $i : M \to \mathbb{R}^n$ be the inclusion map. Then i is an immersion so $g = i^*g_0$ is a Riemannian metric. This metric is just the induced metric we saw before.

Before we continue, we make the following useful definition.

Definition 1.4. Let (U, φ) be a chart on an n-dimensional manifold M, so that U is an open set in M and $\varphi : U \to \varphi(U) \subseteq \mathbb{R}^n$ is a diffeomorphism onto an open subset of $\mathbb{R}^n$. The *coordinate vector fields* in the chart (U, φ) are given by

$$X_i = (\varphi^{-1})_*(\partial_i) \quad \text{for } i = 1, \ldots, n.$$

In other words, we pushforward the standard vector fields on $\mathbb{R}^n$, restricted to $\varphi(U)$, using the diffeomorphism $\varphi^{-1} : \varphi(U) \to U$. We also say $\{X_1, \ldots, X_n\}$ is the *coordinate frame field* in (U, φ).

Remark. As we have seen, it is often convenient to view the chart (U, φ) in terms of a map f from an open set in $\mathbb{R}^n$ into M, so that $f = \varphi^{-1}$. We already saw this explicitly in the case of $\mathcal{S}^2$ and $T^2 \subseteq \mathbb{R}^3$, for example.

Given a chart (U, φ) on (M, g), $\varphi^{-1} : \varphi(U) \to U \subseteq M$ is a diffeomorphism (so in particular an immersion). Hence, $(\varphi^{-1})^* g$ is a Riemannian metric on $\varphi(U) \subseteq \mathbb{R}^n$, so we can write it in terms of a symmetric matrix of functions on $\mathbb{R}^n$. In particular, we see that

$$(\varphi^{-1})^* g(\partial_i, \partial_j) = g((\varphi^{-1})_*(\partial_i), (\varphi^{-1})_*(\partial_j)) = g(X_i, X_j),$$

where X_i are the coordinate vector fields. Thus, the matrix of g with respect to the coordinate vector fields on U is the same as the matrix of $(\varphi^{-1})^* g$ with respect to the standard vector fields on $\mathbb{R}^n$. This means we can easily write down local expressions for Riemannian metrics.

Alternatively, we can also write the Euclidean metric g_0 on $\mathbb{R}^n$ as

$$g_0 = \mathrm{d}x_1^2 + \ldots + \mathrm{d}x_n^2.$$

The rule is that

$$\mathrm{d}x_i \mathrm{d}x_j(\partial_k, \partial_l) = \mathrm{d}x_i \mathrm{d}x_j(\partial_l, \partial_k) = \begin{cases} 1 & \text{if } i = k, j = l \text{ or } i = l, j = k \\ 0 & \text{otherwise} \end{cases}$$

Then any Riemannian metric on $\mathbb{R}^n$ can be written as

$$\sum_{i,j} g_{ij} \mathrm{d}x_i \mathrm{d}x_j$$

where g_{ij} is a positive definite symmetric matrix of functions. We see that if we write

$$(\varphi^{-1})^* g = \sum_{i,j} g_{ij} \mathrm{d}x_i \mathrm{d}x_j$$

then

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

This gives us a way to think about Riemannian metrics on any Riemannian manifold in terms of symmetric positive definite matrices of functions on $\mathbb{R}^n$, at least locally.

Remark. We shall use the notation g_{ij} frequently in the rest of the course for the functions $g(X_i, X_j)$ where $\{X_1, \ldots, X_n\}$ is the coordinate frame field in the chart (U, φ).

Example. Let H^n be the n-dimensional upper half-space

$$H^n = \{(x_1, \ldots, x_n) : x_n > 0\}$$

and define the *hyperbolic metric* g on H^n by

$$g = \frac{\mathrm{d}x_1^2 + \ldots + \mathrm{d}x_n^2}{x_n^2}.$$

This metric plays an important role in geometry and topology.

Example. If $f : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}^2 \setminus \{0\}$ is $f(r, \theta) = (r \cos\theta, r \sin\theta)$ then $X_1 = f_*(\partial_r)$ and $X_2 = f_*(\partial_\theta)$ in our notation before, so

$$\begin{aligned}
f^* g_0(\partial_r, \partial_r) &= g_0(f_*(\partial_r), f_*(\partial_r)) = 1, \\
f^* g_0(\partial_r, \partial_\theta) &= g_0(f_*(\partial_r), f_*(\partial_\theta)) = 0, \\
f^* g_0(\partial_\theta, \partial_\theta) &= g_0(f_*(\partial_\theta), f_*(\partial_\theta)) = r^2.
\end{aligned}$$

Therefore

$$f^*g_0 = \mathrm{d}r^2 + r^2\mathrm{d}\theta^2.$$

Example. Let $f(\theta,\phi) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$ define local coordinates on $\mathcal{S}^2$. The standard induced Riemannian metric g on $\mathcal{S}^2$ is determined in the coordinates (θ,ϕ) by

$$f^*g(\partial_\theta,\partial_\theta) = g_0(f_*\partial_\theta, f_*\partial_\theta) = 1, \quad f^*g(\partial_\theta,\partial_\phi) = 0, \quad f^*g(\partial_\phi,\partial_\phi) = \sin^2\theta$$

so

$$f^*g = \mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2.$$

in these coordinates (which, again, should look familiar).

1.4 Isometries and local isometries

Before we give more examples of Riemannian manifolds, we want to understand when two Riemannian manifolds are the same. Clearly being diffeomorphic is not enough, since we can have many different Riemannian metrics on $\mathcal{S}^2$, for example. The correct notion is the obvious one we now give.

Definition 1.5. A smooth map $f : (M, g) \to (N, h)$ between Riemannian manifolds is an *isometry* if f is a diffeomorphism and $g = f^*h$. (Notice that this makes sense because if f is a diffeomorphism then f^*h is a Riemannian metric by Proposition 1.3).

Clearly, the isometries on (M, g) form a group, in fact a subgroup of $\mathrm{Diff}(M)$, which we denote $\mathrm{Isom}(M, g)$.

We say that f is a *local isometry at p* if there exists open sets $U \ni p$ and $V \ni f(p)$ such that $f : U \to V$ is an isometry, and that f is a *local isometry* if it is a local isometry at all $p \in M$.

Example. The identity map $\mathrm{id} : (M, g) \to (M, g)$ is an isometry.

Example. Recall for a linear map $f(x) = Ax$ on $\mathbb{R}^n$ we have that $f^*g_0 = g_0$ if and only if $A \in \mathrm{O}(n)$.

The reason is that f_* is multiplication by A so

$$f^*g_0(\partial_i,\partial_j) = g_0(f_*\partial_i, f_*\partial_j) = g_0(A\partial_i, A\partial_j) = g_0\left(\sum_{k=1}^n a_{ki}\partial_k, \sum_{l=1}^n a_{lj}\partial_l\right) = \sum_{k=1}^n a_{ki}a_{kj}$$

since $g_0(\partial_i,\partial_j) = \delta_{ij}$.

Thus $f^*g_0 = g_0$ if and only if $\sum_{k=1}^n a_{ki}a_{kj} = \delta_{ij}$, i.e. $A^{\mathrm{T}}A = I$, so $A \in \mathrm{O}(n)$.

Notice that translations $f(x) = x + a$ for any $a \in \mathbb{R}^n$ are also isometries, since $f^* = \mathrm{id}$.

Hence (modulo the fact that you need to prove isometries are linear) we have $\mathrm{Isom}(\mathbb{R}^n, g_0) = \mathrm{O}(n) \ltimes \mathbb{R}^n$.

Example. Clearly, $\mathrm{Isom}(S^n, g) = \mathrm{O}(n+1)$ for the standard round metric by the previous examples (since this is the subgroup of the isometry group of $\mathbb{R}^{n+1}$ which preserves the n-sphere).

Example. Let us consider (H^2, g). Let $z = x_1 + ix_2$, so that

$$g = \frac{\mathrm{d}z\mathrm{d}\bar{z}}{|\mathrm{Im}z|^2}.$$

If $f : H^2 \to H^2$ is holomorphic then

$$f^*\mathrm{d}z = \mathrm{d}(f(z)) = f'(z)\mathrm{d}z$$

and

$$f^*\mathrm{d}\bar{z} = \overline{f'(z)}\mathrm{d}\bar{z}$$

so
$$f^*g = \frac{|f'(z)|^2 \mathrm{d}z\mathrm{d}\bar{z}}{|\mathrm{Im} f(z)|^2}.$$
Hence f is an isometry if and only if it is a diffeomorphism such that
$$|f'(z)|^2 |\mathrm{Im} z|^2 = |\mathrm{Im} f(z)|^2.$$
Now if we let
$$f(z) = \frac{az+b}{cz+d}$$
where $a, b, c, d \in \mathbb{R}$ with $ad - bc = 1$, so identified with a matrix
$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{R}),$$
then
$$\begin{aligned} f(z) &= f(x_1 + ix_2) \\ &= \frac{ax_1 + aix_2 + b}{cx_1 + cix_2 + d} \\ &= \frac{(acx_1^2 + acx_2^2 + bd) + i(ad-bc)x_2}{|cz+d|^2} \\ &= \frac{(ac|z|^2 + bd) + i\mathrm{Im} z}{|cz+d|^2}, \end{aligned}$$
and
$$f'(z) = \frac{ad-bc}{(cz+d)^2} = (cz+d)^{-2}.$$
Hence we see that, since f sends H^2 to H^2 and is smooth with smooth inverse
$$f^{-1}(z) = \frac{dz-b}{-cz+a},$$
we deduce that f is an isometry. In fact, these *Möbius transformations* give all of the orientation preserving isometries of H^2.

Notice the isometries include *dilations*! This is very surprising, but hints as to the nature of hyperbolic geometry.

Example. Recall that for $\mathrm{SU}(n)$ we have
$$T_A \,\mathrm{SU}(n) = \{B \in M_n(\mathbb{C}) \,:\, \bar{A}^{\mathrm{T}} B + \bar{B}^{\mathrm{T}} A = 0,\, \mathrm{tr}(\bar{A}^{\mathrm{T}} B) = 0\} = \{AX \in M_n(\mathbb{C}) \,:\, X + \bar{X}^{\mathrm{T}} = 0, \mathrm{tr}(X) = 0\}.$$

I claim that g given by
$$g_A(B, C) = -\,\mathrm{tr}(\bar{A}^{\mathrm{T}} B \bar{A}^{\mathrm{T}} C) = -\,\mathrm{tr}(XY) = g_A(AX, AY)$$
for all $A \in \mathrm{SU}(n)$, $B = AX, C = AY \in T_A \,\mathrm{SU}(n)$ is a Riemannian metric. Notice that
$$\overline{\mathrm{tr}(XY)} = \mathrm{tr}(\bar{X}\bar{Y}) = \mathrm{tr}(X^{\mathrm{T}} Y^{\mathrm{T}}) = \mathrm{tr}((YX)^{\mathrm{T}}) = \mathrm{tr}(YX) = \mathrm{tr}(XY).$$
It is also positive definite because if we write $\mathbf{x}_1, \ldots, \mathbf{x}_n$ for the columns of X then
$$-\,\mathrm{tr}(X^2) = \mathrm{tr}(\bar{X}^{\mathrm{T}} X) = \sum_{j=1}^n |\mathbf{x}_j|^2.$$
Hence g is a Riemannian metric on $\mathrm{SU}(n)$.

Let $L_C : \mathrm{SU}(n) \to \mathrm{SU}(n)$ be given by $L_C(A) = CA$. I claim that L_C is an isometry. For $AX, AY \in T_A \,\mathrm{SU}(n)$ we have
$$(L_C^* g)_A(AX, AY) = g_{CA}((L_C)_*(AX), (L_C)_*(AY)) = g_{CA}(CAX, CAY) = -\,\mathrm{tr}(XY) = g_A(AX, AY).$$

Hence g is *left-invariant*. Moreover $R_C : \mathrm{SU}(n) \to \mathrm{SU}(n)$ given by $R_C(A) = AC$ is an isometry since

$$\begin{aligned}(R_C^* g)_A(AX, AY) &= g_{AC}(AXC, AYC) = -\operatorname{tr}(\overline{AC}^{\mathrm{T}} AXC\overline{AC}^{\mathrm{T}} AYC) = -\operatorname{tr}(\bar{C}^{\mathrm{T}} XYC)\\ &= -\operatorname{tr}(XY) = g_A(AX, AY)\end{aligned}$$

so g is also *right-invariant*. Hence, we call g a *bi-invariant* Riemannian metric (i.e. both left and right-invariant).

In the special case of $\mathrm{SU}(2) \cong \mathcal{S}^3$, this metric is (up to a multiplicative constant) nothing other than the standard round metric, and left and right-multiplication are rotations which generate $\mathrm{SO}(4)$.

We now make an interesting observation, which shows that the local functions defining a Riemannian metric can help detect whether two Riemannian manifolds are locally isometric.

If we have charts (U, φ) on (M, g) and (V, ψ) on (N, h) such that $\varphi(U) = \psi(V) = W$ and $(\varphi^{-1})^* g = (\psi^{-1})^* h$ on W then

$$(\psi^{-1} \circ \varphi)^* h = \varphi^* \circ (\psi^{-1})^* h = \varphi^* \circ (\varphi^{-1})^* g = g$$

so the map $f = \psi^{-1} \circ \varphi : U \to V$ is an isometry. This is equivalent to saying that

$$g_{ij} = g((\varphi^{-1})_* \partial_i, (\varphi^{-1})_* \partial_j) = (\varphi^{-1})^* g(\partial_i, \partial_j) = (\psi^{-1})^* h(\partial_i, \partial_j) = h_{ij},$$

i.e. the local functions g_{ij} and h_{ij} are equal.

Let us now look at some more sophisticated examples of Riemannian manifolds.

Example. We have minimal surfaces (that is, surfaces whose mean curvature is 0) in $\mathbb{R}^3$ known as the *helicoid*

$$M_1 = \{(s\cos t, s\sin t, t) \ : \ s, t \in \mathbb{R}\}$$

and the *catenoid*

$$M_2 = \{(\cosh z \cos\theta, \cosh z \sin\theta, z) \ : \ z, \theta \in \mathbb{R}\}.$$

Define local coordinates on M_1 by

$$f_1(x_1, x_2) = (\sinh x_1 \cos x_2, \sinh x_1 \sin x_2, x_2)$$

and on M_2 by

$$f_2(x_1, x_2) = (\cosh x_1 \cos x_2, \cosh x_1 \sin x_2, x_1).$$

Then

$$(f_1)_*\partial_1 = \cosh x_1 \cos x_2 \partial_1 + \cosh x_1 \sin x_2 \partial_2 \quad \text{and} \quad (f_1)_*\partial_2 = -\sinh x_1 \sin x_2 \partial_1 + \sinh x_1 \cos x_2 \partial_2 + \partial_3,$$

so

$$(f_1)^* g_0 = \cosh^2 x_1 \mathrm{d}x_1^2 + (1 + \sinh^2 x_1)\mathrm{d}x_2^2 = \cosh^2 x_1(\mathrm{d}x_1^2 + \mathrm{d}x_2^2).$$

Similarly,

$$(f_2)_*\partial_1 = \sinh x_1 \cos x_2 \partial_1 + \sinh x_1 \sin x_2 \partial_2 + \partial_3 \quad \text{and} \quad (f_2)_*\partial_2 = -\cosh x_1 \sin x_2 \partial_1 + \cosh x_1 \cos x_2 \partial_2$$

and hence

$$(f_2)^* g_0 = (1 + \sinh^2 x_1)\mathrm{d}x_1^2 + \cosh^2 x_1 \mathrm{d}x_2^2 = \cosh^2 x_1(\mathrm{d}x_1^2 + \mathrm{d}x_2^2).$$

We deduce that M_1 and M_2 are locally isometric.

Example. If we consider the *pseudo-sphere*

$$M = \left\{\left(t - \tanh t, \frac{\cos\theta}{\cosh t}, \frac{\sin\theta}{\cosh t}\right) \ : \ t, \theta \in \mathbb{R}\right\}$$

and let $f:\mathbb{R}^+\times\mathbb{R}$ be the obvious parametrization of $M\cap\{(x_1,x_2,x_3)\in\mathbb{R}^3 : x_1>0\}$ i.e.

$$f(t,\theta)=\left(t-\tanh t,\frac{\cos\theta}{\cosh t},\frac{\sin\theta}{\cosh t}\right)$$

then

$$f_*(\partial_t)=\tanh^2 t\partial_1-\frac{\cos\theta\sinh t}{\cosh^2 t}\partial_2-\frac{\sin\theta\sinh t}{\cosh^2 t}\partial_3$$

and

$$f_*(\partial_\theta)=-\frac{\sin\theta}{\cosh t}\partial_2+\frac{\cos\theta}{\cosh t}\partial_3.$$

Hence the induced Riemannian metric on M is given in these coordinates by $g=f^*g_0$ so

$$g(\partial_t,\partial_t)=\tanh^4 t+\mathrm{sech}^2 t\tanh^2 t=\tanh^2 t,$$

$g(\partial_t,\partial_\theta)=0$ and

$$g(\partial_\theta,\partial_\theta)=\mathrm{sech}^2 t.$$

So the metric $g=\tanh^2 t\mathrm{d}t^2+\mathrm{sech}^2 t\mathrm{d}\theta^2$ in these coordinates.

Now define $j:\{(x_1,x_2)\in\mathbb{R}^2 : x_2>1\}\to\mathbb{R}^+\times\mathbb{R}$ by $j(x_1,x_2)=(\cosh^{-1}x_2,x_1)$. Then we see that

$$j_*(\partial_1)=\partial_\theta \quad\text{and}\quad j_*(\partial_2)=\frac{1}{\sqrt{x_1^2-1}}\partial_t=\frac{1}{\sinh t}\partial_t.$$

Hence, the metric on M in these coordinates is given by $h=(f\circ j)^*g_0=j^*g$ which satisfies

$$h(\partial_1,\partial_1)=g(\partial_\theta,\partial_\theta)=\mathrm{sech}^2 t=x_2^{-2},$$

$h(\partial_1,\partial_2)=0$ and

$$h(\partial_2,\partial_2)=\frac{1}{\sinh^2 t}g(\partial_t,\partial_t)=\frac{\tanh^2 t}{\sinh^2 t}=\mathrm{sech}^2 t=x_2^{-2}.$$

Thus in these coordinates the metric is given by $h=\frac{\mathrm{d}x_1^2+\mathrm{d}x_2^2}{x_2^2}$. This shows that the pseudosphere (minus a circle) is locally isometric to the upper half-plane with the *hyperbolic metric*, where the local isometry is

$$f\circ j(x_1,x_2)=f(\cosh^{-1}x_2,x_1)=\left(\cosh^{-1}x_2-\frac{\sqrt{x_2^2-1}}{x_2},\frac{\cos x_1}{x_2},\frac{\sin x_1}{x_2}\right)$$

1.5 Group actions

Example. Given a discrete group G acting freely and properly discontinuously on a manifold M, the quotient map $\pi:M\to M/G$ is a local diffeomorphism and hence an immersion. Thus by Proposition 1.3, if we have a Riemannian metric h on M/G we can define a Riemannian metric g on M by $g=\pi^*h$.

In general we cannot pushforward a Riemannian metric, but there is a special case where we can.

Theorem 1.6. *Let G be a discrete group acting freely and properly discontinuously by isometries on a Riemannian manifold (M,g); i.e. suppose that if $x\in G\mapsto f_x\in\mathrm{Diff}(M)$ denotes the group action then $f_x\in\mathrm{Isom}(M,g)$ for all $x\in G$.*

There exists a Riemannian metric h on M/G such that if $\pi:M\to M/G$ is the projection map then $g=\pi^*h$ (so π is a local isometry).

Proof. The idea is to define h so that $g=\pi^*h$ and show that this is well-defined.

First observe that the map $\mathrm{d}\pi_p:T_pM\to T_{\pi(p)}M/G$ is an isomorphism for all $p\in M$ and π is surjective, so we can define h by

$$h_{\pi(p)}(X,Y)=g_p((\mathrm{d}\pi_p)^{-1}X,(\mathrm{d}\pi_p)^{-1}Y).$$

To show that is well-defined we need to show it does not depend on the choice of p, so if $\pi(q) = \pi(p)$ we should get the same answer on the right-hand side for q as for p. So, if we choose q such that $\pi(q) = \pi(p)$ then $q = f_x(p)$ for some $x \in G$ by definition of π so $\pi(p) = \pi \circ f_x(p)$ and hence differentiating gives

$$\mathrm{d}\pi_p = \mathrm{d}(\pi \circ f_x)_p = \mathrm{d}\pi_{f_x(p)} \circ \mathrm{d}(f_x)_p = \mathrm{d}\pi_q \circ \mathrm{d}(f_x)_p.$$

We deduce that

$$\mathrm{d}\pi_q = \mathrm{d}\pi_p \circ \big(\mathrm{d}(f_x)_p\big)^{-1}$$

and thus

$$(\mathrm{d}\pi_q)^{-1} = \mathrm{d}(f_x)_p \circ (\mathrm{d}\pi_p)^{-1}.$$

We deduce that

$$\begin{aligned} g_q((\mathrm{d}\pi_q)^{-1}X, (\mathrm{d}\pi_q)^{-1}Y) &= g_{f_x(p)}(\mathrm{d}(f_x)_p \circ (\mathrm{d}\pi_p)^{-1}X, \mathrm{d}(f_x)_p \circ (\mathrm{d}\pi_p)^{-1}Y) \\ &= (f_x^* g)_p((\mathrm{d}\pi_p)^{-1}X, (\mathrm{d}\pi_p)^{-1}Y) \\ &= g_p((\mathrm{d}\pi_p)^{-1}X, (\mathrm{d}\pi_p)^{-1}Y) \end{aligned}$$

as f_x is an isometry. Morever, h is positive definite because

$$h_{\pi(p)}(X, X) = g_p((\mathrm{d}\pi_p)^{-1}X, (\mathrm{d}\pi_p)^{-1}X) \geq 0$$

and equals zero if and only if $X = 0$ since $\mathrm{d}\pi_p^{-1}$ is an isomorphism.

We also see that, by definition,

$$g_p(U, V) = h_{\pi(p)}(\mathrm{d}\pi_p U, \mathrm{d}\pi_p V) = (\pi^* h)_p(U, V)$$

for all $p \in M$ and $U, V \in T_pM$ so $g = \pi^* h$. (In fact, h is clearly the unique choice of Riemannian metric on M/G such that $\pi^* h = g$.) $\square$

Example. Since id and $-$id are isometries on $\mathbb{R}^{n+1}$, we see that $\mathbb{RP}^n$, the Möbius band and the Klein bottle obtain Riemannian metrics from $\mathcal{S}^n$, the cylinder and the torus in $\mathbb{R}^3$ respectively.

Example. We know that $\mathbb{R}^n/\mathbb{Z}^n$ inherits a Riemannian metric g from $\mathbb{R}^n$ where $\mathbb{Z}^n$ acts on $\mathbb{R}^n$ by $x \mapsto x + 2\pi a$ for $a \in \mathbb{Z}^n$, and that $\pi : \mathbb{R}^n \to \mathbb{R}^n/\mathbb{Z}^n$ is then a local isometry. We also see that if $f : \mathbb{R}^n/\mathbb{Z}^n \to T^n \subseteq \mathbb{R}^{2n}$ is the natural diffeomorphism

$$f([(x_1, \ldots, x_n)]) = \big(\cos(x_1), \sin(x_1), \ldots, \cos(x_n), \sin(x_n)\big)$$

and h is the induced Riemannian metric on T^n, we have that

$$f^* h = g.$$

Example. We recall a construction from the Differentiable Manifolds course to get a Riemannian metric on $\mathbb{CP}^n$. Recall the vector field

$$E = \sum_{j=1}^{n+1} x_{2j-1}\partial_{2j} - x_{2j}\partial_{2j-1}$$

on $\mathcal{S}^{2n+1}$, which satisfies $\pi_*(E) = 0$ where $\pi : \mathcal{S}^{2n+1} \to \mathbb{CP}^n$ is the projection map. For $z \in \mathcal{S}^{2n+1}$ we have $E(z) = iz$ (identifying tangent vectors in $\mathbb{C}^n$ with $\mathbb{C}^n$) and we let

$$H_z = \{X \in T_z\mathcal{S}^{2n+1} \,:\, g(X, E(z)) = 0\}$$

where g is the round metric on $\mathcal{S}^{2n+1}$ and we know

$$\Phi_z = \mathrm{d}\pi_z : H_z \to T_{\pi(z)}\mathbb{CP}^n$$

is invertible so we can define a Riemannian metric h on $\mathbb{CP}^n$ by

$$h_{\pi(z)}(X, Y) = g_z(\Phi_z^{-1}(X), \Phi_z^{-1}(Y)),$$

which is called the *Fubini–Study metric.*

This is related to group actions because $\mathbb{CP}^n = \mathcal{S}^{2n+1}/\,\mathrm{U}(1)$.

2 The Levi-Civita connection

Let (M, g) be a Riemannian manifold. In this section we want to define a fundamental object in Riemannian geometry called the *Levi-Civita connection*.

2.1 Fundamental Theorem of Riemannian Geometry

We recall that the Lie bracket $[X, Y] = X \circ Y - Y \circ X$, which is equal to the Lie derivative $\mathcal{L}_X Y$, gave a means to measure how the vector field Y varies with respect to X. We will now introduce another method for differentiating vector fields, depending on the Riemannian metric, which allows us to "connect" tangent spaces (i.e. compare tangent vectors in different tangent spaces). This is a key idea in Riemannian geometry which we state as the following theorem.

Theorem 2.1 (Fundamental Theorem of Riemannian Geometry). *Let (M, g) be a Riemannian manifold. There exists a unique map $\nabla : \Gamma(TM) \times \Gamma(TM) \to \Gamma(TM)$ denoted by $\nabla : (X, Y) \mapsto \nabla_X Y$ such that, if $X, Y, Z \in \Gamma(TM)$ and a, b are smooth functions on M then:*

(i) $\nabla_{aX+bY} Z = a\nabla_X Z + b\nabla_Y Z$,

(ii) $\nabla_X(Y + Z) = \nabla_X Y + \nabla_X Z$,

(iii) $\nabla_X(aY) = a\nabla_X Y + X(a)Y$,

(iv) $X\big(g(Y, Z)\big) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z)$,

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

We call $\nabla_X Y$ the covariant derivative of Y with respect to X and call ∇ the Levi-Civita connection of g.

Remark. Properties (i)-(iii) say ∇ is a *connection* (on TM). Property (iv) says that the connection is compatible with the Riemannian metric g. Property (v) says that the connection is *torsion-free* (or symmetric).

One can define connections more generally which do not satisfy properties (iv) or (v) and on other vector bundles but we shall not be concerned with them in this course, although they are of importance in differential geometry.

Proof. The proof goes as follows. You first suppose that there is a map ∇ satisfying (i)-(v). You then deduce that you get a formula which defines ∇. So, then any other map ∇' satisfying (i)-(v) is also defined by the same formula so must equal ∇, hence ∇ is unique if it exists. Second, you define a map ∇ by the formula and show that it satisfies (i)-(v), which means you have constructed ∇.

Suppose ∇ exists first and satisfies (i)-(v). Then (iv) implies that

$$\begin{aligned} X\big(g(Y, Z)\big) &= g(\nabla_X Y, Z) + g(Y, \nabla_X Z), \\ Y\big(g(Z, X)\big) &= g(\nabla_Y Z, X) + g(Z, \nabla_Y X), \\ Z\big(g(X, Y)\big) &= g(\nabla_Z X, Y) + g(X, \nabla_Z Y). \end{aligned}$$

We deduce from (v) that

$$X\big(g(Y, Z)\big) + Y\big(g(Z, X)\big) - Z\big(g(X, Y)\big) = 2g(\nabla_X Y, Z) + g(X, [Y, Z]) - g(Y, [Z, X]) - g(Z, [X, Y]).$$

Re-arranging, we see that

$$g(\nabla_X Y, Z) = \frac{1}{2}\Big(X\big(g(Y, Z)\big) + Y\big(g(Z, X)\big) - Z\big(g(X, Y)\big) - g(X, [Y, Z]) + g(Y, [Z, X]) + g(Z, [X, Y])\Big), \quad (*)$$

so $\nabla_X Y$ is uniquely defined by g if it exists using this formula (*) (called the Koszul formula).

Now, we can also define $\nabla_X Y$ by (*) and we just need to check (i)-(v) are satisfied.

For (i), if W is another vector field, we can calculate

$$\begin{aligned} g(\nabla_{aX+bY}Z,W) &= \frac{1}{2}\Big((aX+bY)\big(g(Z,W)\big)+Z\big(g(W,aX+bY)\big)-W\big(g(aX+bY,Z)\big) \\ &\quad - g(aX+bY,[Z,W])+g(Z,[W,aX+bY])+g(W,[aX+bY,Z])\Big) \\ &= g(a\nabla_X Z+b\nabla_Y Z,W)+\frac{1}{2}\Big(Z(a)g(W,X)+Z(b)g(W,Y)-W(a)g(X,Z) \\ &\quad - W(b)g(Y,Z)+g(Z,W(a)X+W(b)Y)-g(W,Z(a)X+Z(b)Y)\Big)\end{aligned}$$

which gives (i).

Property (ii) is obvious as everything on the right-hand side of (*) is linear in its arguments.

For property (iii) we need to make an observation about the Lie bracket. We see that

$$\begin{aligned}[aX,bY] &= (aX)\circ(bY)-(bY)\circ(aX) \\ &= ab(X\circ Y)+aX(b)Y-ab(Y\circ X)-bY(a)X \\ &= ab[X,Y]+aX(b)Y-bY(a)X.\end{aligned}$$

We can then compute:

$$\begin{aligned} g(\nabla_X(aY),Z) &= \frac{1}{2}\Big(X\big(g(aY,Z)\big)+aY\big(g(Z,X)\big)-Z\big(g(X,aY)\big) \\ &\quad - g(X,[aY,Z])+g(aY,[Z,X])+g(Z,[X,aY])\Big) \\ &= g(a\nabla_X Y,Z)+\frac{1}{2}\Big(X(a)g(Y,Z)-Z(a)g(X,Y)+g(X,Z(a)Y)+g(Z,X(a)Y)\Big).\end{aligned}$$

Now for (iv) we see that the last five terms in (*) are anti-symmetric in Y,Z, so $g(\nabla_X Y,Z)+g(\nabla_X Z,Y)=X\big(g(Y,Z)\big)$.

Finally, for (v) we see that the first five terms in (*) are symmetric in X,Y (in particular $g(X,[Z,Y])=-g(X,[Y,Z])$ and $g(Y,[X,Z])=-g(Y,[Z,X])$), so $g(\nabla_X Y-\nabla_Y X,Z)=g(Z,[X,Y])$. ☐

Remark. Recall that, for a diffeomorphism f and vector fields X,Y we have $f_*[X,Y]=[f_*X,f_*Y]$. Therefore, since the standard vector fields on $\mathbb{R}^n$ satisfy

$$[\partial_i,\partial_j]=0,$$

the coordinate vector fields in a chart (U,φ) on M always satisfy

$$[X_i,X_j]=0.$$

This is handy when using the Koszul formula.

Let us try to understand the Levi-Civita connection in simple examples.

Example. On $\mathbb{R}^n$, $[\partial_i,\partial_j]=0$ and $g_0(\partial_i,\partial_j)=\delta_{ij}$ are constant functions (where ∂_i are the standard vector fields as usual), so $g_0(\nabla_{\partial_i}\partial_j,\partial_k)=0$ and hence

$$\nabla_{\partial_i}\partial_j=0.$$

Example. On the standard n-torus $T^n\subseteq\mathbb{R}^{2n}$ we have the standard coordinates

$$f(\theta_1,\ldots,\theta_n)=(\cos\theta_1,\sin\theta_1,\ldots,\cos\theta_n,\sin\theta_n)$$

and coordinate vector fields

$$X_i=f_*(\partial_i)=-\sin\theta_i\partial_{2i-1}+\cos\theta_i\partial_{2i}.$$

We see that $g(X_i,X_j)=\delta_{ij}$ constant and $[X_i,X_j]=0$, so

$$\nabla_{X_i}X_j=0.$$

Example. On $\mathcal{S}^2$ we let $f(\theta,\phi) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$ and let $X_1 = f_*\partial_\theta$ and $X_2 = f_*\partial_\phi$ be the coordinate vector fields on $\mathcal{S}^2$. Then $[X_1, X_2] = 0$. We also have $g(X_1, X_1) = 1$, $g(X_1, X_2) = 0$ and $g(X_2, X_2) = \sin^2\theta$. Of course these are really functions on $\mathcal{S}^2$ but it is clear that we can think of them as functions on $\mathbb{R}^2$ (using the identification f). Under this identification if $h = h(\theta,\phi)$ is a function on $\mathcal{S}^2$ then we have that

$$X_1(h) = \frac{\partial h}{\partial\theta} \quad \text{and} \quad X_2(h) = \frac{\partial h}{\partial\phi}.$$

This is always true whenever we differentiate using the coordinate vector fields. We can therefore calculate by (*)

$$g(\nabla_{X_1}X_1, X_1) = \frac{1}{2}X_1(g(X_1, X_1)) = 0$$

and

$$g(\nabla_{X_1}X_1, X_2) = \frac{1}{2}\big(2X_1(g(X_1, X_2)) - X_2(g(X_1, X_1))\big) = 0$$

since $g(X_1, X_1)$ is constant and $g(X_1, X_2) = 0$. Since X_1, X_2 form a basis when $\sin\theta \neq 0$,

$$\nabla_{X_1}X_1 = 0.$$

However, we see that

$$g(\nabla_{X_2}X_2, X_1) = \frac{1}{2}(2X_2(g(X_2, X_1)) - X_1(g(X_2, X_2))) = -\frac{1}{2}\frac{\partial}{\partial\theta}\sin^2\theta = -\sin\theta\cos\theta$$

and

$$g(\nabla_{X_2}X_2, X_2) = \frac{1}{2}\big(X_2(g(X_2, X_2))\big) = \frac{1}{2}\frac{\partial}{\partial\phi}\sin^2\theta = 0,$$

so the inner product of $\nabla_{X_2}X_2$ with X_1 (which is unit) is non-zero, and thus

$$\nabla_{X_2}X_2 = -\sin\theta\cos\theta X_1.$$

Finally, we compute

$$g(\nabla_{X_1}X_2, X_1) = \frac{1}{2}\Big(X_1\big(g(X_2, X_1)\big) + X_2\big(g(X_1, X_1)\big) - X_1\big(g(X_1, X_2)\big)\Big) = 0$$

and

$$g(\nabla_{X_1}X_2, X_2) = \frac{1}{2}\Big(X_1\big(g(X_2, X_2)\big) + X_2\big(g(X_2, X_1)\big) - X_2\big(g(X_1, X_2)\big)\Big) = \frac{1}{2}\frac{\partial}{\partial\theta}\sin^2\theta = \sin\theta\cos\theta.$$

Since $g(X_2, X_2) = \sin^2\theta$ and $[X_1, X_2] = 0$,

$$\nabla_{X_1}X_2 = \nabla_{X_2}X_1 = \frac{\sin\theta\cos\theta}{\sin^2\theta}X_2 = \cot\theta X_2.$$

We see that this only makes sense for $\sin\theta \neq 0$ as we would expect.

