Part III — Riemannian Geometry

Based on lectures by A. G. Kovalev

Notes taken by Dexter Chua

Lent 2017

This course is a possible natural sequel of the course Differential Geometry offered in Michaelmas Term. We shall explore various techniques and results revealing intricate and subtle relations between Riemannian metrics, curvature and topology. I hope to cover much of the following:

A closer look at geodesics and curvature. Brief review from the Differential Geometry course. Geodesic coordinates and Gauss' lemma. Jacobi fields, completeness and the Hopf–Rinow theorem. Variations of energy, Bonnet–Myers diameter theorem and Synge's theorem.

Hodge theory and Riemannian holonomy. The Hodge star and Laplace–Beltrami operator. The Hodge decomposition theorem (with the 'geometry part' of the proof). Bochner–Weitzenböck formulae. Holonomy groups. Interplays with curvature and de Rham cohomology.

Ricci curvature. Fundamental groups and Ricci curvature. The Cheeger–Gromoll splitting theorem.

Pre-requisites

Manifolds, differential forms, vector fields. Basic concepts of Riemannian geometry (curvature, geodesics etc.) and Lie groups. The course Differential Geometry offered in Michaelmas Term is the ideal pre-requisite.

Contents

1 Basics of Riemannian manifolds

This will be a very brief summary of some topics in the Michaelmas Differential Geometry course.

Definition (Riemannian metric)**.** Let M be a smooth manifold. A *Riemannian metric g* on M is an inner product on the tangent bundle TM varying smoothly with the fibers. Formally, this is a global section of $T^*M \otimes T^*M$ that is fiberwise symmetric and positive definite.

The pair (M, g) is called a *Riemannian manifold*.

On every coordinate neighbourhood with coordinates $x = (x_1, \cdots, x_n)$, we can write

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

and we can find the coefficients g_{ij} by

$$g_{ij} = g\left(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j} \right)$$

and are C^∞ functions.

Example. The manifold $\mathbb{R}^k$ has a canonical metric given by the Euclidean metric. In the usual coordinates, g is given by $g_{ij} = \delta_{ij}$.

Does every manifold admit a metric? Recall

Theorem (Whitney embedding theorem)**.** Every smooth manifold M admits an embedding into $\mathbb{R}^k$ for some k. In other words, M is diffeomorphic to a submanifold of $\mathbb{R}^k$. In fact, we can pick k such that $k \leq 2 \dim M$.

Using such an embedding, we can induce a Riemannian metric on M by restricting the inner product from Euclidean space, since we have inclusions $T_p M \hookrightarrow T_p \mathbb{R}^k \cong \mathbb{R}^k$.

More generally,

Lemma. Let (N, h) is a Riemannian manifold, and $F : M \to N$ is an immersion, then the pullback $g = F^*h$ defines a metric on M.

The condition of immersion is required for the pullback to be non-degenerate.

In Differential Geometry, if we do not have metrics, then we tend to consider diffeomorphic spaces as being the same. With metrics, the natural notion of isomorphism is

Definition (Isometry)**.** Let (M, g) and (N, h) be Riemannian manifolds. We say $f : M \to N$ is an *isometry* if it is a diffeomorphism and $f^*h = g$. In other words, for any $p \in M$ and $u, v \in T_p M$, we need

$$h\big((\mathrm{d}f)_p u, (\mathrm{d}f)_p v\big) = g(u, v).$$

Example. Let G be a Lie group. Then for any x, we have translation maps $L_x, R_x : G \to G$ given by

$$L_x(y) = xy$$
$$R_x(y) = yx$$

3

These maps are in fact diffeomorphisms of G.

We already know that G admits a Riemannian metric, but we might want to ask something stronger — does there exist a *left-invariant metric?* In other words, is there a metric such that each L_x is an isometry?

Recall that we previously had techniques to produce left-invariant vector fields, where a vector field X on G is *left-invariant* if for any $x \in G$, we have $\mathrm{d}(L_x)X = X$.

Given a Lie group G, we can define the *Lie algebra* $\mathfrak{g} = T_e G$. Then we can produce left-invariant vector fields by picking some $X_e \in \mathfrak{g}$, and then setting

$$X_a = \mathrm{d}(L_a)X_e.$$

The resulting vector field is indeed smooth, as shown in the differential geometry course.

Similarly, to construct a left-invariant metric, we can just pick a metric at the identity and the propagating it around using left-translation. More explicitly, given any inner product on $\langle \, \cdot \, , \, \cdot \, \rangle$ on $T_e G$, we can define g by

$$g(u, v) = \langle (\mathrm{d}L_{x^{-1}})_x u, (\mathrm{d}L_{x^{-1}})_x v \rangle$$

for all $x \in G$ and $u, v \in T_x G$. The argument for smoothness is similar to that for vector fields.

Of course, everything works when we replace "left" with "right". A Riemannian metric is said to be *bi-invariant* if it is both left- and right-invariant. These are harder to find, but it is a fact that every compact Lie group admits a bi-invariant metric. The basic idea of the proof is to start from a left-invariant metric, then integrate the metric along right translations of all group elements. Here compactness is necessary for the result to be finite.

We will later see that we cannot drop the compactness condition. There are non-compact Lie groups that do not admit bi-invariant metrics, such as $\mathrm{SL}(2, \mathbb{R})$.

Recall that in order to differentiate vectors, or even tensors on a manifold, we needed a connection on the tangent bundle. There is a natural choice for the connection when we are given a Riemannian metric.

Definition (Levi-Civita connection)**.** Let (M, g) be a Riemannian manifold. The *Levi-Civita connection* is the unique connection $\nabla : \Omega^0_M(TM) \to \Omega^1_M(TM)$ on M satisfying

(i) Compatibility with metric:

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

(ii) Symmetry/torsion-free:

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

With a bit more imagination on what the symbols mean, we can write the first property as
$$\mathrm{d}(g(X, Y)) = g(\nabla X, Y) + g(X, \nabla Y),$$
while the second property can be expressed in coordinate representation by

$$\Gamma^i_{jk} = \Gamma^i_{kj},$$

where Γ^i_{jk} is defined by

$$\nabla_{\partial/\partial x_j} \frac{\partial}{\partial x_k} = \sum_{i=1}^{n} \Gamma^i_{jk} \frac{\partial}{\partial x_i}.$$

In fact, the connection allows us to differentiate many more things, and not just tangent vectors.

Firstly, the connection ∇ induces a unique covariant derivative on T^*M, also denoted ∇, given by

$$X\langle \alpha, Y \rangle = \langle \nabla_X \alpha, Y \rangle + \langle \alpha, \nabla_X Y \rangle$$

for any $X, Y \in \mathrm{Vect}(M)$ and $\alpha \in \Omega^1(M)$.

To extend this to a connection ∇ on tensor bundles $\mathcal{T}^{q,p} \equiv (TM)^{\otimes q} \otimes (T^*M)^{\otimes p}$ for any $p, q \geq 0$, we note the following general construction:

In general, suppose we have vector bundles E and F, and $s_1 \in \Gamma(E)$ and $s_2 \in \Gamma(F)$. If we have connections ∇^E and ∇^F on E and F respectively, then we can define

$$\nabla^{E \otimes F}(s_1 \otimes s_2) = (\nabla^E s_1) \otimes s_2 + s_1 \otimes (\nabla^F s_2).$$

Given this machinery, recall that the Riemannian metric is formally a section $g \in \Gamma(T^*M \otimes T^*M)$. Then the compatibility with the metric can be written in the following even more compact form:

$$\nabla g = 0.$$

2   Riemann curvature

With all those definitions out of the way, we now start by studying the notion of *curvature*. For the sake of sanity, we will always use implicit summation convention, i.e. whenever we write something of the form

$$X_{ijk}Y^{i\ell jk},$$

we mean

$$\sum_{i,j} X_{ijk}Y^{i\ell jk}.$$

We will use upper indices to denote contravariant components, so that, for example, X^i refers to the ith component of an element in TM. Dually, covariant components, i.e. those living in the cotangent space, are denoted with lower indices. Note that we always sum an upper index with a lower index.

We will index the basis elements oppositely, e.g. we write $\mathrm{d}x^k$ instead of $\mathrm{d}x_k$ for a basis element of T^*M, so that the indices in expressions of the form $A_k\,\mathrm{d}x^k$ seem to match up. Whenever we do not follow this convention, we will write out summations explicitly.

The definition of the curvature tensor might not seem intuitive at first, but motivation was somewhat given in the III Differential Geometry course, and we will not repeat that.

Definition (Curvature)**.** Let (M, g) be a Riemannian manifold with Levi-Civita connection ∇. The *curvature* 2-*form* is the section

$$R = -\nabla \circ \nabla \in \Gamma(\Lambda^2 T^*M \otimes T^*M \otimes TM) \subseteq \Gamma(T^{1,3}M).$$

This can be thought of as a 2-form with values in $T^*M \otimes TM = \mathrm{End}(TM)$. Given any $X, Y \in \mathrm{Vect}(M)$, we have

$$R(X, Y) \in \Gamma(\mathrm{End}\,TM).$$

The following formula is a straightforward, and also crucial computation:

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

In local coordinates, we can write

$$R = \left(R^i_{j,k\ell}\mathrm{d}x^k\mathrm{d}x^\ell \right)_{i,j=1,\ldots,\dim M} \in \Omega^2_M(\mathrm{End}(TM)).$$

The comma between j and $k\ell$ is purely for artistic reasons. Then we have

$$R(X, Y)^i_j = R^i_{j,k\ell}X^k Y^\ell,$$

where

$$X = \sum X^k \frac{\partial}{\partial x_k}, \quad Y = \sum Y^\ell \frac{\partial}{\partial x_\ell}.$$

To make life easier, we introduce the following short hands:

Notation. In local coordinates, we will write ∂_k for $\frac{\partial}{\partial x_k}$, and $\nabla_k = \nabla_{\partial_k}$

It is often slightly convenient to consider a different form of the Riemann curvature tensor. Instead of having a tensor of type $(1,3)$, we have one of type $(0,4)$ by

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

for $X, Y, Z, T \in T_p M$. In local coordinates, we write this as

$$R_{ij,k\ell} = g_{iq} R^q_{j,k\ell}.$$

The first thing we want to prove is that $R_{ij,k\ell}$ enjoys some symmetries we might not expect:

Proposition.

(i)
$$R_{ij,k\ell} = -R_{ij,\ell k} = -R_{ji,k\ell}.$$

(ii) The *first Bianchi identity*:
$$R^i_{j,k\ell} + R^i_{k,\ell j} + R^i_{\ell,jk} = 0.$$

(iii)
$$R_{ij,k\ell} = R_{k\ell,ij}.$$

Note that the first Bianchi identity can also be written for the $(0,4)$ tensor as

$$R_{ij,k\ell} + R_{ik,\ell j} + R_{i\ell,jk} = 0.$$

We'll first prove these, and discuss a little bit more afterwards.

Proof.

(i) The first equality is obvious as coefficients of a 2-form. For the second equality, we note that the compatibility of the connection with the metric means
$$\frac{\partial g_{ij}}{\partial x_k} = g(\nabla_k \partial_i, \partial_j) + g(\partial_i, \nabla_k \partial_j).$$

We take a partial derivative, say with respect to x_ℓ, to obtain

$$\frac{\partial^2 g_{ij}}{\partial x_\ell \partial x_k} = g(\nabla_\ell \nabla_k \partial_i, \partial_j) + g(\nabla_k \partial_i, \nabla_\ell \partial_j) + g(\nabla_\ell \partial_i, \nabla_k \partial_j) + g(\partial_i, \nabla_\ell \nabla_k \partial_j).$$

Then we know

$$0 = \frac{\partial^2 g}{\partial x_\ell \partial x_k} - \frac{\partial^2 g}{\partial x_k \partial x_\ell} = g([\nabla_\ell, \nabla_k] \partial_i, \partial_j) + g(\partial_i, [\nabla_\ell, \nabla_k] \partial_j).$$

But we know

$$R(\partial_k, \partial_\ell) = \nabla_{[\partial_k, \partial_\ell]} - [\nabla_k, \nabla_\ell] = -[\nabla_k, \nabla_\ell].$$

Writing $R_{k\ell} = R(\partial_k, \partial_\ell)$, we have

$$0 = g(R_{k\ell} \partial_i, \partial_j) + g(\partial_i, R_{k\ell} \partial_j) = R_{ji,k\ell} + R_{ij,k\ell}.$$

So we are done.

(ii) We now prove the first Bianchi identity. Recall

$$R^i_{j,k\ell} = (R_{k\ell}\partial_j)^i = ([\nabla_\ell, \nabla_k]\partial_j)^i.$$

So we have

$$R^i_{j,k\ell} + R^i_{k,\ell j} + R^i_{\ell,jk}$$
$$= [(\nabla_\ell\nabla_k\partial_j - \nabla_k\nabla_\ell\partial_j) + (\nabla_j\nabla_\ell\partial_k - \nabla_\ell\nabla_j\partial_k) + (\nabla_k\nabla_j\partial_\ell - \nabla_j\nabla_k\partial_\ell)]^i.$$
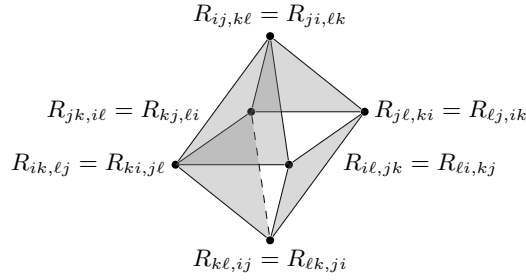
We claim that

$$\nabla_\ell\nabla_k\partial_j - \nabla_\ell\nabla_j\partial_k = 0.$$

Indeed, by definition, we have

$$(\nabla_k\partial_j)^q = \Gamma^q_{kj} = \Gamma^q_{jk} = (\nabla_j\partial_k)^q.$$

The other terms cancel similarly, and we get 0 as promised.

(iii) Consider the following octahedron:

$$R_{ij,k\ell} = R_{ji,\ell k}$$

$$R_{jk,i\ell} = R_{kj,\ell i} \qquad\qquad R_{j\ell,ki} = R_{\ell j,ik}$$

$$R_{ik,\ell j} = R_{ki,j\ell} \qquad\qquad R_{i\ell,jk} = R_{\ell i,kj}$$

$$R_{k\ell,ij} = R_{\ell k,ji}$$

The equalities on each vertex is given by (i). By the first Bianchi identity, for each greyed triangle, the sum of the three vertices is zero.

Now looking at the upper half of the octahedron, adding the two greyed triangles shows us the sum of the vertices in the horizontal square is $(-2)R_{ij,k\ell}$. Looking at the bottom half, we find that the sum of the vertices in the horizontal square is $(-2)R_{k\ell,ij}$. So we must have

$$R_{ij,k\ell} = R_{k\ell,ij}. \qquad\qquad\qquad \square$$

What exactly are the properties of the Levi-Civita connection that make these equality works? The first equality of (i) did not require anything. The second equality of (i) required the compatibility with the metric, and (ii) required the symmetric property. The last one required both properties.

We can express the last property as saying $R_{ij,k\ell}$ is a symmetric bilinear form on $\Lambda^2 T^*_p M$. It is a straightforward exercise to rewrite the statement of this proposition in coordinate free notation using test vector fields. For example, the first equality in (i) can be written as

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

Sectional curvature

The full curvature tensor is rather scary. So it is convenient to obtain some simpler quantities from it. Recall that if we had tangent vectors X, Y, then we can form

$$|X \wedge Y| = \sqrt{g(X, X)g(Y, Y) - g(X, Y)^2},$$

which is the area of the parallelogram spanned by X and Y. We now define

$$K(X, Y) = \frac{R(X, Y, X, Y)}{|X \wedge Y|^2}.$$

Note that this is invariant under (non-zero) scaling of X or Y, and is symmetric in X and Y. Finally, it is also invariant under the transformation $(X, Y) \mapsto (X + \lambda Y, Y)$.

But it is an easy linear algebra fact that these transformations generate all isomorphism from a two-dimensional vector space to itself. So $K(X, Y)$ depends only on the 2-plane spanned by X, Y. So we have in fact defined a function on the Grassmannian of 2-planes, $K : \mathrm{Gr}(2, T_p M) \to \mathbb{R}$. This is called the *sectional curvature* (of g).

It turns out the sectional curvature determines the Riemann curvature tensor completely!

Lemma. Let V be a real vector space of dimension ≥ 2. Suppose $R', R'' : V^{\otimes 4} \to \mathbb{R}$ are both linear in each factor, and satisfies the symmetries we found for the Riemann curvature tensor. We define $K', K'' : \mathrm{Gr}(2, V) \to \mathbb{R}$ as in the sectional curvature. If $K' = K''$, then $R' = R''$.

This is really just linear algebra.

Proof. For any $X, Y, Z \in V$, we know

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

Using linearity of R' and R'', and cancelling equal terms on both sides, we find

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

Now using the symmetry property of R' and R'', this implies

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

Similarly, we replace Y with $Y + T$, and then we get

$$R'(X, Y, Z, T) + R'(X, T, Z, Y) = R''(X, Y, Z, Y) + R'''(X, T, Z, Y).$$

We then rearrange and use the symmetries to get

$$R'(X, Y, Z, T) - R''(X, Y, Z, T) = R'(Y, Z, X, T) - R''(Y, Z, X, T).$$

We notice this equation says $R'(X, Y, Z, T) - R''(X, Y, Z, T)$ is invariant under the cyclic permutation $X \to Y \to Z \to X$. So by the first Bianchi identity, we have

$$3(R'(X, Y, Z, T) - R''(X, Y, Z, T)) = 0.$$

So we must have $R' = R''$.   □

Corollary. Let (M, g) be a manifold such that for all p, the function $K_p : \mathrm{Gr}(2, T_pM) \to \mathbb{R}$ is a constant map. Let

$$R_p^0(X, Y, Z, T) = g_p(X, Z)g_p(Y, T) - g_p(X, T)g_p(Y, Z).$$

Then

$$R_p = K_p R_p^0.$$

Here K_p is just a real number, since it is constant. Moreover, K_p is a smooth function of p.

Proof. We apply the previous lemma as follows: we define $R' = K_p R_p^0$ and $R'' = R_p$. It is a straightforward inspection to see that this R^0 does follow the symmetry properties of R_p, and that they define the same sectional curvature. So $R'' = R'$. We know K_p is smooth in p as both g and R are smooth. $\square$

We can further show that if $\dim M > 2$, then K_p is in fact independent of p under the hypothesis of this function, and the proof requires a second Bianchi identity. This can be found on the first example sheet.

When does the condition of the corollary hold? We pick an *orthonormal* basis, and express R in coordinates $R_{ij,k\ell}$. Then writing $k = K_p$, the hypothesis holds iff

$$R_{ij,k\ell} = k(\delta_{ik}\delta_{j\ell} - \delta_{i\ell}\delta_{jk})$$

iff

$$R_{ij,ij} = -R_{ij,ji} = k,$$

and all other entries all zero.

Other curvatures

There are other quantities we can extract out of the curvature, which will later be useful.

Definition (Ricci curvature). The *Ricci curvature* of g at $p \in M$ is

$$\mathrm{Ric}_p(X, Y) = \mathrm{tr}(v \mapsto R_p(X, v)Y).$$

In terms of coordinates, we have

$$\mathrm{Ric}_{ij} = R_{i,jq}^q = g^{pq}R_{pi,jq},$$

where g^{pq} denotes the inverse of g, and we implicitly sum over repeated indices.

This Ric is a symmetric bilinear form on T_pM. This can be determined by the quadratic form

$$\mathrm{Ric}(X) = \frac{1}{n-1}\mathrm{Ric}_p(X, X).$$

The coefficient $\frac{1}{n-1}$ is just a convention.

There are still two indices we can contract, and we can define

Definition (Scalar curvature). The *scalar curvature* of g is the trace of Ric respect to g. Explicitly, this is defined by

$$s = g^{ij}\mathrm{Ric}_{ij} = g^{ij}R_{i,jq}^q = R^{qi}{}_{iq}.$$

Sometimes a convention is to define the scalar curvature as $\frac{s}{n(n-1)}$ instead. In the case of a constant sectional curvature tensor, we have

$$\mathrm{Ric}_p = (n-1)K_p g_p,$$

and

$$s(p) = n(n-1)K_p.$$

Low dimensions

If $n = 2$, i.e. we have surfaces, then the Riemannian metric g is also known as the *first fundamental form*, and it is usually written as

$$g = E\,\mathrm{d}u^2 + 2F\,\mathrm{d}u\,\mathrm{d}v + G\,\mathrm{d}v^2.$$

Up to the symmetries, the only non-zero component of the curvature tensor is $R_{12,12}$, and using the definition of the scalar curvature, we find

$$R_{12,12} = \frac{1}{2}s(EG - F^2).$$

Thus $s/2$ is also the sectional curvature (there can only be one plane in the tangent space, so the sectional curvature is just a number). One can further check that

$$\frac{s}{2} = K = \frac{LN - M^2}{EG - F^2},$$

the *Gaussian curvature*. We see that in two dimensions, there is only one independent component of the Gaussian curvature. Also, $R_{12,21}$ is the determinant of the second fundamental form.