Example. We saw that on $\mathcal{S}^3$ we have everywhere linearly independent vector fields

$$\begin{aligned}
E_1 &= -x_1\partial_0 + x_0\partial_1 - x_3\partial_2 + x_2\partial_3\\
E_2 &= -x_2\partial_0 + x_3\partial_1 + x_0\partial_2 - x_1\partial_3\\
E_3 &= -x_3\partial_0 - x_2\partial_1 + x_1\partial_2 + x_0\partial_3.
\end{aligned}$$

Clearly, if g is the induced metric then $g(E_i, E_j) = \delta_{ij}$ which are constant. We can also compute

$$\begin{aligned}
[E_1, E_2] &= (-x_1\partial_0 + x_0\partial_1 - x_3\partial_2 + x_2\partial_3)(-x_2\partial_0 + x_3\partial_1 + x_0\partial_2 - x_1\partial_3)\\
&\quad - (-x_2\partial_0 + x_3\partial_1 + x_0\partial_2 - x_1\partial_3)(-x_1\partial_0 + x_0\partial_1 - x_3\partial_2 + x_2\partial_3)\\
&= -x_1\partial_2 - x_0\partial_3 + x_3\partial_0 + x_2\partial_1 - (-x_2\partial_1 - x_3\partial_0 + x_0\partial_3 + x_1\partial_2)\\
&= -2E_3.
\end{aligned}$$

Similarly, $[E_2, E_3] = -2E_1$ and $[E_3, E_1] = -2E_2$, i.e.

$$[E_i, E_j] = -2\epsilon_{ijk}E_k$$

(where ϵ_{ijk} is the permutation symbol). Then,

$$\begin{aligned} g(\nabla_{E_i}E_j, E_k) &= \frac{1}{2}\Big(- g(E_i, [E_j, E_k]) + g(E_j, [E_k, E_i]) + g(E_k, [E_i, E_j])\Big) \\ &= \frac{1}{2}\Big(2\epsilon_{jki} - 2\epsilon_{kij} - 2\epsilon_{kij}\Big) = -\epsilon_{ijk}. \end{aligned}$$

Hence,

$$\nabla_{E_1}E_2 = -\nabla_{E_2}E_1 = -E_3, \quad \nabla_{E_2}E_3 = -\nabla_{E_3}E_2 = -E_1, \quad \nabla_{E_3}E_1 = -\nabla_{E_1}E_3 = -E_2.$$

and $\nabla_{E_1}E_1 = \nabla_{E_2}E_2 = \nabla_{E_3}E_3 = 0$.

2.2 Christoffel symbols

It is useful to encode the Levi-Civita connection locally as follows.

Definition 2.2. Suppose (U, φ) is a coordinate chart on (M, g) and let $X_i = (\varphi^{-1})_*\partial_i \in \Gamma(TU)$.

Since $\{X_i : i = 1, \ldots, n\}$ defines a basis for $\Gamma(TU)$ we can define functions Γ^k_{ij} on U by

$$\nabla_{X_i}X_j = \sum_{k=1}^{n} \Gamma^k_{ij}X_k$$

which are called the *Christoffel symbols* of ∇ (or g) in the chart (U, φ).

The Christoffel symbols depend on the choice of coordinates!

Example. On $\mathbb{R}^n$, $\nabla_{\partial_i}\partial_j = 0$ so $\Gamma^k_{ij} = 0$.

Similarly on T^n with respect to the standard chart we have $\Gamma^k_{ij} = 0$.

Example. For $\mathcal{S}^2$ we see that $\nabla_{X_1}X_1 = 0$ so

$$\Gamma^1_{11} = \Gamma^2_{11} = 0$$

and $\nabla_{X_2}X_2 = -\sin\theta\cos\theta X_1$ so

$$\Gamma^1_{22} = -\sin\theta\cos\theta,\ \ \Gamma^2_{22} = 0.$$

Finally, $\nabla_{X_1}X_2 = \nabla_{X_2}X_1 = \cot\theta X_2$ so

$$\Gamma^1_{12} = \Gamma^1_{21} = 0,\ \ \Gamma^2_{12} = \Gamma^2_{21} = \cot\theta.$$

The following proposition allows us to compute the Levi-Civita connection ∇ locally when using the coordinate vector fields.

Proposition 2.3. *Let (U, φ) be a coordinate chart on (M, g) and let X_i be the coordinate vector fields on U. Let g be given by (g_{ij}) on U (where $g_{ij} = g(X_i, X_j)$).*

Then $\Gamma^k_{ij} = \Gamma^k_{ji}$ and if $(g^{ij}) = g^{-1}$ and we define $\partial_k g_{ij} = X_k(g_{ij})$ then

$$\Gamma^k_{ij} = \frac{1}{2}\sum_{l=1}^{n} g^{kl}(\partial_i g_{jl} + \partial_j g_{il} - \partial_l g_{ij}).$$

Remarks. The fact that the Christoffel symbols are symmetric in i, j justifies the definition of a symmetric (or torsion-free) connection. We also see that the Christoffel symbols are defined by the Riemannian metric and its first derivatives.

Proof. First, $\nabla_{X_i}X_j - \nabla_{X_j}X_i = [X_i, X_j] = 0$ which is equivalent to $\sum_{k=1}^n(\Gamma_{ij}^k - \Gamma_{ji}^k)X_k = 0$ which is then equivalent to the statement that Γ_{ij}^k is symmetric in i, j.

We now calculate

$$\begin{aligned} g(\nabla_{X_i}X_j, X_l) &= \sum_{m=1}^n g(\Gamma_{ij}^m X_m, X_l) = \sum_{m=1}^n \Gamma_{ij}^m g_{ml} \\ &= \frac{1}{2}\Big(X_i\big(g(X_j, X_l)\big) + X_j\big(g(X_i, X_l)\big) - X_l\big(g(X_i, X_j)\big)\Big) \\ &= \frac{1}{2}(\partial_i g_{jl} + \partial_j g_{il} - \partial_l g_{ij}) \end{aligned}$$

using the formula for the Levi-Civita connection and the fact that $[X_i, X_j] = 0$. Finally,

$$\Gamma_{ij}^k = \sum_{l,m=1}^n \Gamma_{ij}^m g_{ml} g^{kl}$$

since $\sum_{l=1}^n g_{ml}g^{kl} = \delta_{km}$. □

Example. If we take the usual coordinate frame on $T^n \subseteq \mathbb{R}^{2n}$; i.e.$X_i = f_*\partial_i$ where $f(\theta_1, \ldots, \theta_n) = (\cos\theta_1, \sin\theta_1, \ldots, \cos\theta_n, \sin\theta_n)$ then $g_{ij} = g(X_i, X_j) = \delta_{ij}$ constant so $\Gamma_{ij}^k = 0$.

Example. For $\mathcal{S}^2$, take $f(\theta, \phi) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$ so $X_1 = f_*\partial_\theta$ and $X_2 = f_*\partial_\phi$. Thus

$$(g_{ij}) = \begin{pmatrix} 1 & 0 \\ 0 & \sin^2\theta \end{pmatrix} \quad \text{and} \quad (g^{ij}) = \begin{pmatrix} 1 & 0 \\ 0 & \operatorname{cosec}^2\theta \end{pmatrix}.$$

If either of i or j is 1 then g_{ij} is constant and $\partial_2 = \partial_\phi$ of anything is zero. Let us calculate Γ_{12}^1 and Γ_{12}^2 using the formula. We see that

$$\Gamma_{12}^1 = \frac{1}{2}\sum_{l=1}^2 g^{1l}(\partial_1 g_{2l} + \partial_2 g_{1l} - \partial_l g_{12}) = \frac{1}{2}g^{11}(\partial_1 g_{21} + \partial_2 g_{11} - \partial_1 g_{12}) = 0$$

and

$$\Gamma_{12}^2 = \frac{1}{2}\sum_{l=1}^2 g^{2l}(\partial_1 g_{2l} + \partial_2 g_{1l} - \partial_l g_{12}) = \frac{1}{2}g^{22}\partial_1 g_{22} = \frac{1}{2\sin^2\theta}\frac{\partial}{\partial\theta}(\sin^2\theta) = \frac{2\sin\theta\cos\theta}{2\sin^2\theta} = \cot\theta.$$

Example. Let $f : \mathbb{R}^2 \to \mathbb{R}^3$ be given by $f(\theta, \phi) = ((2 + \cos\theta)\cos\phi, (2 + \cos\theta)\sin\phi, \sin\theta)$ so that $f(\mathbb{R}^2) = T^2 \subseteq \mathbb{R}^3$ and let $X_1 = f_*\partial_\theta$, $X_2 = f_*\partial_\phi$. We saw that we can identify g with the matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & (2 + \cos\theta)^2 \end{pmatrix}$$

and hence g^{-1} is given by

$$\begin{pmatrix} 1 & 0 \\ 0 & (2 + \cos\theta)^{-2} \end{pmatrix}.$$

Then if i or j is 1 then $\partial_k g_{ij} = 0$ and $\partial_2 = \partial_\phi$ of anything is zero and both g and g^{-1} are diagonal.

Hence,

$$\Gamma_{11}^1 = \frac{1}{2}\sum_{l=1}^{2} g^{1l}(\partial_1 g_{1l} + \partial_1 g_{1l} - \partial_l g_{11}) = 0$$

$$\Gamma_{22}^1 = \frac{1}{2}\sum_{l=1}^{2} g^{1l}(\partial_2 g_{2l} + \partial_2 g_{2l} - \partial_l g_{22}) = -\frac{1}{2}\partial_\theta(2+\cos\theta)^2 = (2+\cos\theta)\sin\theta$$

$$\Gamma_{12}^1 = \frac{1}{2}\sum_{l=1}^{2} g^{1l}(\partial_1 g_{2l} + \partial_2 g_{1l} - \partial_l g_{12}) = 0 = \Gamma_{21}^1$$

$$\Gamma_{11}^2 = \frac{1}{2}\sum_{l=1}^{2} g^{2l}(\partial_1 g_{1l} + \partial_1 g_{1l} - \partial_l g_{11}) = 0$$

$$\Gamma_{22}^2 = \frac{1}{2}\sum_{l=1}^{2} g^{2l}(\partial_2 g_{2l} + \partial_2 g_{2l} - \partial_l g_{22}) = 0$$

$$\Gamma_{12}^2 = \frac{1}{2}\sum_{l=1}^{2} g^{2l}(\partial_1 g_{2l} + \partial_2 g_{1l} - \partial_l g_{12}) = \frac{1}{2}(2+\cos\theta)^{-2}\partial_\theta(2+\cos\theta)^2 = -\frac{\sin\theta}{2+\cos\theta} = \Gamma_{21}^2.$$

Remark. We will see that at any point $p \in (M,g)$ we can choose local coordinates (i.e. a chart (U,φ) containing p) so that

$$g_{ij} = \delta_{ij} \quad \text{and} \quad \Gamma_{ij}^k = 0 \quad \text{at } p.$$

A set of coordinates that ensure this are called *geodesic normal coordinates*, which, unsurprisingly, involve using geodesics - the subject of the next section.

2.3 Parallel transport

Let (M,g) be an n-dimensional Riemannian manifold.

Now we want to make sense of the term "connection" as a means to "connect" tangent spaces. We do this using curves and distinguished vector fields.

Recall that if we have a curve $\alpha : (-\epsilon,\epsilon) \to M$ then we have the tangent vector field α' along α so that $t \mapsto \alpha'(t) \in T_{\alpha(t)}M$ is smooth. Further, if we have a function $f = f(\alpha(t))$ defined along the curve α in M, then

$$\alpha'(f) = (f\circ\alpha)'(t) = \frac{\mathrm{d}}{\mathrm{d}t}f(\alpha(t))$$

since we can also view $\alpha' = \alpha_*(\partial_t)$.

Definition 2.4. Let α be a curve in M and X be a vector field along α (i.e. $X(\alpha(t)) \in T_{\alpha(t)}M$ and $t \mapsto X(\alpha(t))$ is smooth). We say that X is *parallel* (along α) if $\nabla_{\alpha'}X = 0$.

Remark. If we have a curve $\alpha : (-\epsilon,\epsilon) \to (M,g)$ and a vector field X along α, then one can write

$$X' = \nabla_{\alpha'}X$$

for ease of notation. You may also see in some textbooks (e.g. do Carmo) the notation $\frac{D}{Dt}$ for $\nabla_{\alpha'}$, but we will avoid this.

Suppose that α is contained in a chart (U,φ), then write $(\varphi\circ\alpha)(t) = (x_1(t),\ldots,x_n(t))$ and, if $X_i = (\varphi^{-1})_*\partial_i$ as usual then write $X = \sum_{i=1}^n a_i X_i$. We see that

$$\alpha' = \sum_{i=1}^{n} x_i'(\varphi_i^{-1})_*\partial_i = \sum_{i=1}^{n} x_i' X_i.$$

Moreover,

$$\begin{aligned}\nabla_{\alpha'}\sum_{i=1}^n a_iX_i &= \sum_{i=1}^n \alpha'(a_i)X_i + a_i\nabla_{\alpha'}X_i\\ &= \sum_{i=1}^n a_i'X_i + \sum_{i=1}^n a_i\nabla_{\sum_{j=1}^n x_j'X_j}X_i = \sum_{i=1}^n a_i'X_i + \sum_{i,j=1}^n a_ix_j'\nabla_{X_j}X_i\\ &= \sum_{k=1}^n a_k'X_k + \sum_{i,j,k=1}^n a_ix_j'\Gamma_{ij}^kX_k\\ &= \sum_{k=1}^n\left(a_k' + \sum_{i,j=1}^n \Gamma_{ij}^k a_i x_j'\right)X_k. \qquad (**)\end{aligned}$$

So the condition to be parallel is a first order ODE on X in local coordinates.

Let us see how this works in examples.

Example. On $\mathbb{R}^n$, $\Gamma_{ij}^k = 0$ so

$$\nabla_{\alpha'}X = \sum_{k=1}^n a_k'\partial_k,$$

so the parallel vector fields are given by $a_k' = 0$, which means a_k is constant and so $X \in \mathrm{Span}\{\partial_1, \ldots, \partial_n\}$. Notice that the parallel vector fields are independent of the choice of curve.

Similarly, on T^n, $\Gamma_{ij}^k = 0$ so the parallel vectors fields are (constant) linear combinations of the coordinate vector fields X_i.

Example. Let X_1, X_2 be the usual coordinate vector fields on $\mathcal{S}^2$ which are the pushforwards of $\partial_\theta, \partial_\phi$ under $f(\theta,\phi) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$. Then

$$\Gamma_{11}^1 = \Gamma_{11}^2 = 0, \quad \Gamma_{22}^1 = -\sin\theta\cos\theta, \quad \Gamma_{22}^2 = 0, \quad \Gamma_{12}^1 = 0, \quad \Gamma_{12}^2 = \cot\theta.$$

So, if $X = a_1X_1 + a_2X_2$, then given any curve $\alpha(t) = f(\theta(t),\phi(t))$ we have that

$$\nabla_{\alpha'}X = (a_1' - (\sin\theta\cos\theta)a_2\phi')X_1 + (a_2' + \cot\theta(a_1\phi' + a_2\theta'))X_2.$$

We see that if α is a curve with ϕ constant (so a line of longitude) so we can take for example $\theta = t$ and $\phi' = 0$ then we have

$$\nabla_{\alpha'}X = a_1'X_1 + (a_2' + \cot t a_2)X_2$$

so X_1 is parallel along α (as $a_1 = 1$ and $a_2 = 0$) but X_2 is not (as $a_2 = 1$ and $a_1 = 0$), in fact

$$\nabla_{\alpha'}X_1 = 0,\ \nabla_{\alpha'}X_2 = \cot tX_2$$

If α is a curve with θ constant (so a line of latitude) then if we take $\phi = t$ we have

$$\nabla_{\alpha'}X = (a_1' - \sin\theta\cos\theta a_2)X_1 + (a_2' + \cot\theta a_1)X_2.$$

Hence, X_1 and X_2 are both parallel along α (which means $a_1' = 0$ and $a_2' = 0$ solve $\nabla_{\alpha'}X = 0$) if and only if $\theta = \frac{\pi}{2}$ (so the equator). We see that

$$\nabla_{\alpha'}X_1 = \cot\theta X_2 \quad \text{and} \quad \nabla_{\alpha'}X_2 = -\sin\theta\cos\theta X_1.$$

Example. Suppose we have the 2-torus in $\mathbb{R}^3$ parametrised as usual by $f(\theta,\phi)$, with vector fields $X_1 = f_*\partial_\theta$ and $X_2 = f_*\partial_\phi$, so $[X_1, X_2] = 0$ and

$$\Gamma_{22}^1 = (2+\cos\theta)\sin\theta,\ \Gamma_{12}^2 = -\frac{\sin\theta}{2+\cos\theta}$$

and otherwise $\Gamma_{ij}^k = 0$. Hence, if $X = a_1X_1 + a_2X_2$ and $\alpha(t) = f(\theta(t), \phi(t))$ is a curve on T^2, then we compute

$$\nabla_{\alpha'}X = (a_1' + (2+\cos\theta)\sin\theta a_2\phi')X_1 + (a_2' - \frac{\sin\theta}{2+\cos\theta}(a_1\phi' + a_2\theta'))X_2.$$

We see that if α is a curve with ϕ constant so we take $\alpha(t) = f(t, \phi)$ then

$$\nabla_{\alpha'}X = a_1'X_1 + (a_2' - \frac{\sin t}{2+\cos t}a_2)X_2.$$

Again, we see that X_1 is parallel along α but X_2 is not. Similarly, if $\alpha(t) = f(\theta, t)$ then

$$\nabla_{\alpha'}X = (a_1' + (2+\cos\theta)\sin\theta a_2)X_1 + (a_2' - \frac{\sin\theta}{2+\cos\theta}a_1)X_2,$$

so we see that

$$\nabla_{\alpha'}X_1 = -\frac{\sin\theta}{2+\cos\theta}X_2 \quad \text{and} \quad \nabla_{\alpha'}X_2 = (2+\cos\theta)\sin\theta X_1,$$

so we see that X_1 and X_2 are both parallel if and only if $\theta = 0$ or $\theta = \pi$, which is the inner and outer ring on the torus.

We can now define a fundamental notion in Riemannian geometry which is *parallel transport*. This is how we "connect" tangent spaces using the Levi-Civita connection.

Theorem 2.5. *Let $p, q \in M$ and let $\alpha : [0, L] \to M$ be a curve between p and q.*
Given $X_0 \in T_pM$ there exists a unique parallel vector field X along α such that $X(p) = X_0$.
The map $\tau_\alpha : T_pM \to T_qM$ given by $\tau_\alpha(X_0) = X(q)$ is an isometry, so an isomorphism such that

$$g_p(X_0, Y_0) = g_q(\tau_\alpha(X_0), \tau_\alpha(Y_0)),$$

called the parallel transport along α.

Proof. It is enough to show that the result holds for curves α contained in a chart (U, φ) since using compactness of $[0, L]$ we can cover it with a finite number of intersecting open intervals $I_1, \ldots, I_m$ so that each $\alpha(I_j)$ is contained in a coordinate chart, and the uniqueness result will prove that the vector field is well-defined along α (as it agrees on the overlap of any of the intervals).

As we saw above, by (**) we see that X is parallel if and only if the right-hand side of (**) is zero. This is n linear first order ODEs in n unknowns $(a_1, \ldots, a_n)$, together with the n initial conditions that $(a_1(0), \ldots, a_n(0)) = X_0$, so a solution exists on all of $[0, L]$ and is unique as claimed.

To see that τ_α is an isomorphism, let $\beta(t) = \alpha(L - t)$ and consider $\tau_\beta : T_qM \to T_pM$. There exists a unique parallel vector field Y along β such that $Y(q) = X(q)$. However, $\beta'(t) = -\alpha'(L - t)$ so $\nabla_{\alpha'}X = 0$ implies that $\nabla_{\beta'}X = 0$, so X is also parallel along β. The uniqueness of Y means that $Y(p) = X_0$. We deduce that $\tau_\beta \circ \tau_\alpha = \mathrm{id}$ so τ_α is an isomorphism (as it is clearly linear).

Let X, Y be vector fields along α. Then, since $\alpha' = \alpha^*(\partial_t)$, along α we have that

$$\frac{\mathrm{d}}{\mathrm{d}t}g(X, Y) = \alpha'\big(g(X, Y)\big) = g(\nabla_{\alpha'}X, Y) + g(X, \nabla_{\alpha'}Y).$$

If $X_0, Y_0 \in T_pM$, then let X, Y be the unique parallel vector fields along α such that $X(p) = X_0$ and $Y(p) = Y_0$. Then $\frac{\mathrm{d}}{\mathrm{d}t}g(X, Y) = 0$ as X, Y are parallel so $g(X, Y)(\alpha(t))$ is independent of $t \in [0, L]$. We deduce that

$$g_p(X_0, Y_0) = g_p(X(p), Y(p)) = g(X, Y)(\alpha(0)) = g(X, Y)(\alpha(L)) = g_q(X(q), Y(q)) = g_q(\tau_\alpha(X_0), \tau_\alpha(Y_0)).$$

Thus τ_α is an isometry as claimed. □

Example. On $\mathbb{R}^n$ for any curve α from p to q and $X_0 \in T_p\mathbb{R}^n = \mathbb{R}^n$, the parallel vector field X along α is constant, so $\tau_\alpha(X_0) = X_0$; i.e. parallel transport is just translation along the curve and is thus effectively the identity.

The same will be true for $T^n \subseteq \mathbb{R}^{2n}$ since again the Christoffel symbols vanish.

Example. Suppose α in $\mathcal{S}^2$ is given by $\alpha(t) = (\sin\theta\cos t, \sin\theta\sin t, \cos\theta) = f(\theta, t)$ in our usual parametrisation for some $\theta \in (0, \pi)$. Then $\alpha' = X_2$ (again in our usual notation) and if we write $X = a_1X_1 + a_2X_2$ then X is parallel along α if and only if

$$\nabla_{\alpha'}X = (a_1' - \sin\theta\cos\theta a_2)X_1 + (a_2' + \cot\theta a_1)X_2 = 0,$$

so

$$a_1' = \sin\theta\cos\theta a_2 \quad \text{and} \quad a_2' = -\cot\theta a_1.$$

Differentiating again we see that $a_i'' = -\cos^2\theta a_i$, hence we see that

$$a_1(t) = a_1(0)\cos(t\cos\theta) + a_2(0)\sin\theta\sin(t\cos\theta) \quad \text{and} \quad a_2(t) = a_2(0)\cos(t\cos\theta) - \frac{a_1(0)}{\sin\theta}\sin(t\cos\theta)$$

for $\theta \neq \frac{\pi}{2}$ and for $\theta = \frac{\pi}{2}$ we have that $a_1(t) = a_1(0)$ and $a_2(t) = a_2(0)$.

Hence the parallel transport map

$$\tau_\alpha : T_{\alpha(0)}\mathcal{S}^2 \to T_{\alpha(t)}\mathcal{S}^2$$

is the map $\tau_\alpha : X(0) \to X(t)$ where X is parallel. Then for any $a_1, a_2 \in \mathbb{R}$ we have

$$\tau_\alpha(a_1X_1 + a_2X_2) = (a_1\cos(t\cos\theta) + a_2\sin\theta\sin(t\cos\theta))X_1 + (-\frac{a_1}{\sin\theta}\sin(t\cos\theta) + a_2\cos(t\cos\theta))X_2,$$

which is the identity when $\theta = \frac{\pi}{2}$ for any t. Therefore, with respect to the orthonormal basis $E_1 = X_1$ and $E_2 = \frac{X_2}{\sin\theta}$ the matrix of τ_α is

$$\begin{pmatrix} \cos(t\cos\theta) & \sin(t\cos\theta) \\ -\sin(t\cos\theta) & \cos(t\cos\theta) \end{pmatrix},$$

which is clearly a rotation (and thus an isometry as we expected). Notice that the size of the rotation around a loop (so for $t = 2\pi$) depends on θ: this is related to the idea of holonomy.

Example. We can perform a similar calculation for $T^2 \subseteq \mathbb{R}^3$. In this case, if we take parallel transport around a loop where ϕ is constant, we get the identity map. Instead, parallel transport around a loop where θ is constant will give a rotation by $2\pi\sin\theta$, which will be the identity when $\theta = 0$ or π (i.e. the inner and outer circle). Notice this is in marked contrast to $T^2 \subseteq \mathbb{R}^4$.

Remark. In fact, we can recover the Levi-Civita connection from the parallel transport maps: we can define the derivative using a similar formula to the Lie derivative, but using the parallel transport instead of the flow of the vector field.

3 Geodesics

We now move on to one of the central ideas in Riemannian geometry, that of *geodesics*. You already saw geodesics as critical curves for the length functional in the Differentiable Manifolds course. We shall now give an alternative definition which does not involve a variational characterisation (that, in principle, would involve us comparing a potential geodesic against all other curves with the same endpoints). However, a key result we shall prove is that the notion we give is the same as before; i.e. that they are locally the shortest paths between points.

3.1 Definition

We begin with the formal definition.

Definition 3.1. A curve γ in (M,g) is a *geodesic* if

$$\nabla_{\gamma'}\gamma' = 0.$$

Since

$$\frac{\mathrm{d}}{\mathrm{d}t}g(\gamma',\gamma') = \gamma'\big(g(\gamma',\gamma')\big) = 2g(\nabla_{\gamma'}\gamma',\gamma') = 0$$

it follows that $|\gamma'| = \sqrt{g(\gamma',\gamma')}$ is constant along the curve γ. We say γ is *normalised* (or parameterized by arclength) if $|\gamma'| = 1$.

Remark. Using our simplify notation from earlier, as γ' is a vector field along γ we could write the geodesic equation as

$$\gamma'' = \nabla_{\gamma'}\gamma' = 0.$$

This may be reminiscent of formulae you have seen elsewhere, or the idea that geodesics are curves with "zero acceleration".

In a coordinate chart (U,φ) we can write $\varphi\circ\gamma = (x_1,\ldots,x_n)$ and

$$(\varphi\circ\gamma)' = \sum_{i=1}^n x_i'\partial_i = \varphi_*(\gamma')$$

by the Chain rule. Hence,

$$\gamma' = \sum_{i=1}^n x_i'(\varphi_*)^{-1}(\partial_i) = \sum_{i=1}^n x_i'X_i$$

where $X_i = (\varphi^{-1})_*\partial_i$ are the coordinate vector fields.

We therefore see from the properties of the Levi-Civita connection from the Fundamental Theorem of Riemannian Geometry, and the definition of the Christoffel symbols, that

$$\begin{aligned}
\nabla_{\gamma'}\gamma' &= \sum_{i=1}^n \nabla_{\gamma'}(x_i'X_i)\\
&= \sum_{i=1}^n \gamma'(x_i')X_i + x_i'\nabla_{\gamma'}X_i\\
&= \sum_{i=1}^n x_i''X_i + x_i'\sum_{j=1}^n x_j'\nabla_{X_j}X_i\\
&= \sum_{i=1}^n x_i''X_i + \sum_{j,k=1}^n x_i'x_j'\Gamma_{ji}^k X_k\\
&= \sum_{k=1}^n (x_k'' + \sum_{i,j=1}^n \Gamma_{ij}^k x_i'x_j')X_k.
\end{aligned}$$

Remark. This result of course follows from the more general formula for $\nabla_{\alpha'}X$ we derived during our discussion of parallel transport.

We deduce the following result.

Proposition 3.2. *Let (U,φ) be a chart on (M,g) and let γ be a curve in U. If we write $\varphi\circ\gamma=(x_1,\ldots,x_n)$ then γ is a geodesic if and only if*

$$x_k''+\sum_{i,j=1}^n\Gamma_{ij}^k x_i'x_j'=0$$

for all k, which are called the geodesic equations.

These equations should be familiar to you if you have studied surfaces in $\mathbb{R}^3$ or general relativity. Notice again that these equations are dependent on the coordinate chart we choose.

3.2 Examples

Let us calculate the geodesics in some examples.

Example. For $\mathbb{R}^n$, we saw that $\Gamma_{ij}^k=0$ so the geodesic equations for $\gamma=(x_1,\ldots,x_n)$ are simply

$$x_k''=0$$

which define straight lines $x_k(t)=a_kt+b_k$. We see that the condition for γ to be normalised is: $\sum_{i=1}^n a_i^2=1$.

Example. If we let $f:\mathbb{R}^+\times\mathbb{R}\to\mathbb{R}^2\setminus\{0\}$ be $f(r,\theta)=(r\cos\theta,r\sin\theta)$, then we saw that the pullback metric $g=f^*g_0$ was given by

$$(g_{ij})=\begin{pmatrix}1&0\\0&r^2\end{pmatrix}$$

so

$$(g^{ij})=\begin{pmatrix}1&0\\0&\frac{1}{r^2}\end{pmatrix}.$$

Therefore, we see that the Christoffel symbols are (recalling that in this setting $X_1=\partial_r$ and $X_2=\partial_\theta$):

$$\Gamma_{11}^1=0,\ \Gamma_{22}^1=-r,\ \Gamma_{12}^1=0,\ \Gamma_{11}^2=0,\ \Gamma_{22}^2=0,\ \Gamma_{12}^2=\frac{1}{r};$$

so

$$\nabla_{X_1}X_1=0,\ \nabla_{X_1}X_2=\nabla_{X_2}X_1=\frac{1}{r}X_2,\ \nabla_{X_2}X_2=-rX_1.$$

We see that the geodesic equations now are (since $x_1=r$ and $x_2=\theta$)

$$r''-r(\theta')^2=0\quad\text{and}\quad\theta''+\frac{2}{r}r'\theta'=0.$$

We see straight away that $\theta'=0$ and $r''=0$ gives a solution, which corresponds to a ray emanating from the origin.

However, it is now not as easy to see that all geodesics are just straight lines. This shows how important it is to choose the right coordinates!

Example. On the standard n-torus $T^n\subseteq\mathbb{R}^{2n}$ we saw that $\Gamma_{ij}^k=0$ when we choose $f(\theta_1,\ldots,\theta_n)=(\cos\theta_1,\sin\theta_1,\ldots,\cos\theta_n,\sin\theta_n)$ and thus the geodesic equations are given by

$$\theta_k''=0.$$

We deduce that $\theta_k=a_kt+b_k$, so the geodesics are

$$\gamma(t)=(\cos(a_1t+b_1),\sin(a_1t+b_1),\ldots,\cos(a_nt+b_n),\sin(a_nt+b_n)),$$

the images in T^n of the straight lines in $\mathbb{R}^n$.

Example. For $\mathcal{S}^2$ suppose we take a normalised geodesic

$$\gamma(t) = (\sin\theta(t)\cos\phi(t), \sin\theta(t)\sin\phi(t), \cos\theta(t)).$$

So here $f = f(\theta,\phi)$ is as usual and $X_1 = f_*\partial_\theta$ and $X_2 = f_*\partial_\phi$. Since

$$\Gamma^1_{11} = \Gamma^1_{12} = 0,\ \Gamma^1_{22} = -\sin\theta\cos\theta,\ \Gamma^2_{11} = \Gamma^2_{22} = 0,\ \Gamma^2_{12} = \cot\theta,$$

we see that the geodesic equations are

$$\theta'' - \sin\theta\cos\theta(\phi')^2 = 0 \quad \text{and} \quad \phi'' + 2\cot\theta\theta'\phi' = 0.$$

and

$$|\gamma'|^2 = (\theta')^2 + \sin^2\theta(\phi')^2 = 1.$$

We see that $\phi' = 0$ and $\theta'' = 0$ gives a solution if $\theta' = 1$, which is

$$\gamma(t) = \big(\sin(t+\theta_0)\cos\phi_0, \sin(t+\theta_0)\sin\phi_0, \cos(t+\theta_0)\big)$$

with θ_0, ϕ_0 constant, called a great circle.

It is useful to compare this to our discussion of parallel transport on $(\mathcal{S}^2, g)$.

We shall see that all geodesics are great circles in $\mathcal{S}^2$ (and in fact in $\mathcal{S}^n$) are great circles.

Example. Suppose we take the upper half-plane plane $H^2 = \{(x_1,x_2) \in \mathbb{R}^2 : x_2 > 0\}$ and the hyperbolic metric

$$g = \frac{\mathrm{d}x_1^2 + \mathrm{d}x_2^2}{x_2^2}.$$

We can compute that the Christoffel symbols are:

$$\Gamma^1_{11} = \Gamma^1_{22} = 0,\ \Gamma^2_{11} = -\Gamma^2_{22} = \frac{1}{x_2},\ \Gamma^1_{12} = -\frac{1}{x_2},\ \Gamma^2_{12} = 0.$$

We have the geodesic equations for $\gamma(t) = (x_1(t), x_2(t))$ are given by

$$x_1'' - \frac{2}{x_2}x_1'x_2' = 0, \quad x_2'' + \frac{1}{x_2}((x_1')^2 - (x_2')^2) = 0.$$

There is clearly a solution given by x_1 is constant and $x_2 = e^t$, so vertical half-lines are geodesics (and notice that they are defined for all $t \in \mathbb{R}$).

We have seen that it is quite laborious to compute the Christoffel symbols, but there is a much faster way as follows.

Proposition 3.3. *Let (U,φ) be a chart on (M,g) and let*

$$L = \frac{1}{2}\sum_{i,j} g_{ij}x_i'x_j'.$$

Then γ given by $\varphi\circ\gamma = (x_1,\ldots,x_n)$ is a geodesic if and only if, for all k,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial x_k'}\right) - \frac{\partial L}{\partial x_k} = 0.$$

This is a straightforward calculation, but we will prove this formula later without computation because what it says is that γ is a geodesic if and only if the Euler–Lagrange equations for the function L are satisfied, which means that γ must be critical point for $\int L$, which we will see is the *energy* of the curve. So, if we show that geodesics are critical for energy (which will follow from the fact that they are locally length minimizing), then this formula is a direct consequence.