If $n = 3$, one can check that $R(g)$ is determined by the Ricci curvature.

3 Geodesics

3.1 Definitions and basic properties

We will eventually want to talk about geodesics. However, the set up we need to write down the definition of geodesics can be done in a much more general way, and we will do that.

The general setting is that we have a vector bundle $\pi : E \to M$.

Definition (Lift)**.** Let $\pi : E \to M$ be a vector bundle with typical fiber V. Consider a curve $\gamma : (-\varepsilon, \varepsilon) \to M$. A *lift* of γ is a map $\gamma^E : (-\varepsilon, \varepsilon) \to E$ if $\pi \circ \gamma^E = \gamma$, i.e. the following diagram commutes:

$$
\begin{array}{ccc}
 & & E \\
 & \overset{\gamma^E}{\nearrow} & \downarrow \pi \\
(-\varepsilon, \varepsilon) & \overset{\gamma}{\longrightarrow} & M
\end{array} \;.
$$

For $p \in M$, we write $E_p = \pi^{-1}(\{p\}) \cong V$ for the fiber above p. We can think of E_p as the space of some "information" at p. For example, if $E = TM$, then the "information" is a tangent vector at p. In physics, the manifold M might represent our universe, and a point in E_p might be the value of the electromagnetic field at p.

Thus, given a path γ in M, a lift corresponds to providing that piece of "information" at each point along the curve. For example, if $E = TM$, then we can canonically produce a lift of γ, given by taking the derivative of γ at each point.

Locally, suppose we are in some coordinate neighbourhood $U \subseteq M$ such that E is trivial on U. After picking a trivialization, we can write our lift as Then γ^E is of the form

$$
\gamma^E(t) = (\gamma(t), a(t))
$$

for some function $a : (-\varepsilon, \varepsilon) \to V$.

One thing we would want to do with such lifts is to differentiate them, and see how it changes along the curve. When we have a *section* of E on the whole of M (or even just an open neighbourhood), rather than just a lift along a curve, the connection provides exactly the information needed to do so. It is not immediately obvious that the connection also allows us to differentiate curves along paths, but it does.

Proposition. Let $\gamma : (-\varepsilon, \varepsilon) \to M$ be a curve. Then there is a uniquely determined operation $\frac{\nabla}{\mathrm{d}t}$ from the space of all lifts of γ to itself, satisfying the following conditions:

(i) For any $c, d \in \mathbb{R}$ and lifts $\tilde{\gamma}^E, \gamma^E$ of γ, we have.

$$
\frac{\nabla}{\mathrm{d}t}(c\gamma^E + d\tilde{\gamma}^E) = c\frac{\nabla\gamma^E}{\mathrm{d}t} + d\frac{\nabla\tilde{\gamma}^E}{\mathrm{d}t}
$$

(ii) For any lift γ^E of γ and function $f : (-\varepsilon, \varepsilon) \to \mathbb{R}$, we have

$$
\frac{\nabla}{\mathrm{d}t}(f\gamma^E) = \frac{\mathrm{d}f}{\mathrm{d}t} + f\frac{\nabla\gamma^E}{\mathrm{d}t}.
$$

(iii) If there is a local section s of E and a local vector field V on M such that

$$\gamma^E(t) = s(\gamma(t)), \quad \dot{\gamma}(t) = V(\gamma(t)),$$

then we have

$$\frac{\nabla \gamma^E}{\mathrm{d}t} = (\nabla_V s) \circ \gamma.$$

Locally, this is given by

$$\left(\frac{\nabla \gamma^E}{\mathrm{d}t} \right)^i = \dot{a}^i + \Gamma^i_{jk} a^j \dot{x}^k.$$

The proof is straightforward — one just checks that the local formula works, and the three properties force the operation to be locally given by that formula.

Definition (Covariant derivative)**.** The uniquely defined operation in the proposition above is called the *covariant derivative.*

In some sense, lifts that have vanishing covariant derivative are "constant" along the map.

Definition (Horizontal lift)**.** Let ∇ be a connection on E with $\Gamma^i_{jk}(x)$ the coefficients in a local trivialization. We say a lift γ^E is *horizontal* if

$$\frac{\nabla \gamma^E}{\mathrm{d}t} = 0.$$

Since this is a linear first-order ODE, we know that for a fixed γ, given any initial $a(0) \in E_{\gamma(0)}$, there is a unique way to obtain a horizontal lift.

Definition (Parallel transport)**.** Let $\gamma : [0, 1] \to M$ be a curve in M. Given any $a_0 \in E_{\gamma(0)}$, the unique horizontal lift of γ with $\gamma^E(0) = (\gamma(0), a_0)$ is called the *parallel transport* of a_0 along $\gamma(0)$. We sometimes also call $\gamma^E(1)$ the parallel transport.

Of course, we want to use this general theory to talk about the case where M is a Riemannian manifold, $E = TM$ and ∇ is the Levi-Civita connection of g. In this case, each curve $\gamma(t)$ has a canonical lift independent of the metric or connection given simply by taking the derivative $\dot{\gamma}(t)$.

Definition (Geodesic)**.** A curve $\gamma(t)$ on a Riemannian manifold (M, g) is called a *geodesic curve* if its canonical lift is horizontal with respect to the Levi-Civita connection. In other words, we need

$$\frac{\nabla \dot{\gamma}}{\mathrm{d}t} = 0.$$

In local coordinates, we write this condition as

$$\ddot{x}_i + \Gamma^i_{jk} \dot{x}^j \dot{x}^k = 0.$$

This time, we obtain a second-order ODE. So a geodesic is uniquely specified by the initial conditions $p = x(0)$ and $a = \dot{x}(0)$. We will denote the resulting geodesic as $\gamma_p(t; a)$, where t is the time coordinate as usual.

Since we have a non-linear ODE, existence is no longer guaranteed on all time, but just for some interval $(-\varepsilon, \varepsilon)$. Of course, we still have uniqueness of solutions.

We now want to prove things about geodesics. To do so, we will need to apply some properties of the covariant derivative we just defined. Since we are lazy, we would like to reuse results we already know about the covariant derivative for vector fields. The trick is to notice that locally, we can always extend $\dot\gamma$ to a vector field.

Indeed, we work in some coordinate chart around $\gamma(0)$, and we wlog assume

$$\dot\gamma(0) = \frac{\partial}{\partial x_1}.$$

By the inverse function theorem, we note that $x_1(t)$ is invertible near 0, and we can write $t = t(x_1)$ for small x_1. Then in this neighbourhood of 0, we can view x_k as a function of x_1 instead of t. Then we can define the vector field
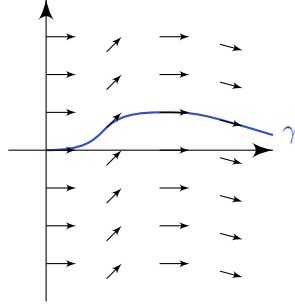
$$\underline{\dot\gamma}(x_1, \cdots, x_k) = \dot\gamma(x_1, x_2(x_1), \cdots, x_k(x_1)).$$

By construction, this agrees with $\dot\gamma$ along the curve.

Using this notation, the geodesic equation can be written as

$$\nabla_{\underline{\dot\gamma}} \underline{\dot\gamma} \Big|_{\gamma(t)} = 0,$$

where the ∇ now refers to the covariant derivative of vector fields, i.e. the connection itself.

Using this, a lot of the desired properties of geodesics immediately follow from well-known properties of the covariant derivative. For example,

Proposition. If γ is a geodesic, then $|\dot\gamma(t)|_g$ is constant.

Proof. We use the extension $\underline{\dot\gamma}$ around $p = \gamma(0)$, and stop writing the underlines. Then we have
$$\dot\gamma(g(\dot\gamma, \dot\gamma)) = g(\nabla_{\dot\gamma} \dot\gamma, \dot\gamma) + g(\dot\gamma, \nabla_{\dot\gamma} \dot\gamma) = 0,$$
which is valid at each $q = \gamma(t)$ on the curve. But at each q, we have

$$\dot\gamma(g(\dot\gamma, \dot\gamma)) = \dot x^k \frac{\partial}{\partial x_k} g(\dot\gamma, \dot\gamma) = \frac{\mathrm{d}}{\mathrm{d}t} |\dot\gamma(t)|_g^2$$

by the chain rule. So we are done. $\qquad\square$

At this point, it might be healthy to look at some examples of geodesics.

Example. In $\mathbb{R}^n$ with the Euclidean metric, we have $\Gamma^i_{jk} = 0$. So the geodesic equation is

$$\ddot{x}_k = 0.$$

So the geodesics are just straight lines.

Example. On a sphere S^n with the usual metric induced by the standard embedding $S^n \hookrightarrow \mathbb{R}^{n+1}$. Then the geodesics are great circles.

To see this, we may wlog $p = e_0$ and $a = e_1$, for a standard basis $\{e_i\}$ of $\mathbb{R}^{n+1}$. We can look at the map

$$\varphi : (x_0, \cdots, x_n) \mapsto (x_0, x_1, -x_2, \cdots, -x_n),$$

and it is clearly an isometry of the sphere. Therefore it preserves the Riemannian metric, and hence sends geodesics to geodesics. Since it also preserves p and a, we know $\varphi(\gamma) = \gamma$ by uniqueness. So it must be contained in the great circle lying on the plane spanned by e_0 and e_1.

Lemma. Let $p \in M$, and $a \in T_p M$. As before, let $\gamma_p(t, a)$ be the geodesic with $\gamma(0) = p$ and $\dot{\gamma}(0) = p$. Then

$$\gamma_p(\lambda t, a) = \gamma_p(t, \lambda a),$$

and in particular is a geodesic.