Proof. The proof is an easy calculation, which I give for completeness. We see that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}g_{ij} &= \gamma'(g(X_i, X_j))\\ &= g(\nabla_{\gamma'}X_i, X_j) + g(X_i, \nabla_{\gamma'}X_j)\\ &= g(\sum_l x_l'\nabla_{X_l}X_i, X_j) + g(X_i, \sum_l x_l'\nabla_{X_l}X_j)\\ &= g(\sum_{l,m} x_l'\Gamma_{li}^m X_m, X_j) + g(X_i, \sum_{l,m} x_l'\Gamma_{lj}^m X_m)\\ &= \sum_{l,m}(x_l'\Gamma_{li}^m g_{mj} + x_l'\Gamma_{lj}^m g_{im}).\end{aligned}$$

Therefore

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial x_k'} &= \frac{\mathrm{d}}{\mathrm{d}t}(\sum_i g_{ik}x_i')\\ &= \sum_i g_{ik}x_i'' + \sum_{i,l,m}(x_l'(\Gamma_{li}^m g_{mk} + \Gamma_{lk}^m g_{im})x_i')\\ &= \sum_i \big(g_{ik}x_i'' + \sum_{l,m}(\Gamma_{li}^m g_{mk} + \Gamma_{lk}^m g_{im})x_i'x_l'\big).\end{aligned}$$

We then see that

$$\begin{aligned}\frac{\partial L}{\partial x_k} &= \frac{1}{2}\sum_{i,j} X_k(g_{ij})x_i'x_j'\\ &= \frac{1}{2}\sum_{i,j}\big(g(\nabla_{X_k}X_i, X_j) + g(X_i, \nabla_{X_k}X_j)\big)x_i'x_j'\\ &= \frac{1}{2}\sum_{i,j,l}(\Gamma_{ki}^l g_{lj} + \Gamma_{kj}^l g_{il})x_i'x_j'\\ &= \sum_{i,j,l}\Gamma_{ki}^l g_{lj}x_i'x_j'\\ &= \sum_{i,l,m}\Gamma_{ki}^m g_{ml}x_i'x_l'\end{aligned}$$

since the sum is symmetric in i, j. Multiplying by g^{ka} and notice that $\sum_k g^{ka}g_{ki} = \delta_{ia}$ we deduce that the equation in the proposition holds if and only if

$$x_a'' + \sum_{i,l}\big(\Gamma_{li}^a + \sum_m(g^{ka}\Gamma_{lk}^m g_{im} - g^{ka}\Gamma_{ki}^m g_{ml})\big)x_i'x_l' = 0$$

The last two terms cancel, since the sum is symmetric in i, l and the Christoffel symbols are symmetric in the lower indices, and hence

$$x_k'' + \sum_{i,j}\Gamma_{ij}^k x_i'x_j' = 0$$

upon relabeling. □

Remark. The proof shows that the Euler–Lagrange equations for the energy given in Proposition 3.3 are *equivalent* to the geodesic equations. This means that we can *compute* the Christoffel symbols using the Euler–Lagrange equations! This is computationally expedient.

Example. For the case of polar coordinates, $g = \mathrm{d}r^2 + r^2\mathrm{d}\theta^2$ so $L = \frac{1}{2}\big((r')^2 + r^2(\theta')^2\big)$. We see that

$$\frac{\partial L}{\partial r'} = r' \quad \text{and} \quad \frac{\partial L}{\partial r} = r(\theta')^2$$

so we have a geodesic equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial r'}\right) - \frac{\partial L}{\partial r} = r'' - r(\theta')^2 = 0.$$

Even more simply,

$$\frac{\partial L}{\partial\theta'} = r^2\theta' \quad\text{and}\quad \frac{\partial L}{\partial\theta} = 0$$

so the other geodesic equation is:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial\theta'}\right) - \frac{\partial L}{\partial\theta} = (r^2\theta')' = r^2\theta'' + 2rr'\theta' = 0.$$

Example. For S^2 the answer is similar since $g = \mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2$ so $L = \frac{1}{2}\big((\theta')^2 + \sin^2\theta(\phi')^2\big)$. Then

$$\frac{\partial L}{\partial\theta'} = \theta' \quad\text{and}\quad \frac{\partial L}{\partial\theta} = \sin\theta\cos\theta(\phi')^2,$$

so the first geodesic equation is:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial\theta'}\right) - \frac{\partial L}{\partial\theta} = \theta'' - \sin\theta\cos\theta(\phi')^2 = 0.$$

We also have

$$\frac{\partial L}{\partial\phi'} = \sin^2\theta\phi' \quad\text{and}\quad \frac{\partial L}{\partial\phi} = 0$$

so the other geodesic equation is

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial\phi'}\right) - \frac{\partial L}{\partial\phi} = (\sin^2\theta\phi')' = \sin^2\theta\phi'' + 2\sin\theta\cos\theta\theta'\phi' = 0.$$

Example. Finally, we do the case of (H^2, g), where

$$L = \frac{1}{2}\frac{(x_1')^2 + (x_2')^2}{x_2^2}.$$

Therefore we easily compute:

$$\frac{\partial L}{\partial x_1'} = \frac{x_1'}{x_2^2} \quad\text{and}\quad \frac{\partial L}{\partial x_1} = 0,$$
$$\frac{\partial L}{\partial x_2'} = \frac{x_2'}{x_2^2} \quad\text{and}\quad \frac{\partial L}{\partial x_2} = -\frac{(x_1')^2 + (x_2')^2}{x_2^3},$$

so we have the geodesic equations:

$$\left(\frac{x_1'}{x_2^2}\right)' = 0 \quad\text{and}\quad \left(\frac{x_2'}{x_2^2}\right)' + \frac{(x_1')^2 + (x_2')^2}{x_2^3} = 0,$$

which is equivalent to our previous equations.

3.3 Isometries

Suppose we want to understand all geodesics in (H^2, g). Clearly solving the general equations is hard, but one trick that will be useful more generally is to use isometries.

It is important to note that an isometry $f : (M, g) \to (N, h)$ identifies the metrics on M and N, and since the metric uniquely determines the Levi-Civita connection, it identifies the Levi-Civita connections and hence the geodesics. Moreover, the condition to be a geodesic is a local one, so if f is only a local isometry then we may restrict to open subsets U of M and V of N where it is an isometry and see that any part of a geodesic in U will be mapped to a geodesic in V. Therefore, we have the following important result.

Lemma 3.4. *A local isometry $f : (M, g) \to (N, h)$ between Riemannian manifolds maps geodesics in (M, g) to geodesics in (N, h).*

Example. Recall that the orientation-preserving isometries of (H^2, g) are the Möbius transformations

$$f(z) = \frac{az+b}{cz+d}$$

where $a, b, c, d \in \mathbb{R}$ with $ad - bc = 1$.

We can then see what happens to the vertical half-line under a Möbius transformation. Well,

$$f(ix_2) = \frac{acx_2^2 + bd + ix_2}{c^2x_2^2 + d^2} = u + iv$$

and we see that

$$(2cdu - (ad+bc))^2 + (2cdv)^2 = 1$$

which defines a circle centered at a point on the line $x_2 = 0$ if $cd \neq 0$. If $cd = 0$ then either $c = 0$ or $d = 0$ (but not both) and in both cases we just get back the vertical half-line we started with. Hence, we have geodesics of H^2 given by vertical half-lines and semi-circles centered at points on $x_2 = 0$ (equivalently, the circles which meet the x_1-axis at right angles). Notice again that v never reaches zero as $t \to \pm\infty$ because x_2 is an exponential in t.

3.4 Existence and uniqueness

We now want to see later how geodesics give a distinguished way to "move" inside the manifold. To see this, we need an existence and uniqueness result for geodesics, which is important in its own right. Here, for $p \in (M, g)$ we introduce the notation $B_\epsilon(0) \subseteq T_pM$ for

$$B_\epsilon(0) = \{X \in T_pM \,:\, |X| < \epsilon\}$$

where $|X| = \sqrt{g_p(X, X)}$.

Theorem 3.5. *Let $p \in M$. There exist an open set $U \ni p$, $\epsilon > 0$ and a smooth map $\Gamma : (-2, 2) \times V \to M$ where*

$$V = \{(q, X) : q \in U, X \in B_\epsilon(0) \subseteq T_qM\}$$

such that $\gamma_{(q,X)}(t) = \Gamma(t, q, X)$ is the unique geodesic in M with $\gamma_{(q,X)}(0) = q$ and $\gamma'_{(q,X)}(0) = X$.

Proof. The geodesic equations are a system of second order ODEs, which are linear in the second derivatives, so standard ODE theory states that there exist an open set $U \ni p$, $\epsilon' > 0$ and $\delta > 0$ such that for all $q \in U$ and $Y \in B_{\epsilon'}(0)$ there exists a unique geodesic $\alpha_{(q,Y)} : (-\delta, \delta) \to M$ with $\alpha_{(q,Y)}(0) = q$ and $\alpha'_{(q,Y)}(0) = Y$. Moreover, the map $(t, q, Y) \mapsto \alpha_{(q,Y)}(t)$ is smooth.

If $\delta \geq 2$, we are done, but if $\delta < 2$ we define a curve

$$\gamma_{(q,X)}(t) = \alpha_{(q, \frac{2X}{\delta})}\left(\frac{\delta t}{2}\right)$$

where $X \in B_{\frac{\delta\epsilon'}{2}}(0) \subseteq B_{\epsilon'}(0)$ (so $Y = \frac{2X}{\delta} \in B_{\epsilon'}(0)$) and $t \in (-2, 2)$ (so $|\frac{\delta t}{2}| < \epsilon$).

Now $\gamma_{(q,X)}(0) = q$, $\gamma'_{(q,X)}(0) = \frac{\delta}{2}\alpha'_{(q,\frac{2X}{\delta})}(0) = X$ and

$$\nabla_{\gamma'_{(q,X)}}\gamma'_{(q,X)} = \frac{\delta^2}{4}\nabla_{\alpha'_{(q,\frac{2X}{\delta})}}\alpha'_{(q,\frac{2X}{\delta})} = 0$$

so $\gamma_{(q,X)}$ is a geodesic. By the uniqueness result, $\gamma_{(q,X)}$ is the unique geodesic with the given initial conditions, so the result follows with $\epsilon = \frac{\delta\epsilon'}{2}$. □

The uniqueness result has important consequences. In particular, we can describe all of the geodesics in simple examples.

Example. Given $q \in \mathcal{S}^2$ and unit $Y \in T_q\mathcal{S}^2$, let γ be the unique geodesic such that $\gamma(0) = q$ and $\gamma'(0) = Y$. There exists $T \in \mathrm{SO}(3)$ such that $T(0, 0, 1) = q$ and $T(0, 1, 0) = Y$. We have a geodesic

$\alpha(t) = (0, \sin t, \cos t)$ such that $\alpha(0) = (0, 0, 1)$ and $\alpha'(0) = (0, 1, 0)$. Since T is an isometry it takes geodesics to geodesics. Therefore $\beta(t) = T(\alpha(t))$ is a geodesic with $\beta(0) = q$ and $\beta'(0) = Y$, so uniqueness of geodesics means that $\gamma = \beta$. Therefore, every geodesic in $\mathcal{S}^2$ is a great circle.

Remark. For the next example, and for the rest of the course, we let $\mathbf{e}_i$ denote the vector in $\mathbb{R}^n$ with 1 in the ith place and 0 otherwise.

Example. Let $p \in \mathcal{S}^n$ and unit vector $X \in T_p\mathcal{S}^n$. As in the argument for $\mathcal{S}^2$, there exists $T \in \mathrm{SO}(n+1)$ such that $T(0, \ldots, 0, 1) = p$ and $T(0, \ldots, 0, 1, 0) = X$, so the unique geodesic γ through p with tangent vector X at p is given by $T(\alpha)$ where α is the geodesic through $\mathbf{e}_{n+1}$ with tangent vector $\mathbf{e}_n$.

Let $\rho(x_1, \ldots, x_{n+1}) = (-x_1, \ldots, -x_{n-1}, x_n, x_{n+1})$. This is an isometry of $\mathcal{S}^n$ with $\rho(\mathbf{e}_{n+1}) = \mathbf{e}_{n+1}$ and $\rho(\mathbf{e}_n) = \mathbf{e}_n$. Therefore $\rho(\alpha) = \alpha$ by the uniqueness of geodesics, so $\alpha \in \mathrm{Span}\{\mathbf{e}_n, \mathbf{e}_{n+1}\}$ and hence $\alpha = \sin t\mathbf{e}_n + \cos t\mathbf{e}_{n+1}$, a great circle. Therefore all of the geodesics in $\mathcal{S}^n$ are great circles; that is $\Pi \cap \mathcal{S}^n$ for 2-planes Π through 0.

Example. Theorem 3.5 shows that the geodesics we found earlier on the hyperbolic upper-half plane (H^2, g), i.e. the vertical half-lines and semi-circles centred on the horizontal axis, comprise all of the geodesics on (H^2, g).

Example. By Theorem 1.6, the projection map $\pi : \mathcal{S}^n \to \mathbb{RP}^n = \mathcal{S}^n/\mathbb{Z}_2$ is a local isometry. Therefore, if $\tilde{\gamma}$ is a normalized geodesic in $\mathcal{S}^n$ then $\gamma = \pi \circ \tilde{\gamma}$ is a normalized geodesic in $\mathbb{RP}^n$.

Notice that although $\tilde{\gamma}(t + 2\pi) = \tilde{\gamma}(t)$ (as the circumference of a great circle is 2π), we have that $\gamma(t + \pi) = \gamma(t)$ since $\tilde{\gamma}(t + \pi) = -\tilde{\gamma}(t)$ and so $\pi(\tilde{\gamma}(t + \pi)) = \pi(\tilde{\gamma}(t))$.

Let $[p] \in \mathbb{RP}^n$ and $X \in T_{[p]}\mathbb{RP}^n$ with $|X| = 1$. Since we can take $p \in \mathcal{S}^n$ and $\mathrm{d}\pi_p : T_p\mathcal{S}^n \to T_{[p]}\mathbb{RP}^n$ is an isomorphism, there exists a unique great circle α through p with $\mathrm{d}\pi_p(\alpha'(0)) = X$ and the projection $\pi \circ \alpha$ of α is a geodesic through $[p]$ with $(\pi \circ \alpha)'(0) = X$. By Theorem 3.5, there exists a unique geodesic γ in $\mathbb{RP}^n$ through $[p]$ with $\gamma'(0) = X$. Hence $\gamma = \pi \circ \alpha$.

In other words, every geodesic in $\mathbb{RP}^n$ is the projection of a great circle in $\mathcal{S}^n$, as we might expect.

3.5 Exponential map

Definition 3.6. We define a smooth map $\exp_p : V \to M$ by $\exp_p(q, X) = \gamma_{(q,X)}(1)$, which we call the *exponential map*. We often restrict to $\exp_p : B_\epsilon(0) \subseteq T_pM \to M$ by $\exp_p(X) = \gamma_{(p,X)}(1)$, which we still call the exponential map.

Notice that

$$\gamma_{(p,tX)}(1) = \exp_p(tX) = \gamma_{(p,X)}(t),$$

when both sides make sense, so the exponential map moves points along geodesics emanating from p. This fact also means we usual care about the exponential map acting on unit tangent vectors, if this is well-defined.

Example. In $(\mathbb{R}^n, g_0)$ we have $\gamma_{(p,X)}(t) = p + tX$, so $\exp_p(X) = p + X$. This clearly makes sense for any vector X, regardless of size. In other words the exponential map $\exp_p : \mathbb{R}^n \to \mathbb{R}^n$ is just translation by p. Since it is a translation, it is a diffeomorphism.

Example. A similar example occurs in $T^n \subseteq \mathbb{R}^{2n}$. If $p = (\cos\theta_1, \sin\theta_1, \ldots, \cos\theta_n, \sin\theta_n)$ and $X = f_*(\sum_{i=1}^n a_i\partial_i)$ where $f(x_1, \ldots, x_n) = (\cos x_1, \sin x_1, \ldots, \cos x_n, \sin x_n)$ as usual, then we have

$$\gamma_{(p,X)}(t) = (\cos(a_1t + \theta_1), \sin(a_1t + \theta_1), \ldots, \cos(a_nt + \theta_n), \sin(a_nt + \theta_n))$$

so

$$\exp_p(X) = (\cos(a_1 + \theta_1), \sin(a_1 + \theta_1), \ldots, \cos(a_n + \theta_n), \sin(a_n + \theta_n))$$

which is again "translation" along the circles in T^n by p. Notice that if $a_1 = 2\pi$ and $a_i = 0$ otherwise, then $\exp_p(X) = p$, so unlike the $\mathbb{R}^n$ case the exponential map is not injective, and so it is only a local diffeomorphism.

Example. On $\mathcal{S}^2$ we recall we have geodesics, for $c \in \mathbb{R}$ given by

$$\gamma(t) = (\sin(ct + \theta_0)\cos\phi_0, \sin(ct + \theta_0)\sin\phi_0, \cos(ct + \theta_0).$$

These all start from the same point $p = \gamma(0) = f(\theta_0, \phi_0)$, in the usual notation, but $\gamma'(0) = cX_1$, so we see that

$$\exp_p(cX_1) = \gamma(1) = (\sin(c + \theta_0)\cos\phi_0, \sin(c + \theta_0)\sin\phi_0, \cos(c + \theta_0)).$$

We see that $\exp_p(2\pi X_1) = \exp_p(X_1)$ so again the exponential map is not injective.

Example. We saw on the hyperbolic plane (H^2, g) that we have geodesics for $c \in \mathbb{R}$ given by

$$\gamma(t) = (x_1, x_2e^{ct}).$$

so $\gamma(0) = (x_1, x_2) = p$ and $\gamma'(0) = c\partial_2$. Hence,

$$\exp_p(c\partial_2) = \gamma(1) = (x_1, x_2e^c).$$

Now we see that, at least in this direction, the exponential map is defined for all $c \in \mathbb{R}$ (since $e^c > 0$ for all c) and it is injective. So there is a question: is the exponential map a diffeomorphism or not in this setting?

Example. For $\mathrm{SU}(n)$ (as well as other compact matrix Lie groups) with its bi-invariant Riemannian metric, the exponential map $\exp_I : T_I\,\mathrm{SU}(n) \to \mathrm{SU}(n)$ is $\exp_I(X) = \exp(X)$, recalling that

$$\mathfrak{su}(n) = T_I\,\mathrm{SU}(n) = \{X \in M_n(\mathbb{C}) \,:\, X + \bar{X}^{\mathrm{T}} = 0,\ \mathrm{tr}(X) = 0\}.$$

Notice that since $X \in \mathfrak{su}(n)$ is skew-Hermitian, $\exp(X) \in \mathrm{SU}(n)$. This motivates the name "exponential map". We also see that geodesics through I are 1-parameter subgroups $\gamma(t) = \exp(tX)$. More generally, geodesics through $A \in \mathrm{SU}(n)$ will be $\gamma(t) = A\exp(tX)$ for $X \in \mathfrak{su}(n)$. This is all part of the more general theory relating *Lie algebras* and *Lie groups*

So we have a suspicion, based on these examples, that the exponential map might "fill out" a neighbourhood of the point we care about. This turns out to be correct. The exponential map thus allows us to locally identify a open neighbourhood in the tangent space of the manifold (and thus in $\mathbb{R}^n$) with an open neighbourhood of the manifold. Moreover, this identification encodes the behaviour of the geodesics, unlike say random choices of coordinate charts. We will see the importance of this later.

We can even do slightly better as the next theorem shows.

Theorem 3.7. *Given $p \in M$ there exist an open set $W \ni p$ and $\delta > 0$ such that for all $q \in W$*

$$\exp_q : B_\delta(0) \subseteq T_qM \to \exp_q(B_\delta(0)) \supseteq W$$

is a diffeomorphism onto its image.

Proof. The key to the proof is to calculate the differential of the exponential map at 0. The exponential map $\exp_p : T_pM \to M$ so its differential $\mathrm{d}(\exp_p)_0 : T_0(T_pM) \to T_{\exp_p(0)}M = T_pM$. Since T_pM is just a vector space, we may identify $T_0(T_pM) = T_pM$, which means $\mathrm{d}(\exp_p)_0 : T_pM \to T_pM$. If $X \in T_pM$ then

$$\mathrm{d}(\exp_p)_0(X) = \frac{\mathrm{d}}{\mathrm{d}t}\exp_p(tX)|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t}\gamma_{(p,X)}(t)|_{t=0} = \gamma'_{(p,X)}(0) = X.$$

Thus $\mathrm{d}(\exp_p)_0 = \mathrm{id}$, the identity.

The rest of the proof now follows from the inverse function theorem. Let U, V, ϵ be as in Theorem 3.5. Define $F : V \subseteq TM \to M \times M$ by $F(q, X) = (q, \exp_q(X))$. Hence,

$$\mathrm{d}F_{(p,0)} : T_{(p,0)}TM \cong T_pM \times T_0(T_pM) \cong T_pM \times T_pM \to T_pM \times T_pM$$

can be written

$$\mathrm{d}F_{(p,0)} = \begin{pmatrix} I & 0 \\ A & I \end{pmatrix}$$

for some matrix A. Hence, $\mathrm{d}F_{(p,0)}$ is an isomorphism so F is a local diffeomorphism. Thus there exist $\epsilon > \delta > 0$ and open sets $\tilde{U} \subseteq U$ and $\tilde{W} \subseteq M \times M$ such that $(p, p) \in \tilde{W}$ and if

$$\tilde{V} = \{(q, X) : q \in \tilde{U}, X \in B_\delta(0) \subseteq T_qM\} \subseteq V$$

then $F : \tilde{V} \to \tilde{W}$ is a diffeomorphism. Choose an open set $W \ni p$ such that $W \times W \subseteq \tilde{W}$.

Then if $q \in W$ we have that $W \subseteq \exp_q(B_\delta(0))$ as required. $\square$

Remark. Using the exponential map, for any $p \in M$ we can define local coordinates on (M, g) called *geodesic normal coordinates* (U, φ) at p which have the property that the functions g_{ij} and Christoffel symbols Γ_{ij}^k in (U, φ)

$$\varphi(p) = 0, \quad g_{ij}(p) = \delta_{ij}, \quad \Gamma_{ij}^k = 0.$$

This means that if we call the geodesic normal coordinates $x = (x_1, \ldots, x_n)$ then every Riemannian metric equals the Euclidean metric "to first order" in geodesic normal coordinates x.

3.6 Length and normal neighbourhoods

We want to show geodesics are locally length minimizing; i.e. if the geodesic is sufficiently short then it minimizes the distance between the two endpoints amongst all nearby curves. We first define what we mean by the length.

Definition 3.8. The *length* of a curve $\alpha : [0, L] \to M$ is

$$L(\alpha) = \int_0^L |\alpha'(t)|\mathrm{d}t = \int_0^L \sqrt{g\big(\alpha'(t), \alpha'(t)\big)}\mathrm{d}t.$$

For normalised geodesics $\gamma : [0, L] \to M$, $L(\gamma) = L$ since $|\gamma'| = 1$.

The curve α is (length) *minimizing* if $L(\alpha) \leq L(\beta)$ for all curves $\beta : [0, L] \to M$ such that $\alpha(0) = \beta(0)$ and $\alpha(L) = \beta(L)$.

Example. For the normalised geodesics $\gamma(t) = (\sin(t + \theta_0)\cos\phi_0, \sin(t + \theta_0)\sin\phi_0, \cos(t + \theta_0))$ in $\mathcal{S}^2$ for $t \in [0, L]$ we see that $\gamma(2\pi) = \gamma(0)$, which means that if $L = 2\pi$ then γ is a full circle, which is has length 2π as we expect. Similarly, $\gamma(\pi) = -\gamma(0)$, so if $L = \pi$ we get a half-circle whose length is π.

Since every geodesic in $\mathcal{S}^n$ is contained in some $\mathcal{S}^2 \subseteq \mathcal{S}^n$, we see that the same argument works for normalised geodesics in $\mathcal{S}^n$: full great circles have length 2π and half great circles have length π.

Example. On $\mathbb{RP}^n$ we observed that normalised geodesics are given by $\gamma = \pi \circ \alpha$ where α is a normalised geodesic in $\mathcal{S}^n$ and π is the projection map. Since $\gamma(\pi) = \pi \circ \alpha(\pi) = \pi(-\alpha(0)) = \pi \circ \alpha(0) = \gamma(0)$, as we saw before, this shows that normalised geodesics with length π in $\mathbb{RP}^n$ are loops (rather than 2π).

Notice that geodesics cannot be length minimizing globally in general. For example, if we take geodesics longer than π on the sphere then they are no longer minimizing (because it is longer than a half-circle), whereas straight lines in $\mathbb{R}^n$ are always minimizing.

How small the geodesic should be so that it is minimizing will depend on the Riemannian metric g on M. In order to understand this we make the next definition.

Definition 3.9. An open set $U \subseteq M$ with $U \ni p$ is called a *normal neighbourhood* of p if there exists an open set $V \subseteq T_pM$ such that $\exp_p : V \to U$ is a diffeomorphism.

If $\overline{B_\epsilon(0)} \subseteq V$ we define $B_\epsilon(p) = \exp_p(B_\epsilon(0))$ to be the *geodesic ball* of radius ϵ centered at p and $\partial\overline{B_\epsilon(p)} = \mathcal{S}_\epsilon(p)$ to be the *geodesic sphere* of radius ϵ around p.

An open set $W \subseteq M$ is a *totally normal neighbourhood* if it is a normal neighbourhood of every $q \in W$.

Remark. Theorem 3.7 ensures the existence of totally normal neighbourhoods. The geodesics in the normal neighbourhood at p which emanate from p are often called *radial geodesics*. Notice that given a point $q \in B_\epsilon(p)$ radial geodesics from p to q are the unique geodesics from p to q contained in $B_\epsilon(p)$ by Theorem 3.5.

The idea should be that geodesics in normal neighbourhoods are minimizing. Let us see that this gives us the correct notion in the examples we understand.

Example. Given $p \in \mathbb{R}^n$ and $X \in T_p\mathbb{R}^n = \mathbb{R}^n$, then $\exp_p(X) = p + X$ so $\exp_p$ is defined for all $X \in T_p\mathbb{R}^n = \mathbb{R}^n$ and $\exp_p(T_p\mathbb{R}^n) = \mathbb{R}^n$, so $\exp_p$ defines a diffeomorphism between $T_p\mathbb{R}^n$ and $\mathbb{R}^n$.

Hence $\mathbb{R}^n$ is a totally normal neighbourhood and so geodesics of any length should be minimizing in $\mathbb{R}^n$: this is true because the shortest path between two points is the unique straight line between them.

Observe that the geodesic ball $B_\epsilon(p)$ is the usual metric ball of radius ϵ about p in $\mathbb{R}^n$.

Example. Given the North pole $N \in \mathcal{S}^n$, (you can think of $n = 2$ if it helps) $\exp_N$ is a map which follows a great circle, and if $X \in T_N\mathcal{S}^n$ such that $|X| = \pi$ then $\exp_N(X) = S$, the South pole.

Hence

$$\exp_N : B_\pi(0) \subseteq T_N\mathcal{S}^n \to \mathcal{S}^n \setminus \{S\}$$

is a diffeomorphism, so $\mathcal{S}^n \setminus \{S\}$ is a normal neighbourhood of N. Generally, given $p \in \mathcal{S}^n$, $\mathcal{S}^n \setminus \{-p\}$ is a normal neighbourhood of p diffeomorphic to $B_\pi(0)$.

We deduce that geodesics starting at p of length less than π are minimizing, as we would expect.

3.7 Geodesics are locally length minimizing

To prove our result about geodesics begin locally length minimizing we need a key lemma, called the *Gauss Lemma*, which is a little bit tricky. We observe again that if $p \in M$ and $X \in T_pM$ then we can identify $T_X(T_pM)$ with T_pM since there are both just copies of $\mathbb{R}^n$ and based at the same point in M.

Lemma 3.10. (Gauss Lemma). *Let $p \in M$ and $X \in T_pM$ such that $\exp_p(X)$ defined. If $Y \in T_X(T_pM) \cong T_pM$ then*

$$g_{\exp_p(X)}\left(\mathrm{d}(\exp_p)_X(X), \mathrm{d}(\exp_p)_X(Y)\right) = g_p(X, Y).$$

Remark. The Gauss Lemma says that (normalised) radial geodesics $r \mapsto \exp_p(rZ)$ in $B_\epsilon(p)$ for $Z \in T_pM$ with $|Z| = 1$ and $r \in (0, \epsilon)$ are orthogonal to the geodesic spheres $\mathcal{S}_\delta(p)$ for $\delta \in (0, \epsilon)$. In other words, we have "*geodesic polar coordinates*" near p given by $r \in (0, \epsilon)$ (the "radial" coordinate) and $Z \in T_pM$ with $|Z| = 1$ (which is the "angle" coordinate since Z lies in the unit sphere in T_pM).

Proof. Write $Y = Y^{\mathrm{T}} + Y^\perp$ where $Y^{\mathrm{T}} \in \mathrm{Span}\{X\}$ and $Y^\perp \in \mathrm{Span}\{X\}^\perp$.

The geodesic $\gamma_{(p,X)}$ so that $\gamma_{(p,X)}(0) = p$ and $\exp_p(X) = \gamma_{(p,X)}(1)$ satisfies $\gamma_{(p,X)}(t) = \exp_p(tX)$ so

$$\gamma'_{(p,X)}(t) = \mathrm{d}(\exp_p)_{tX}(X).$$

Notice that this means that

$$\gamma'_{(p,X)}(0) = \mathrm{d}(\exp_p)_0(X) = X$$

(since $\mathrm{d}(\exp_p)_0 = \mathrm{id}$) as we know but also that

$$\gamma'_{(p,X)}(1) = \mathrm{d}(\exp_p)_X(X).$$

Moreover,

$$|\gamma'_{(p,X)}(t)|^2 = g_{\gamma_{(p,X)}(t))}\big(\gamma'_{(p,X)}(t), \gamma'_{(p,X)}(t)\big) = g_{\exp_p(tX)}\big(\mathrm{d}(\exp_p)_{tX}(X), \mathrm{d}(\exp_p)_{tX}(X)\big)$$

is constant by Definition 3.1 because γ is a geodesic. Thus, choosing $t = 0$ and $t = 1$ gives

$$g_{\exp_p(X)}\big(\mathrm{d}(\exp_p)_X(X), \mathrm{d}(\exp_p)_X(X)\big) = g_p(X, X).$$

Hence, since $Y^{\mathrm{T}} \in \mathrm{Span}\{X\}$ (so $Y^{\mathrm{T}} = \lambda X$ for some constant λ) we see that

$$g_{\exp_p(X)}\big(\mathrm{d}(\exp_p)_X(X), \mathrm{d}(\exp_p)_X(Y^{\mathrm{T}})\big) = g_p(X, Y^{\mathrm{T}})$$

as $\mathrm{d}(\exp_p)_X$ is linear and g is bilinear.

So since $g_p(X, Y^\perp) = 0$ by definition it is now enough to show that

$$g_{\exp_p(X)}\big(\mathrm{d}(\exp_p)_X(X), \mathrm{d}(\exp_p)_X(Y^\perp)\big) = 0.$$

There exists $\epsilon > 0$ such that if

$$X(t) = X\cos t + Y^\perp \sin t$$

then $\exp_p(sX(t))$ is well-defined for $s \in [0,1]$ and $t \in (-\epsilon, \epsilon)$. (In other words, if $\exp_p(X)$ is defined, then so is $\exp_p(Z)$ for all Z in a little sector based at 0 containing X.) Let

$$f(s,t) = \exp_p(sX(t))$$

so $s \mapsto f(s,t) = \exp_p(sX(t))$ are radial geodesics. We can differentiate f to get

$$\frac{\partial f}{\partial s} = \mathrm{d}(\exp_p)_{sX(t)}(X(t)) \quad \text{and} \quad \frac{\partial f}{\partial t} = \mathrm{d}(\exp_p)_{sX(t)}(sX'(t)).$$

(Notice that $\frac{\partial f}{\partial s}$ and $\frac{\partial f}{\partial t}$ are vector fields on the image of f which are tangent to curves where t and s are constant respectively.) Hence, since $f(1,0) = \exp_p(X)$ and $X'(0) = Y^\perp$ we see that

$$g_{\exp_p(X)}\left(\mathrm{d}(\exp_p)_X(X), \mathrm{d}(\exp_p)_X(Y^\perp)\right) = g_{\exp_p(X)}\left(\frac{\partial f}{\partial s}(1,0), \frac{\partial f}{\partial t}(1,0)\right)$$

Now the covariant derivative along curves where t is constant (i.e. the radial geodesics $s \mapsto f(s,t)$) is

$$\nabla_{\frac{\partial f}{\partial s}}\frac{\partial f}{\partial s} = 0,$$

since $s \mapsto f(s,t)$ is a geodesic and $\frac{\partial f}{\partial s}$ is the tangent vector field to this geodesic.

To continue the proof we will need the following important lemma which we will use a number of times in the course.

Lemma 3.11 (Symmetry Lemma)**.** *Let $V \subseteq \mathbb{R}^2$ be open, let $A \subseteq \mathbb{R}^2$ be connected such that $V \subseteq A \subseteq \bar{V}$ and ∂A is a curve with vertex angles $\neq \pi$. Let $f : A \to (M, g)$ be smooth and let (u,v) be coordinates on A. Then*

$$\nabla_{\frac{\partial f}{\partial u}}\frac{\partial f}{\partial v} = \nabla_{\frac{\partial f}{\partial v}}\frac{\partial f}{\partial u}.$$

Here, $\frac{\partial f}{\partial u}$ and $\frac{\partial f}{\partial v}$ are the vector fields in $f(A) \subseteq M$ which are tangent to the curves where v and u are constant respectively.

Remark. The conditions on A ensure that the notion of smooth map on A is well-defined. The vertex angle at a point x where α is not smooth is the angle between the two rays meeting at x.

Proof. Let (U, φ) be a coordinate chart at $p \in f(A)$ and write $\varphi \circ f(u, v) = (x_1(u, v), \ldots, x_n(u, v))$. Then

$$\nabla_{\frac{\partial f}{\partial u}} \frac{\partial f}{\partial v} = \nabla_{\frac{\partial f}{\partial u}} \sum_{j=1}^{n} \frac{\partial x_j}{\partial v} X_j = \sum_{j=1}^{n} \frac{\partial^2 x_j}{\partial u \partial v} X_j + \sum_{j,k=1}^{n} \frac{\partial x_j}{\partial u} \frac{\partial x_k}{\partial v} \nabla_{X_k} X_j.$$

The first term is clearly symmetric in u, v and since $\nabla_{X_k} X_j = \sum_{i=1}^{n} \Gamma_{kj}^{i} X_i$ is symmetric in j, k the second term is symmetric in j, k and hence u, v also. ☐

Applying the Symmetry Lemma (Lemma 3.11) in our situation gives

$$\nabla_{\frac{\partial f}{\partial s}} \frac{\partial f}{\partial t} = \nabla_{\frac{\partial f}{\partial t}} \frac{\partial f}{\partial s}.$$

Now we see that, using our earlier calculation in proving Theorem 2.5,

$$\begin{aligned}
\frac{\partial}{\partial s} g\left(\frac{\partial f}{\partial s}, \frac{\partial f}{\partial t}\right) &= g\left(\nabla_{\frac{\partial f}{\partial s}} \frac{\partial f}{\partial s}, \frac{\partial f}{\partial t}\right) + g\left(\frac{\partial f}{\partial s}, \nabla_{\frac{\partial f}{\partial s}} \frac{\partial f}{\partial t}\right) \\
&= g\left(\nabla_{\frac{\partial f}{\partial t}} \frac{\partial f}{\partial s}, \frac{\partial f}{\partial s}\right) \\
&= \frac{1}{2} \frac{\partial}{\partial t} \left|\frac{\partial f}{\partial s}\right|^2 \\
&= \frac{1}{2} \frac{\partial}{\partial t} |X(t)|^2 = 0
\end{aligned}$$

as $|X(t)|^2 = |X|^2 + |Y^\perp|^2$ is constant.