Proof. We apply the chain rule to get

$$\frac{\mathrm{d}}{\mathrm{d}t}\gamma(\lambda t, a) = \lambda \dot{\gamma}(\lambda t, a)$$
$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\gamma(\lambda t, a) = \lambda^2 \ddot{\gamma}(\lambda t, a).$$

So $\gamma(\lambda t, a)$ satisfies the geodesic equations, and have initial velocity λa. Then we are done by uniqueness of ODE solutions. $\qquad\square$

Thus, instead of considering $\gamma_p(t, a)$ for arbitrary t and a, we can just fix $t = 1$, and look at the different values of $\gamma_p(1, a)$. By ODE theorems, we know this depends smoothly on a, and is defined on some open neighbourhood of $0 \in T_p M$.

Definition (Exponential map)**.** Let (M, g) be a Riemannian manifold, and $p \in M$. We define $\exp_p$ by

$$\exp_p(a) = \gamma(1, a) \in M$$

for $a \in T_p M$ whenever this is defined.

We know this function has domain at least some open ball around $0 \in T_p M$, and is smooth. Also, by construction, we have $\exp_p(0) = p$.

In fact, the exponential map gives us a chart around p locally, known as *geodesic local coordinates*. To do so, it suffices to note the following rather trivial proposition.

Proposition. We have
$$(\mathrm{d}\exp_p)_0 = \mathrm{id}_{T_pM},$$
where we identify $T_0(T_pM) \cong T_pM$ in the natural way.

All this is saying is if you go in the direction of $a \in T_pM$, then you go in the direction of a.

Proof.
$$(\mathrm{d}\exp_p)_0(v) = \frac{\mathrm{d}}{\mathrm{d}t}\exp_p(tv) = \frac{\mathrm{d}}{\mathrm{d}t}\gamma(1, tv) = \frac{\mathrm{d}}{\mathrm{d}t}\gamma(t, v) = v.$$

□

Corollary. $\exp_p$ maps an open ball $B(0, \delta) \subseteq T_pM$ to $U \subseteq M$ diffeomorphically for some $\delta > 0$.

Proof. By the inverse mapping theorem. □

This tells us the inverse of the exponential map gives us a chart of M around p. These coordinates are often known as *geodesic local coordinates*.

In these coordinates, the geodesics from p have the very simple form
$$\gamma(t, a) = ta$$
for all $a \in T_pM$ and t sufficiently small that this makes sense.

Now consider a linear isometry $(T_pN, g(p)) \cong (\mathbb{R}^n, \mathrm{eucl})$, and under this identification, we have a map
$$(r, \mathbf{v}) \in (0, \delta) \times S^{n-1} \mapsto \exp_p(r\mathbf{v}) \in M^n.$$

This chart is known as *geodesic polar coordinates*. For each fixed r, the image of this map is called a *geodesic sphere* of geodesic radius r, written Σ_r. This is an embedded submanifold of M.

Note that in geodesic local coordinates, the metric at $0 \in T_pN$ is given by the Euclidean metric. However, the metric at other points can be complicated. Fortunately, Gauss' lemma says it is not *too* complicated.

Theorem (Gauss' lemma). The geodesic spheres are perpendicular to their radii. More precisely, $\gamma_p(t, a)$ meets every Σ_r orthogonally, whenever this makes sense. Thus we can write the metric in geodesic polars as
$$g = \mathrm{d}r^2 + h(r, \mathbf{v}),$$
where for each r, we have
$$h(r, \mathbf{v}) = g|_{\Sigma_r}.$$
In matrix form, we have
$$g = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & h & \\ 0 & & & \end{pmatrix}$$

The proof is not hard, but it involves a few subtle points.

Proof. We work in geodesic coordinates. It is clear that $g(\partial_r, \partial_r) = 1$.

Consider an arbitrary vector field $X = X(\mathbf{v})$ on S^{n-1}. This induces a vector field on some neighbourhood $B(0, \delta) \subseteq T_p M$ by

$$\tilde{X}(r\mathbf{v}) = X(\mathbf{v}).$$

Pick a direction $\mathbf{v} \in T_p M$, and consider the unit speed geodesic γ in the direction of $\mathbf{v}$. We define

$$G(r) = g(\tilde{X}(r\mathbf{v}), \dot{\gamma}(r)) = g(\tilde{X}, \dot{\gamma}(r)).$$

We begin by noticing that

$$\nabla_{\partial_r} \tilde{X} - \nabla_{\tilde{X}} \partial_r = [\partial_r, \tilde{X}] = 0.$$

So we have

$$\frac{\mathrm{d}}{\mathrm{d}r} G(r) = g(\nabla_{\dot{\gamma}} \tilde{X}, \dot{\gamma}) + g(\tilde{X}, \nabla_{\dot{\gamma}} \dot{\gamma}).$$

We know the second term vanishes, since γ is a geodesic. Noting that $\dot{\gamma} = \frac{\partial}{\partial r}$, we know the first term is equal to

$$g(\nabla_{\tilde{X}} \partial_r, \partial_r) = \frac{1}{2} \Big(g(\nabla_{\tilde{X}} \partial_r, \partial_r) + g(\partial_r, \nabla_{\tilde{X}} \partial_r) \Big) = \frac{1}{2} \tilde{X}(g(\partial_r, \partial_r)) = 0,$$

since we know that $g(\partial_r, \partial_r) = 1$ constantly.

Thus, we know $G(r)$ is constant. But $G(0) = 0$ since the metric at 0 is the Euclidean metric. So G vanishes everywhere, and so ∂_r is perpendicular to Σ_g. $\qquad \square$

Corollary. Let $a, w \in T_p M$. Then

$$g((\mathrm{d} \exp_p)_a a, (\mathrm{d} \exp_p)_a w) = g(a, w)$$

whenever a lives in the domain of the geodesic local neighbourhood.

3.2   Jacobi fields

Fix a Riemannian manifold M. Let's imagine that we have a "manifold" of all smooth curves on M. Then this "manifold" has a "tangent space". Morally, given a curve γ, a "tangent vector" at γ in the space of curve should correspond to providing a tangent vector (in M) at each point along γ:
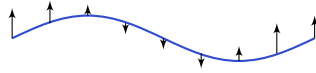

Since we are interested in the geodesics only, we consider the "submanifold" of geodesics curves. What are the corresponding "tangent vectors" living in this "submanifold"?

In rather more concrete terms, suppose $f_s(t) = f(t, s)$ is a family of geodesics in M indexed by $s \in (-\varepsilon, \varepsilon)$. What do we know about $\frac{\partial f}{\partial s}\big|_{s=0}$, a vector field along f_0?

We begin by considering such families that fix the starting point $f(0, s)$, and then derive some properties of $\frac{\partial f}{\partial s}$ in these special cases. We will then define a

Jacobi field to be any vector field along a curve that satisfies these properties. We will then prove that these are exactly the variations of geodesics.

Suppose $f(t, s)$ is a family of geodesics such that $f(0, s) = p$ for all s. Then in geodesics local coordinates, it must look like this:
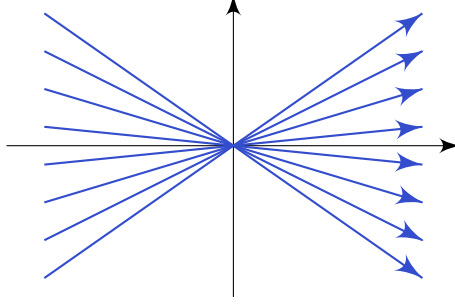
For a fixed p, such a family is uniquely determined by a function

$$a(s) : (-\varepsilon, \varepsilon) \to T_p M$$

such that

$$f(t, s) = \exp_p(ta(s)).$$

The initial conditions of this variation can be given by $a(0) = a$ and

$$\dot{a}(0) = w \in T_a(T_p M) \cong T_p M.$$

We would like to know the "variation field" of $\gamma(t) = f(t, 0) = \gamma_p(t, a)$ this induces. In other words, we want to find $\frac{\partial f}{\partial s}(t, 0)$. This is not hard. It is just given by

$$(\mathrm{d}\exp_p)_{ta_0}(tw) = \frac{\partial f}{\partial s}(t, 0),$$

As before, to prove something about f, we want to make good use of the properties of ∇. Locally, we extend the vectors $\frac{\partial f}{\partial s}$ and $\frac{\partial f}{\partial t}$ to vector fields $\frac{\partial}{\partial t}$ and $\frac{\partial}{\partial s}$. Then in this set up, we have

$$\dot{\gamma} = \frac{\partial f}{\partial t} = \frac{\partial}{\partial t}.$$

Note that in $\frac{\partial f}{\partial t}$, we are differentiating f with respect to t, whereas the $\frac{\partial}{\partial t}$ on the far right is just a formal expressions.

By the geodesic equation, we have

$$0 = \frac{\nabla}{\mathrm{d}t}\dot{\gamma} = \Delta_{\partial_t}\partial_t.$$

Therefore, using the definition of the curvature tensor R, we obtain

$$0 = \nabla_{\partial_s}\nabla_{\partial_t}\frac{\partial}{\partial t} = \nabla_{\partial_t}\nabla_{\partial_s}\partial_t - R(\partial_s, \partial_t)\partial_t$$

$$= \nabla_{\partial_t}\nabla_{\partial_s}\partial_t + R(\partial_t, \partial_s)\partial_t$$

We let this act on the function f. So we get

$$0 = \frac{\nabla}{\mathrm{d}t}\frac{\nabla}{\mathrm{d}s}\frac{\partial f}{\partial t} + R(\partial_t, \partial_s)\frac{\partial f}{\partial t}.$$

We write

$$J(t) = \frac{\partial f}{\partial s}(t, 0),$$

which is a vector field along the geodesic γ. Using the fact that

$$\frac{\nabla}{\mathrm{d}s}\frac{\partial f}{\partial t} = \frac{\nabla}{\mathrm{d}t}\frac{\partial f}{\partial s},$$

we find that J must satisfy the ordinary differential equation

$$\frac{\nabla^2}{\mathrm{d}t^2}J + R(\dot{\gamma}, J)\dot{\gamma} = 0.$$

This is a linear second-order ordinary differential equation.

Definition (Jacobi field)**.** Let $\gamma : [0, L] \to M$ be a geodesic. A *Jacobi field* is a vector field J along γ that is a solution of the *Jacobi equation* on $[0, L]$

$$\frac{\nabla^2}{\mathrm{d}t^2}J + R(\dot{\gamma}, J)\dot{\gamma} = 0. \tag{$\dagger$}$$

We now embark on a rather technical journey to prove results about Jacobi fields. Observe that $\dot{\gamma}(t)$ and $t\dot{\gamma}(t)$ both satisfy this equation, rather trivially.