Thus,

$$g\left(\frac{\partial f}{\partial s}, \frac{\partial f}{\partial t}\right)(1, 0) = g\left(\frac{\partial f}{\partial s}, \frac{\partial f}{\partial t}\right)(s, 0)$$

for all s. Now

$$\frac{\partial f}{\partial t}(s, 0) = \mathrm{d}(\exp_p)_{sX}(sY^\perp) \to 0 \quad \text{as } s \to 0$$

so

$$g\left(\frac{\partial f}{\partial s}, \frac{\partial f}{\partial t}\right)(1, 0) = g_{\exp_p(X)}\left(\mathrm{d}(\exp_p)_X(X), \mathrm{d}(\exp_p)_X(Y^\perp)\right) = 0$$

as required. ☐

We can now state our main result about geodesics.

Theorem 3.12. *Geodesics $\gamma : [0, L] \to (M, g)$ in $B_\epsilon(p)$ with $\gamma(0) = p$ are minimizing. Moreover if $\alpha : [0, L] \to M$ is a curve such that $\alpha(0) = \gamma(0)$, $\alpha(L) = \gamma(L)$ and $L(\alpha) = L(\gamma)$ then $\alpha([0, L]) = \gamma([0, L])$.*

Proof. Suppose α is a comparison curve to γ and suppose without loss of generality that $\gamma(0) \neq \gamma(L)$ (otherwise γ is simply the constant geodesic which is clearly minimizing). Since we are in $B_\epsilon(p)$ the unique geodesic from $\gamma(0)$ to $\gamma(L)$ is the radial geodesic.

If $\alpha([0, L]) \not\subseteq B_\epsilon(p)$ then let $T \in (0, L]$ be least such that $\alpha(T) \in \mathcal{S}_\epsilon(p)$. Thus $L(\alpha) \geq L(\alpha|_{[0,T]})$ and $\alpha|_{[0,T]}$ is a curve contained in $\overline{B_\epsilon(p)}$. Reparameterise such that $\alpha|_{[0,T]}$ is defined on $[0, L]$ (this does not change its length), and call this α. If we can show that this new α is at least as long as the radial geodesic from $\alpha(0) = p$ to $\alpha(L) = q$ then we are done, since we will have shown that any curve connecting p to any other point in the geodesic ball is at least as long as a radial geodesic, so radial geodesics are minimizing.

Hence we now assume that α is contained in $B_\epsilon(p)$. Without loss of generality we can assume $\alpha(t) \neq p$ for $t > 0$ (since otherwise it stays at p for a while before moving away from p). Thus we can write

$$\alpha(t) = \exp_p(r(t)X(t))$$

for $t \in (0, L]$ where $r : (0, L] \to \mathbb{R}^+$ is piecewise smooth and $X(t)$ is a curve in T_pM with $|X(t)| = 1$. Notice that

$$q = \gamma(L) = \alpha(L) = \exp_p\left(r(L)X(L)\right)$$

so we can write the geodesic $\gamma : [0, L] \to M$ from p to q as

$$\gamma(s) = \exp_p\left(\frac{sr(L)X(L)}{L}\right)$$

since it is a radial geodesic.

In the notation of the proof of the Gauss Lemma, $\alpha(t) = f(r(t), t)$ so

$$\alpha'(t) = r'(t)\frac{\partial f}{\partial s}(r(t), t) + \frac{\partial f}{\partial t}(r(t), t).$$

The Gauss Lemma implies that

$$g\left(\frac{\partial f}{\partial s}, \frac{\partial f}{\partial s}\right)(r(t), t) = g_{\exp_p(r(t)X(t))}\big(\mathrm{d}(\exp_p)_{r(t)X(t)}(X(t)), \mathrm{d}(\exp_p)_{r(t)X(t)}(X(t))\big) = g_p(X(t), X(t)) = 1$$

and

$$\begin{aligned} g\left(\frac{\partial f}{\partial s}, \frac{\partial f}{\partial t}\right)(r(t), t) &= g_{\exp_p(r(t)X(t))}\big(\mathrm{d}(\exp_p)_{r(t)X(t)}(X(t)), \mathrm{d}(\exp_p)_{r(t)X(t)}(X'(t))\big) \\ &= g_p(X(t), X'(t)) = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}|X(t)|^2 = 0 \end{aligned}$$

since $|X(t)| = 1$ for all t.

Hence

$$|\alpha'|^2 = |r'|^2 + \left|\frac{\partial f}{\partial t}\right|^2 \geq |r'|^2.$$

Integrating we see that

$$L(\alpha) = \int_0^L |\alpha'(t)|\mathrm{d}t \geq \int_0^L |r'(t)|\mathrm{d}t \geq \int_0^L r'(t)\mathrm{d}t = r(L) = L(\gamma)$$

since we observed that, as γ is a radial geodesic, $L(\gamma) = L.\frac{|r(L)X(L)|}{L} = r(L)$ as $r(L) > 0$ and $|X(L)| = 1$. We deduce that γ is minimizing.

Moreover $L(\alpha) = L(\gamma)$ only if $\frac{\partial f}{\partial t} = 0$, so $X'(t) = 0$ which means $X(t) = X$ is constant, and $|r'| = r' > 0$. Therefore α is a monotonic reparametrization of γ (as $\gamma(s) = \exp_p(\frac{sr(L)X}{L})$), so $\alpha([0, L]) = \gamma([0, L])$. □

We have shown geodesics are *locally minimizing* (i.e. in a neighbourhood of each point it is minimizing). We now show that we can ensure a locally minimizing curve has to be a geodesic.

Proposition 3.13. *If $\gamma : [0, L] \to M$ is a curve with $|\gamma'|$ constant and it is locally minimizing then γ is a geodesic.*

Proof. Let $t \in [0, L]$ and let W be a totally normal neighbourhood of $\gamma(t)$. Then there exists $\delta > 0$ such that if we let $\alpha = \gamma|_{[t-\delta,t+\delta]\cap[0,L]}$ then α is minimizing and and α is contained in W. Therefore α is a curve from p to q in a geodesic ball centred at p as W is a totally normal neighbourhood.

By Theorem 3.12 $L(\alpha)$ is the length of the radial geodesic $\beta(s) = \exp_p(sX)$ from p to q, so α is a monotonic parametrisation of β; i.e. $\alpha(s) = \exp_p(r(s)X)$ for some positive increasing function r such that $r(0) = 0$. However, $|\alpha'|^2 = |r'|^2$ by the proof of Theorem 3.12 so $|r'| = r'$ (as r is increasing) is constant which means that r is a multiple of s. Hence α is a radial geodesic and thus γ is a geodesic on $[t - \delta, t + \delta] \cap [0, L]$ and in particular at t.

Since t was arbitrary, γ satisfies the geodesic equation at t for all t and so is a geodesic. □

3.8 First variation formula

We now want to take an alternative approach to studying the minimizing properties of geodesics using the viewpoint of the Calculus of Variations. We first define what we mean by a variation of a curve.

Definition 3.14. Let $\alpha : [0, L] \to (M, g)$ be a curve. A *variation* of α is a smooth map $f : (-\epsilon, \epsilon) \times [0, L] \to M$, for some $\epsilon > 0$, such that $f(0, t) = \alpha(t)$ for all $t \in [0, L]$. A variation f is *proper* if $f(s, 0) = \alpha(0)$ and $f(s, L) = \alpha(L)$ for all $s \in (-\epsilon, \epsilon)$.

The *variation field* of f is the vector field $V_f(t) = \frac{\partial f}{\partial s}(0, t)$ along α.

Writing down a variation explicitly can be quite challenging, whereas writing down a vector field along a curve (the variation field) is very easy. This motivates us to prove that given a vector field along a curve we can construct a variation which realises that vector field as the variation field.

Proposition 3.15. *Given a vector field V along a curve $\alpha : [0, L] \to (M, g)$, there exists a variation f of α such that $V = V_f$, the variation field of f. Moreover, if $V(0) = V(L) = 0$ we can choose f to be proper.*

Proof. For each $t \in [0, L]$ let $W_t = \{\exp_{\alpha(t)}(X) : |X| < \delta_t\}$ be a totally normal neighbourhood of $\alpha(t)$, where $\delta_t > 0$. Since $\{W_t : t \in [0, L]\}$ covers $\alpha([0, L])$, which is compact, there is a finite subcover $\{W_{t_1}, \ldots, W_{t_k}\}$. Let $\delta = \min_i\{\delta_{t_i}\}$, let $0 < \epsilon < \delta / \max_{t\in[0,L]} |V(t)|$ and let $f(s, t) = \exp_{\alpha(t)}(sV(t))$ for $s \in (-\epsilon, \epsilon)$, $t \in [0, L]$. We may then calculate

$$V_f(t) = \frac{\partial f}{\partial s}(0, t) = \frac{\mathrm{d}}{\mathrm{d}s}\big(\exp_{\alpha(t)}(sV(t))\big)|_{s=0} = \mathrm{d}(\exp_{\alpha(t)})_0(V(t)) = V(t)$$

as required. Moreover, if $V(0) = V(L) = 0$ then $f(s, 0) = \alpha(0)$ and $f(s, L) = \alpha(L)$ so f is proper. □

From the point of view of the Calculus of Variations it is more convenient to work with a different functional than the length functional when studying curves and geodesics; namely, the energy functional.

Definition 3.16. The *energy* of a curve $\alpha : [0, L] \to (M, g)$ is

$$E(\alpha) = \int_0^L |\alpha'(t)|^2 \mathrm{d}t.$$

The *energy* of a variation f of α is

$$E_f(s) = \int_0^L |\frac{\partial f}{\partial t}(s, t)|^2 \mathrm{d}t.$$

Our next result shows the relationship between length and energy.

Lemma 3.17. *Let $\alpha : [0, L] \to (M, g)$ be a curve. Then*

$$L(\alpha)^2 \leq LE(\alpha)$$

with equality if and only if $|\alpha'|$ is constant.

Proof. By the Cauchy–Schwarz inequality, $(\int_0^L ab\mathrm{d}t)^2 \leq \int_0^L a^2\mathrm{d}t \int_0^L b^2\mathrm{d}t$ with equality if and only if $b = \lambda a$ for some constant λ. Setting $a = 1$ and $b = |\alpha'|$ gives the result. □

We know that geodesics locally minimize length. We now show that they locally minimize energy as well.

Lemma 3.18. *Let $p, q \in (M, g)$ and let $\gamma : [0, L] \to M$ be a minimizing geodesic between p and q. For all curves $\alpha : [0, L] \to M$ such that $\alpha(0) = p$, $\alpha(L) = q$, we have that $E(\gamma) \leq E(\alpha)$ and equality holds if and only if α is a minimizing geodesic.*

Proof. From Lemma 3.17 we deduce that

$$LE(\gamma) = L(\gamma)^2 \leq L(\alpha)^2 \leq LE(\alpha)$$

with equality if and only if $L(\alpha) = L(\gamma)$ and $|\alpha'|$ constant. Proposition 3.13 gives the result. □

Theorem 3.19 (First Variation Formula)**.** *Let $\alpha : [0, L] \to (M, g)$ be a curve and let f be a variation of α. Then the energy E_f of f satisfies*

$$\frac{1}{2}E_f'(0) = -\int_0^L g(V_f, \nabla_{\alpha'}\alpha')\mathrm{d}t - g(V_f(0), \alpha'(0)) + g(V_f(L), \alpha'(L)).$$

Proof. Using the Symmetry Lemma (Lemma 3.11)

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\int_0^L g(\frac{\partial f}{\partial t}, \frac{\partial f}{\partial t})\,\mathrm{d}t &= 2\int_0^L g(\nabla_{\frac{\partial f}{\partial s}}\frac{\partial f}{\partial t}, \frac{\partial f}{\partial t})\mathrm{d}t \\ &= 2\int_0^L g(\nabla_{\frac{\partial f}{\partial t}}\frac{\partial f}{\partial s}, \frac{\partial f}{\partial t})\mathrm{d}t \\ &= 2\int_0^L \frac{\mathrm{d}}{\mathrm{d}t}g(\frac{\partial f}{\partial s}, \frac{\partial f}{\partial t})\mathrm{d}t - 2\int_0^L g(\frac{\partial f}{\partial s}, \nabla_{\frac{\partial f}{\partial t}}\frac{\partial f}{\partial t})\mathrm{d}t.\end{aligned}$$

We deduce that

$$\frac{1}{2}E_f'(s) = g(\frac{\partial f}{\partial s}, \frac{\partial f}{\partial t})|_0^L - \int_0^L g(\frac{\partial f}{\partial s}, \nabla_{\frac{\partial f}{\partial t}}\frac{\partial f}{\partial t})\mathrm{d}t.$$

Setting $s = 0$ gives the result. □

We now have a new characterisation of geodesics.

Corollary 3.20. *A curve $\alpha : [0, L] \to (M, g)$ is a geodesic if and only if for all proper variations f of α, $E_f'(0) = 0$.*

Proof. Suppose α is a geodesic and f is a proper variation of α. Then $\nabla_{\alpha'}\alpha' = 0$, $V_f(0) = V_f(L) = 0$ and α is smooth so $E_f'(0) = 0$ by Theorem 3.19.

Now suppose that $E_f'(0) = 0$ for all proper variations f of α. Let $h : [0, L] \to \mathbb{R}$ be a smooth function such that $h(t) > 0$ for $t \in (0, L)$ and $h(0) = h(L) = 0$ and let $V(t) = h(t)\nabla_{\alpha'}\alpha'$. Proposition 3.15 implies there exists a proper variation f such that $V_f = V$. Theorem 3.19 then implies that

$$E_f'(0) = -\int_0^L h|\nabla_{\alpha'}\alpha'|^2\mathrm{d}t = 0$$

so α is a geodesic. □

4 Curvature

We now move on to the other key idea in the course: namely, curvature. Curvature is something we understand intuitively but if you think about it our usual notions involve the way the object sits inside Euclidean space: we see the ellipsoid as being a different curvature from the round sphere exactly this way. We therefore need to think about curvature "intrinsically".

So far, we have seen objects that are defined by at most first derivatives of the Riemannian metric, for example geodesics are determined by the Christoffel symbols, which depend on the Riemannian metric and its first derivatives. As a result, we have never truly noticed the curvature of the manifold: for example, we have seen that geodesics are just (locally at least) the images of straight lines.

To understand curvature we need to look at second derivatives of the Riemannian metric and, just as the Riemannian metric is an operator on vector fields, curvature is defined in a similar way. Initially, it will look a bit abstract, but we will soon see how to give a natural intuitive interpretation of curvature.

Let (M, g) be a Riemannian manifold with Levi-Civita connection ∇ throughout.

4.1 Riemann curvature

Proposition 4.1. *For vector fields X, Y, Z on (M, g) we define*

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

which is a vector field on M. Then $R(.,.)$ is bilinear in its arguments, $R(X,Y)$ is a linear operator and $R(X,Y)Z(p) \in T_pM$ only depends on $X(p), Y(p), Z(p) \in T_pM$.

The operator $R(X,Y)$ which sends vector fields to vector fields given $X, Y \in \Gamma(TM)$ is called the Riemann curvature operator. Notice that $R(X,Y) = -R(Y,X)$.

Remark. We can informally think about this as pushing the vector Z around a parallelogram determined by the vector fields X and Y. The outcome of this procedure is a new tangent vector which may be different from Z. The limit of this procedure as the sides of the parallelogram goes to 0 is the operator $R(X,Y)$ (when $[X,Y] = 0$).

Proof. Clearly $R(X_1 + X_2, Y) = R(X_1, Y) + R(X_2, Y)$, $R(X, Y_1 + Y_2) = R(X, Y_1) + R(X, Y_2)$ and $R(X,Y)(Z_1 + Z_2) = R(X,Y)Z_1 + R(X,Y)Z_2$ by Theorem 2.1.

Let $f : M \to \mathbb{R}$ be a smooth function. By Theorem 2.1 and the fact that

$$[fX, Y] = (fX)Y - Y(fX) = f(XY - YX) - Y(f)X$$

we see that

$$\begin{aligned} R(fX,Y)Z &= \nabla_{fX}\nabla_Y Z - \nabla_Y \nabla_{fX} Z - \nabla_{[fX,Y]} Z = f\nabla_X \nabla_Y Z - \nabla_Y (f\nabla_X Z) - \nabla_{f[X,Y]-Y(f)X} Z \\ &= f\nabla_X \nabla_Y Z - f\nabla_Y \nabla_X Z - Y(f)\nabla_X Z - f\nabla_{[X,Y]} Z + Y(f)\nabla_X Z = fR(X,Y)Z \end{aligned}$$

so $R(.,.)$ is bilinear in its arguments as $R(Y,X) = -R(X,Y)$.

A similar argument works for $R(X,Y)(fZ)$:

$$\begin{aligned} R(X,Y)(fZ) &= (\nabla_X \nabla_Y - \nabla_Y \nabla_X - \nabla_{[X,Y]})(fZ) \\ &= \nabla_X (f\nabla_Y Z + Y(f)Z) - \nabla_Y (f\nabla_X Z + X(f)Z) - f\nabla_{[X,Y]} Z - [X,Y](f)Z \\ &= f\nabla_X \nabla_Y Z + X(f)\nabla_Y Z + Y(f)\nabla_X Z + XY(f)Z - f\nabla_Y \nabla_X Z - Y(f)\nabla_X Z \\ &\quad - X(f)\nabla_Y Z - YX(f)Z - f\nabla_{[X,Y]} Z - [X,Y](f)Z = fR(X,Y)Z. \end{aligned}$$

Thus $R(X,Y)$ is linear as claimed.

For the final part, if we let $\{X_1, \ldots, X_n\}$ be a coordinate frame field in a chart (U, φ) at p and write $X = \sum_{i=1}^n a_i X_i$, $Y = \sum_{i=1}^n b_i X_i$ and $Z = \sum_{i=1}^n c_i X_i$ then a direct computation shows that

$$R(X,Y)Z = \sum_{i,j,k=1}^{n} a_i b_j c_k R(X_i, X_j) X_k$$

and since $R(X_i, X_j)X_k$ is independent of X, Y, Z, this shows that $R(X,Y)Z(p)$ only depends on $X(p)$, $Y(p)$ and $Z(p)$. $\square$

Example. Suppose $M = \mathbb{R}^n$ with the Euclidean metric g_0. If ∂_i are the standard vector fields on $\mathbb{R}^n$, then we know that $[\partial_i, \partial_j] = 0$ and $\nabla_{\partial_i}\partial_j = 0$ so

$$R(\partial_i, \partial_j)\partial_k = 0.$$

Since R is linear, we see that $R(X,Y) = 0$ for all vector fields X, Y on $\mathbb{R}^n$.

Example. Just in case we do not believe the previous example, let us calculate things in polar coordinates instead on $\mathbb{R}^2 \setminus \{0\}$. In this case, we recall that $X_1 = f_*\partial_r$, $X_2 = f_*\partial_\theta$ so $[X_1, X_2] = 0$ and

$$\nabla_{X_1} X_1 = 0, \quad \nabla_{X_1} X_2 = \nabla_{X_2} X_1 = \frac{1}{r} X_2, \quad \nabla_{X_2} X_2 = -r X_1.$$

Hence, we see that

$$\begin{aligned}
R(X_1, X_2)X_1 &= \nabla_{X_1}\nabla_{X_2} X_1 - \nabla_{X_2}\nabla_{X_1} X_1 \\
&= \nabla_{X_1}(\frac{1}{r} X_2) - \nabla_{X_2}(0) \\
&= X_1(\frac{1}{r}) X_2 + \frac{1}{r}\nabla_{X_1} X_2 \\
&= -\frac{1}{r^2} X_2 + \frac{1}{r^2} X_2 = 0.
\end{aligned}$$

Similarly, we find that

$$\begin{aligned}
R(X_1, X_2)X_2 &= \nabla_{X_1}\nabla_{X_2} X_2 - \nabla_{X_2}\nabla_{X_1} X_2 \\
&= \nabla_{X_1}(-r X_1) - \nabla_{X_2}(\frac{1}{r} X_2) \\
&= X_1(-r) X_1 - \frac{1}{r}\nabla_{X_2} X_2 \\
&= -X_1 + \frac{1}{r} r X_1 = 0.
\end{aligned}$$

So $R(X_1, X_2) = 0$ as we would expect.

It is more usual to think about Riemann curvature in the following way.

Definition 4.2. We define R by

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

for vector fields X, Y, Z, W on M. This is well-defined because at $p \in M$ it only depends on g_p and the values of X, Y, Z, W at p. We call R the *Riemann curvature tensor*.

Remark. If X_i are coordinate vector fields we let $R_{ijkl} = R(X_i, X_j, X_k, X_l)$. If we take geodesic normal coordinates $(x_1, \ldots, x_n)$ at p so that $g_{ij} = \delta_{ij}$ and $\Gamma_{ij}^k = 0$ at $p = (0, \ldots, 0)$ as mentioned earlier, we find that

$$g_{ij} = \delta_{ij} - \sum_{k,l} \frac{1}{3} R_{ijkl} x_k x_l + O(|x|^3),$$

so R measures the true first difference between the Riemannian metric on M and the Euclidean metric on M.

Example. We see that $R = 0$ on $\mathbb{R}^n$. We call Riemannian manifolds for which $R = 0$ *flat*.

Example. Since $\nabla_{X_i} X_j = [X_i, X_j] = 0$ for the standard vector fields on $T^n \subseteq \mathbb{R}^{2n}$ we see that $R = 0$ and hence T^n is flat.

Example. On $\mathcal{S}^2$, if $X_1 = f_*\partial_\theta$ and $X_2 = f_*\partial_\phi$ for $f(\theta, \phi) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$ as usual, we have $[X_1, X_2] = 0$ and

$$\nabla_{X_1} X_1 = 0, \quad \nabla_{X_2} X_2 = -\sin\theta\cos\theta X_1, \quad \text{and} \quad \nabla_{X_2} X_1 = \nabla_{X_1} X_2 = \cot\theta X_2.$$

Therefore,

$$R(X_1, X_2)X_1 = \nabla_{X_1}\nabla_{X_2}X_1 - \nabla_{X_2}\nabla_{X_1}X_1 = \nabla_{X_1}(\cot\theta X_2) = -\mathrm{cosec}^2\theta X_2 + \cot\theta\nabla_{X_1}X_2 = -X_2$$

and

$$R(X_1, X_2)X_2 = \nabla_{X_1}\nabla_{X_2}X_2 - \nabla_{X_2}\nabla_{X_1}X_2 = -\nabla_{X_1}(\sin\theta\cos\theta X_1) - \nabla_{X_2}(\cot\theta X_2) = \sin^2\theta X_1.$$

Therefore, we see that on $\mathcal{S}^2$ with the usual X_1, X_2 we have that

$$R(X_1, X_2, X_1, X_1) = 0, \quad R(X_1, X_2, X_1, X_2) = -g(X_2, X_2) = -\sin^2\theta,$$

$$R(X_1, X_2, X_2, X_1) = \sin^2\theta g(X_1, X_1) = \sin^2\theta, \quad R(X_1, X_2, X_2, X_2) = 0.$$

If we let $E_1 = X_1$ and $E_2 = \frac{X_2}{\sin\theta}$ then E_1, E_2 are orthonormal and by linearity we see that

$$R(E_1, E_2, E_2, E_1) = 1$$

which suggests maybe this should be some definition of having curvature 1, as we would expect the sphere to have.

Example. Remember on $\mathcal{S}^3$ we had vector fields E_1, E_2, E_3 which are orthonormal and

$$\nabla_{E_i} E_j = \frac{1}{2}[E_i, E_j] = -\epsilon_{ijk} E_k.$$

We see that

$$\begin{aligned} R(E_1, E_2)E_2 &= \nabla_{E_1}\nabla_{E_2}E_2 - \nabla_{E_2}\nabla_{E_1}E_2 - \nabla_{[E_1,E_2]}E_2 \\ &= \nabla_{E_1}0 + \nabla_{E_2}E_3 + 2\nabla_{E_3}E_2 \\ &= -E_1 + 2E_1 = E_1. \end{aligned}$$

Hence,

$$R(E_1, E_2, E_2, E_1) = g(E_1, E_1) = 1.$$

Similarly,

$$R(E_2, E_3, E_3, E_2) = R(E_3, E_1, E_1, E_3) = 1.$$

We also see that

$$\begin{aligned} R(E_1, E_2)E_3 &= \nabla_{E_1}\nabla_{E_2}E_3 - \nabla_{E_2}\nabla_{E_1}E_3 - \nabla_{[E_1,E_2]}E_3 \\ &= -\nabla_{E_1}E_1 - \nabla_{E_2}E_2 + 2\nabla_{E_3}E_3 = 0. \end{aligned}$$

We deduce that

$$R(E_1, E_2, E_3, E_i) = 0$$

for all i.

Before we continue with examples, we notice that R has various symmetries which we now derive.

Proposition 4.3. *Let X, Y, Z, W be vector fields on (M, g).*

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

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

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

(d) (Bianchi identity) $R(X, Y, Z, W) + R(Y, Z, X, W) + R(Z, X, Y, W) = 0$.

Proof. (a) is immediate as $R(X, Y) = -R(Y, X)$. (b) is the same as showing $R(X, Y, Z, Z) = 0$ since then

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

as $R(X, Y, Z, Z) = R(X, Y, W, W) = 0$, which gives the result. First,

$$g(\nabla_X\nabla_Y Z, Z) = X\big(g(\nabla_Y Z, Z)\big) - g(\nabla_Y Z, \nabla_X Z) = \frac{1}{2}X\big(Y\big(g(Z, Z)\big)\big) - g(\nabla_Y Z, \nabla_X Z)$$

and

$$g(\nabla_{[X,Y]} Z, Z) = \frac{1}{2}[X, Y]\big(g(Z, Z)\big).$$

So

$$R(X, Y, Z, Z) = g(R(X, Y)Z, Z) = \frac{1}{2}\Big(X\big(Y\big(g(Z, Z)\big)\big) - Y\big(X\big(g(Z, Z)\big)\big)\Big) - \frac{1}{2}[X, Y]\big(g(Z, Z)\big) = 0.$$

(d) is really just a restatement of the Jacobi identity for the Lie bracket:

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

by the Jacobi identity, so taking the inner product with W using g gives the result.

Finally, for (c), we use (d):

$$\begin{aligned}
R(X, Y, Z, W) + R(Y, Z, X, W) + R(Z, X, Y, W) &= 0 \\
R(Y, Z, W, X) + R(Z, W, Y, X) + R(W, Y, Z, X) &= 0 \\
R(Z, W, X, Y) + R(W, X, Z, Y) + R(X, Z, W, Y) &= 0 \\
R(W, X, Y, Z) + R(X, Y, W, Z) + R(Y, W, X, Z) &= 0.
\end{aligned}$$

Adding plus using (a) and (b) gives $2R(Z, X, Y, W) + 2R(W, Y, Z, X) = 0$ and the result follows from using (a) again. □

As ever, it is good to have a local understanding of the Riemann curvature tensor.

Proposition 4.4. *Let (U, φ) be a coordinate chart and X_i be the coordinate vector fields. We have that*

$$R(X_i, X_j)X_k = \sum_{l=1}^{n} R^l_{ijk} X_l$$

where, letting $\partial_l \Gamma^k_{ij} = X_l(\Gamma^k_{ij})$:

$$R^l_{ijk} = \partial_i \Gamma^l_{jk} - \partial_j \Gamma^l_{ki} + \sum_{m=1}^{n} \Gamma^l_{im}\Gamma^m_{jk} - \sum_{m=1}^{n} \Gamma^l_{jm}\Gamma^m_{ki}.$$

Moreover,

$$R_{ijkl} = R(X_i, X_j, X_k, X_l) = \sum_{m=1}^{n} R^m_{ijk} g_{lm}.$$

Proof. Since the X_i form a basis for the vector fields on U, we can write $R(X_i,X_j)X_k$ as claimed. Since $[X_i,X_j]=0$ we know that

$$R(X_i,X_j)X_k = (\nabla_{X_i}\nabla_{X_j} - \nabla_{X_j}\nabla_{X_i})X_k = \sum_{l=1}^n R^l_{ijk}X_l.$$

The formula for R^l_{ijk} then follows from Proposition 2.3.

The final claim follows from the calculation:

$$R_{ijkl} = R(X_i,X_j,X_k,X_l) = g(R(X_i,X_j)X_k,X_l) = g(\sum_{m=1}^n R^m_{ijk}X_m,X_l) = \sum_{m=1}^n R^m_{ijk}g_{lm}.$$

□

Example. Let $\mathcal{H}^2 = \{(x_1,x_2,x_3)\in\mathbb{R}^3 : x_1^2+x_2^2-x_3^2=-1,\ x_3>0\}$ with the Riemannian metric given by the restriction g of $\mathrm{d}x_1^2+\mathrm{d}x_2^2-\mathrm{d}x_3^2$. Let

$$f(\theta,\phi) = (\sinh\theta\cos\phi, \sinh\theta\sin\phi, \cosh\theta)$$

parameterize $\mathcal{H}^2$ and let

$$X_1 = f_*\partial_\theta = \cosh\theta\cos\phi\partial_1 + \cosh\theta\sin\phi\partial_2 + \sinh\theta\partial_3$$

and

$$X_2 = f_*\partial_\phi = -\sinh\theta\sin\phi\partial_1 + \sinh\theta\cos\phi\partial_2.$$

Then

$$g(X_1,X_1) = \cosh^2\theta\cos^2\phi + \cosh^2\theta\sin^2\phi - \sinh^2\theta = \cosh^2\theta - \sinh^2\theta = 1,$$
$$g(X_1,X_2)=0 \quad\text{and}\quad g(X_2,X_2)=\sinh^2\theta.$$

Thus

$$f^*g = \mathrm{d}\theta^2 + \sinh^2\theta\mathrm{d}\phi^2$$

in these coordinates so we can calculate the Christoffel symbols by looking at

$$L = \frac{1}{2}(\theta')^2 + \frac{1}{2}\sinh^2\theta(\phi')^2.$$

We compute

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial\theta'}\right) - \frac{\partial L}{\partial\theta} &= \theta'' - \sinh\theta\cosh\theta(\phi')^2\\
&= x_1'' + \sum_{i,j}\Gamma^1_{ij}x_i'x_j',\\
\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial\phi'}\right) - \frac{\partial L}{\partial\phi} &= (\sinh^2\theta\phi')'\\
&= \sinh^2\theta\phi'' + 2\sinh\theta\cosh\theta\theta'\phi'\\
&= \sinh^2\theta(x_2'' + \sum_{i,j}\Gamma^2_{ij}x_i'x_j').
\end{aligned}$$

Hence,

$$\Gamma^1_{11}=\Gamma^2_{11}=0,\quad \Gamma^1_{22}=-\sinh\theta\cosh\theta,\quad \Gamma^2_{22}=0,\quad \Gamma^1_{12}=0,\quad \Gamma^2_{12}=\coth\theta$$

and hence that

$$\nabla_{X_1}X_1=0,\quad \nabla_{X_2}X_2 = -\sinh\theta\cosh\theta X_1,\quad\text{and}\quad \nabla_{X_2}X_1=\nabla_{X_1}X_2=\coth\theta X_2.$$

Thus, since $[X_1, X_2] = 0$,

$$\begin{aligned}
R(X_1, X_2)X_2 &= \nabla_{X_1}\nabla_{X_2}X_2 - \nabla_{X_2}\nabla_{X_1}X_2 \\
&= \nabla_{X_1}(-\sinh\theta\cosh\theta X_1) - \nabla_{X_2}(\coth\theta X_2) \\
&= X_1(-\sinh\theta\cosh\theta)X_1 - \coth\theta\nabla_{X_2}X_2 \\
&= (-\cosh^2\theta - \sinh^2\theta)X_1 + \cosh^2\theta X_1 \\
&= -\sinh^2\theta X_1.
\end{aligned}$$

Hence,

$$R(X_1, X_2, X_2, X_1) = -\sinh^2\theta g(X_1, X_1) = -\sinh^2\theta.$$

If we let $E_1 = X_1$ and $E_2 = \frac{X_2}{\sinh\theta}$ which are orthonormal, then by linearity

$$R(E_1, E_2, E_2, E_1) = -1.$$

Notice this is the opposite sign to the $\mathcal{S}^2$ case, and suggests the definition of curvature -1.

4.2 Sectional curvature

As we have seen the Riemann curvature tensor is some complicated object which is a little difficult to understand. However, the essential idea is that one can restrict to looking at pieces of the Riemann curvature tensor which together tell you everything.

Definition 4.5. Let $\sigma = \mathrm{Span}\{X, Y\} \subseteq T_pM$ be a 2-plane. The *sectional curvature* of σ is given by

$$K(\sigma) = K(X, Y) = \frac{R(X, Y, Y, X)}{g(X, X)g(Y, Y) - g(X, Y)^2}.$$

Notice that this is well-defined; i.e. $K(\sigma)$ is independent of the choice of basis for σ.

Any other basis is of the form $\{aX + bY, cX + dY\}$ where $(ad - bc)^2 \neq 0$ so that the vectors are linearly independent. Clearly

$$R(aX + bY, cX + dY, cX + dY, aX + bY) = (ad - bc)^2 gR(X, Y, Y, X)$$

using the properties of R in Proposition 4.3. Moreover,

$$\begin{aligned}
&g(aX + bY, aX + bY)g(cX + dY, cX + dY) - g(aX + bY, cX + dY)^2 \\
&= \big(a^2 g(X, X) + 2ab g(X, Y) + b^2 g(Y, Y)\big)\big(c^2 g(X, Y) + 2cd g(X, Y) + d^2 g(Y, Y)\big) \\
&\quad - \big(ac g(X, X) + (ad + bc)g(X, Y) + bd g(Y, Y)\big)^2 \\
&= (ad - bc)^2\big(g(X, X)g(Y, Y) - g(X, Y)^2\big).
\end{aligned}$$

Therefore the factor of $(ad - bc)^2 \neq 0$ cancels and $K(\sigma)$ is independent of the choice of basis.

Now we want to show the claim that the sectional curvature actually contains all of the useful information.