Theorem. Let $\gamma : [0, L] \to N$ be a geodesic in a Riemannian manifold (M, g). Then

(i) For any $u, v \in T_{\gamma(0)}M$, there is a unique Jacobi field J along Γ with

$$J(0) = u, \quad \frac{\nabla J}{\mathrm{d}t}(0) = v.$$

If

$$J(0) = 0, \quad \frac{\nabla J}{\mathrm{d}t}(0) = k\dot{\gamma}(0),$$

then $J(t) = kt\dot{\gamma}(t)$. Moreover, if both $J(0), \frac{\nabla J}{\mathrm{d}t}(0)$ are orthogonal to $\dot{\gamma}(0)$, then $J(t)$ is perpendicular to $\dot{\gamma}(t)$ for all $[0, L]$.

In particular, the vector space of all Jacobi fields along γ have dimension $2n$, where $n = \dim M$.

The subspace of those Jacobi fields pointwise perpendicular to $\dot{\gamma}(t)$ has dimensional $2(n-1)$.

(ii) $J(t)$ is independent of the parametrization of $\dot{\gamma}(t)$. Explicitly, if $\tilde{\gamma}(t) = \tilde{\gamma}(\lambda t)$, then $\tilde{J}$ with the same initial conditions as J is given by

$$\tilde{J}(\tilde{\gamma}(t)) = J(\gamma(\lambda t)).$$

This is the kind of theorem whose statement is longer than the proof.

Proof.

(i) Pick an orthonormal basis $e_1, \cdots, e_n$ of $T_p M$, where $p = \gamma(0)$. Then parallel transports $\{X_i(t)\}$ via the Levi-Civita connection preserves the inner product.

We take e_1 to be parallel to $\dot\gamma(0)$. By definition, we have

$$X_i(0) = e_i, \quad \frac{\nabla X_i}{\mathrm{d}t} = 0.$$

Now we can write

$$J = \sum_{i=1}^{n} y_i X_i.$$

Then taking $g(X_i, \cdot)$ of (†) , we find that

$$\ddot{y}_i + \sum_{j=2}^{n} R(\dot\gamma, X_j, \dot\gamma, X_i) y_j = 0.$$

Then the claims of the theorem follow from the standard existence and uniqueness of solutions of differential equations.

In particular, for the orthogonality part, we know that $J(0)$ and $\frac{\nabla J}{\mathrm{d}t}(0)$ being perpendicular to $\dot\gamma$ is equivalent to $y_1(0) = \dot{y}_1(0) = 0$, and then Jacobi's equation gives

$$\ddot{y}_1(t) = 0.$$

(ii) This follows from uniqueness.   □

Our discussion of Jacobi fields so far has been rather theoretical. Now that we have an explicit equation for the Jacobi field, we can actually produce some of them. We will look at the case where we have constant sectional curvature.

Example. Suppose the sectional curvature is constantly $K \in \mathbb{R}$, for $\dim M \geq 3$. We wlog $|\dot\gamma| = 1$. We let J along γ be a Jacobi field, normal to $\dot\gamma$.

Then for any vector field T along γ, we have

$$\langle R(\dot\gamma, J)\dot\gamma, T \rangle = K(g(\dot\gamma, \dot\gamma)g(J, T) - g(\dot\gamma, J)g(\dot\gamma, T)) = Kg(J, T).$$

Since this is true for all T, we know

$$R(\dot\gamma, J)\dot\gamma = KJ.$$

Then the Jacobi equation becomes

$$\frac{\nabla^2}{\mathrm{d}t^2} J + KJ = 0.$$

So we can immediately write down a collection of solutions

$$J(t) = \begin{cases} \frac{\sin(t\sqrt{K})}{\sqrt{K}} X_i(t) & K > 0 \\ t X_i(t) & K = 0 \\ \frac{\sinh(t\sqrt{-K})}{\sqrt{-K}} X_i(t) & K < 0 \end{cases}.$$

for $i = 2, \cdots, n$, and this has initial conditions

$$J(0) = 0, \quad \frac{\nabla J}{\mathrm{d}t}(0) = e_i.$$

Note that these Jacobi fields vanishes at 0.

We can now deliver our promise, proving that Jacobi fields are precisely the variations of geodesics.

Proposition. Let $\gamma : [a, b] \to M$ be a geodesic, and $f(t, s)$ a variation of $\gamma(t) = f(t, 0)$ such that $f(t, s) = \gamma_s(t)$ is a geodesic for all $|s|$ small. Then

$$J(t) = \frac{\partial f}{\partial s}$$

is a Jacobi field along $\dot{\gamma}$.

Conversely, every Jacobi field along γ can be obtained this way for an appropriate function f.

Proof. The first part is just the exact computation as we had at the beginning of the section, but for the benefit of the reader, we will reproduce the proof again.

$$\begin{aligned}
\frac{\nabla^2 J}{\mathrm{d}t} &= \nabla_t \nabla_t \frac{\partial f}{\partial s} \\
&= \nabla_t \nabla_s \frac{\partial f}{\partial t} \\
&= \nabla_s \left(\nabla_t \frac{\partial f}{\partial t} \right) - R(\partial_t, \partial_s)\dot{\gamma}_s.
\end{aligned}$$

We notice that the first term vanishes, because $\nabla_t \frac{\partial f}{\partial t} = 0$ by definition of geodesic. So we find

$$\frac{\nabla^2 J}{\mathrm{d}t} = -R(\dot{\gamma}, J)\dot{\gamma},$$

which is the Jacobi equation.

The converse requires a bit more work. We will write $J'(0)$ for the covariant derivative of J along γ. Given a Jacobi field J along a geodesic $\gamma(t)$ for $t \in [0, L]$, we let $\tilde{\gamma}$ be another geodesic such that

$$\tilde{\gamma}(0) = \gamma(0), \quad \dot{\tilde{\gamma}}(0) = J(0).$$

We take parallel vector fields X_0, X_1 along $\tilde{\gamma}$ such that

$$X_0(0) = \dot{\gamma}(0), \quad X_1(0) = J'(0).$$

We put $X(s) = X_0(s) + sX_1(s)$. We put

$$f(t, s) = \exp_{\tilde{\gamma}(s)}(tX(s)).$$

In local coordinates, for each fixed s, we find

$$f(t, s) = \tilde{\gamma}(s) + tX(s) + O(t^2)$$

as $t \to 0$. Then we define

$$\gamma_s(t) = f(t, s)$$

whenever this makes sense. This depends smoothly on s, and the previous arguments say we get a Jacobi field

$$\hat{J}(t) = \frac{\partial f}{\partial s}(t, 0)$$

We now want to check that $\hat{J} = J$. Then we are done. To do so, we have to check the initial conditions. We have

$$\hat{J}(0) = \frac{\partial f}{\partial s}(0,0) = \frac{\mathrm{d}\tilde{\gamma}}{\mathrm{d}s}(0) = J(0),$$

and also

$$\hat{J}'(0) = \frac{\nabla}{\mathrm{d}t}\frac{\partial f}{\partial s}(0,0) = \frac{\nabla}{\mathrm{d}s}\frac{\partial f}{\partial t}(0,0) = \frac{\nabla X}{\mathrm{d}s}(0) = X_1(0) = J'(0).$$
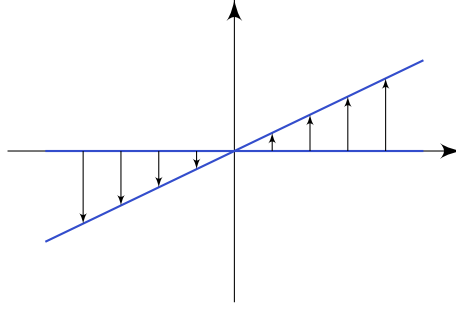
So we have $\hat{J} = J$. $\qquad\square$

Corollary. Every Jacobi field J along a geodesic γ with $J(0) = 0$ is given by

$$J(t) = (\mathrm{d}\exp_p)_{t\dot\gamma(0)}(tJ'(0))$$

for all $t \in [0, L]$.

This is just a reiteration of the fact that if we pull back to the geodesic local coordinates, then the variation must look like this:

But this corollary is stronger, in the sense that it holds even if we get out of the geodesic local coordinates (i.e. when $\exp_p$ no longer gives a chart).

Proof. Write $\dot\gamma(0) = a$, and $J'(0) = w$. By above, we can construct the variation by

$$f(t, s) = \exp_p(t(a + sw)).$$

Then

$$(\mathrm{d}\exp_p)_{t(a+sw)}(tw) = \frac{\partial f}{\partial s}(t, s),$$

which is just an application of the chain rule. Putting $s = 0$ gives the result. $\quad\square$

It can be shown that in the situation of the corollary, if $a \perp w$, and $|a| = |w| = 1$, then

$$|J(t)| = t - \frac{1}{3!}K(\sigma)t^3 + o(t^3)$$

as $t \to 0$, where σ is the plane spanned by a and w.

3.3 Further properties of geodesics

We can now use Jacobi fields to prove interesting things. We now revisit the Gauss lemma, and deduce a stronger version.

Lemma (Gauss' lemma). Let $a, w \in T_p M$, and

$$\gamma = \gamma_p(t, a) = \exp_p(ta)$$

a geodesic. Then

$$g_{\gamma(t)}((\mathrm{d}\exp_p)_{ta} a, (\mathrm{d}\exp_p)_{ta} w) = g_{\gamma(0)}(a, w).$$

In particular, γ is orthogonal to $\exp_p\{v \in T_p M : |v| = r\}$. Note that the latter need not be a submanifold.

This is an improvement of the previous version, which required us to live in the geodesic local coordinates.