Proposition 4.6. *Let $\bar{R}$ be such that it has the same properties as R given in Proposition 4.3. Suppose that for all $p \in M$ and for all 2-dimensional subspaces $\sigma = \mathrm{Span}\{X, Y\} \subseteq T_pM$ we have that*

$$\bar{K}(\sigma) = \frac{\bar{R}(X, Y, Y, X)}{g(X, X)g(Y, Y) - g(X, Y)^2} = K(\sigma).$$

Then $R = \bar{R}$.

This result can be paraphrased as "the sectional curvature determines the Riemann curvature".

Proof. Suppose that $K = \bar{K}$, which is equivalent to saying that $R(X,Y,Y,X) = \bar{R}(X,Y,Y,X)$ for all X,Y. Then $R(X+Z,Y,Y,X+Z) = \bar{R}(X+Z,Y,Y,X+Z)$ for X,Y,Z so

$$\begin{aligned}&R(X,Y,Y,X) + R(Z,Y,Y,Z) + 2R(X,Y,Y,Z)\\ &\qquad\qquad = \bar{R}(X,Y,Y,X) + \bar{R}(Z,Y,Y,Z) + 2\bar{R}(X,Y,Y,Z),\end{aligned}$$

using the fact that $\bar{R}(Z,Y,Y,X) = \bar{R}(Y,X,Z,Y) = \bar{R}(X,Y,Y,Z)$ and the same is true for R. Thus

$$R(X,Y,Y,Z) = \bar{R}(X,Y,Y,Z)$$

for all X,Y,Z. So, $R(X,Y+W,Y+W,Z) = \bar{R}(X,Y+W,Y+W),Z)$ for all X,Y,Z,W which implies

$$R(X,Y,W,Z) + R(X,W,Y,Z) = \bar{R}(X,Y,W,Z) + \bar{R}(X,W,Y,Z)$$

and thus

$$\begin{aligned}R(X,Y,Z,W) - \bar{R}(X,Y,Z,W) &= R(X,W,Y,Z) - \bar{R}(X,W,Y,Z)\\ &= R(Y,Z,X,W) - \bar{R}(Y,Z,X,W) \qquad (***)\end{aligned}$$

using the symmetry properties of $R, \bar{R}$. We see that cyclic permutations of X,Y,Z leave (***) invariant so

$$(R-\bar{R})(X,Y,Z,W) = (R-\bar{R})(Y,Z,X,W) = (R-\bar{R})(Z,X,Y,W).$$

The Bianchi identity then means that

$$2(R-\bar{R})(X,Y,Z,W) = (R-\bar{R})(Y,Z,X,W) + (R-\bar{R})(Z,X,Y,W) = -(R-\bar{R})(X,Y,Z,W)$$

so $R(X,Y,Z,W) = \bar{R}(X,Y,Z,W)$ for all X,Y,Z,W. $\square$

Example. For $\mathbb{R}^n$, $K = 0$ since $R = 0$. The same is true for any flat manifold such as $T^n \subseteq \mathbb{R}^{2n}$ or the cylinder $\mathcal{S}^1 \times \mathbb{R}$.

Example. For $\mathcal{S}^2$, we that $T_p\mathcal{S}^2 = \mathrm{Span}\{X_1, X_2\}$ where $g(X_1,X_1) = 1$ and $g(X_2,X_2) = \sin^2\theta$ and $g(X_1,X_2) = 0$ so that

$$K(X_1,X_2) = \frac{R(X_1,X_2,X_2,X_1)}{g(X_1,X_1)g(X_2,X_2) - g(X_1,X_2)^2} = \frac{\sin^2\theta}{\sin^2\theta} = 1$$

so $K(T_p\mathcal{S}^2) = 1$ for all $p \in \mathcal{S}^2$.

Example. We see that on $\mathcal{S}^3$, $K(E_i,E_j) = 1$ for $i \neq j$, so all of the sectional curvatures are 1.

Example. For $\mathcal{H}^2$ with the hyperbolic metric g, we see that $T_p\mathcal{H}^2 = \mathrm{Span}\{X_1,X_2\}$ where $g(X_1,X_1) = 1$ and $g(X_2,X_2) = \sinh^2\theta$ and $g(X_1,X_2) = 0$ so that

$$K(X_1,X_2) = \frac{R(X_1,X_2,X_2,X_1)}{g(X_1,X_1)g(X_2,X_2) - g(X_1,X_2)^2} = \frac{-\sinh^2\theta}{\sinh^2\theta} = -1$$

so $K(T_p\mathcal{H}^2) = -1$ for all $p \in \mathcal{H}^2$.

Since we have an isometry from $(\mathcal{H}^2, g)$ to the hyperbolic upper half-plane (H^2, h), we deduce that the hyperbolic upper half-plane has constant sectional curvature -1. We can also see this explicitly using our earlier formula for the Christoffel symbols on (H^2, h)

You have probably seen the symbol K being used for curvature before when studying surfaces in $\mathbb{R}^3$, and this is no coincidence.

Proposition 4.7. *Let M be an oriented surface in $\mathbb{R}^3$. Then the sectional curvature $K(T_pM) = K(p)$, the Gaussian curvature of M at p.*

Proof. **(Not examinable).** The result holds by the Christoffel symbol formula for the curvature given in Proposition 4.4. □

Example. You have probably seen in a earlier course that the Gaussian curvature of $T^2 \in \mathbb{R}^3$ at a point $p = ((2+\cos\theta)\cos\phi, (2+\cos\theta)\sin\phi, \sin\theta)$ is given by

$$K(p) = \frac{\cos\theta}{2+\cos\theta},$$

which is certainly not constant.

We can confirm this since we know that when $X_1 = f_*\partial_\theta$ and $X_2 = f_*\partial_\phi$, then $[X_1, X_2] = 0$ and

$$\nabla_{X_1}X_1 = 0, \quad \nabla_{X_1}X_2 = \nabla_{X_2}X_1 = -\frac{\sin\theta}{2+\cos\theta}X_2, \quad \nabla_{X_2}X_2 = (2+\cos\theta)\sin\theta X_1.$$

Hence,

$$\begin{aligned}
R(X_1, X_2)X_1 &= \nabla_{X_1}\nabla_{X_2}X_1 - \nabla_{X_2}\nabla_{X_1}X_1 \\
&= \nabla_{X_1}(-\frac{\sin\theta}{2+\cos\theta}X_2) \\
&= -\frac{2\cos\theta+1}{(2+\cos\theta)^2}X_2 + \frac{\sin^2\theta}{(2+\cos\theta)^2}X_2 \\
&= \frac{-2\cos\theta - \cos^2\theta}{(2+\cos\theta)^2}X_2 \\
&= -\frac{\cos\theta}{2+\cos\theta}X_2.
\end{aligned}$$

Therefore,

$$R(X_1, X_2, X_2, X_1) = -R(X_1, X_2, X_1, X_2) = \cos\theta(2+\cos\theta)$$

Hence,

$$K(X_1, X_2) = \frac{R(X_1, X_2, X_2, X_1)}{g(X_1, X_1)g(X_2, X_2) - g(X_1, X_2)^2} = \frac{\cos\theta}{2+\cos\theta}.$$

Hence the torus in $\mathbb{R}^3$ is not flat and not isometric to the 2-torus in $\mathbb{R}^4$.

Notice on the inner circle ($\cos\theta = -1$) we see that $K < 0$ and on the outer circle ($\cos\theta = 1$) we see that $K > 0$ and on the middle circles ($\cos\theta = 0$) we see that $K = 0$ (flat). This fits with the discussion we had right at the beginning of the course.

Remark. It is important to note that the sectional curvature is a local quantity and so is preserved by local isometries.

Example. Since the pseudosphere minus a circle is locally isometric to the hyperbolic upper half-plane, it has a metric with constant sectional curvature -1.

Example. Recall that $\mathbb{RP}^2$, the Möbius band and the Klein bottle obtain Riemannian metrics from $\mathcal{S}^2$, the cylinder and the torus in $\mathbb{R}^3$ respectively.

Since the projection map is a local isometry, we know that $\mathbb{RP}^2$ has constant sectional curvature 1, the Möbius band is flat (since the cylinder is flat) and the Klein bottle has areas of both positive and negative curvature like T^2 in $\mathbb{R}^3$.

Example. Recall that we defined an action of $\mathbb{Z}^n$ on $(\mathbb{R}^n, g_0)$ by isometries inducing a metric g on $\mathbb{R}^n/Z^n$ such that $(\mathbb{R}^n/Z^n, g)$ is isometric to (T^n, h) where $T^n \subseteq \mathbb{R}^{2n}$ has the induced metric h.

The metric g must be flat since $\pi : \mathbb{R}^n \to \mathbb{R}^n/\mathbb{Z}^n$ is a local isometry. This reconfirms that $\mathbb{R}^n/\mathbb{Z}^n$ is flat and shows that (T^n, h) is flat. In particular, $(\mathbb{R}^2/\mathbb{Z}^2, g)$ is *not* isometric to $T^2 \subseteq \mathbb{R}^3$ with its induced metric, as we know.

4.3 Ricci and scalar curvature

We now introduce further curvature quantities which are kinds of "average" curvatures. They play a crucial role in geometry and mathematical physics.

Definition 4.8. We define the Ricci curvature tensor $\mathrm{Ric} \in \Gamma(S^2T^*M)$ by

$$\mathrm{Ric}(X,Y) = \sum_{i=1}^n R(E_i, X, Y, E_i)$$

for all $p \in M$, $X, Y \in T_pM$, where $\{E_1, \ldots, E_n\}$ is an orthonormal frame for T_pM. Notice that

$$\mathrm{Ric}(Y,X) = \sum_{i=1}^n R(E_i, Y, X, E_i) = \sum_{i=1}^n R(X, E_i, E_i, Y) = \sum_{i=1}^n R(E_i, X, Y, E_i) = \mathrm{Ric}(X,Y)$$

so it is symmetric as claimed.

We can interpret the Ricci curvature as a trace as follows. Given vector fields X, Y, Z on M we have a map $Z \mapsto R(Z,X)Y$ which sends vector fields to vector fields. At each point this only depends on the value of X, Y, Z at p, so gives a well-defined map from T_pM to T_pM, and thus may be viewed as a matrix once we choose a basis for T_pM. The Ricci curvature is then given by

$$\mathrm{Ric}(X,Y) = \mathrm{tr}\big(Z \mapsto R(Z,X)Y\big).$$

Notice that this does not depend on the choice of basis for T_pM just by linear algebra (the trace is invariant under coordinate transformations).

Let us try to understand the Ricci curvature tensor locally.

If (U, φ) is a coordinate chart on M and $\{X_1, \ldots, X_n\}$ is the coordinate frame then we can write an orthonormal frame $\{E_1, \ldots, E_n\}$ on U as $E_k = \sum_{l=1}^n a_{lk}X_l$ for an invertible matrix of functions $A = (a_{ij})$. Notice that

$$\delta_{kl} = g(E_k, E_l) = g\left(\sum_{i=1}^n a_{ik}X_i, \sum_{j=1}^n a_{jk}X_j\right) = \sum_{i,j=1}^n a_{ik}g_{ij}a_{jk}$$

which is (in matrix notation) $A^{\mathrm{T}}gA = I$. Thus $g = (A^{\mathrm{T}})^{-1}A^{-1} = (AA^{\mathrm{T}})^{-1}$ so $g^{-1} = AA^{\mathrm{T}}$ and hence

$$\begin{aligned}\mathrm{Ric}(X_i, X_j) &= \sum_{k=1}^n R(X_i, E_k, E_k, X_j) = \sum_{k,l,m=1}^n R(X_i, a_{lk}X_l, a_{mk}X_m, X_j) \\ &= \sum_{k,l,m=1}^n R_{ilmj}a_{lk}a_{mk} = \sum_{l,m=1}^n R_{ilmj}g^{lm}\end{aligned}$$

because $g^{ij} = \sum_{k=1}^n a_{ik}a_{jk}$.

Remark. If we take geodesic normal coordinates and locally let Ω be the Riemannian volume form so that $(\varphi^{-1})^*\Omega = \sqrt{\det(g_{ij})}\Omega_0$ where Ω_0 is the standard volume form on $\mathbb{R}^n$, then

$$(\varphi^{-1})^*\Omega = (1 - \frac{1}{6}\sum_{i,j} R_{ij}x_ix_j + O(|x|^3))\Omega_0.$$

So the Ricci curvature measures the first difference between the Riemannian volume form and the Euclidean volume form.

Example. For a 2-dimensional Riemannian manifold M, take $\{E_1, E_2\}$ to be an orthonormal basis for T_pM. We see that

$$K(T_pM) = R(E_1, E_2, E_2, E_1) = \mathrm{Ric}(E_1, E_1) = \mathrm{Ric}(E_2, E_2).$$

Hence the sectional curvatures are given by the Ricci curvatures on unit vectors and hence the Ricci curvature defines the Riemann curvature tensor. In particular the Ricci curvatures are just the Gaussian curvature if M is a surface in $\mathbb{R}^3$.

Remark. The Ricci curvature also determines the Riemann curvature tensor in 3 dimensions, but in higher dimensions they are different.

Remark. Now the Ricci curvature tensor is a symmetric $(0,2)$-tensor, and the same is true of the Riemannian metric so we can compare them. We say that (M,g) is *Einstein* if $\mathrm{Ric} = \lambda g$ for some constant λ. The equation $\mathrm{Ric} = \lambda g$ is the Riemannian version of Einstein's field equations from General Relativity in the absence of matter.

Of particular geometric interest is the case where $\lambda = 0$, where M is *Ricci-flat*. Notice that Ricci-flat is definitely not the same thing as being flat!

Definition 4.9. The *scalar curvature* S of M is a smooth function on M given by

$$S(p) = \sum_{i,j=1}^{n} R(E_i, E_j, E_j, E_i) = \sum_{j=1}^{n} \mathrm{Ric}(E_j, E_j)$$

for $p \in M$ where $\{E_1, \ldots, E_n\}$ is an orthonormal frame for T_pM. We can view this as a "trace" of the Ricci curvature tensor.

Remark. We see that for $\epsilon > 0$ small and $p \in M$ that we can relate the volume of the geodesic ball $B_\epsilon(p)$ with that of the Euclidean ball $B_\epsilon(0)$ by

$$\mathrm{vol}(B_\epsilon(p)) = \left(1 - \frac{S}{6(n+2)}\epsilon^2 + O(\epsilon^4)\right) \mathrm{vol}(B_\epsilon(0)),$$

so the scalar curvature measures the differences in these volumes.

If $\{E_1, \ldots, E_n\}$ is an orthonormal frame on a chart (U, φ), then by our above calculation we can write

$$S = \sum_{i,j=1}^{n} R(E_i, E_j, E_j, E_i) = \sum_{i,j,k,l=1}^{n} R_{ijkl} g^{il} g^{jk} = \sum_{i,j=1}^{n} \mathrm{Ric}(X_i, X_j) g^{ij}.$$

Scalar curvature is a rather weak invariant of (M,g). In particular, being scalar flat ($S = 0$) definitely does not mean that the manifold is flat. It is now known by the solution of the *Yamabe problem* that all manifolds admit Riemannian metrics with constant scalar curvature, and such metrics still form an active research area.

Example. For oriented surfaces M in $\mathbb{R}^3$, we see that for $p \in M$

$$K(p) = K(T_pM) = \mathrm{Ric}(E_1, E_1) = \mathrm{Ric}(E_2, E_2),$$

where $\{E_1, E_2\}$ is an orthonormal basis for T_pM and thus

$$S(p) = \mathrm{Ric}(E_1, E_1) + \mathrm{Ric}(E_2, E_2) = 2K(p),$$

so the scalar curvature is just twice the sectional (or Gaussian) curvature.

5 Riemannian submanifolds

The aim of this section is to extend many of the ideas from surfaces in $\mathbb{R}^3$ to a more general setting. To do this we consider Riemannian submanifolds.

For this section we let $\iota : (M, g_M) \to (N, g)$, where (M, g_M) is an n-dimensional Riemannian manifold and N is an $(n+m)$-dimensional Riemannian manifold, be an isometric embedding, i.e. ι is an injective immersion such that $\iota : M \to \iota(M)$ is a homemorphism and $\iota^* g = g_M$.

We are assuming that ι is an embedding so that M is an (embedded) submanifold of N with the induced metric, but much of what we can say can be generalised to when ι is simply an immersion. We identify M with $\iota(M)$.

We also denote the Levi-Civita connections on M and N by ∇^M and ∇^N.

Remark. Typically we will be given M as a submanifold in (N, g) and ι will be the inclusion map.

5.1 Tangential and normal vector fields

Definition 5.1. For $p \in M$ we define the *normal space* $(T_pM)^\perp$ to M at p (in N) to be

$$(T_pM)^\perp = \{X \in T_pN : g_p(X, Y) = 0 \;\forall Y \in T_pM\}.$$

We can then write $T_pN = T_pM \oplus (T_pM)^\perp$ and therefore any $X \in T_pN$ can be written uniquely as $X = X^{\mathrm{T}} + X^\perp$ where $X^{\mathrm{T}} \in T_pM$ and $X^\perp \in (T_pM)^\perp$.

Remark. The *normal bundle* $(TM)^\perp$ of M (in N) is given by

$$(TM)^\perp = \bigcup_{p\in M} (T_pM)^\perp$$

and is a rank m vector bundle over M.

Example. We can view the round n-sphere $(\mathcal{S}^n, g)$ as a Riemannian submanifold of $(\mathbb{R}^{n+1}, g_0)$. For $x \in \mathcal{S}^n$, since $T_x\mathcal{S}^n = (\mathrm{Span}\{x\})^\perp$, we then have that the normal space at x is

$$(T_x\mathcal{S}^n)^\perp = \mathrm{Span}\{x\}.$$

(Notice that the normal bundle has rank 1 as we would expect.)

Definition 5.2. A vector field X along (or on) M is an assignment of $X(p) \in T_pN$ for all $p \in M$ such that the map $p \mapsto X(p)$ for $p \in M$ is smooth.

We can uniquely write a vector field X along M as $X = X^{\mathrm{T}} + X^\perp$ where $X^{\mathrm{T}}(p) \in T_pM$ and $X^\perp(p) \in (T_pM)^\perp$ for all $p \in M$. Note that the projection maps $X \mapsto X^{\mathrm{T}}$ and $X \mapsto X^\perp$ are smooth.

We call X^{T} a *tangent vector field* on M and $X^\perp$ a *normal vector field* on M, and X^{T} and $X^\perp$ the tangential and normal components of X.

Remark. A normal vector field on M is nothing other than a section of the normal bundle $(TM)^\perp$.

Example. If $M = \mathbb{R}^n \subseteq N = \mathbb{R}^{n+m}$ and we take the Euclidean metric on N (and hence the Euclidean metric on M), then we see that the tangent vector fields along M are spanned by $\{\partial_1, \ldots, \partial_n\}$ and the normal vector fields along M are spanned by $\{\partial_{n+1}, \ldots, \partial_{n+m}\}$.

Example. A normal vector field $X^\perp$ on $\mathcal{S}^n \subseteq (\mathbb{R}^{n+1}, g_0)$ must satisfy

$$X^\perp(x) = f(x)x$$

for some smooth function $f : \mathcal{S}^n \to \mathbb{R}$.

Equivalently, if we use polar coordinates on $\mathbb{R}^{n+1} \setminus \{0\}$ and let r denote distance from 0 then

$$X^\perp = fr\partial_r$$

5.2 Second fundamental form

Suppose X and Y are tangent vector fields on M. Choose any extensions $\bar{X}$ and $\bar{Y}$ of X to Y to vector fields on N. We can then calculate

$$\nabla^M_X Y = (\nabla^N_{\bar{X}} \bar{Y})^{\mathrm{T}}$$

(The fact that $(\nabla^N_{\bar{X}} \bar{Y})^{\mathrm{T}}$ is independent of the choice of extensions $\bar{X}$ and $\bar{Y}$ of X and Y is fairly obvious, but we shall not prove it.) Since $\bar{X}$ and $\bar{Y}$ are arbitrary we simply write $\bar{X} = X$ and $\bar{Y} = Y$ for simplicity.

The fact that we have tangential and normal components for the covariant derivative of tangent vector fields on M leads us to the following natural definition.

Definition 5.3. The *second fundamental form* B of M (in N) is defined for tangent vector fields X, Y on M by

$$B(X,Y) = \nabla^N_X Y - \nabla^M_X Y = (\nabla^N_X Y)^{\perp}.$$

Thus $B(X,Y)(p) \in (T_pM)^{\perp}$ for all $p \in M$. In fact, $B(X,Y)(p)$ only depends on $X(p), Y(p) \in T_pM$.

The second fundamental form helps us to understand how M "sits inside" N.

Remark. This definition of the second fundamental form extends the one you will have seen for surfaces in $\mathbb{R}^3$.

Example. If we let $M = \mathbb{R}^n \subseteq (N,g) = (\mathbb{R}^{n+m}, g_0)$ as before, we see that

$$\nabla^N_{\partial_i} \partial_j = 0 = \nabla^M_{\partial_i} \partial_j.$$

We deduce that $\nabla^N_X Y = \nabla^M_X Y$ for all tangent vector fields X, Y along M. Hence the second fundamental form of M is $B = 0$ in this case.

We now examine some of the properties of B.

Proposition 5.4. *Let X, Y, Z be tangent vector fields on M and a, b be smooth functions on M. The second fundamental form satisfies*

$$B(Y,X) = B(X,Y) \quad \text{and} \quad B(aX + bY, Z) = aB(X,Z) + bB(Y,Z).$$

Hence, B is a quadratic form on tangent vector fields with values in normal vector fields.

Proof. Using property (v) of the Levi-Civita connection, we see that

$$B(X,Y) - B(Y,X) = \nabla^N_X Y - \nabla^M_X Y - \nabla^N_Y X + \nabla^M_Y X = [X,Y] + [Y,X] = 0,$$

since the Lie bracket is skew-symmetric.

Clearly $B(X+Y, Z) = B(X,Z) + B(Y,Z)$ and then linearity in the second entry follows by symmetry. We can easily calculate

$$\begin{aligned} B(aX, bY) &= \nabla^N_{aX}(bY) - \nabla^M_{aX}(bY) = a(\nabla^N_X(bY) - \nabla^M_X(bY)) \\ &= a(X(b)Y + b\nabla^N_X Y - X(b)Y - b\nabla^M_X Y) = abB(X,Y). \end{aligned}$$

This completes the proof. □

Remark. One can use the ideas from the proof of the previous proposition to show that $B(X,Y)(p)$ only depends on $X(p)$, $Y(p)$ and so B can be viewed as a symmetric 2-tensor on M with values in the normal bundle.

Example. Let us compute the second fundamental form of $\mathcal{S}^3$ in $(\mathbb{R}^4, g_0)$ from the definition, so $\mathcal{S}^3$ has the round metric. Recall the vector fields

$$\begin{aligned} E_1 &= -x_1\partial_0 + x_0\partial_1 - x_3\partial_2 + x_2\partial_3 \\ E_2 &= -x_2\partial_0 + x_3\partial_1 + x_0\partial_2 - x_1\partial_3 \\ E_3 &= -x_3\partial_0 - x_2\partial_1 + x_1\partial_2 + x_0\partial_3 \end{aligned}$$

on $\mathbb{R}^4$ that restrict to give orthonormal tangent vector fields on $\mathcal{S}^3$. We also have the vector field

$$E_0 = x_0\partial_0 + x_1\partial_1 + x_2\partial_2 + x_3\partial_3$$

which is orthogonal to E_1, E_2, E_3 and is unit length when restricted to $\mathcal{S}^3$. (Note that $E_0 = r\partial_r$ if r is the distance from 0 in $\mathbb{R}^4$.) Thus E_0 is a unit normal vector field along $\mathcal{S}^3$.

We already computed $\nabla^{\mathcal{S}^3}_{E_i} E_j$ for $i, j \in \{1, 2, 3\}$ so we just need to calculate $\nabla^{\mathcal{R}^4}_{E_i} E_j$ for $i, j = 1, 2, 3$. We see that

$$\begin{aligned} \nabla^{\mathbb{R}^4}_{E_1} E_1 &= \nabla_{-x_1\partial_0 + x_0\partial_1 - x_3\partial_2 + x_2\partial_3}(-x_1\partial_0 + x_0\partial_1 - x_3\partial_2 + x_2\partial_3) \\ &= (-x_1\nabla_{\partial_0} + x_0\nabla_{\partial_1} - x_3\nabla_{\partial_2} + x_2\nabla_{\partial_3})(-x_1\partial_0 + x_0\partial_1 - x_3\partial_2 + x_2\partial_3) \\ &= -x_0\partial_0 - x_1\partial_1 - x_2\partial_2 - x_3\partial_3 \\ &= -E_0. \end{aligned}$$

Similarly, $\nabla^{\mathbb{R}^4}_{E_2} E_2 = \nabla^{\mathbb{R}^4}_{E_3} E_3 = -E_0$. Therefore,

$$B(E_i, E_i) = -E_0 \quad \text{for } i = 1, 2, 3.$$

We similarly can calculate

$$\begin{aligned} \nabla^{\mathbb{R}^4}_{E_1} E_2 &= \nabla_{-x_1\partial_0 + x_0\partial_1 - x_3\partial_2 + x_2\partial_3}(-x_2\partial_0 + x_3\partial_1 - x_0\partial_2 - x_1\partial_3) \\ &= (-x_1\nabla_{\partial_0} + x_0\nabla_{\partial_1} - x_3\nabla_{\partial_2} + x_2\nabla_{\partial_3})(-x_2\partial_0 + x_3\partial_1 - x_0\partial_2 - x_1\partial_3) \\ &= x_3\partial_0 + x_2\partial_1 + x_1\partial_2 - x_0\partial_3 \\ &= -E_3 = \nabla^{\mathcal{S}^3}_{E_1} E_2. \end{aligned}$$

We similarly see that $\nabla^{\mathbb{R}^4}_{E_i} E_j = \nabla^{\mathcal{S}^3}_{E_i} E_j$ whenever $i \neq j$. Hence,

$$B(E_i, E_j) = 0 \quad \text{for } i \neq j.$$

Notice that we have the relation

$$B(E_i, E_j) = -g_0(E_i, E_j)E_0 = -\delta_{ij}E_0.$$

Associated to the second fundamental form are a collection of operators called shape operators, which we now define.

Definition 5.5. For a normal vector field ξ on M we define the *shape operator* $S_\xi : \Gamma(TM) \to \Gamma(TM)$ by

$$g(S_\xi(X), Y) = g(B(X, Y), \xi)$$

for all tangent vector fields X, Y on M. Since B is symmetric we see that S_ξ is a self-adjoint operator on tangent vector fields, i.e.

$$g(S_\xi(X), Y) = g(X, S_\xi(Y))$$

for tangent vector fields X, Y on M.

Example. For $\mathcal{S}^3$ in $(\mathbb{R}^4, g_0)$ we gave a unit normal vector field E_0 along $\mathcal{S}^3$ and tangent vector fields along $\mathcal{S}^3$ so that

$$B(E_i, E_j) = -g_0(E_i, E_j)E_0.$$

We therefore see that the shape operator S_{E_0} must satisfy

$$g_0(S_{E_0}(E_i), E_j) = g_0(B(E_i, E_j), E_0) = -g_0(E_i, E_j).$$

Hence, $S_{E_0} = -\,\mathrm{id}$.

We now give an alternative way to calculate the shape operators which may sometimes prove useful.

Proposition 5.6. *If X and ξ are tangent and normal vector fields along M then*

$$S_\xi(X) = -(\nabla^N_X \xi)^{\mathrm{T}}.$$

Proof. We notice that

$$g(S_\xi(X), Y) = g(B(X,Y),\xi) = g(\nabla^N_X Y, \xi) = X(g(Y,\xi)) - g(Y, \nabla^N_X\xi) = -g((\nabla^N_X\xi)^{\mathrm{T}}, Y).$$

Here we used the fact that $g(\nabla^M_X Y, \xi) = 0$ and $g(Y,\xi) = 0$. □

5.3 The Fundamental Equations

To gain more of an understand of the geometry of M inside N we need to measure how "curved" the normal directions to M are, so we define the following, which is clearly analogous to the definition of the Riemann curvature.

Definition 5.7. We define the *normal connection* $\nabla^\perp$ on M by

$$\nabla^\perp_X\xi = (\nabla^N_X\xi)^\perp = \nabla^N_X\xi - (\nabla^N_X\xi)^{\mathrm{T}} = \nabla^N_X\xi + S_\xi(X)$$

for tangent vector fields X and normal vector fields ξ on M. Notice that $\nabla^\perp_X\xi$ is a normal vector field.

We then define the *normal curvature* $R^\perp$ of M by

$$R^\perp(X,Y)\xi = \nabla^\perp_X\nabla^\perp_Y\xi - \nabla^\perp_Y\nabla^\perp_X\xi - \nabla^\perp_{[X,Y]}\xi$$

for tangent vector fields X, Y and normal vector fields ξ on M. Notice that $R^\perp(X,Y)\xi$ is a normal vector field.

Remark. Just as for Riemann curvature, for each $p \in M$ we have that $\big(R^\perp(X,Y)\xi\big)(p)$ only depends on $X(p)$, $Y(p)$, $\xi(p)$.

Now that we have all of these various objects describing the geometry of submanifolds we can relate them all using the so-called *Fundamental Equations* of Gauss, Codazzi and Ricci.

Theorem 5.8 (The Fundamental Equations)**.** *Let X, Y, Z, W be tangent vector fields and let ξ, ζ be normal vector fields on M. Let R^M, R^N be Riemann curvature associated with ∇^M, ∇^N respectively.*

(a) (Gauss) $g(R^N(X,Y)Z, W) = g(R^M(X,Y)Z, W) + g(B(X,Z), B(Y,W)) - g(B(X,W), B(Y,Z))$.

(b) (Codazzi) $g(R^N(X,Y)Z, \xi) = g\big((\nabla^N_X B)(Y,Z), \xi\big) - g\big((\nabla^N_Y B)(X,Z), \xi\big)$ *where*

$$(\nabla^N_X B)(Y,Z) = \nabla^\perp_X(B(Y,Z)) - B(\nabla^M_X Y, Z) - B(Y, \nabla^M_X Z).$$

(c) (Ricci) $g(R^N(X,Y)\xi, \zeta) = g(R^\perp(X,Y)\xi, \zeta) - g([S_\xi, S_\zeta]X, Y)$, *where* $[S_\xi, S_\zeta] = S_\xi \circ S_\zeta - S_\zeta \circ S_\xi$.

Proof. Recall that

$$\nabla^N_X Y = (\nabla^N_X Y)^{\mathrm{T}} + (\nabla^N_X Y)^\perp = \nabla^M_X Y + B(X,Y),$$
$$\nabla^N_X \xi = (\nabla^N_X \xi)^{\mathrm{T}} + (\nabla^N_X \xi)^\perp = -S_\xi X + \nabla^\perp_X \xi.$$

Using this, we observe that

$$\begin{aligned}R^N(X,Y)Z &= \nabla^N_X\nabla^N_Y Z - \nabla^N_Y\nabla^N_X Z - \nabla^N_{[X,Y]}Z\\ &= \nabla^N_X(\nabla^M_Y Z + B(Y,Z)) - \nabla^N_Y(\nabla^M_X Z + B(X,Z)) - \nabla^M_{[X,Y]}Z - B([X,Y],Z)\\ &= (\nabla^M_X\nabla^M_Y - \nabla^M_Y\nabla^M_X - \nabla^M_{[X,Y]})Z + B(X,\nabla^M_Y Z) - S_{B(Y,Z)}X + \nabla^\perp_X B(Y,Z)\\ &\quad - B(Y,\nabla^M_X Z) + S_{B(X,Z)}Y - \nabla^\perp_Y B(X,Z) - B([X,Y],Z)\\ &= R^M(X,Y)Z + B(X,\nabla^M_Y Z) - S_{B(Y,Z)}X + \nabla^\perp_X B(Y,Z)\\ &\quad - B(Y,\nabla^M_X Z) + S_{B(X,Z)}Y - \nabla^\perp_Y B(X,Z) - B([X,Y],Z).\end{aligned}$$

Therefore,

$$g(R^N(X,Y)Z,W) = g(R^M(X,Y)Z,W) - g(S_{B(Y,Z)}X,W) + g(S_{B(X,Z)}Y,W),$$

from which (a) follows as $g(S_\xi Z,W) = g(B(Z,W),\xi)$ by definition.

Using the above calculation, we have that

$$\begin{aligned}g(R^N(X,Y)Z,\xi) &= g(B(X,\nabla^M_Y Z),\xi) - g(B(Y,\nabla^M_X Z),\xi) + g(\nabla^\perp_X B(Y,Z),\xi) - g(\nabla^\perp_Y B(X,Z),\xi)\\ &\quad - g(B([X,Y],Z),\xi)\\ &= g(\nabla^\perp_X B(Y,Z) - B(\nabla^M_X Y,Z) - B(Y,\nabla^M_X Z),\xi)\\ &\quad - g(\nabla^\perp_Y B(X,Z) - B(\nabla^M_Y X,Z) - B(X,\nabla^M_Y Z),\xi)\\ &= g(\nabla^N_X B(Y,Z),\xi) - g(\nabla^N_Y B(X,Z),\xi),\end{aligned}$$

using the fact that $[X,Y] = \nabla^M_X Y - \nabla^M_Y X$, from which (b) follows.

For (c), we calculate

$$\begin{aligned}R^N(X,Y)\xi &= \nabla^N_X\nabla^N_Y\xi - \nabla^N_Y\nabla^N_X\xi - \nabla^N_{[X,Y]}\xi\\ &= \nabla^N_X(-S_\xi Y + \nabla^\perp_Y\xi) - \nabla^N_Y(-S_\xi X + \nabla^\perp_X\xi) + S_\xi([X,Y]) - \nabla^\perp_{[X,Y]}\xi\\ &= -\nabla^M_X(S_\xi(Y)) - B(S_\xi Y,X) - S_{\nabla^\perp_Y\xi}X + \nabla^M_Y(S_\xi X) + B(S_\xi X,Y) + S_{\nabla^\perp_X\xi}Y + S_\xi[X,Y]\\ &\quad + (\nabla^\perp_X\nabla^\perp_Y - \nabla^\perp_Y\nabla^\perp_X - \nabla^\perp_{[X,Y]})\xi.\end{aligned}$$

Therefore,

$$\begin{aligned}g(R^N(X,Y)\xi,\zeta) &= g(R^\perp(X,Y)\xi,\zeta) - g(B(S_\xi Y,X),\zeta) + g(B(S_\xi X,Y),\zeta)\\ &= g(R^\perp(X,Y)\xi,\zeta) - g(S_\zeta\circ S_\xi Y,X) + g(S_\zeta\circ S_\xi X,Y),\end{aligned}$$

from which (c) follows by the self-adjointness of the shape operator. □

5.4 Hypersurfaces

We now see how to relate the theory of surfaces in $\mathbb{R}^3$ to more general Riemannian Geometry.