Proof. We fix any $r > 0$, and consider the Jacobi field J satisfying

$$J(0) = 0, \quad J'(0) = \frac{w}{r}.$$

Then by the corollary, we know the Jacobi field is

$$J(t) = (\mathrm{d}\exp_p)_{ta}\left(\frac{tw}{r}\right).$$

We may write

$$\frac{w}{r} = \lambda a + u,$$

with $a \perp u$. Then since Jacobi fields depend linearly on initial conditions, we write

$$J(t) = \lambda t \dot\gamma(t) + J_n(t)$$

for a Jacobi field J_n a normal vector field along γ. So we have

$$g(J(r), \dot\gamma(r)) = \lambda r |\dot\gamma(r)|^2 = g(w, a).$$

But we also have

$$g(w, a) = g(\lambda a r + u, a) = \lambda r |a|^2 = \lambda r |\dot\gamma(0)|^2 = \lambda r |\dot\gamma(r)|^2.$$

Now we use the fact that

$$J(r) = (\mathrm{d}\exp_p)_{ra} w$$

and

$$\dot\gamma(r) = (\mathrm{d}\exp_p)_{ra} a,$$

and we are done. □

Corollary (Local minimizing of length). Let $a \in T_p M$. We define $\varphi(t) = ta$, and $\psi(t)$ a piecewise C^1 curve in $T_p M$ for $t \in [0, 1]$ such that

$$\psi(0) = 0, \quad \psi(1) = a.$$

Then

$$\mathrm{length}(\exp_p \circ \psi) \geq \mathrm{length}(\exp_p \circ \varphi) = |a|.$$

It is important to interpret this corollary precisely. It only applies to curves with the same end point *in* T_pM. If we have two curves in T_pM whose end points have the same image in M, then the result need not hold (the torus would be a counterexample).

Proof. We may of course assume that ψ never hits 0 again after $t = 0$. We write

$$\psi(t) = \rho(t)\mathbf{u}(t),$$

where $\rho(t) \geq 0$ and $|\mathbf{u}(t)| = 1$. Then

$$\psi' = \rho'\mathbf{u} + \rho\mathbf{u}'.$$

Then using the extended Gauss lemma, and the general fact that if $\mathbf{u}(t)$ is a unit vector for all t, then $\mathbf{u} \cdot \mathbf{u}' = \frac{1}{2}(\mathbf{u} \cdot \mathbf{u})' = 0$, we have

$$\left|\frac{\mathrm{d}}{\mathrm{d}x}(\exp_p \circ \psi)(t)\right|^2 = \left|(\mathrm{d}\exp_p)_{\psi(t)}\psi'(t)\right|^2$$
$$= \rho'(t)^2 + 2g(\rho'(t)\mathbf{u}(t), \rho(t)\mathbf{u}'(t)) + \rho(t)^2|(\mathrm{d}\exp_p)_{\psi(t)}\mathbf{u}'(t)|^2$$
$$= \rho'(t)^2 + \rho(t)^2|(\mathrm{d}\exp_p)_{\psi(t)}\mathbf{u}'(t)|^2,$$

Thus we have

$$\text{length}(\exp_p \circ \psi) \geq \int_0^1 \rho'(t)\,\mathrm{d}t = \rho(1) - \rho(0) = |a|.$$

$\square$

Notation. We write $\Omega(p, q)$ for the set of all piecewise C^1 curves from p to q.

We now wish to define a metric on M, in the sense of metric spaces.

Definition (Distance)**.** Suppose M is connected, which is the same as it being path connected. Let $(p, q) \in M$. We define

$$d(p, q) = \inf_{\xi \in \Omega(p,q)} \text{length}(\xi),$$

where

To see this is indeed a metric, All axioms of a metric space are obvious, apart from the non-negativity part.

Theorem. Let $p \in M$, and let ε be such that $\exp_p|_{B(0,\varepsilon)}$ is a diffeomorphism onto its image, and let U be the image. Then

– For any $q \in U$, there is a unique geodesic $\gamma \in \Omega(p, q)$ with $\ell(\gamma) < \varepsilon$. Moreover, $\ell(\gamma) = d(p, q)$, and is the unique curve that satisfies this property.

– For any point $q \in M$ with $d(p, q) < \varepsilon$, we have $q \in U$.

– If $q \in M$ is any point, $\gamma \in \Omega(p, q)$ has $\ell(\gamma) = d(p, q) < \varepsilon$, then γ is a geodesic.

Proof. Let $q = \exp_p(a)$. Then the path $\gamma(t) = \exp_p(ta)$ is a geodesic from p to q of length $|a| = r < \varepsilon$. This is clearly the only such geodesic, since $\exp_p|_{B(0,\varepsilon)}$ is a diffeomorphism.

Given any other path $\tilde{\gamma} \in \Omega(p, q)$, we want to show $\ell(\tilde{\gamma}) > \ell(\gamma)$. We let

$$\tau = \sup\left\{ t \in [0, 1] : \gamma([0, t]) \subseteq \exp_p(\overline{B(0, r)}) \right\}.$$

Note that if $\tau \neq 1$, then we must have $\gamma(\tau) \in \Sigma_r$, the geodesic sphere of radius r, otherwise we can continue extending. On the other hand, if $\tau = 1$, then we certainly have $\gamma(\tau) \in \Sigma_r$, since $\gamma(\tau) = q$. Then by local minimizing of length, we have

$$\ell(\tilde{\gamma}) \geq \ell(\tilde{\gamma}_{[0,\tau]}) \geq r.$$

Note that we can always lift $\tilde{\gamma}[0, \tau]$ to a curve from 0 to a in T_pM, since $\exp_p$ is a diffeomorphism in $B(0, \varepsilon)$.

By looking at the proof of the local minimizing of length, and using the same notation, we know that we have equality iff $\tau = 1$ and

$$\rho(t)^2|(\mathrm{d}\exp_p)_{\psi(t)}\psi(t)\mathbf{u}'(t)|^2 = 0$$

for all t. Since $\mathrm{d}\exp_p$ is regular, this requires $\mathbf{u}'(t) = 0$ for all t (since $\rho(t) \neq 0$ when $t \neq 0$, or else we can remove the loop to get a shorter curve). This implies $\tilde{\gamma}$ lifts to a straight line in T_pM, i.e. is a geodesic.

Now given any $q \in M$ with $r = d(p, q) < \varepsilon$, we pick $r' \in [r, \varepsilon)$ and a path $\gamma \in \Omega(p, q)$ such that $\ell(\gamma) = r'$. We again let

$$\tau = \sup\left\{ t \in [0, 1] : \gamma([0, t]) \subseteq \exp_p(\overline{B(0, r')}) \right\}.$$

If $\tau \neq 1$, then we must have $\gamma(\tau) \in \Sigma_{r'}$, but lifting to T_pM, this contradicts the local minimizing of length.

The last part is an immediate consequence of the previous two. $\square$

Corollary. The distance d on a Riemannian manifold is a metric, and induces the same topology on M as the C^∞ structure.

Definition (Minimal geodesic)**.** A *minimal geodesic* is a curve $\gamma : [0, 1] \to M$ such that

$$d(\gamma(0), \gamma(1)) = \ell(\gamma).$$

One would certainly want a minimal geodesic to be an actual geodesic. This is an easy consequence of what we've got so far, using the observation that a sub-curve of a minimizing geodesic is still minimizing.

Corollary. Let $\gamma : [0, 1] \to M$ be a piecewise C^1 minimal geodesic with constant speed. Then γ is in fact a geodesic, and is in particular C^∞.

Proof. We wlog γ is unit speed. Let $t \in [0, 1]$, and pick $\varepsilon > 0$ such that $\exp_p|_{B(0,\varepsilon)}$ is a diffeomorphism. Then by the theorem, $\gamma_{[t,t+\frac{1}{2}\varepsilon]}$ is a geodesic. So γ is C^∞ on $(t, t + \frac{1}{2}\varepsilon)$, and satisfies the geodesic equations there.

Since we can pick ε continuously with respect to t by ODE theorems, any $t \in (0, 1)$ lies in one such neighbourhood. So γ is a geodesic. $\square$

While it is not true that geodesics are always minimal geodesics, this is locally true:

Corollary. Let $\gamma : [0,1] \subseteq \mathbb{R} \to M$ be a C^2 curve with $|\dot{\gamma}|$ constant. Then this is a geodesic iff it is locally a minimal geodesic, i.e. for any $t \in [0,1)$, there exists $\delta > 0$ such that
$$d(\gamma(t), \gamma(t+\delta)) = \ell(\gamma|_{[t,t+\delta]}).$$

*Proof.* This is just carefully applying the previous theorem without getting confused.

To prove $\Rightarrow$, suppose γ is a geodesic, and $t \in [0,1)$. We wlog γ is unit speed. Then pick U and ε as in the previous theorem, and pick $\delta = \frac{1}{2}\varepsilon$. Then $\gamma|_{[t,t+\delta]}$ is a geodesic with length $< \varepsilon$ between $\gamma(t)$ and $\gamma(t+\delta)$, and hence must have minimal length.

To prove the converse, we note that for each t, the hypothesis tells us $\gamma|_{[t,t+\delta]}$ is a minimizing geodesic, and hence a geodesic, but the previous corollary. By continuity, γ must satisfy the geodesic equation at t. Since t is arbitrary, γ is a geodesic. $\qquad\square$

There is another sense in which geodesics are locally length minimizing. Instead of chopping up a path, we can say it is minimal "locally" in the space $\Omega(p,q)$. To do so, we need to give $\Omega(p,q)$ a topology, and we pick the topology of uniform convergence.

Theorem. Let $\gamma(t) = \exp_p(ta)$ be a geodesic, for $t \in [0,1]$. Let $q = \gamma(1)$. Assume ta is a regular point for $\exp_p$ for all $t \in [0,1]$. Then there exists a neighbourhood of γ in $\Omega(p,q)$ such that for all ψ in this neighbourhood, $\ell(\psi) \geq \ell(\gamma)$, with equality iff $\psi = \gamma$ up to reparametrization.

Before we prove the result, we first look at why the two conditions are necessary. To see the necessity of ta being regular, we can consider the sphere and two antipodal points:
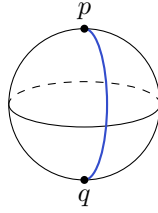

Then while the geodesic between them does minimize distance, it does not do so strictly.

We also do not guarantee global minimization of length. For example, we can consider the torus
$$T^n = \mathbb{R}^n / \mathbb{Z}^n.$$

This has a flat metric from $\mathbb{R}^n$, and the derivative of the exponential map is the "identity" on $\mathbb{R}^n$ at all points. So the geodesics are the straight lines in $\mathbb{R}^n$. Now consider any two $p, q \in T^n$, then there are infinitely many geodesics joining them, but typically, only one of them would be the shortest.
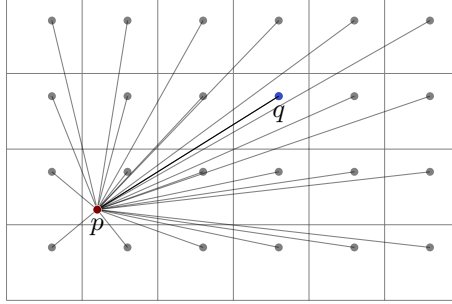
Proof. The idea of the proof is that if ψ is any curve close to γ, then we can use the regularity condition to lift the curve back up to T_pM, and then apply our previous result.

Write $\varphi(t) = ta \in T_pM$. Then by the regularity assumption, for all $t \in [0,1]$, we know $\exp_p$ is a diffeomorphism of some neighbourhood $W(t)$ of $\varphi(t) = at \in T_pM$ onto the image. By compactness, we can cover $[0,1]$ by finitely many such covers, say $W(t_1), \cdots, W(t_n)$. We write $W_i = W(t_i)$, and we wlog assume

$$0 = t_0 < t_1 < \cdots < t_k = 1.$$

By cutting things up, we may assume

$$\gamma([t_i, t_{i+1}]) \subseteq W_i.$$

We let

$$U = \bigcup \exp_p(W_i).$$

Again by compactness, there is some $\varepsilon < 0$ such that for all $t \in [t_i, t_{i+1}]$, we have $B(\gamma(t), \varepsilon) \subseteq W_i$.

Now consider any curve ψ of distance ε away from γ. Then $\psi([t_i, t_{i+1}]) \subseteq W_i$. So we can lift it up to T_pM, and the end point of the lift is a. So we are done by local minimization of length. $\qquad\square$

Note that the tricky part of doing the proof is to make sure the lift of ψ has the same end point as γ in T_pM, which is why we need to do it neighbourhood by neighbourhood.

3.4   Completeness and the Hopf–Rinow theorem

There are some natural questions we can ask about geodesics. For example, we might want to know if geodesics can be extended to exist for all time. We might also be interested if distances can always be realized by geodesics. It turns out these questions have the same answer.

Definition (Geodesically complete)**.** We say a manifold (M, g) is *geodesically complete* if each geodesic extends for all time. In other words, for all $p \in M$, $\exp_p$ is defined on *all* of T_pM.

Example. The upper half plane

$$H^2 = \{(x, y) : y > 0\}$$

under the induced Euclidean metric is not geodesically complete. However, H^2 and $\mathbb{R}^2$ are diffeomorphic but $\mathbb{R}^2$ is geodesically complete.

The first theorem we will prove is the following:

Theorem. Let (M, g) be geodesically complete. Then any two points can be connected by a minimal geodesic.

In fact, we will prove something stronger — let $p \in M$, and suppose $\exp_p$ is defined on all of T_pM. Then for all $q \in M$, there is a minimal geodesic between them.

To prove this, we need a lemma

Lemma. Let $p, q \in M$. Let

$$S_\delta = \{x \in M : d(x, p) = \delta\}.$$

Then for all sufficiently small δ, there exists $p_0 \in S_\delta$ such that

$$d(p, p_0) + d(p_0, q) = d(p, q).$$

Proof. For $\delta > 0$ small, we know $S_\delta = \Sigma_\delta$ is a geodesic sphere about p, and is compact. Moreover, $d(\,\cdot\,, q)$ is a continuous function. So there exists some $p_0 \in \Sigma_\delta$ that minimizes $d(\,\cdot\,, q)$.

Consider an arbitrary $\gamma \in \Omega(p, q)$. For the sake of sanity, we assume $\delta < d(p, q)$. Then there is some t such that $\gamma(t) \in \Sigma_\delta$, and

$$\ell(\gamma) \geq d(p, \gamma(t)) + d(\gamma(t), q) \geq d(p, p_0) + d(p_0, q).$$

So we know

$$d(p, q) \geq d(p, p_0) + d(p_0, p).$$

The triangle inequality gives the opposite direction. So we must have equality. $\square$

We can now prove the theorem.

Proof of theorem. We know $\exp_p$ is defined on T_pM. Let $q \in M$. Let $q \in M$. We want a minimal geodesic in $\Omega(p, q)$. By the first lemma, there is some $\delta > 0$ and p_0 such that

$$d(p, p_0) = \delta, \quad d(p, p_0) + d(p_0, q) = d(p, q).$$

Also, there is some $v \in T_pM$ such that $\exp_p v = p_0$. We let

$$\gamma_p(t) = \exp_p\left(t\frac{v}{|v|}\right).$$

We let

$$I = \{t \in \mathbb{R} : d(q, \gamma_p(t)) + t = d(p, q)\}.$$

Then we know

(i) $\delta \in I$

(ii) I is closed by continuity.

Let
$$T = \sup\{I \cap [0, d(p,q)]\}.$$
Since I is closed, this is in fact a maximum. So $T \in I$. We claim that $T = d(p,q)$. If so, then $\gamma_p \in \Omega(p,q)$ is the desired minimal geodesic, and we are done.

Suppose this were not true. Then $T < d(p,q)$. We apply the lemma to $\tilde{p} = \gamma_p(T)$, and q remains as before. Then we can find $\varepsilon > 0$ and some $p_1 \in M$ with the property that
$$\begin{aligned}
d(p_1, q) &= d(\gamma_p(T), q) - d(\gamma_p(T), p_1) \\
&= d(\gamma_p(T), q) - \varepsilon \\
&= d(p,q) - T - \varepsilon
\end{aligned}$$
Hence we have
$$d(p, p_1) \geq d(p,q) - d(q, p_1) = T + \varepsilon.$$
Let γ_1 be the radial (hence minimal) geodesic from $\gamma_p(T)$ to p_1. Now we know
$$\ell(\gamma_p|_{[0,T]}) + \ell(\gamma_1) = T + \varepsilon.$$
So γ_1 concatenated with $\gamma_p|_{[0,T]}$ is a length-minimizing geodesic from p to p_1, and is hence a geodesic. So in fact p_1 lies on γ_p, say $p_1 = \gamma_p(T + s)$ for some s. Then $T + s \in I$, which is a contradiction. So we must have $T = d(p,q)$, and hence
$$d(q, \gamma_p(T)) + T = d(p,q),$$
hence $d(q, \gamma_p(T)) = 0$, i.e. $q = \gamma_p(T)$. $\qquad\square$

Corollary (Hopf–Rinow theorem)**.** For a connected Riemannian manifold (M, g), the following are equivalent:

(i) (M, g) is geodesically complete.

(ii) For all $p \in M$, $\exp_p$ is defined on all $T_p M$.

(iii) For some $p \in M$, $\exp_p$ is defined on all $T_p M$.

(iv) Every closed and bounded subset of (M, d) is compact.

(v) (M, d) is complete as a metric space.

Proof. (i) and (ii) are equivalent by definition. (ii) $\Rightarrow$ (iii) is clear, and we proved (iii) $\Rightarrow$ (ii).

– (iii) $\Rightarrow$ (iv): Let $K \subseteq M$ be closed and bounded. Then by boundedness, K is contained in $\exp_p(\overline{B(0, R)})$. Let K' be the pre-image of K under $\exp_p$. Then it is a closed and bounded subset of $\mathbb{R}^n$, hence compact. Then K is the continuous image of a compact set, hence compact.

– (iv) $\Rightarrow$ (v): This is a general topological fact.

– (v) $\Rightarrow$ (i): Let $\gamma(t) : I \to \mathbb{R}$ be a geodesic, where $I \subseteq \mathbb{R}$. We wlog $|\dot{\gamma}| \equiv 1$. Suppose $I \neq \mathbb{R}$. We wlog $\sup I = a < \infty$. Then $\lim_{t \to a} \gamma(t)$ exist by completeness, and hence $\gamma(a)$ exists. Since geodesics are locally defined near a, we can pick a geodesic in the direction of $\lim_{t \to a} \gamma'(t)$. So we can extend γ further, which is a contradiction.

$\qquad\square$

3.5 Variations of arc length and energy

This section is mostly a huge computation. As we previously saw, geodesics are locally length-minimizing, and we shall see that another quantity, namely the energy is also a useful thing to consider, as minimizing the energy also forces the parametrization to be constant speed.

To make good use of these properties of geodesics, it is helpful to compute explicitly expressions for how length and energy change along variations. The computations are largely uninteresting, but it will pay off.

Definition (Energy)**.** The *energy function* $E : \Omega(p, q) \to \mathbb{R}$ is given by

$$E(\gamma) = \frac{1}{2} \int_0^T |\dot{\gamma}|^2 \, dt,$$

where $\gamma : [0, T] \to M$.

Recall that $\Omega(p, q)$ is defined as the space of piecewise C^1 curves. Often, we will make the simplifying assumption that all curves are in fact C^1. It doesn't really matter.

Note that the length of a curve is independent of parametrization. Thus, if we are interested in critical points, then the critical points cannot possibly be isolated, as we can just re-parametrize to get a nearby path with the same length. On the other hand, the energy E *does* depend on parametrization. This does have isolated critical points, which is technically very convenient.

Proposition. Let $\gamma_0 : [0, T] \to M$ be a path from p to q such that for all $\gamma \in \Omega(p, q)$ with $\gamma : [0, T] \to M$, we have $E(\gamma) \geq E(\gamma_0)$. Then γ_0 must be a geodesic.

Recall that we already had such a result for length instead of energy. The proof is just the application of Cauchy-Schwartz.

Proof. By the Cauchy-Schwartz inequality, we have

$$\int_0^T |\dot{\gamma}|^2 \, dt \geq \left(\int_0^T |\dot{\gamma}(t)| \, dt \right)^2$$

with equality iff $|\dot{\gamma}|$ is constant. In other words,

$$E(\gamma) \geq \frac{\ell(\gamma)^2}{2T}.$$

So we know that if γ_0 minimizes energy, then it must be constant speed. Now given any γ, if we just care about its length, then we may wlog it is constant speed, and then

$$\ell(\gamma) = \sqrt{2E(\gamma)T} \geq \sqrt{2E(\gamma_0)T} = \ell(\gamma_0).$$

So γ_0 minimizes length, and thus γ_0 is a geodesic. $\qquad\square$

We shall consider smooth variations $H(t, s)$ of $\gamma_0(t) = H(t, 0)$. We require that $H : [0, T] \times (-\varepsilon, \varepsilon) \to M$ is smooth. Since we are mostly just interested

in what happens "near" $s = 0$, it is often convenient to just consider the corresponding vector field along γ:

$$Y(t) = \left.\frac{\partial H}{\partial s}\right|_{s=0} = (\mathrm{d}H)_{(t,0)}\frac{\partial}{\partial s},$$

Conversely, given any such vector field Y, we can generate a variation H that gives rise to Y. For example, we can put

$$H(t,s) = \exp_{\gamma_0(t)}(sY(t)),$$

which is valid on some neighbourhood of $[0,T] \times \{0\}$. If $Y(0) = 0 = Y(T)$, then we can choose H fixing end-points of γ_0.