Definition 5.9. An n-dimensional submanifold of an $(n+1)$-dimensional manifold N is called a *hypersurface* in N.

For this subsection we let $m = 1$, i.e. M is a hypersurface in (N, g) with the induced metric. Furthermore, we suppose that M and N are both oriented. In this situation we can define the principal curvatures as for surfaces in $\mathbb{R}^3$.

Definition 5.10. Let $p \in M$ and let $\{E_1, \dots, E_n\}$ be any positively oriented orthonormal basis of T_pM. We can then define a unit normal vector $\nu(p) \in (T_pM)^\perp$ uniquely by requiring that $\{E_1, \dots, E_n, \nu(p)\}$ be a positively oriented orthonormal basis for T_pN. This then defines a unit normal vector field ν on M.

For each $p \in M$ the shape operator defines a self-adjoint map $S_\nu : T_pM \to T_pM$, so has a positively oriented orthonormal basis of eigenvectors, which we also denote $\{E_1, \ldots, E_n\}$, i.e.

$$S_\nu E_i = \lambda_i E_i.$$

We call λ_i the *principal curvatures* of M at p and E_i the *principal directions* at p.

We call the function K_M given by

$$K_M(p) = \prod_{i=1}^{n} \lambda_i(p)$$

the *Gaussian curvature* of M and H_M given by

$$H_M(p) = \frac{1}{n}\sum_{i=1}^{n} \lambda_i(p)$$

the *mean curvature* of M.

We are most interested in the case where $N = \mathbb{R}^{n+1}$.

Definition 5.11. If $(N, g) = (\mathbb{R}^{n+1}, g_0)$ the *Gauss map* of the oriented hypersurface M is given by $\nu : M \to \mathcal{S}^n$ where $\nu(p) \in (T_pM)^\perp \subseteq T_p\mathbb{R}^{n+1} = \mathbb{R}^{n+1}$ is unit for all $p \in M$.

Since $T_{\nu(p)}\mathcal{S}^n$ and T_pM are naturally parallel planes in $\mathbb{R}^{n+1}$, we can view $\mathrm{d}\nu_p : T_pM \to T_pM$ for all $p \in M$.

Remark. The Gauss map is the natural generalisation of the Gauss map for surfaces in $\mathbb{R}^3$.

One may define the second fundamental form of a surface in $\mathbb{R}^3$ using the derivative of the Gauss map. We see from the definition of the shape operator S_ν that it plays the role of the derivative of the Gauss map in the definition of the second fundamental form above. Therefore, the next result is no surprise.

Lemma 5.12. *Let $(N, g) = (\mathbb{R}^{n+1}, g_0)$ and let ν be the Gauss map on the oriented hypersurface M. Then $\mathrm{d}\nu_p = -S_{\nu(p)}$ for all $p \in M$.*

Proof. If $X \in T_pM$, then there exists a curve $\alpha : (-\epsilon, \epsilon) \to M$ such that $\alpha(0) = p$ and $\alpha'(0) = X$. Note that $g(\nu, \nu) = 1$ implies that

$$\alpha'(g(\nu, \nu)) = 2g(\nabla^{\mathbb{R}^{n+1}}_{\alpha'}\nu, \nu) = 0$$

and hence $\nabla^{\mathbb{R}^{n+1}}_{\alpha'}\nu = (\nabla^{\mathbb{R}^{n+1}}_{\alpha'}\nu)^{\mathrm{T}}$. We may therefore calculate

$$\mathrm{d}\nu_p(X) = \frac{\mathrm{d}}{\mathrm{d}t}(\nu \circ \alpha(t))|_{t=0} = (\nabla^{\mathbb{R}^{n+1}}_{\alpha'}\nu)|_{t=0} = (\nabla^{\mathbb{R}^{n+1}}_{\alpha'}\nu)^{\mathrm{T}}|_{t=0} = -S_\nu(\alpha')|_{t=0} = -S_{\nu(p)}(X)$$

using the fact that $\nu \circ \alpha : (-\epsilon, \epsilon) \to \mathcal{S}^n \subseteq \mathbb{R}^{n+1}$ and $\nabla^{\mathbb{R}^{n+1}}$ is just the usual derivative on $\mathbb{R}^{n+1}$. $\square$

Remark. Observe that $K_M(p) = \det(-\mathrm{d}\nu_p)$ and $H_M(p) = -\frac{1}{n}\mathrm{tr}(\mathrm{d}\nu_p)$, which extends the formulae for surfaces in $\mathbb{R}^3$.

We now want to think about the curvature of hypersurfaces. From the Gauss equation we have that, if $X, Y \in T_pM$ are orthonormal,

$$K^M(X, Y) - K^N(X, Y) = g(B(X, X), B(Y, Y)) - g(B(X, Y), B(X, Y)).$$

By definition of the principal curvatures, $S_\nu(E_i) = \lambda_i E_i$ so

$$g(B(E_i, E_j), \nu) = g(S_\nu E_i, E_j) = \lambda_i \delta_{ij}.$$

Therefore, we have that, for $i \neq j$,

$$K^M(E_i, E_j) - K^N(E_i, E_j) = \lambda_i \lambda_j. \tag{\dagger}$$

In particular, if M is a surface in $\mathbb{R}^3$, then $K^M(E_1, E_2) = \lambda_1\lambda_2 = K_M$, which is the Gaussian curvature of M. We deduce Gauss's Theorem Egregium: the Gaussian curvature of a surface in $\mathbb{R}^3$ is invariant under local isometries.

We also have the following neat result.

Proposition 5.13. *The sectional curvatures of the round metric on $\mathcal{S}^n$ are all* 1.

Proof. For the standard orientation on $\mathcal{S}^n$, the Gauss map $\nu(p) = p$ for $p \in \mathcal{S}^n$, so $\nu = \mathrm{id}$. Lemma 5.12 implies that $S_\nu = -\mathrm{id}$ (as we saw explicitly for $\mathcal{S}^3$), so all its eigenvalues are -1. Hence, $K = 1$ on $\mathcal{S}^n$ by (†) as $\mathbb{R}^{n+1}$ has constant sectional curvature 0. $\square$

5.5 Totally geodesic and minimal submanifolds

We conclude this section by returning to the general situation of a Riemannian submanifold, but now we focus on some special types of submanifold which occur.

To motivate the first class of submanifolds, consider a curve $\alpha : (-\epsilon, \epsilon) \to (N, g)$ such that $|\alpha'|$ is constant (which we can always assume by reparametrizing the curve). Then $M = \alpha(-\epsilon, \epsilon)$ is obviously a Riemannian submanifold of (N, g). We then see that

$$0 = \frac{\mathrm{d}}{\mathrm{d}t} g(\alpha', \alpha') = 2g(\nabla^N_{\alpha'}\alpha', \alpha').$$

Since α' spans the space of tangent vector fields along α, we deduce that

$$\nabla^N_{\alpha'}\alpha' = (\nabla^N_{\alpha'}\alpha')^\perp = B(\alpha', \alpha'),$$

where B is the second fundamental form of $M = \alpha(-\epsilon, \epsilon)$. Therefore, α is a geodesic if and only if $B = 0$.

We can generalize this idea as follows.

Definition 5.14. We say that M is *geodesic at* p if $B = 0$ at p. We say that M is *totally geodesic* if $B = 0$ everywhere.

Example. We saw that $M = \mathbb{R}^n \subseteq (\mathbb{R}^{n+m}, g_0)$ is totally geodesic.

The point of this definition is given by the following result.

Proposition 5.15. *M is geodesic at p if and only if every geodesic in M starting from p is a geodesic in N at p.*

Proof. Let γ be a geodesic in M and let $X = \gamma'$. Then $\nabla^M_X X = 0$ implies that $(\nabla^N_X X)^T = \nabla^M_X X = 0$. Therefore $B(X, X) = (\nabla^N_X X)^\perp = 0$ at $t = 0$ if and only if $\nabla^N_X X = 0$ at $t = 0$. Therefore, since every element in T_pM can be realised as the tangent vector to some geodesic starting at p, the result follows. $\square$

This result gives us one of the nicest ways to think about sectional curvature, as follows. Let V be a normal neighbourhood of $p \in N$ and let $V = \exp_p(U)$. Let σ be a 2-plane in T_pN. Then $M = \exp_p(\sigma \cap U)$ is a 2-dimensional submanifold of N with $p \in M$. Proposition 5.15 implies that M is geodesic at p, so $B = 0$ at p. By the Gauss equation we have that $K^M(\sigma) = K^N(\sigma)$ and we know that the sectional curvature $K^M(\sigma)$ is the same as the Gaussian curvature of M at p. We conclude that the sectional curvature $K(\sigma)$ of N at p is the Gaussian curvature of a small surface in N containing p.

Example. Suppose $\Pi \subseteq \mathbb{R}^{n+1}$ is a $(k+1)$-plane through 0. Then the k-sphere $\mathcal{S}^k \cong \Pi \cap \mathcal{S}^n$ is totally geodesic in the round n-sphere $(\mathcal{S}^n, g)$, since every geodesic in the k-sphere $\Pi \cap \mathcal{S}^n$ is clearly a geodesic in $(\mathcal{S}^n, g)$.

Remark. Being totally geodesic is obviously a strong condition. In particular, a result due to Cartan states that if for every $p \in M$ and for every 2-plane $\sigma \subseteq T_pM$ there exists a totally geodesic submanifold $M \subseteq N$ such that $p \in M$ and $T_pM = \sigma$ then N must have constant sectional curvature.

Definition 5.16. We say that M is a *minimal* submanifold if for all $p \in M$ we have $\sum_{i=1}^n B(E_i, E_i) = 0$ for an orthonormal frame $\{E_1, \ldots, E_n\}$ of T_pM. Equivalently, M is minimal if $\mathrm{tr}(S_\xi) = 0$ for all normal vector fields ξ on M.

Minimal submanifolds are stationary points for the area functional (like geodesics are stationary points for the length functional) so can locally minimize area. However, since they are just stationary points they could be local maxima as well, for example!

Example. A k-plane in $\mathbb{R}^{n+1}$ is minimal since its second fundamental form is $B = 0$. We can see this geometrically as any perturbation of any bounded portion of a 2-plane in $\mathbb{R}^3$ say clearly has larger area.

Example. Any totally geodesic submanifold is minimal.

Example. The sphere $\mathcal{S}^n$ in $\mathbb{R}^{n+1}$ is not minimal since we can simply make the sphere slightly smaller or larger and it will have smaller or larger area. Alternatively, we see that $S_\nu = -\,\mathrm{id}$ so its trace is clearly non-zero.

Example. The catenoid and helicoid we saw earlier are minimal surfaces in $\mathbb{R}^3$.

Example. If $f : \mathbb{C} \to \mathbb{C}$ is holomorphic then $\mathrm{Graph}(f) = \{(z, f(z)) \,:\, z \in \mathbb{C}\}$ is minimal in $\mathbb{C}^2 = \mathbb{R}^4$.

Minimal submanifolds form a significant part of modern research in Riemannian Geometry.

6 Jacobi fields

In this section we define certain vector fields along geodesics which enable us to relate our fundamental objects in this course; that is, geodesics and curvature. The tools we develop in this section will be invaluable for later parts of the course.

As usual we assume that (M, g) is an n-dimensional Riemannian manifold.

We begin with a technical result which is, in some sense, an improved version of the Symmetry Lemma (Lemma 3.11).

Proposition 6.1. *Recall the notation of Lemma 3.11. In particular we have a smooth map $f : A \subseteq \mathbb{R}^2 \to M$ with coordinates (u, v) on A. We say that X is a vector field along f if X assigns a tangent vector $X(f(u, v)) \in T_{f(u,v)}M$ for all $(u, v) \in A$ such that the map $(u, v) \mapsto X(f(u, v))$ is smooth. Then*

$$\left(\nabla_{\frac{\partial f}{\partial u}}\nabla_{\frac{\partial f}{\partial v}} - \nabla_{\frac{\partial f}{\partial v}}\nabla_{\frac{\partial f}{\partial u}}\right) X = R(\frac{\partial f}{\partial u}, \frac{\partial f}{\partial v})X.$$

Proof. In chart (U, φ) write $X = \sum_i a_i X_i$ with $\{X_1, \ldots, X_n\}$ coordinate frame field. Then

$$\nabla_{\frac{\partial f}{\partial u}}\nabla_{\frac{\partial f}{\partial v}} X = \sum_i a_i \nabla_{\frac{\partial f}{\partial u}}\nabla_{\frac{\partial f}{\partial v}} X_i + \sum_i \frac{\partial a_i}{\partial v}\nabla_{\frac{\partial f}{\partial u}} X_i + \sum_i \frac{\partial a_i}{\partial u}\nabla_{\frac{\partial f}{\partial v}} X_i + \sum_i \frac{\partial^2 a_i}{\partial u \partial v} X_i.$$

Thus, by the symmetry in the last three terms in the above equation in u, v we see that

$$\left(\nabla_{\frac{\partial f}{\partial u}}\nabla_{\frac{\partial f}{\partial v}} - \nabla_{\frac{\partial f}{\partial v}}\nabla_{\frac{\partial f}{\partial u}}\right) X = \sum_i a_i \left(\nabla_{\frac{\partial f}{\partial u}}\nabla_{\frac{\partial f}{\partial v}} - \nabla_{\frac{\partial f}{\partial v}}\nabla_{\frac{\partial f}{\partial u}}\right) X_i.$$

Writing $f = (f_1, \ldots, f_n)$ we have that

$$\nabla_{\frac{\partial f}{\partial u}}\nabla_{\frac{\partial f}{\partial v}} X_i = \sum_j \frac{\partial^2 f_j}{\partial u \partial v}\nabla_{X_j} X_i + \sum_{j,k} \frac{\partial f_j}{\partial u}\frac{\partial f_k}{\partial v}\nabla_{X_j}\nabla_{X_k} X_i.$$

Again, by the symmetry in the first term in u, v in the above equation, we see that

$$\begin{aligned}\left(\nabla_{\frac{\partial f}{\partial u}}\nabla_{\frac{\partial f}{\partial v}} - \nabla_{\frac{\partial f}{\partial v}}\nabla_{\frac{\partial f}{\partial u}}\right) X_i &= \sum_{j,k} \frac{\partial f_j}{\partial u}\frac{\partial f_k}{\partial v}(\nabla_{X_j}\nabla_{X_k} - \nabla_{X_k}\nabla_{X_j})X_i \\ &= \sum_{j,k} \frac{\partial f_j}{\partial u}\frac{\partial f_k}{\partial v} R(X_j, X_k)X_i.\end{aligned}$$

The result follows by the linearity properties of the Riemann curvature tensor. □

6.1 The Jacobi equation

Recall that, for a vector field X along a curve α we write

$$X' = \nabla_{\alpha'} X.$$

Since X' is again a vector field along α we can write

$$X'' = \nabla_{\alpha'} X' = \nabla_{\alpha'}\nabla_{\alpha'} X.$$

Definition 6.2. Given a geodesic $\gamma : [0, L] \to (M, g)$, the *Jacobi equation* for a vector field J along γ is

$$J'' + R(J, \gamma')\gamma' = 0.$$

A solution J to the Jacobi equation is called a *Jacobi field* (along γ).

Our goal is to try to understand the space of Jacobi fields. We start with some simple examples.

Example. Clearly γ' is a Jacobi field along γ which is everywhere non-zero since $R(\gamma', \gamma') = 0$ and $(\gamma')'' = (\gamma'')' = 0$.

Example. Define a vector field $J(s) = s\gamma'(s)$ along a geodesic γ. Then

$$J'(s) = \gamma'(s) + s\gamma''(s) = \gamma'(s) \quad \text{and} \quad J'' = \gamma'' = 0.$$

Moreover, $R(J, \gamma') = 0$ as J is proportional to γ' at each point of γ. Hence J is also a Jacobi field along γ whose covariant derivative $J' = \gamma'$ along γ is everywhere non-zero.

Remark. On the basis of the previous two examples, we often consider J perpendicular to γ', and sometimes also with J' perpendicular to γ'.

Consider the map $f(s,t) = \exp_p(sX(t))$ for a curve $X(t)$ in T_pM with $X(0) = X$, $X'(0) = Y$ and $s \in [0,1]$ (and assume that $X(t)$ is chosen such that f is well-defined). Then

$$\mathrm{d}(\exp_p)_{sX}(sY) = \frac{\partial f}{\partial t}(s,0) = J(s)$$

is a vector field along the geodesic $\gamma(s) = \exp_p(sX)$ which measures how geodesics "spread" from γ. Moreover, notice that $J(0) = 0$ and zeros of J for $s > 0$ correspond to critical points of $\exp_p$, i.e. where $\exp_p$ fails to be an immersion.

Example. If X, Y are orthonormal vectors in $(\mathbb{R}^n, g_0)$ and $p = 0$, we can take $X(t) = X\cos t + Y \sin t$ and see that

$$\gamma(s) = \exp_p(sX(t)) = sX\cos t + sY\sin t$$

and

$$J(s) = -sX\sin t + sY\cos t|_{t=0} = sY,$$

so J grows linearly with s.

We will see now that J is a Jacobi field along γ.

Lemma 6.3. *Let $p \in (M,g)$, let $X(t)$ be a curve in T_pM with $X(0) = X$, $X'(0) = Y$ and let J be the vector field along $\gamma(s) = \exp_p(sX)$ given by*

$$J(s) = \mathrm{d}(\exp_p)_{sX}(sY).$$

Then J is a Jacobi field along γ.

Proof. Since $\gamma_t(s) = \exp_p(sX(t))$ is a geodesic we have, using the notation before the lemma,

$$\nabla_{\frac{\partial f}{\partial s}} \frac{\partial f}{\partial s} = 0.$$

Therefore, by Lemma 3.11 and Proposition 6.1,

$$\begin{aligned}
0 &= \nabla_{\frac{\partial f}{\partial t}} \left(\nabla_{\frac{\partial f}{\partial s}} \frac{\partial f}{\partial s} \right) \\
&= \nabla_{\frac{\partial f}{\partial s}} \left(\nabla_{\frac{\partial f}{\partial t}} \frac{\partial f}{\partial s} \right) - R(\frac{\partial f}{\partial s}, \frac{\partial f}{\partial t})\frac{\partial f}{\partial s} \\
&= \nabla_{\frac{\partial f}{\partial s}} \left(\nabla_{\frac{\partial f}{\partial s}} \frac{\partial f}{\partial t} \right) - R(\frac{\partial f}{\partial s}, \frac{\partial f}{\partial t})\frac{\partial f}{\partial s} \\
&= \nabla_{\frac{\partial f}{\partial s}} \nabla_{\frac{\partial f}{\partial s}} (\frac{\partial f}{\partial t}) + R(\frac{\partial f}{\partial t}, \frac{\partial f}{\partial s})\frac{\partial f}{\partial s}.
\end{aligned}$$

Evaluating this at $t = 0$ gives us that J is a Jacobi field. □

Lemma 6.4. *A Jacobi field J along a geodesic $\gamma : [0, L] \to (M,g)$ is uniquely determined by $J(0)$ and $J'(0)$. Hence, on an n-dimensional (M,g), there are $2n$ linearly independent Jacobi fields along γ.*

Proof. Observe that J is determined by $J(0)$ and $J'(0)$ as the Jacobi equation is a second order ODE.

In fact, if $\{E_1, \ldots, E_n\}$ are parallel orthonormal vector fields along γ, then we can write

$$J = \sum_{j=1}^{n} a_j E_j \quad \text{and} \quad b_{jk} = g(R(E_j, \gamma')\gamma', E_k).$$

Then the Jacobi equation becomes

$$a_k'' + \sum_{j=1}^{n} a_j b_{jk} = 0 \quad \text{for } k = 1, \ldots, n,$$

so there are $2n$ Jacobi fields defined by the $a_j(0)$ and $a_j'(0)$. $\square$

Example. Let M have constant sectional curvature K, let γ be a normalized geodesic and suppose that J is Jacobi field along γ with $|J| = 1$ so that $g(J, \gamma') = 0$. Then $R(J, \gamma')\gamma' = KJ$ as $\{J, \gamma'\}$ are orthonormal vector fields. Therefore the Jacobi equation is

$$J'' + KJ = 0.$$

Let X be a unit parallel vector field along γ such that $g(X, \gamma') = 0$. Then the Jacobi fields with $J(0) = 0$ and $J'(0) = X(0)$ are given by:

$$J(s) = \begin{cases} \lambda^{-1}\sin(\lambda s)X(s), & K = \lambda^2 > 0; \\ sX(s), & K = 0; \\ \lambda^{-1}\sinh(\lambda s)X(s), & K = -\lambda^2 < 0. \end{cases}$$

In particular, we see that if $K = 1$ then J has a zero at $s = \pi$ (think of the round 2-sphere for example), and if $K = -1$ then J is nowhere vanishing for $s > 0$ and grows exponentially rather than linearly as in the $K = 0$ case.

We now show that the earlier construction we gave for Jacobi fields is, in some sense, the only way to construct Jacobi fields.

Proposition 6.5. *Let $\gamma : [0, 1] \to M$ be a geodesic, let $p = \gamma(0)$ and let J be a Jacobi field along γ with $J(0) = 0$. Let $X(t)$ be a curve in T_pM with $X(0) = \gamma'(0)$ and $X'(0) = J'(0) \in T_pM$. If*

$$f(s, t) = \exp_p(sX(t)) \quad \textit{and} \quad \bar{J}(s) = \frac{\partial f}{\partial t}(s, 0)$$

then $J = \bar{J}$.

Proof. We first note that

$$f(s, 0) = \exp_p(sX(0)) = \exp_p(s\gamma'(0)),$$

so the geodesic $\bar{\gamma}(s) = f(s, 0)$ satisfies $\bar{\gamma}(0) = p = \gamma(0)$ and $\bar{\gamma}'(0) = \gamma'(0)$, so $\bar{\gamma} = \gamma$. We also see that

$$\bar{J}(s) = \mathrm{d}(\exp_p)_{sX(t)}(sX'(t))\,|_{t=0} = \mathrm{d}(\exp_p)_{s\gamma'(0)}(sJ'(0)) = s\mathrm{d}(\exp_p)_{s\gamma'(0)}\big(J'(0)\big).$$

Notice that $\bar{J}(0) = 0$ and that

$$\begin{aligned} \bar{J}'(s) &= \nabla_{\gamma'}\big(s\mathrm{d}(\exp_p)_{s\gamma'(0)}(J'(0))\big) \\ &= \mathrm{d}(\exp_p)_{s\gamma'(0)}(J'(0)) + s\nabla_{\gamma'}\big(\mathrm{d}(\exp_p)_{s\gamma'(0)}(J'(0))\big). \end{aligned}$$

Therefore,

$$\bar{J}'(0) = \mathrm{d}(\exp_p)_0(J'(0)) = J'(0) \quad \text{and} \quad \bar{J}(0) = 0 = J(0),$$

so $J = \bar{J}$. $\square$

The real utility of this result is given by the following corollary, which is immediately obvious.

Corollary 6.6. *If J is a Jacobi field with $J(0) = 0$ along a geodesic $\gamma : [0, L] \to (M, g)$ with $\gamma(0) = p$ then*

$$J(s) = \mathrm{d}(\exp_p)_{s\gamma'(0)}(sJ'(0)).$$

We now make an important observation concerning Jacobi fields which reveals the relation between Jacobi fields and curvature.

Proposition 6.7. *Let $\gamma : [0, L] \to (M, g)$ be a geodesic and let J be a Jacobi field along γ with $J(0) = 0$ and $|J'(0)| = 1$. Then*

$$|J(s)|^2 = s^2 - \frac{1}{3}R(\gamma'(0), J'(0), J'(0), \gamma'(0))s^4 + o(s^4)$$

for s near 0. Hence, if γ is a normalized geodesic and $g(J'(0), \gamma'(0)) = 0$, then

$$|J(s)|^2 = s^2 - \frac{1}{3}K(\gamma'(0), J'(0))s^4 + o(s^4)$$

for s near 0.

Proof. We are computing the first few terms of the Taylor expansion of the function

$$f(s) = |J(s)|^2 = g(J(s), J(s)).$$

Since $J(0) = 0$, we see that $f(0) = 0$. We may then see that

$$f'(0) = 2g(J(0), J'(0)) = 0$$

as well. We may also see that

$$\frac{f''(0)}{2} = g(J(0), J''(0)) + g(J'(0), J'(0)) = |J'(0)|^2 = 1$$

by assumption, which then gives us the s^2 term in the expansion.

For the possible s^3 term we have that

$$\frac{f'''(0)}{3!} = \frac{1}{3}g(J(0), J'''(0)) + g(J'(0), J''(0)) = g(J'(0), J''(0)).$$

Now we can use the Jacobi equation $J'' + R(J, \gamma')\gamma' = 0$ to deduce that $J''(0) = 0$ as $J(0) = 0$. Hence, there is no s^3 term in the Taylor expansion.

We now turn to the s^4 term. We see that this will be governed by

$$\frac{f''''(0)}{4!} = \frac{1}{12}g(J(0), J''''(0)) + \frac{1}{3}g(J'(0), J'''(0)) + \frac{1}{4}g(J''(0), J''(0)) = \frac{1}{3}g(J'(0), J'''(0)).$$

By the Jacobi equation, we know that

$$J''' = -\nabla_{\gamma'}\big(R(J, \gamma')\gamma'\big).$$

To compute this we observe that, for any vector field X along γ, we have that

$$\begin{aligned} g(\nabla_{\gamma'}\big(R(J, \gamma')\gamma'\big), X) &= \nabla_{\gamma'}\big(g(R(J, \gamma')\gamma', X)\big) - g(R(J, \gamma')\gamma', \nabla_{\gamma'}X) \\ &= \nabla_{\gamma'}\big(g(R(X, \gamma')\gamma', J)\big) - g(R(J, \gamma')\gamma', X'), \end{aligned}$$

using the symmetries of the Riemann curvature tensor. Thus,

$$g(\nabla_{\gamma'}\big(R(J, \gamma')\gamma'\big), X) = g\big(\nabla'_{\gamma}\big(R(X, \gamma')\gamma'\big), J\big) + g(R(X, \gamma')\gamma', J') - g(R(J, \gamma')\gamma', X').$$

Setting $s = 0$ gives $J(0) = 0$, so the first and last terms on the right-hand side in the equation vanish. We deduce that, choosing $X = J'$,

$$\begin{aligned} g(J'''(0), J'(0)) &= -g(\nabla_{\gamma'}\big(R(J, \gamma')\gamma'\big), J')(0) \\ &= -g(R(J', \gamma')\gamma', J')(0). \end{aligned}$$

This gives the claimed Taylor expansion and the result after follows immediately from the definition of sectional curvature. □

As we shall see, this result gives a useful computational tool in some cases. In particular, when we know the geodesics in (M, g), we can use the Taylor expansion to compute the curvature of (M, g).

6.2 Conjugate points

We now continue our discussion of Jacobi fields by analysing so-called conjugate points. These will be very important for many of the results in later parts of the course.

Definition 6.8. Let $\gamma : [0, L] \to (M, g)$ be a geodesic. We say that $\gamma(T)$ is *conjugate* to $\gamma(0)$ (along γ) if $T \in (0, L]$ and there exists a Jacobi field $J \neq 0$ such that $J(0) = J(T) = 0$.

Given $\gamma(T)$ conjugate to $\gamma(0)$, the maximum number $\mathrm{mult}(\gamma(T))$ of linearly independent Jacobi fields vanishing at 0 and T is called the *multiplicity* of $\gamma(T)$.

Remark. There at most n Jacobi fields along γ with $J(0) = 0$, defined by $J'(0)$. However, $s\gamma'(s)$ is nowhere vanishing except at 0, so $\mathrm{mult}(\gamma(T)) \leq n - 1$.

Example. On $(\mathbb{R}^n, g_0)$ given any point p and any geodesic γ with $\gamma(0) = p$, there can never be any conjugate points along γ. We can see this because the set of Jacobi fields vanishing at p along the straight line γ are spanned by $\{s\partial_1, \ldots, s\partial_n\}$, which are clearly all nowhere vanishing for $s > 0$.

Example. For $\mathcal{S}^n$ with the round metric g , let γ be a normalized geodesic with $\gamma(0) = p$. Recall the formulae for the Jacobi fields on $\mathcal{S}^n$ in our earlier example, using the fact that $(\mathcal{S}^n, g)$ has constant sectional curvature 1.

From these formulae we see that $\gamma(\pi) = -p$ is conjugate to p and $\mathrm{mult}(-p) = n - 1$. We also see that $p = \gamma(2\pi)$ is conjugate to p.

The examples suggest we should try to understand the set of conjugate points to $\gamma(0)$ for geodesics γ, and we should restrict ourselves to the *first* conjugate points.

Definition 6.9. The *conjugate locus* $C(p)$ of $p \in (M, g)$ is the set of first conjugate points for all geodesics from p.

Example. Clearly, for all points p in Euclidean space $(\mathbb{R}^n, g_0)$, the conjugate locus $C(p) = \emptyset$.

Example. We see explicitly that for all p in the hyperbolic upper half-plane (H^2, g), $C(p) = \emptyset$.

Example. If we take any point p in the round n-sphere $(\mathcal{S}^n, g)$, we see from our earlier example that $C(p) = \{-p\}$.

The point of the conjugate locus will become apparent later, though it is clear that there is a link to the curvature of the ambient manifold. For now, we make the following crucial observation.

Proposition 6.10. *Let $\gamma : [0, L] \to (M, g)$ be a geodesic with $\gamma(0) = p$. The point $\gamma(T)$ is conjugate to $\gamma(0) = p$ if and only if $T\gamma'(0)$ is a critical point of $\exp_p$. Moreover,*

$$\mathrm{mult}(\gamma(T)) = \dim \mathrm{Ker}(\mathrm{d}(\exp_p)_{T\gamma'(0)}).$$

Proof. Suppose J is a Jacobi field along γ such that $J \neq 0$ and $J(0) = J(T) = 0$. By Corollary 6.6 we have that

$$J(T) = \mathrm{d}(\exp_p)_{T\gamma'(0)}(TJ'(0)) = 0.$$

However, this equation is equivalent to saying that $T\gamma'(0)$ is a critical point of $\exp_p$ since $\gamma'(0) \neq 0$.

Now suppose $\mathrm{mult}(\gamma(T)) = k$, so we have linearly independent Jacobi fields $J_1, \ldots, J_k$ along γ such that $J_i(0) = J_i(T) = 0$. The linear independence of the J_i is equivalent to the linear independence of the $J_i'(0)$. Moreover, $J_i'(0)$ lies in $\mathrm{Ker}(\mathrm{d}(\exp_{\gamma(0)})_{T\gamma'(0)})$ for all i. The result follows. $\square$

We now derive some further facts concerning Jacobi fields.

Proposition 6.11. *Let J be a Jacobi field along a geodesic γ in (M, g). Then*

$$g(J, \gamma')(s) = sg(J', \gamma')(0) + g(J, \gamma')(0).$$

Proof. First, from the Jacobi equation (and the fact that γ is a geodesic),

$$\frac{\mathrm{d}}{\mathrm{d}s} g(J', \gamma') = g(J'', \gamma') = -g(R(J, \gamma')\gamma', \gamma') = 0.$$

Therefore, $g(J', \gamma') = g(J', \gamma')(0)$ and thus constant.

Since

$$\frac{\mathrm{d}}{\mathrm{d}s} g(J, \gamma') = g(J', \gamma') = g(J', \gamma')(0),$$

we deduce the formula as claimed by integrating. □

We deduce the following corollaries quickly from this proposition.

Corollary 6.12. *Let J be a Jacobi field along a geodesic $\gamma : [0, L] \to (M, g)$ such that $J(0) = J(L) = 0$. Then $g(J, \gamma') = 0$.*

Proof. Since $J(0) = 0$, we have that $g(J, \gamma')(s) = sg(J', \gamma')(0)$, but putting $s = L$ the left-hand side is zero, so $g(J', \gamma')(0) = 0$. □

Corollary 6.13. *Let J be a Jacobi field along a geodesic $\gamma : [0, L] \to (M, g)$ such that $J(0) = 0$.*

Then $g(J', \gamma')(0) = 0$ if and only if $g(J, \gamma') = 0$, so

$$\dim\{\textit{Jacobi fields } J \textit{ along } \gamma : J(0) = 0,\ g(J, \gamma') = 0\} = n - 1.$$

Proof. Since the condition $J(0) = 0$ imposes n conditions on J and the condition $g(J', \gamma')(0) = 0$ imposes one further condition on J, the result follows. □

7 Completeness

We now move on to another important notion in Riemannian geometry where geodesics play an essential role, namely *completeness*.

7.1 Definitions

We shall assume throughout that (M, g) is a connected Riemannian manifold.

Definition 7.1. (M, g) is *(geodesically) complete* if $\exp_p(X)$ is defined for all $X \in T_pM$ and for all $p \in M$.

Equivalently, normalised geodesics $\gamma_{(p,X)}(t) = \exp_p(tX)$ are defined for all $X \in T_pM$ with $|X| = 1$, for all $t \in \mathbb{R}$ and for all $p \in M$.

Let us see some examples of Riemannian manifolds which are both complete and not complete (we usually say *incomplete*).

Example. We see that $(\mathbb{R}^2, g_0)$ (and therefore $\mathbb{R}^n$) is complete because straight lines

$$\gamma(t) = (x_1 + ty_1, x_2 + ty_2)$$

are defined for all $t \in \mathbb{R}$ and any $y_1, y_2 \in \mathbb{R}$.

The same is obviously true for $T^n \subseteq \mathbb{R}^{2n}$, for example on $T^2 \subseteq \mathbb{R}^4$ geodesics are

$$\gamma(t) = \big(\cos(\theta_1 + a_1t), \sin(\theta_1 + a_1t), \cos(\theta_2 + a_2t), \sin(\theta_2 + a_2t)\big)$$

which are clearly defined for all $t \in \mathbb{R}$ and all $a_1, a_2 \in \mathbb{R}$.

Example. If we look at straight lines

$$\gamma(t) = (x_1, x_2 + t)$$

on H^2 we see γ is only defined for $t > -x_2$, and hence (H^2, g_0) is not complete.

The corresponding normalised geodesic on H^2 with the hyperbolic metric is

$$\gamma(t) = (x_1, x_2e^t)$$

which is now defined for all $t \in \mathbb{R}$. It actually follows from this and the isometries of the hyperbolic upper half-plane that H^2 with the hyperbolic metric is complete.

Example. On $\mathcal{S}^2$ (and therefore $\mathcal{S}^n$) with the round metric g normalised geodesics are great circles, for example

$$\gamma(t) = (\sin(t + \theta_0)\cos\phi_0, \sin(t + \theta_0)\sin\phi_0, \cos(t + \theta_0))$$

which is defined for all $t \in \mathbb{R}$, and so are certainly defined for all points and tangent vectors, hence $(\mathcal{S}^2, g)$ is complete.

However, if we remove a point from $\mathcal{S}^2$ (or $\mathcal{S}^n$), say the South pole, then the geodesics that passed through that point are now no longer defined for all $t \in \mathbb{R}$ (for example, normalised geodesics $\gamma(t)$ with $\gamma(0) = N$ are now only defined for $|t| < \pi$ in the usual parameterization since $\gamma(\pm\pi) = S$).

In fact, we see that if we take any Riemannian manifold and remove a point then it cannot be complete with the induced Riemannian metric.

Example. We see from the description of geodesics in $\mathbb{RP}^n$ that it is complete.

You will have come across the concept of completeness before in the study of metric spaces, so you may ask if the two concepts are related. The answer is yes, but first we need to understand how we should view (M, g) as a metric space in a way which is compatible with g.

Proposition 7.2. *If $p, q \in (M, g)$, define*

$$d(p,q) = \inf\{L(\alpha) : \alpha \text{ is a curve from } p \text{ to } q\}.$$

Then (M, d) is a metric space.

Proof. The metric balls $B^d_\epsilon(p)$ in (M, d) for ϵ sufficiently small are the geodesic balls $B_\epsilon(p)$ in (M, g) by Theorem 3.12. Any geodesic ball is an open set in (M, g) by definition. Moreover, given any open set U in (M, g), then for all $p \in U$ there exists $\epsilon(p) > 0$ such that $B_{\epsilon(p)}(p) \subseteq U$ by the existence of normal neighbourhoods. Hence U can be written as the union of geodesic balls. Thus the metric d induces the given topology on M.

Clearly $d(p, p) = 0$ for all $p \in M$ by taking α to be the constant curve $\alpha(t) = p$ for all t.

Let $p, q \in M$, then $d(p, q) = d(q, p)$ since given any curve $\alpha : [0, L] \to M$ from p to q the backwards curve $\beta : [0, L] \to M$ given by $\beta(t) = \alpha(L - t)$ satisfies $\beta'(t) = -\alpha'(L - t)$ so $|\beta'(t)| = |\alpha'(L - t)|$ and thus $L(\alpha) = L(\beta)$.

If $p, q, r \in M$ and α, β are any curves from p to q and q to r, then the curve γ given by joining α and β is a curve from p to r with $L(\gamma) = L(\alpha) + L(\beta)$ so we have that $d(p, r) \le L(\alpha) + L(\beta)$. Since this is true for all α, β we can take the infimum over all α, β and deduce that $d(p, r) \le d(p, q) + d(q, r)$.

Now suppose $p \ne q$. There exists an open set $U \ni p$ in M such that $q \notin U$. Since $\exp_p$ is continuous, there exists $\delta > 0$ such that $\exp_p(\overline{B_\delta(0)})$ is well-defined and contained in U. Hence, $q \notin \exp_p(\overline{B_\delta(0)})$. Let α be a curve from p to q. Then the portion β of α contained in $\exp_p(\overline{B_\delta(0)})$ must meet the geodesic sphere $\mathcal{S}_\delta(p)$. However, since geodesics are locally length minimizing by Theorem 3.12, we must have that $L(\beta) \ge \delta$ which then means that $L(\alpha) \ge L(\beta) \ge \delta$. Therefore $d(p, q) \ge \delta > 0$.

Hence, (M, d) is a metric space. □

7.2 Hopf–Rinow

We now state and prove one of the main theorems in the course, the *Hopf–Rinow Theorem*, which says that the notion of geodesic completeness agrees with our previous idea of metric space completeness. However, the real key to this result (and its proof) is that if a Riemannian manifold is complete then any two points can be joined by a minimizing geodesic.

Theorem 7.3 (Hopf–Rinow Theorem). *Let (M, g) be a connected Riemannian manifold. The following are equivalent:*

(a) (M, g) is (geodesically) complete;

(b) $\exp_p$ is defined on all of T_pM for some $p \in M$;

(c) closed bounded subsets of M are compact;

(d) (M, d) is a complete metric space.

Moreover, if (M, g) is complete then for all $p, q \in M$ there exists a geodesic γ from p to q such that $d(p, q) = L(\gamma)$.

Proof. (a) ⇒ (b) is trivial by definition.

(b) ⇒ (c). We first show that for any $q \in M$ there exists a geodesic $\gamma : [0, L] \to M$ such that $\gamma(0) = p$ and $\gamma(L) = q$.

Let $q \in M$ and let $d(p, q) = L$. Let $\delta > 0$ be such that $B_\delta(p)$ is a well-defined geodesic ball around p and let $S_\delta(p) = \partial\overline{B_\delta(p)}$ be the usual geodesic sphere. The map $x \mapsto d(q, x)$ is continuous on $S_\delta(p)$ so $d(q, x_0)$ is a minimum for some $x_0 \in S_\delta(p)$. Since $x_0 \in S_\delta(p)$, $x_0 = \exp_p(\delta X)$ for some $X \in T_pM$ with $|X| = 1$.

Let $\gamma(s) = \exp_p(sX)$ which is defined for all $s \in \mathbb{R}$ by assumption. The idea is to show that this is the geodesic we want. It is pointing in the right direction (since it is minimizing the distance to q in $B_\delta(p)$) so we just need to show that it extends all the way to q.

Let $A = \{s \in [0, L] : d(\gamma(s), q) = L - s\}$. We see that A is non-empty because $d(p, q) = d(\gamma(0), q) = L$ so $0 \in A$ and A is closed because the metric d is continuous. We want to show that A is open because then A is closed and open and non-empty in the connected interval $[0, L]$ so it must equal $[0, L]$. This means that in particular that $L \in A$ so $d(\gamma(L), q) = L - L = 0$ so $\gamma(L) = q$ and hence γ is a geodesic from p to q and $L(\gamma) = L|X| = L = d(p, q)$ as desired.

Suppose that $s_0 < L$. We need to show that $s_0 + \delta_0 \in A$ for some $\delta_0 > 0$ to show that A is open. Let $\delta_0 > 0$ be such that $B_{\delta_0}(\gamma(s_0))$ is a well-defined geodesic ball. Let $y_0 \in S_{\delta_0}(\gamma(s_0))$ be a point where $y \mapsto d(y, q)$ has a minimum (which exists as d is continuous). Then since $s_0 \in A$,

$$L - s_0 = d(\gamma(s_0), q) = \delta_0 + \min_{y \in S_{\delta_0}(q)} d(y, q) = \delta_0 + d(y_0, q).$$

Hence

$$d(y_0, q) = L - (s_0 + \delta_0).$$

If we can show that $y_0 = \gamma(s_0 + \delta_0)$ then

$$d(\gamma(s_0 + \delta_0), q) = d(y_0, q) = L - (s_0 + \delta_0)$$

so $s_0 + \delta_0 \in A$ and thus A is open.

Now

$$d(p, y_0) \geq |d(p, q) - d(q, y_0)| = |L - (L - (s_0 + \delta_0))| = s_0 + \delta_0.$$

However, the curve α given by following γ from p to $\gamma(s_0)$ and then the radial geodesic in $B_{\delta_0}(\gamma(s_0))$ from $\gamma(s_0)$ to y_0 has length $L(\alpha) = s_0 + \delta_0$. Since α is a curve from p to y_0 we have that

$$d(p, y_0) \leq L(\alpha) \leq s_0 + \delta_0$$

so we deduce that

$$d(p, y_0) = s_0 + \delta_0.$$

Furthermore, α is minimizing and $|\alpha'|$ is constant (as it is a union of geodesics) and thus is a geodesic by Proposition 3.13. Therefore, by uniqueness of geodesics, $\alpha = \gamma$ and thus $y_0 = \gamma(s_0 + \delta_0)$ as required.

We conclude that there is always a minimizing geodesic from p to q.

Now if $C \subseteq M$ is closed and bounded then $C \subseteq B^d_R(p) \subseteq \exp_p(\overline{B_{R'}(0)})$ for some $R, R' > 0$ by what we have just shown (i.e. we can connect p by a radial geodesic to any point $q \in C$ so that $d(p, q)$ is the length of that geodesic). Since $\overline{B_{R'}(0)}$ is compact and $\exp_p$ is continuous we see that $\exp_p(\overline{B_{R'}(0)})$ is compact and thus C is compact as desired.

(c) $\Rightarrow$ (d). Let (p_n) be a Cauchy sequence in M with respect to d. Then (p_n) is bounded so $C = \overline{\{p_n : n \in \mathbb{N}\}}$ is closed and bounded and thus C is compact by assumption. We deduce by metric space theory that (p_n) has a convergent subsequence and thus (M, d) is complete by definition.

(d) $\Rightarrow$ (a). This time we argue by contradiction. Suppose M is not (geodesically) complete. That means that there exists a normalized geodesic γ which is defined for $s < s_0$ but not for $s = s_0$.

Let (s_n) be a strictly increasing sequence in $[0, s_0)$ converging to s_0. Then (s_n) is convergent so it is Cauchy and thus $\big(\gamma(s_n)\big)$ is Cauchy as

$$d(\gamma(s_n), \gamma(s_m)) = |s_n - s_m| \to 0 \quad \text{as } n, m \to \infty.$$

We are assuming that (M, d) is complete so there exists $p_0 \in M$ and a subsequence of (s_n) which we still call (s_n) for simplicity such that

$$d(\gamma(s_n), p_0) \to 0 \text{ as } n \to \infty.$$

If W is a totally normal neighbourhood of p_0, there exists $\delta > 0$ such that $\exp_q : B_\delta(0) \to M$ is a diffeomorphism onto an open set containing W for all $q \in W$. Let N be sufficiently large that if $n, m > N$ then $\gamma(s_n) \in W$ for all $n > N$ and $d(\gamma(s_n), \gamma(s_m)) < \delta$ for $m, n > N$. Then choose $m, n > N$. There exists a unique geodesic $\alpha : [0, L] \to W$ such that $\alpha(0) = \gamma(s_n)$, $\alpha(L) = \gamma(s_m)$. Necessarily α and γ

coincide where they are both defined by uniqueness. Since $\exp_{\gamma(s_n)}$ is a diffeomorphism on $B_\delta(0)$ and its image contains W we see that α, which is radial geodesic from $\gamma(s_n)$, extends γ beyond s_0 (as α passes through p_0 for example), giving our required contradiction.

The final conclusion is obvious given that (b) implies the existence of a minimizing geodesic from p to any point q. $\square$

Remark. The minimizing geodesic is not necessarily unique: if we take the North and South poles $N, S \in \mathcal{S}^2$, then there are infinitely many minimizing geodesics between them given by the lines of longitude.

Moreover, we see that the upper half-space or the upper hemisphere has the property that there is a minimizing geodesic between any two points, but these manifolds are not complete.

Example. Since any closed bounded subset of a compact metric space is compact, Theorem 7.3 implies that any compact Riemannian manifold is complete. In particular, T^n, $\mathcal{S}^n$, $\mathbb{RP}^n$ and $\mathbb{CP}^n$ are complete.

7.3 Cartan–Hadamard

We can now state one of the fundamental theorems in Riemannian Geometry which shows the interaction between curvature and topology. Recall that we say that a manifold is *simply connected* if every loop in M can be continuously deformed to a point.

Example. $\mathbb{R}^n$ is simply connected.

However, $\mathbb{R}^2 \setminus \{0\}$ is not simply connected: there is a "hole" at the origin, which means a loop around 0 cannot be deformed to a point. For $n > 2$, $\mathbb{R}^n \setminus \{0\}$ is simply connected because you can now find room to move your curve encircling the origin and shrink it to a point.

Example. $\mathcal{S}^n$ is simply connected for $n \geq 2$.

However, $\mathcal{S}^1$ is not simply connected: the reason is the same as $\mathbb{R}^2 \setminus \{0\}$.

Example. T^n is never simply connected.

Example. $\mathbb{RP}^n$ is not simply connected but $\mathbb{CP}^n$ is simply connected.

Example. Many matrix Lie groups, like $\mathrm{SL}(n, \mathbb{R})$, $\mathrm{O}(n)$, $\mathrm{SO}(n)$ and $\mathrm{U}(n)$ are not simply connected but $\mathrm{SU}(n)$ is simply connected.

When M is not simply connected it will be useful to have the following definition.

Definition 7.4. If M is a connected n-dimensional manifold then there is a unique (up to diffeomorphism) connected and simply connected n-dimensional manifold $\widetilde{M}$ covering M called the *universal cover* of M.

Note that the *fundamental group* $\pi_1(M)$ of M acts freely and properly discontinuously on $\widetilde{M}$ by diffeomorphisms, since $\pi_1(M)$ is isomorphic to the covering (or deck) transformations on $\widetilde{M}$.

We will have the following useful lemma.

Lemma 7.5. *Let (M, g) be connected, let $\widetilde{M}$ be the universal cover of M and let $\pi : \widetilde{M} \to M$ be the covering map. There exists a unique Riemannian metric $\tilde{g}$ on $\widetilde{M}$ so that π is a local isometry. The metric $\tilde{g}$ is called the covering metric on $\widetilde{M}$.*

Moreover, the fundamental group $\pi_1(M)$ acts on $(\tilde{M}, \tilde{g})$ by isometries.

Remark. Since $\pi : (\widetilde{M}, \tilde{g}) \to (M, g)$ is a local isometry, it maps geodesics to geodesics and $(\widetilde{M}, \tilde{g})$ has the same curvature as (M, g).

Theorem 7.6 (Cartan–Hadamard). *Let (M, g) be a simply connected, connected and complete n-dimensional Riemannian manifold with sectional curvature $K \leq 0$. Then $\exp_p : T_pM \to M$ is a diffeomorphism, so M is diffeomorphic to $\mathbb{R}^n$.*

Remark. If M is not simply connected but otherwise satisfies the hypotheses of the theorem, then there M is diffeomorphic to $\mathbb{R}^n/\pi_1(M)$.

To prove this result we need two lemmas.

Lemma 7.7. *If (M, g) is complete and has sectional curvature $K \leq 0$ then the conjugate locus $C(p) = \emptyset$ for all $p \in M$. Therefore $\exp_p : T_pM \to M$ is a surjective local diffeomorphism for all $p \in M$.*

Proof. Let $p \in M$ and let $\gamma : [0, \infty) \to M$ be a geodesic such that $\gamma(0) = p$. Let J be a Jacobi field along γ such that $J(0) = 0$ but $J \neq 0$ (so $J'(0) \neq 0$). Thus

$$g(J, J)'(0) = 2g(J, J')(0) = 0 \quad \text{and} \quad g(J, J)''(0) = 2|J'(0)|^2 > 0,$$

so $g(J, J)'(t) > 0$ for all $t > 0$ sufficiently small. By the Jacobi equation,

$$\begin{aligned} g(J, J)'' &= 2g(J', J') + 2g(J'', J) \\ &= 2|J'|^2 - 2g(R(J, \gamma')\gamma', J) \\ &= 2|J'|^2 - 2K(J, \gamma')(|J|^2|\gamma'|^2 - g(J, \gamma')^2) \geq 0. \end{aligned}$$

Thus, $g(J, J)'$ is increasing. Therefore $g(J, J)'(t) > 0$ for all $t > 0$ which means that $g(J, J)(t)$ is strictly increasing for $t > 0$. Since $g(J, J)(0) = 0$ we deduce that $g(J, J)(t) > 0$ for all $t > 0$. Therefore $J(t) \neq 0$ for all $t > 0$, so $\gamma(t)$ is not conjugate to $\gamma(0)$ for all $t > 0$. Therefore $C(p) = \emptyset$. The conclusion follows from Proposition 6.10 and Theorem 7.3. □

Lemma 7.8. *If (M, g) is complete, (N, h) is a Riemannian manifold and $f : M \to N$ is a surjective local diffeomorphism such that $|\mathrm{d}f_p(X)| \geq |X|$ for all $X \in T_pM$ and for all $p \in M$, then f is a covering map.*

Proof. We recall from topology that f is a covering map if and only if f has the curve-lifting property (i.e. given a curve α in N and $p \in f^{-1}(\alpha(0))$ there exists a curve β in M such that $\beta(0) = p$ and $f \circ \beta = \alpha$).

Let $\alpha : [0, 1] \to N$ be a curve and $p \in f^{-1}(\alpha(0))$. Since f is a local diffeomorphism there exists $\epsilon > 0$ and a curve $\beta : [0, \epsilon) \to M$ such that $\beta(0) = p$ and $f \circ \beta(t) = \alpha(t)$ for $t \in [0, \epsilon)$.

Let

$$A = \{T \in [0, 1] : \text{there exists a curve}\, \beta : [0, T] \to M \text{ such that}\, \beta(0) = p,\ f \circ \beta(t) = \alpha(t)\, \text{for all}\, t \in [0, T]\}.$$

We have shown that $A \supseteq [0, \epsilon)$. Since f is a local diffeomorphism, one sees that A is half-open on the right. Suppose for a contradiction that $A = [0, t_0)$ for some $t_0 \leq 1$. Then there exists an increasing sequence (t_n) in A such that $t_n \to t_0$ as $n \to \infty$.

Using the properties of the differential of f and the fact that $f \circ \beta = \alpha$,

$$\begin{aligned} d(\beta(t_n), \beta(0)) &\leq \int_0^{t_n} |\beta'(t)|\mathrm{d}t \leq \int_0^{t_n} |\mathrm{d}f_{\beta(t)}(\beta'(t))|\mathrm{d}t \\ &= \int_0^{t_n} |\alpha'(t)|\mathrm{d}t = L(\alpha|_{[0,t_n]}) \leq L(\alpha|_{[0,t_0]}). \end{aligned}$$

Therefore $(\beta(t_n))$ is a bounded sequence. Since M is complete, using the Hopf–Rinow Theorem, we deduce that after passing to a subsequence we have that $\beta(t_n) \to q \in M$ as $n \to \infty$.

Let V be an open neighbourhood of $q \in M$ such that $f|_V$ is a diffeomorphism onto its image. Then $f \circ \beta(t_n) = \alpha(t_n) \to \alpha(t_0)$ and $f \circ \beta(t_n) \to f(q)$. Therefore $\alpha(t_0) = f(q) \in f(V)$. By the continuity of α there exists $\delta > 0$ such that $\alpha\big((t_0 - \delta, t_0 + \delta)\big) \subseteq f(V)$. There certainly exists n such that $t_n \in (t_0 - \delta, t_0 + \delta)$ so $\beta\big((t_0 - \delta, t_n]\big) \subseteq V$.

Since $f|_V$ is a diffeomorphism we have that there exists a curve $\bar{\beta} : (t_0 - \delta, t_0 + \delta) \to M$ such that $f \circ \bar{\beta}(t) = \alpha(t)$ for $t \in (t_0 - \delta, t_0 + \delta)$. Moreover, $f \circ \beta = f \circ \bar{\beta}$ on $(t_0 - \delta, t_n]$ so $\beta = \bar{\beta}$ on $(t_0 - \delta, t_n]$ as $f|_V$ is a diffeomorphism. Thus, $\bar{\beta}$ extends β to $[0, t_0 + \delta)$ which implies that $A \supseteq [0, t_0 + \delta)$, which contradicts $t_0 \leq 1$. Therefore $A = [0, 1]$ and f has the curve-lifting property and so is a covering map. □

Proof of Theorem 7.6. Since M is complete, $\exp_p : T_pM \to M$ is surjective. Lemma 7.7 implies that $\exp_p$ is a local diffeomorphism. Define a Riemannian metric h on T_pM such that $\exp_p$ is a local isometry; i.e. for $X \in T_pM$,

$$h_X(Y, Z) = g_{\exp_p(X)}(\mathrm{d}(\exp_p)_X(Y), \mathrm{d}(\exp_p)_X(Z)).$$

Geodesics in T_pM through 0 are straight lines so by Theorem 7.3, h is complete. By Lemma 7.8, $\exp_p$ is a covering map. Since T_pM and M are simply connected we conclude that $\exp_p$ is a diffeomorphism. □

Remark. We actually proved that if there exists $p \in (M, g)$ such that $C(p) = \emptyset$, when (M, g) is complete and simply connected, then $\exp_p : T_pM \to M$ is a diffeomorphism and so M is diffeomorphic to $\mathbb{R}^n$.

The Cartan–Hadamard Theorem has some simple corollaries.

Corollary 7.9. *If (M, g) is complete and has sectional curvature $K \leq 0$ then the universal cover $\widetilde{M}$ is diffeomorphic to $\mathbb{R}^n$.*

Proof. This is immediate because if (M, g) is complete then $(\widetilde{M}, \tilde{g})$ is complete and if $K \leq 0$ then the sectional curvature of $(\widetilde{M}, \tilde{g})$ is non-negative also. Applying Cartan–Hadamard to $(\widetilde{M}, \tilde{g})$ gives the result. □

Corollary 7.10. *If (M, g) is complete, simply connected and has sectional curvature $K \leq 0$ then M is non-compact.*

Proof. The hypotheses mean that M is diffeomorphic to $\mathbb{R}^n$ by Cartan–Hadamard, so non-compact. □

Example. A trivial example is a simply connected, connected and complete flat n-dimensional Riemannian manifold (M, g) must be diffeomorphic to $\mathbb{R}^n$.

A less trivial example is that if (M, g) is connected and complete and flat, then M must be diffeomorphic to $\mathbb{R}^n/G$ for some group G.

A special case is $\mathcal{S}^1$ which is trivially flat and, as we know, it is diffeomorphic to $\mathbb{R}/\mathbb{Z}$. The fundamental group of $\mathcal{S}^1$ is $\mathbb{Z}$. Similarly, T^n with its standard metric is flat and this is diffeomorphic to $\mathbb{R}^n/\mathbb{Z}^n$ so the fundamental group of T^n is $\mathbb{Z}^n$. Notice that T^n is obviously not diffeomorphic to $\mathbb{R}^n$ which shows why we need simply connected in the statement of Cartan–Hadamard.

Example. The hyperboloid model of hyperbolic space $(\mathcal{H}^2, g)$ is simply connected, connected and has constant curvature -1. We know that it is diffeomorphic to $\mathbb{R}^2$ since it is diffeomorphic to H^2, the upper-half plane, which is then diffeomorphic to $\mathbb{R}^2$.

Example. We know that $\mathcal{S}^n$ is connected and simply connected for $n \geq 2$, so it cannot have a complete metric with $K \leq 0$ as it is not diffeomorphic to $\mathbb{R}^n$. This is an extension of what you know is true for $\mathcal{S}^2$ by the Gauss–Bonnet theorem. We know that $\mathcal{S}^2$ can have a metric which has areas of negative curvature (consider the dumbbell) but it must always have areas of positive curvature.

Similarly, $\mathbb{CP}^n$ and $\mathrm{SU}(n)$ cannot have complete metrics with $K \leq 0$.

We also see that $\mathbb{RP}^n$ cannot have a complete metric with $K \leq 0$ either for $n \geq 2$ since its universal cover is $\mathcal{S}^n$.

8 Constant curvature

We want to study complete manifolds with constant sectional curvature (that is, $K(\sigma) = K$ for all 2-planes σ in any tangent space) and try to understand their geometry.

8.1 Basic formulae

The first thing we can do is describe the Riemann curvature tensor.

Proposition 8.1. *A Riemannian manifold (M, g) has constant sectional curvature K if and only if for all vector fields X, Y, Z, W on M*

$$R(X,Y,Z,W) = K\big(g(X,W)g(Y,Z) - g(X,Z)g(Y,W)\big).$$

Proof. Suppose that (M, g) has constant sectional curvature K. Define

$$\bar{R}(X,Y,Z,W) = K\big(g(X,W)g(Y,Z) - g(X,Z)g(Y,W)\big).$$

Then

$$R(X,Y,Y,X) = K(g(X,X)g(Y,Y) - g(X,Y)^2) = \bar{R}(X,Y,Y,X).$$

Since $\bar{R}$ has the same symmetries as R (which is easy to check), Proposition 4.6 implies that $R = \bar{R}$.

Suppose that R is as given. Then for all independent X, Y we have

$$K(X,Y) = \frac{R(X,Y,Y,X)}{g(X,X)g(Y,Y) - g(X,Y)^2} = \frac{K(g(X,X)g(Y,Y) - g(X,Y)^2)}{g(X,X)g(Y,Y) - g(X,Y)^2} = K$$

so we see that $K(X,Y) = K$ for all X, Y. $\square$

We can also describe the Ricci and scalar curvatures of Riemannian manifolds with constant sectional curvature.

Proposition 8.2. *If (M, g) has constant sectional curvature K then* $\mathrm{Ric} = (n-1)Kg$ *and* $S = n(n-1)K$.

Proof. By Proposition 8.1 we see that if $p \in M$, $\{E_1, \ldots, E_n\}$ is an orthonormal frame for T_pM and $X, Y \in T_pM$, then

$$\mathrm{Ric}(X,Y) = \sum_{k=1}^{n} R(X,E_k,E_k,Y) = K\sum_{k=1}^{n}(g(X,Y)g(E_k,E_k) - g(X,E_k)g(Y,E_k)) = K(n-1)g(X,Y).$$

Thus

$$S = \sum_{i,j=1}^{n} R(E_i,E_j,E_j,E_i) = K\sum_{i,j=1}^{n}(g(E_i,E_i)g(E_j,E_j) - g(E_i,E_j)^2) = K(n^2-n) = Kn(n-1).$$

We have used the fact that $g(E_i, E_j) = \delta_{ij}$ and $g(X,Y) = \sum_{i=1}^{n} g(X,E_i)g(Y,E_i)$. $\square$

So Riemannian manifolds with constant sectional curvature are Einstein manifolds and have constant scalar curvature.

Example. $(\mathbb{R}^n, g_0)$ has constant sectional curvature 0. The same is true of $\mathbb{R}^n/\mathbb{Z}^n \cong T^n$. So their Ricci and scalar curvatures are also 0.

Example. We saw that $\mathcal{S}^2$ with the round metric has constant sectional curvature 1. The same is also true of $\mathbb{RP}^2$. Their Ricci curvature tensors are $\mathrm{Ric} = g$ and scalar curvature $S = 2$.

Example. We saw that $\mathcal{H}^2$ with the hyperbolic metric has constant sectional curvature -1. Its Ricci curvature tensor is $\mathrm{Ric} = -g$ and scalar curvature $S = -2$.

We now observe that the Fundamental Equations for Riemannian submanifolds of Riemannian manifolds with constant sectional curvature are particularly simple and useful.

Proposition 8.3. *If (N, g) has constant sectional curvature K and (M, g_M) is a Riemannian submanifold of (N, g) then*

$$\begin{aligned} K(|X|^2|Y|^2 - g(X,Y)^2) &= K^M(X,Y)(|X|^2|Y|^2 - g(X,Y)^2) + |B(X,Y)|^2 - g(B(X,X),B(Y,Y)); \\ (\nabla^N_X B)(Y,Z) &= (\nabla^N_Y B)(X,Z); \\ g(R^\perp(X,Y)\xi,\zeta) &= g([S_\xi, S_\zeta]X, Y), \end{aligned}$$

for tangent vector fields X, Y, Z and normal vector fields ξ, ζ on M.

Proof. The first equation is immediate from the Gauss equation and Proposition 8.1. The second and third equations come from seeing that the left-hand side in the Codazzi and Ricci equations must be zero by Proposition 8.1. □

8.2 Model spaces

If (M, g) has constant sectional curvature K, we can always rescale the metric so that $K \in \{-1, 0, 1\}$, since if we multiply the metric by t then the sectional curvature changes by a factor of t^{-1}. So, a 2-sphere of radius r has constant sectional curvature $\frac{1}{r^2}$. We have seen that $\mathbb{R}^n$ is complete with constant sectional curvature 0 and $\mathrm{O}(n) \ltimes \mathbb{R}^n$ give the isometries, but what about $K = 1$ and $K = -1$?

We begin with the easier and familiar case of $K = 1$.

Theorem 8.4. *The unit n-sphere $(\mathcal{S}^n, g)$ in $\mathbb{R}^{n+1}$,*

$$\mathcal{S}^n = \{(x_1, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} : \sum_{i=1}^{n+1} x_i^2 = 1\},$$

with round metric g is

- *complete,*
- *its geodesics are the great circles given by $\Pi \cap \mathcal{S}^n$ for 2-planes Π in $\mathbb{R}^{n+1}$ through the origin,*
- *it has constant sectional curvature* 1
- *and* $\mathrm{Isom}(\mathcal{S}^n, g) = \mathrm{O}(n+1) = \{A \in M_n(\mathbb{R}) : A^{\mathrm{T}}A = I\}$.

Proof. We already saw that $\mathcal{S}^n$ is complete by the Hopf–Rinow Theorem and that the geodesics are as described. We know that $\mathrm{O}(n+1)$ defines isometries of $\mathbb{R}^{n+1}$. Moreover, it is also clear that $\mathrm{O}(n+1)$ defines the only linear maps of $\mathbb{R}^{n+1}$ that preserve $\mathcal{S}^n$, so these give the isometries of $\mathcal{S}^n$.

We already saw that $(\mathcal{S}^n, g)$ had constant sectional curvature 1 using the fact that it was a hypersurface, but now we give an intrinsic proof. Let $p \in \mathcal{S}^n$ and σ a 2-plane in $T_p\mathcal{S}^n = \mathrm{Span}\{p\}^\perp$. Since $\mathrm{O}(n+1)$ gives isometries we can rotate so that $p = \mathbf{e}_1$ and $\sigma = \mathrm{Span}\{X_1 = \mathbf{e}_3, X_2 = \mathbf{e}_2\}$. Define

$$f(\theta, \phi) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta, 0, \ldots, 0)$$

for $\theta \in (0, \pi)$ and $\phi \in (0, 2\pi)$ so that $f(\theta, \phi) = p$ if and only if $\theta = \frac{\pi}{2}$ and $\phi = 0$. Then $f_*\partial_\theta(p) = X_1$ and $f_*\partial_\phi(p) = X_2$. Hence, by our previous calculations for $\mathcal{S}^2$ we see that $K(\sigma) = 1$.

This is related to our exponential map discussion because $\theta = \frac{\pi}{2}$ and $\phi = 0$ are geodesics in $\mathcal{S}^2$ and $\mathcal{S}^n$. □

We now see that this extends in a natural way to the case of $K = -1$.

Theorem 8.5. *The hyperbolic n-space $(\mathcal{H}^n, g)$ where*

$$\mathcal{H}^n = \{(x_1, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} : \sum_{i=1}^{n} x_i^2 - x_{n+1}^2 = -1,\, x_{n+1} > 0\},$$

and g is the restriction of

$$\sum_{i=1}^{n} \mathrm{d}x_i^2 - \mathrm{d}x_{n+1}^2,$$

is

- *complete,*
- *the geodesics are given by $\Pi \cap \mathcal{H}^n$ for 2-planes Π in $\mathbb{R}^{n+1}$ through the origin which meet $\mathcal{H}^n$ (these are called Lorentz planes),*
- *it has constant sectional curvature -1*
- *and $\mathrm{Isom}(\mathcal{H}^n, g) = \mathrm{O}^+(n,1) = \{A \in M_{n+1}(\mathbb{R}) : A^{\mathrm{T}}GA = G,\ a_{n+1,n+1} > 0\}$, where*

$$G = \begin{pmatrix} I & 0 \\ 0 & -1 \end{pmatrix}.$$

Proof. The proof is very similar to the one for $\mathcal{S}^n$.

Clearly, the isometries are as stated because $\mathrm{O}^+(n,1)$ is the group which preserves G on $\mathbb{R}^{n+1}$ and $\mathcal{H}^n$.

Given $p = (0, \ldots, 0, 1) \in \mathcal{H}^n$ and $X \in T_p\mathcal{H}^n$, let γ be the unique geodesic through p with tangent vector X. If we define $\rho \in \mathrm{O}(n,1)$ to be the reflection in the plane $\Pi = \mathrm{Span}\{p, X\}$, since ρ is an isometry we see that $\rho \circ \gamma$ is another geodesic with the same properties as γ. Thus, by uniqueness of geodesics, $\rho \circ \gamma = \gamma$, which means that $\gamma = \Pi \cap \mathcal{H}^n$.

Concretely, since $p = (0, \ldots, 0, 1)$, if we take $X = (0, 0, \ldots, 1, 0) \in T_p\mathcal{H}^n$ (we can always achieve by using an isometry) then $\gamma(t) = (0, \ldots, 0, \sinh t, \cosh t)$. Clearly, these geodesics are defined for all $t \in \mathbb{R}$ so $\mathcal{H}^n$ is complete and uniqueness implies that these are all the geodesics as claimed.

By a similar argument to the previous theorem, we can restrict to calculating the sectional curvature of $\mathcal{H}^2$, which we know is -1, so the result follows. $\square$

The manifold $(\mathcal{H}^n, g)$ is called the *hyperboloid model* of hyperbolic n-space. We have other models for the hyperbolic space, which we record here.

Example. We have an isometry $f : (\mathcal{H}^n, g) \to (B^n, h)$ where

$$B^n = \{(y_1, \ldots, y_n) \in \mathbb{R}^n : \sum_{i=1}^{n} y_i^2 < 1\}$$

is the unit ball in $\mathbb{R}^n$ and

$$h = \sum_{i=1}^{n} \frac{4\mathrm{d}y_i^2}{(1 - \sum_{i=1}^{n} y_i^2)^2}$$

given by

$$f(x_1, \ldots, x_{n+1}) = \frac{(x_1, \ldots, x_n)}{1 + x_{n+1}}.$$

We call (B^n, h) the *Poincaré disk model* of hyperbolic n-space.

Example. We have an isometry $f : (\mathcal{H}^n, g) \to (H^n, h)$ where

$$H^n = \{(z_1, \ldots, z_n) \in \mathbb{R}^n : z_n > 0\}$$

is the upper half-space and

$$h = \sum_{i=1}^{n} \frac{\mathrm{d}z_i^2}{z_n^2}$$

given by

$$f(x_1, \ldots, x_{n+1}) = \frac{(x_1, \ldots, x_{n-1}, 1)}{x_n + x_{n+1}}.$$

This is the *upper-half space model* of hyperbolic n-space.

8.3 The curvature determines the metric

We start with the following result of Cartan which is often paraphrased as "the curvature determines the metric".

Theorem 8.6 (Cartan). *Let (M, g_M), (N, g_N) be Riemannian manifolds.*

(i) Let $p \in M$, $q \in N$, let $\iota : T_pM \to T_qN$ be an isometry and let $\exp_p$ and $\widetilde{\exp}_q$ be the exponential maps.

(ii) Let V be a normal neighbourhood of p such that $\widetilde{\exp}_q$ is defined on $U = \iota \circ \exp_p^{-1}(V)$.

(iii) Let $f = \widetilde{\exp}_q \circ \iota \circ \exp_p^{-1} : V \to \widetilde{\exp}_q(U)$.

(iv) For $r \in V$ let $\gamma^r : [0, L] \to (M, g_M)$ be the (unique) normalised geodesic such that $\gamma^r(0) = p$ and $\gamma^r(L) = r$ and let $\tilde{\gamma}^r : [0, L] \to (N, g_N)$ be the (unique) normalised geodesic such that $\tilde{\gamma}^r(0) = q$ and $(\tilde{\gamma}^r)'(0) = \iota((\gamma^r)'(0))$.

(v) For $r \in V$ and $s \in [0, L]$ let $\tau_s^r : T_pM \to T_{\gamma^r(s)}M$ be parallel transport along γ^r and let $\tilde{\tau}_s^r : T_qN \to T_{\tilde{\gamma}^r(s)}N$ be parallel transport along $\tilde{\gamma}^r$.

(vi) For $r \in V$ and $s \in [0, L]$, let $\phi_s^r = \tilde{\tau}_s^r \circ \iota \circ (\tau_s^r)^{-1} : T_{\gamma^r(s)}M \to T_{\tilde{\gamma}^r(s)}N$. Notice that $\phi_L^r : T_rM \to T_{f(r)}N$.

If

$$R^M(X, Y, Z, W) = R^N(\phi_L^r(X), \phi_L^r(Y), \phi_L^r(Z), \phi_L^r(W))$$

for all $X, Y, Z, W \in T_rM$ and $r \in V$ then $f : V \to f(V) \subseteq N$ is a local isometry and $\mathrm{d}f_p = \iota$.

Proof. Let $r \in V$, $X \in T_rM$, and let J be the unique Jacobi field along γ^r such that $J(0) = 0$ and $J(L) = X$. Let $\{E_1, \ldots, E_n\}$ be an orthonormal basis of T_pM with $E_n = (\gamma^r)'(0)$ and let $E_i(s) = \tau_s^r(E_i)$.

Write $J(s) = \sum_i y_i(s)E_i(s)$, so the Jacobi equation is equivalent to

$$y_i'' + \sum_j R^M(E_j, E_n, E_n, E_i)y_j = 0.$$

Let $\tilde{J}(s) = \phi_s^r(J(s))$, which gives a vector field $\tilde{J}$ along $\tilde{\gamma}^r$, and let $\tilde{E}_i(s) = \phi_s^r(E_i(s))$. Since $\tilde{J}(s) = \sum_i y_i(s)\tilde{E}_i(s)$ and we have that

$$y_i'' + \sum_j R^N(\tilde{E}_j, \tilde{E}_n, \tilde{E}_n, \tilde{E}_i)y_j = 0$$

by hypothesis, we deduce that $\tilde{J}$ is a Jacobi field along $\tilde{\gamma}^r$ with $\tilde{J}(0) = 0$.

Since ϕ_s^r is an isometry we have that $|\tilde{J}(s)| = |J(s)|$. We want to show that $\tilde{J}(L) = \mathrm{d}f_r(X) = \mathrm{d}f_r(J(L))$, since then $\mathrm{d}f_r$ is an isometry for all $r \in V$, and thus f is a local isometry on V as claimed.

Since the E_i and $\tilde{E}_i$ are parallel vector fields along γ^r and $\tilde{\gamma}^r$, we see that $\tilde{J}'(s) = \phi_s^r(J'(s))$, so $\tilde{J}'(0) = \iota(J'(0))$. Moreover, we know that

$$J(s) = \mathrm{d}(\exp_p)_{s(\gamma^r)'(0)}(sJ'(0)),$$
$$\tilde{J}(s) = \mathrm{d}(\widetilde{\exp}_q)_{s(\tilde{\gamma}^r)'(0)}(s\tilde{J}'(0)).$$

Therefore

$$\tilde{J}(L) = \mathrm{d}(\widetilde{\exp}_q)_{L(\tilde{\gamma}^r)'(0)}(L\iota(J'(0)) = \mathrm{d}(\widetilde{\exp}_q)_{L(\tilde{\gamma}^r)'(0)} \circ \iota \circ (\mathrm{d}(\exp_p)_{L(\gamma^r)'(0)})^{-1}(J(L)) = \mathrm{d}f_r(J(L))$$

as required. □

Cartan's Theorem produces a local isometry between Riemannian manifolds with the same Riemann curvature tensor, but a natural question is: is this local isometry unique? The answer is provided by the following useful uniqueness result.

Lemma 8.7. *If M is connected and $F, G : (M, g_M) \to (N, g_N)$ are local isometries with $F(p) = G(p)$ and $\mathrm{d}F_p = \mathrm{d}G_p$ for some $p \in M$ then $F = G$.*

Proof. Let $q \in M$. Then since M is connected there exists a curve $\alpha : [0,1] \to M$ such that $\alpha(0) = p$ and $\alpha(1) = q$. Let

$$A = \{t \in [0,1] : F(\alpha(t)) = G(\alpha(t)),\, \mathrm{d}F_{\alpha(t)} = \mathrm{d}G_{\alpha(t)}\},$$

which is clearly closed.

Let U, V be normal neighbourhoods of p such that $F|_U$ and $G|_V$ are isometries and $F(U) = G(V)$ (this is possible because we can simply intersect the open sets $F(U)$ and $G(V)$ if necessary). Then $f = G^{-1} \circ F : U \to V$ is an isometry such that $f(p) = p$ and $\mathrm{d}f_p = \mathrm{id}$. If $r \in U$ then there exists a unique $X \in T_pM$ such that $\exp_p(X) = r$. The curve $\gamma(s) = f(\exp_p(sX))$ for $s \in [0,1]$ is a geodesic (as f is an isometry) with $\gamma(0) = f(p)$ and $\gamma'(0) = \mathrm{d}f_p(X)$, so it must be equal to the geodesic given by $s \mapsto \exp_{f(p)}(s\mathrm{d}f_p(X))$ by uniqueness. Hence

$$f(r) = f(\exp_p(X)) = \exp_{f(p)}(\mathrm{d}f_p(X)) = \exp_p(X) = r.$$

Therefore $f(r) = r$ for all $r \in U$ so $U = V$ and $F = G$ on U. Thus $\sup A > 0$ since there exist $T \in (0,1)$ such that $\alpha(t) \in V$ for all $t \in [0,T]$.

If $\sup A = T < 1$ then, since $T \in A$ as A is closed, we can repeat the argument above for the point $\alpha(T)$ and get a contradiction, so $1 \in A$ so $F(q) = G(q)$ and thus $F = G$. □

Corollary 8.8. *If (M, g_M) and (N, g_N) have the same constant curvature, $p \in M$, $q \in N$ and $\iota : T_pM \to T_qN$ is an isometry then there exist normal neighbourhoods $V \ni p$ and $W \ni q$ and a unique isometry $f : V \to W$ such that $f(p) = q$ and $\mathrm{d}f_p = \iota$.*

Proof. Since M and N have the same constant curvature we can apply Cartan's Theorem (Theorem 8.6) and deduce the existence of V, W and f as claimed. Uniqueness then follows from the fact that V and W are connected and Lemma 8.7. □

8.4 Space forms

We now want to classify *space forms*: complete Riemannian manifolds (M, g) with constant curvature K. As we noted, we can always rescale the metric simply by multiplying by a constant so that $K \in \{-1, 0, 1\}$. We therefore restrict our attention to this situation.

Theorem 8.9. *Let (M, g) be a complete and simply connected n-dimensional Riemannian manifold with constant sectional curvature $K \in \{-1, 0, 1\}$. Then (M, g) is isometric to*

- *$\mathcal{S}^n$ with the round metric if $K = 1$,*
- *$\mathbb{R}^n$ with the Euclidean metric if $K = 0$, or*
- *$\mathcal{H}^n$ with the hyperbolic metric if $K = -1$.*

Remark. For general constant $K < 0$ we write $\mathcal{H}^n(K)$ for hyperbolic n-space with constant curvature K, and for $K > 0$ we write $\mathcal{S}^n(K)$ for the n-sphere with constant curvature K.

Proof. Suppose $K = -1$ or 0 and (N, h) is either $\mathcal{H}^n$ or $\mathbb{R}^n$ with its constant curvature metric. Let $p \in M$, $q \in N$ and let $\iota : T_pM \to T_qN$ be an isometry. Then $f = \exp_q \circ \iota \circ \exp_p^{-1} : M \to N$ is well-defined and surjective by the Cartan–Hadamard Theorem (Theorem 7.6). Cartan's Theorem (Theorem 8.6) and Corollary 8.8 imply that f is a local isometry. Lemma 7.8 implies that f is a covering map so f is a diffeomorphism as M and N are simply connected. We conclude that f is an isometry.

Now suppose $K = 1$ and let $p \in M$, $q \in \mathcal{S}^n$ and let $\iota : T_q\mathcal{S}^n \to T_pM$ be an isometry. Then $F = \exp_p \circ \iota \circ \exp_q^{-1} : \mathcal{S}^n \setminus \{-q\} \to M$ is well-defined. Cartan's Theorem and Corollary 8.8 imply that

F is a local isometry. Take $r \in \mathcal{S}^n \setminus \{q, -q\}$. Let $s = F(r)$ and $\jmath = \mathrm{d}f_r : T_r\mathcal{S}^n \to T_sM$. Define $G = \exp_s \circ \jmath \circ \exp_r^{-1} : \mathcal{S}^n \setminus \{-r\} \to M$.

Now $N = \mathcal{S}^n \setminus \{-q, -r\}$ is connected (as $n > 1$), $r \in N$, $F(r) = s = G(r)$ and $\mathrm{d}F_r = \jmath = \mathrm{d}G_r$. Applying Lemma 8.7 we deduce that $F = G$ on N. Let

$$H(t) = \begin{cases} F(t) & \text{if } t \neq -q \\ G(t) & \text{if } t \neq -r. \end{cases}$$

Then H is a local isometry from $\mathcal{S}^n$ with the round metric to (M, g) so as before we have that H is a diffeomorphism and hence an isometry. □

We now have our main result classifying the space forms, utilizing the following elementary result.

Proposition 8.10. *Let (M, g) be complete with constant sectional curvature. There exists a discrete subgroup G of $\mathrm{Isom}(\widetilde{M}, \tilde{g})$, acting freely and properly discontinuously on $\widetilde{M}$, such that (M, g) is isometric to $\widetilde{M}/G$ with the quotient metric.*

Proof. Let $\pi : \widetilde{M} \to M$ be the covering map and $G = \pi_1(M)$ be the group of covering transformations (which acts freely and properly discontinuously by definition of universal cover). Then $\pi(p) = \pi(q)$ if and only if there exists some $\phi \in G$ such that $q = \phi(p)$, which is if and only if $\xi(p) = \xi(q)$ where $\xi : \widetilde{M} \to \widetilde{M}/G$ is the projection map. Therefore there exists a bijection $f : M \to \widetilde{M}/G$ such that $f \circ \pi = \xi$. Since ξ and π are local isometries we see that f is a local isometry, but since f is also a bijection we deduce that f is an isometry. □

From this result, we have our classification of space forms, which is one of the main results in Riemannian geometry.

Theorem 8.11. *Let (M, g) be a complete n-dimensional Riemannian manifold with constant sectional curvature $K \in \{-1, 0, 1\}$. Then there exists a discrete group G acting freely and properly discontinuous by isometries such that (M, g) is isometric to*

- $\mathcal{S}^n/G$ *if* $K = 1$,
- $\mathbb{R}^n/G$ *if* $K = 0$,
- $\mathcal{H}^n/G$ *if* $K = -1$.

This result has many fascinating consequences which I encourage you to explore. One of these is the following.

Proposition 8.12. *Let (M, g) be a complete $2n$-dimensional Riemannian manifold with constant sectional curvature 1. Then (M, g) is isometric to $\mathcal{S}^{2n}$ or $\mathbb{RP}^{2n}$ with their standard Riemannian metrics.*

Proof. We know that M is isometric to $\mathcal{S}^{2n}/G$ where G acts freely and properly discontinuously by isometries by Theorem 8.11. Hence $G \subseteq \mathrm{O}(2n + 1)$. Let $x \in G$ and f_x be the corresponding isometry. Then $\det f_x = \pm 1$.

If $\det f_x = 1$ then f_x has 1 as an eigenvalue because the eigenvalues (which may be complex) all have modulus 1 as $f_x \in \mathrm{O}(2n+1)$ and they cannot all be non-real because complex eigenvalues occur in complex conjugate pairs and $2n + 1$ is odd. Thus f_x has a fixed point on $\mathcal{S}^{2n}$ (corresponding to a unit eigenvector p with eigenvalue 1, so $f_x(p) = p$). But this contradicts the assumption that the action is free unless $f_x = \mathrm{id}$, so this must be the case.

Suppose instead that $\det f_x = -1$. Then $\det(f_x^2) = 1$ so $f_x^2 = \mathrm{id}$ and hence $f_x = \pm\,\mathrm{id}$.

Therefore either $f_x = \mathrm{id}$ for all $x \in G$, so $\mathcal{S}^{2n}/G = \mathcal{S}^{2n}$, or there exists $x \in G$ such that $f_x = -\,\mathrm{id}$ and all isometries are $\pm\,\mathrm{id}$, so $\mathcal{S}^{2n}/G = \mathbb{RP}^{2n}$. □

Example. Proposition 8.12 is definitely false in odd dimensions. For example, any cyclic subgroup $\mathbb{Z}_k$ for $k \geq 2$ acts freely and properly discontinuous by isometries on $\mathcal{S}^3 \subseteq \mathbb{R}^4$ by setting

$$f_x = \begin{pmatrix} \cos\frac{2\pi}{k} & -\sin\frac{2\pi}{k} & 0 & 0 \\ \sin\frac{2\pi}{k} & \cos\frac{2\pi}{k} & 0 & 0 \\ 0 & 0 & \cos\frac{2\pi}{k} & \sin\frac{2\pi}{k} \\ 0 & 0 & -\sin\frac{2\pi}{k} & \cos\frac{2\pi}{k} \end{pmatrix}$$

where x is a generator of $\mathbb{Z}_k$ (so $Z_k = \{e, x, x^2, \ldots, x^{k-1}\}$). Then $\mathcal{S}^3/\mathbb{Z}_k$ has a metric with constant curvature 1 and is called a Lens space. There are also more complicated subgroups of O(4) that can act on $\mathcal{S}^3$ in the appropriate way, such as the tetrahedral group (which is of order 24 and describes the symmetries of a tetrahedron).

Example. If we look at compact orientable surfaces, then $\mathcal{S}^2$ has a metric with constant sectional curvature 1 and $T^2 \cong \mathbb{R}^2/\mathbb{Z}^2$ so has a flat metric (though this is not the induced metric on $\mathbb{R}^3$). Now it is possible to realise every compact orientable surface of genus at least 2 as $\mathcal{H}^2/G$ for some G acting freely and properly discontinuously by isometries (by choosing an appropriate geodesic polygon in the Poincaré disk for example and identifying sides), so every such surface has a *hyperbolic metric* (i.e. a metric with *constant sectional curvature* -1).

The previous example hints at the fact there are very many groups G which can occur in the $\mathcal{H}^n/G$ case. These hyperbolic manifolds are of significant interest and in the case of $n = 3$ are related to work towards the resolution of the *Poincaré Conjecture* and *Thurston's Geometrization Conjecture*.

9 Second variation formula and applications

Recall that when we were discussing geodesics we derived the first variation formula for the energy of a variation. We saw that geodesics corresponded to zeroes of the derivative of the energy (i.e. stationary points) for proper variations. However, we know that geodesics are local minima, so the second derivative at zero must be non-negative. This observation turns out to have very powerful consequence, resulting in arguably the most satisfying and surprising theorems in the course. These results really show the power of Riemannian Geometry as a way of imposing global topological conditions by specifying local curvature constraints.

9.1 Second variation formula

We derive our key formula which has even wider applications than we have time to explore here.

Theorem 9.1 (Second Variation Formula)**.** *Let $\gamma : [0, L] \to (M, g)$ be a geodesic, let $f : (-\epsilon, \epsilon) \times [0, L] \to M$ be a variation of γ, let V_f be the variation field of f and let E_f be the energy of f. Then*

$$\begin{aligned}\frac{1}{2}E_f''(0) = &-\int_0^L g(V_f'' + R(V_f,\gamma')\gamma', V_f)\mathrm{d}t\\ &- g(\nabla_{\frac{\partial f}{\partial s}}\frac{\partial f}{\partial s}(0,0),\gamma'(0)) + g(\nabla_{\frac{\partial f}{\partial s}}\frac{\partial f}{\partial s}(0,L),\gamma'(L)) - g(V_f(0),V_f'(0)) + g(V_f(L),V_f'(L))\\ =& \int_0^L g(V_f',V_f') - R(V_f,\gamma',\gamma',V_f)\mathrm{d}t - g(\nabla_{\frac{\partial f}{\partial s}}\frac{\partial f}{\partial s}(0,0),\gamma'(0)) + g(\nabla_{\frac{\partial f}{\partial s}}\frac{\partial f}{\partial s}(0,L),\gamma'(L)).\end{aligned}$$

Proof. Recall from the derivation of the First variation formula:

$$\frac{1}{2}E_f'(s) = [g(\frac{\partial f}{\partial s},\frac{\partial f}{\partial t})]_0^L - \int_0^L g(\frac{\partial f}{\partial s},\nabla_{\frac{\partial f}{\partial t}}\frac{\partial f}{\partial t})\mathrm{d}t.$$

Therefore

$$\begin{aligned}\frac{1}{2}E_f''(s) =& [g(\nabla_{\frac{\partial f}{\partial s}}\frac{\partial f}{\partial s},\frac{\partial f}{\partial t})]_0^L + [g(\frac{\partial f}{\partial s},\nabla_{\frac{\partial f}{\partial s}}\frac{\partial f}{\partial t})]_0^L\\ &- \int_0^L g(\nabla_{\frac{\partial f}{\partial s}}\frac{\partial f}{\partial s},\nabla_{\frac{\partial f}{\partial t}}\frac{\partial f}{\partial t})\mathrm{d}t - \int_0^L g(\frac{\partial f}{\partial s},\nabla_{\frac{\partial f}{\partial s}}\nabla_{\frac{\partial f}{\partial t}}\frac{\partial f}{\partial t})\mathrm{d}t.\end{aligned}\tag{‡}$$

At $s = 0$, $\frac{\partial f}{\partial s} = V_f$, $\frac{\partial f}{\partial t} = \gamma'$ and $\nabla_{\frac{\partial f}{\partial t}}\frac{\partial f}{\partial t} = \nabla_{\gamma'}\gamma' = 0$. Thus the third term on the right-hand side of (‡) is

$$\int_0^L g(\nabla_{\frac{\partial f}{\partial s}}\frac{\partial f}{\partial s},\nabla_{\frac{\partial f}{\partial t}}\frac{\partial f}{\partial t})\mathrm{d}t = 0.$$

For the fourth term in (‡), using Proposition 6.1, we calculate

$$\left(\nabla_{\frac{\partial f}{\partial s}}\nabla_{\frac{\partial f}{\partial t}} - \nabla_{\frac{\partial f}{\partial t}}\nabla_{\frac{\partial f}{\partial s}}\right)\frac{\partial f}{\partial t} = R(\frac{\partial f}{\partial s},\frac{\partial f}{\partial t})\frac{\partial f}{\partial t}.$$

Thus, at $s = 0$, we can use Lemma 3.11 to deduce that

$$\nabla_{\frac{\partial f}{\partial s}}\nabla_{\frac{\partial f}{\partial t}}\frac{\partial f}{\partial t} = V_f'' + R(V_f,\gamma')\gamma'.$$

For the first term in (‡) at $s = 0$ we see that:

$$[g(\nabla_{\frac{\partial f}{\partial s}}\frac{\partial f}{\partial s},\frac{\partial f}{\partial t})]_0^L = -g(\nabla_{\frac{\partial f}{\partial s}}\frac{\partial f}{\partial s}(0,0),\gamma'(0)) + g(\nabla_{\frac{\partial f}{\partial s}}\frac{\partial f}{\partial s}(0,L),\gamma'(L)).$$

For the second term in (‡) at $s = 0$ we see that

$$[g(\frac{\partial f}{\partial s},\nabla_{\frac{\partial f}{\partial s}}\frac{\partial f}{\partial t})]_0^L = -g(V_f(0),V_f'(0)) + g(V_f(L),V_f'(L)).$$

Putting together all these observations yields the first line in the Second variation formula.

To deduce the second line we observe that

$$\frac{\mathrm{d}}{\mathrm{d}t}g(V,V') = g(V,V'') + g(V',V').$$

Hence applying the Fundamental Theorem of Calculus gives the result. □

As for the First variation formula we are most interested in the case when f is a proper variation.

Corollary 9.2. *Suppose that f is a proper variation in Theorem 9.1. Then*

$$\begin{aligned}\frac{1}{2}E_f''(0) &= -\int_0^L g(V_f''+R(V_f,\gamma')\gamma',V_f)\mathrm{d}t\\ &= \int_0^L g(V_f',V_f')-R(V_f,\gamma',\gamma',V_f)\mathrm{d}t.\end{aligned}$$

Proof. This is immediate from the definition of a proper variation (i.e. $f(s,0)=\gamma(0)$ and $f(s,L)=\gamma(L)$ for all s). □

9.2 Bonnet–Myers Theorem

We now prove one of the nicest theorems in this Riemannian Geometry course.

Theorem 9.3 (Bonnet–Myers). *Let (M,g) be complete and n-dimensional with $\mathrm{Ric}\geq\frac{n-1}{r^2}>0$. Then M is compact and diam$(M)\leq\pi r$.*

Remark. Here, we say that $\mathrm{Ric}\geq\delta$ if for all unit length tangent vectors X we have $\mathrm{Ric}(X,X)\geq\delta$. We denote the diameter of M as

$$\mathrm{diam}(M)=\sup\{d(p,q):p,q\in M\}$$

when this exists.

Proof. Let $p,q\in M$. By Theorem 7.3, there exists a minimizing geodesic $\gamma:[0,1]\to M$ from p to q.

It is enough to show that $L(\gamma)\leq\pi r$ (since then $\mathrm{diam}(M)\leq\pi r$ which means that M is complete and bounded so M is compact by Theorem 7.3). So, for a contradiction, suppose that $L(\gamma)>\pi r$.

There exist orthonormal parallel vector fields $X_1,\ldots,X_{n-1}$ along γ such that $g(X_j,\gamma')=0$ for all j. Define vector fields V_j along γ by $V_j(t)=\sin(\pi t)X_j(t)$. Then there exist proper variations f_j of γ with variation field V_j and energy E_j.

Let $X_n=\gamma'/L(\gamma)$, which is a unit vector field orthogonal to $X_1,\ldots,X_{n-1}$. Corollary 9.2 and the fact that X_j is parallel and unit length implies that

$$\begin{aligned}\frac{1}{2}E_j''(0) &= \int_0^L -g(V_j,V_j'')-R(V_j,\gamma',\gamma',V_j)\mathrm{d}t\\ &= -\int_0^L g(X_j(t)\sin(\pi t),-\pi^2X_j(t)\sin(\pi t))-L(\gamma)^2R(X_j,X_n,X_n,X_j)\sin^2(\pi t)\mathrm{d}t\\ &= \int_0^1\sin^2(\pi t)\big(\pi^2-L(\gamma)^2K(X_n,X_j)\big)\mathrm{d}t.\end{aligned}$$

Therefore

$$\frac{1}{2}\sum_{j=1}^{n-1}E_j''(0)=\int_0^1\sin^2(\pi t)\big((n-1)\pi^2-L(\gamma)^2\,\mathrm{Ric}(X_n,X_n)\big)\mathrm{d}t<0$$

since $\mathrm{Ric}(X_n,X_n)\geq\frac{n-1}{r^2}$ and $L(\gamma)>\pi r$, so

$$L(\gamma)^2\,\mathrm{Ric}(X_n,X_n)>\pi^2r^2\frac{n-1}{r^2}=(n-1)\pi^2.$$

Thus $E_j''(0)<0$ for some j, so γ is not a local minimum for the energy E_j, but this contradicts Lemma 3.18. □

Remark. The sphere satisfies the critical case of Myers Theorem. In fact, if (M,g) is complete and n-dimensional with $\mathrm{Ric}\geq\frac{n-1}{r^2}$ then $\mathrm{diam}(M)=\pi r$ if and only if (M,g) isometric to a n-sphere of the appropriate curvature.

We now provide some simple applications of Myers Theorem.

Corollary 9.4. *If (M,g) is complete with* $\mathrm{Ric} \geq \delta > 0$ *then the universal cover $\widetilde{M}$ is compact and the fundamental group $\pi_1(M)$ is finite.*

Proof. Endow $\widetilde{M}$ with the covering metric $\tilde{g}$. Then the covering map π is a local isometry so $\widetilde{M}$ is complete with $\mathrm{Ric} \geq \delta > 0$. Applying the Bonnet–Myers Theorem (Theorem 9.3) we see that $\widetilde{M}$ compact. Therefore $\pi^{-1}(p)$ is finite for all $p \in M$ so $\pi_1(M)$ is finite. □

Remark. It is important to note that for *compact* Riemannian manifolds (M,g), the statements that $\mathrm{Ric} > 0$ and that there exists $\delta > 0$ such that $\mathrm{Ric} \geq \delta > 0$ are equivalent. This is because the bundle of unit tangent vectors over M is compact. For similar reasons, the statements that $K > 0$ and there exists $\delta > 0$ such that $K \geq \delta > 0$ are equivalent for compact Riemannian manifolds.

Remark. Corollary 9.4 shows that compact manifolds which admit metrics with positive Ricci curvature have greatly constrained topology. In particular the n-torus T^n cannot admit a metric with $\mathrm{Ric} > 0$. This generalizes the result from Gauss–Bonnet that T^2 cannot admit a metric with positive Gauss curvature.

However, in contrast, it is known that *every* compact manifold of dimension $n \geq 3$ admits a metric with negative Ricci curvature $\mathrm{Ric} < 0$.

Remark. In fact, the n-torus cannot admit a metric with positive *scalar curvature* $S > 0$. This is a much more challenging result, which is related to the study of minimal hypersurfaces, and to the study of spin geometry.

The general problem of which manifolds admit metrics of positive sectional/Ricci/scalar curvature is of fundamental importance in modern Riemannian geometry, as well as in topology and mathematical physics.

We conclude with a consequence of Bonnet–Myers for manifolds with positive sectional curvature.

Corollary 9.5. *If (M,g) is complete with sectional curvature $K \geq \frac{1}{r^2}$ then M is compact, diam$(M) \leq \pi r$ and $\pi_1(M)$ is finite.*

Proof. If $K \geq \frac{1}{r^2}$ then $\mathrm{Ric} \geq \frac{n-1}{r^2}$. Applying Theorem 9.3 and Corollary 9.4 gives the result. □

Remark. The paraboloid

$$\{(x_1,x_2,x_3) \in \mathbb{R}^3 \,:\, x_3 = x_1^2 + x_2^2\}$$

with its induced metric has positive sectional curvature and is complete but non-compact. So we cannot replace $K \geq \delta > 0$ with $K > 0$ in Corollary 9.5, even though these statements are equivalent for compact Riemannian manifolds.

9.3 Synge Theorem

We conclude the course with the following surprising result.

Theorem 9.6 (Synge). *Let (M,g) be compact and n-dimensional with sectional curvature $K > 0$.*

(a) If n is even and M is orientable then M is simply connected.

(b) If n is odd then M is orientable.

Example. $\mathbb{RP}^2$ is not orientable and not simply connected (as it is non-trivial quotient of $\mathcal{S}^2$ which is simply connected) but has a positive sectional curvature metric, so we need orientable in Theorem 9.6(a) and we need n odd in (b).

Example. $\mathbb{RP}^3$ is orientable and has a positive sectional curvature metric but it is not simply connected, so we need n even in Theorem 9.6(a).

To prove Synge's Theorem we apply the following more powerful result.

Theorem 9.7 (**Synge–Weinstein).** *If (M,g) is compact, oriented, n-dimensional and has sectional curvature $K>0$ and $f:M\to M$ is an isometry such that*

$$\det(\mathrm{d}f_p)=(-1)^n \quad \textit{for all } p\in M$$

(i.e. f is orientation preserving/reversing for n even/odd), then f has a fixed point.

Example. Notice that if we take the round n-sphere $(\mathcal{S}^n,g)$ then it is compact, oriented with positive sectional curvature $K=1$. The antipodal map $f=-\mathrm{id}$ is an isometry on $(\mathcal{S}^n,g)$ with *no fixed points*, but notice that

$$\det(\mathrm{d}f_p)=(-1)^{n+1}$$

We therefore see the importance of the sign in the Synge–Weinstein Theorem.

Proof of Theorem 9.6 given Theorem 9.7.

(a) Endow the universal cover $\widetilde{M}$ with the covering metric $\tilde{g}$. Since M is compact, (M,g) is complete and there exists $\delta>0$ such that $K\geq\delta$. Since the covering map is a local isometry, $(\widetilde{M},\tilde{g})$ is complete and $K\geq\delta$. The Bonnet–Myers Theorem then implies that $\widetilde{M}$ is compact.

Let f be a covering transformation on $\widetilde{M}$. We can apply the Synge–Weinstein Theorem since f is an orientation preserving isometry, so f has a fixed point. Since the covering transformations act freely we conclude that $f=\mathrm{id}$, so M is diffeomorphic to its universal cover and so is simply connected.

(b) For a contradiction, suppose M is non-orientable. Then $\overline{M}$, the oriented double cover of M, with the covering metric is compact and has $K\geq\delta>0$.

Let $f\neq\mathrm{id}$ be a covering transformation on $\overline{M}$. Since f is an orientation reversing isometry we deduce from the Synge–Weinstein Theorem that f has a fixed point. Again $f=\mathrm{id}$ as it is a covering transformation, which gives us our required contradiction.

□

Example. Since $\mathcal{S}^2$ ($\mathcal{S}^{2n}$) is compact and orientable with constant curvature 1 it must be simply connected (as we know).

Example. T^2 (T^{2n}) is compact and orientable but not simply connected, and so cannot have a metric with sectional curvature $K>0$.

Proof of Theorem 9.7. For a contradiction suppose there is no fixed point of f and let p minimise the function $q\mapsto d\big(q,f(q)\big)$ for $q\in M$. Since (M,g) is complete, Theorem 7.3 implies that there exists a normalised minimizing geodesic $\gamma:[0,L]\to(M,g)$ from p to $f(p)$.

The idea is to build a variation h of γ and then use the second variation formula to obtain a contradiction to the fact that γ minimizes the energy. A complication arises because we cannot build a proper variation, so our variation has to be carefully chosen so that, in particular, the boundary terms in the second variation formula vanish. We see that it will be useful to find a geodesic β starting at p which is initially orthogonal to γ. This will then give us the direction in which to vary γ.

We start by letting $\tau_\gamma^{-1}:T_{f(p)}M\to T_pM$ be parallel transport back along γ and let $A=\tau_\gamma^{-1}\circ\mathrm{d}f_p:T_pM\to T_pM$, which is an isometry by assumption. We want to show that $\gamma'(0)$ is a fixed point of A. We calculate

$$A(\gamma'(0))=(\tau_\gamma^{-1}\circ\mathrm{d}f_p)(\gamma'(0))=\tau_\gamma^{-1}((f\circ\gamma)'(0))$$

so, applying τ_γ to both sides, we see that $A(\gamma'(0))=\gamma'(0)$ if and only if $(f\circ\gamma)'(0)=\gamma'(L)$.

To show that $(f\circ\gamma)'(0)=\gamma'(L)$ we show that $\gamma\cup(f\circ\gamma)$ is a geodesic so we must have that the final velocity of γ and initial velocity of $f\circ\gamma$ agree. To do this, we let $q=\gamma(t)$ for some $t\in(0,L)$ and show that $d(q, f(p) + d(f(p), f(q)) = d(q, f(q))$, which shows that $\gamma \cup (f \circ \gamma)$ is inded a geodesic. Using the triangle inequality, the fact that f is an isometry, γ is a geodesic and the definition of p, we see that

$$d(q, f(q)) \leq d(q, f(p)) + d(f(p), f(q)) = d(q, f(p)) + d(p, q) = d(p, f(p)) \leq d(q, f(q)).$$

Therefore $d(q, f(p)) + d(f(p), f(q)) = d(q, f(q))$, so $\gamma \cup (f \circ \gamma)$ is a geodesic as we wanted. We deduce that $\gamma'(L) = (f \circ \gamma)'(0)$. Hence, $\gamma'(0)$ is a fixed point of A.

Let $B = A : \langle \gamma'(0) \rangle^{\perp} \to \langle \gamma'(0) \rangle^{\perp}$. Then B is an orthogonal transformation and

$$\det B = \det A = \det \mathrm{d}f_p = (-1)^n.$$

We deduce that B has 1 as an eigenvalue (consider even and odd dimensions separately and the fact that eigenvalues occur in complex conjugate pairs and real eigenvalues are ± 1). Therefore, there exists a unit parallel vector field $X(t)$ along γ such that $A(X(0) = B(X(0)) = X(0)$ and $g(X(t), \gamma'(t)) = 0$ for all t. Notice that this means that $\mathrm{d}f_p(X(0)) = X(L)$ since $A = \tau_\gamma^{-1} \circ \mathrm{d}f_p$ and X is parallel along γ.

There exists a unique geodesic $\beta : (-\epsilon, \epsilon) \to M$ such that $\beta(0) = p$ and $\beta'(0) = X(0)$. This is the geodesic we want to use to build our variation. Moreover, observe that $f \circ \beta$ is a geodesic (since f is an isometry) such that $(f \circ \beta)(0) = f(p)$ and $(f \circ \beta)'(0) = X(L)$. We therefore define a smooth variation of γ by

$$h(s, t) = \exp_{\gamma(t)} \big(sX(t) \big).$$

Then $h(s, 0) = \beta(s)$, $h(s, L) = (f \circ \beta)(s)$ and the variation field $V_h(t) = \frac{\partial h}{\partial s}(0, t) = X(t)$ so $V_h' = V_h'' = 0$.

Applying the general Second variation formula we see that all of the boundary terms vanish (as β and $f \circ \beta$ are geodesics and $V_h' = 0$) and the term involving V_h'' vanishes, so the energy E_h satisfies

$$\frac{1}{2} E_h''(0) = -\int_0^L K(X(t), \gamma'(t)) \mathrm{d}t < 0.$$

Therefore $E_h'(s)$ is decreasing near 0 so there exists a curve α in the variation such that $E(\alpha) < E(\gamma)$. Therefore

$$L(\alpha)^2 \leq LE(\alpha) < LE(\gamma) = L(\gamma)^2$$

(using Lemma 3.17). Now $\alpha(0) = q$ and $\alpha(L) = f(q)$ so $d(q, f(q)) < d(p, f(p))$, but this contradicts the choice of p. □

This concludes the course on Riemannian Geometry. There are many more beautiful results in the subject, particularly involving the interaction of curvature and topology, such as the Sphere Theorem mentioned at the beginning of the course, which we have not been able to cover here in the time. I encourage you to read and learn more about this fantastic topic which is very much at the forefront of current research.