Theorem (First variation formula)**.**

(i) For any variation H of γ, we have

$$\left.\frac{\mathrm{d}}{\mathrm{d}s}E(\gamma_s)\right|_{s=0} = g(Y(t),\dot\gamma(t))|_0^T - \int_0^T g\left(Y(t),\frac{\nabla}{\mathrm{d}t}\dot\gamma(t)\right)\,\mathrm{d}t. \qquad (*)$$

(ii) The critical points, i.e. the γ such that

$$\left.\frac{\mathrm{d}}{\mathrm{d}s}E(\gamma_s)\right|_{s=0}$$

for all (end-point fixing) variation H of γ, are geodesics.

(iii) If $|\dot\gamma_s(t)|$ is constant for each fixed $s \in (-\varepsilon,\varepsilon)$, and $|\dot\gamma(t)| \equiv 1$, then

$$\left.\frac{\mathrm{d}}{\mathrm{d}s}E(\gamma_s)\right|_{s=0} = \left.\frac{\mathrm{d}}{\mathrm{d}s}\ell(\gamma_s)\right|_{s=0}$$

(iv) If γ is a critical point of the length, then it must be a reparametrization of a geodesic.

This is just some calculations.

Proof. We will assume that we can treat $\frac{\partial}{\partial s}$ and $\frac{\partial}{\partial t}$ as vector fields on an embedded submanifold, even though H is not necessarily a local embedding.

The result can be proved without this assumption, but will require more technical work.

(i) We have

$$\begin{aligned}
\frac{1}{2}\frac{\partial}{\partial s}g(\dot\gamma_s(t),\dot\gamma_s(t)) &= g\left(\frac{\nabla}{\mathrm{d}s}\dot\gamma_s(t),\dot\gamma_s(t)\right) \\
&= g\left(\frac{\nabla}{\mathrm{d}t}\frac{\partial H}{\partial s}(t,s),\frac{\partial H}{\partial t}(t,s)\right) \\
&= \frac{\partial}{\partial t}g\left(\frac{\partial H}{\partial s},\frac{\partial H}{\partial t}\right) - g\left(\frac{\partial H}{\partial s},\frac{\nabla}{\mathrm{d}t}\frac{\partial H}{\partial t}\right).
\end{aligned}$$

Comparing with what we want to prove, we see that we get what we want by integrating $\int_0^T\,\mathrm{d}t$, and then putting $s = 0$, and then noting that

$$\left.\frac{\partial H}{\partial s}\right|_{s=0} = Y, \qquad \left.\frac{\partial H}{\partial t}\right|_{s=0} = \dot\gamma.$$

(ii) If γ is a geodesic, then
$$\frac{\nabla}{\mathrm{d}t}\dot{\gamma}(t) = 0.$$

So the integral on the right hand side of $(*)$ vanishes. Also, we have $Y(0) = 0 = Y(T)$. So the RHS vanishes.

Conversely, suppose γ is a critical point for E. Then choose H with
$$Y(t) = f(t)\frac{\nabla}{\mathrm{d}t}\dot{\gamma}(t)$$

for some $f \in C^{\infty}[0,T]$ such that $f(0) = f(T) = 0$. Then we know
$$\int_0^T f(t)\left|\frac{\nabla}{\mathrm{d}t}\dot{\gamma}(t)\right|^2 \mathrm{d}t = 0,$$

and this is true for all f. So we know
$$\frac{\nabla}{\mathrm{d}t}\dot{\gamma} = 0.$$

(iii) This is evident from the previous proposition. Indeed, we fix $[0,T]$, then for all H, we have
$$E(\gamma_s) = \frac{\ell(\gamma_s)^2}{2T},$$

and so
$$\left.\frac{\mathrm{d}}{\mathrm{d}s}E(\gamma_s)\right|_{s=0} = \frac{1}{T}\ell(\gamma_s)\left.\frac{\mathrm{d}}{\mathrm{d}s}\ell(\gamma_s)\right|_{s=0},$$

and when $s = 0$, the curve is parametrized by arc-length, so $\ell(\gamma_s) = T$.

(iv) By reparametrization, we may wlog $|\dot{\gamma}| \equiv 1$. Then γ is a critical point for ℓ, hence for E, hence a geodesic.

$\square$

Often, we are interested in more than just whether the curve is a critical point. We want to know if it maximizes or minimizes energy. Then we need more than the "first derivative". We need the "second derivative" as well.

Theorem (Second variation formula)**.** Let $\gamma(t) : [0,T] \to M$ be a geodesic with $|\dot{\gamma}| = 1$. Let $H(t,s)$ be a variation of γ. Let
$$Y(t,s) = \frac{\partial H}{\partial s}(t,s) = (\mathrm{d}H)_{(t,s)}\frac{\partial}{\partial s}.$$

Then

(i) We have
$$\left.\frac{\mathrm{d}^2}{\mathrm{d}s^2}E(\gamma_s)\right|_{s=0} = g\left(\frac{\nabla Y}{\mathrm{d}s}(t,0), \dot{\gamma}\right)\Big|_0^T + \int_0^T \left(|Y'|^2 - R(Y,\dot{\gamma},Y,\dot{\gamma})\right)\mathrm{d}t.$$

(ii) Also

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2}\ell(\gamma_s)\bigg|_{s=0} = g\left(\frac{\nabla Y}{\mathrm{d}s}(t,0),\dot{\gamma}(t)\right)\bigg|_0^T$$
$$+ \int_0^T \left(|Y'|^2 - R(Y,\dot{\gamma},Y,\dot{\gamma}) - g(\dot{\gamma},Y')^2\right)\,\mathrm{d}t,$$

where R is the $(4,0)$ curvature tensor, and

$$Y'(t) = \frac{\nabla Y}{\mathrm{d}t}(t,0).$$

Putting

$$Y_n = Y - g(Y,\dot{\gamma})\dot{\gamma}$$

for the normal component of Y, we can write this as

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2}\ell(\gamma_s)\bigg|_{s=0} = g\left(\frac{\nabla Y_n}{\mathrm{d}s}(t,0),\dot{\gamma}(t)\right)\bigg|_0^T + \int_0^T \left(|Y_n'|^2 - R(Y_n,\dot{\gamma},Y_n,\dot{\gamma})\right)\,\mathrm{d}t.$$

Note that if we have fixed end points, then the first terms in the variation formulae vanish.

Proof. We use

$$\frac{\mathrm{d}}{\mathrm{d}s}E(\gamma_s) = g(Y(t,s),\dot{\gamma}_s(t))|_{t=0}^{t=T} - \int_0^T g\left(Y(t,s),\frac{\nabla}{\mathrm{d}t}\dot{\gamma}_s(t)\right)\,\mathrm{d}t.$$

Taking the derivative with respect to s again gives

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2}E(\gamma_s) = g\left(\frac{\nabla Y}{\mathrm{d}s},\dot{\gamma}\right)\bigg|_{t=0}^{t=T} + g\left(Y,\frac{\nabla}{\mathrm{d}s}\dot{\gamma}_s\right)\bigg|_{t=0}^T$$
$$- \int_0^T \left(g\left(\frac{\nabla Y}{\mathrm{d}s},\frac{\nabla}{\mathrm{d}t}\dot{\gamma}_s\right) + g\left(Y,\frac{\nabla}{\mathrm{d}s}\frac{\nabla}{\mathrm{d}t}\dot{\gamma}\right)\right)\,\mathrm{d}t.$$

We now use that

$$\frac{\nabla}{\mathrm{d}s}\frac{\nabla}{\mathrm{d}t}\dot{\gamma}_s(t) = \frac{\nabla}{\mathrm{d}t}\frac{\nabla}{\mathrm{d}s}\dot{\gamma}_s(t) + R\left(\frac{\partial H}{\partial s},\frac{\partial H}{\partial t}\right)\dot{\gamma}_s$$
$$= \left(\frac{\nabla}{\mathrm{d}t}\right)^2 Y(t,s) + R\left(\frac{\partial H}{\partial s},\frac{\partial H}{\partial t}\right)\dot{\gamma}_s.$$

We now set $s=0$, and then the above gives

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2}E(\gamma_s)\bigg|_{s=0} = g\left(\frac{\nabla Y}{\mathrm{d}s},\dot{\gamma}\right)\bigg|_0^T + g\left(Y,\frac{\nabla\dot{\gamma}}{\mathrm{d}s}\right)\bigg|_0^T$$
$$- \int_0^T \left[g\left(Y,\left(\frac{\nabla}{\mathrm{d}t}\right)^2 Y\right) + R(\dot{\gamma},Y,\dot{\gamma},Y)\right]\,\mathrm{d}t.$$

Finally, applying integration by parts, we can write

$$-\int_0^T g\left(Y, \left(\frac{\nabla}{\mathrm{d}t}\right)^2 Y\right)\,\mathrm{d}t = -\left.g\left(Y, \frac{\nabla}{\mathrm{d}t}Y\right)\right|_0^T + \int_0^T \left|\frac{\nabla Y}{\mathrm{d}t}\right|^2\,\mathrm{d}t.$$

Finally, noting that

$$\frac{\nabla}{\mathrm{d}s}\dot\gamma(s) = \frac{\nabla}{\mathrm{d}t}Y(t,s),$$

we find that

$$\left.\frac{\mathrm{d}^2}{\mathrm{d}s^2}E(\gamma_s)\right|_{s=0} = \left.g\left(\frac{\nabla Y}{\mathrm{d}s}, \dot\gamma\right)\right|_0^T + \int_0^T \left(|Y'|^2 - R(Y, \dot\gamma, Y, \dot\gamma)\right)\,\mathrm{d}t.$$

It remains to prove the second variation of length. We first differentiate

$$\frac{\mathrm{d}}{\mathrm{d}s}\ell(\gamma_s) = \int_0^T \frac{1}{2\sqrt{g(\dot\gamma_s, \dot\gamma_s)}}\frac{\partial}{\partial s}g(\dot\gamma_s, \dot\gamma_s)\,\mathrm{d}t.$$

Then the second derivative gives

$$\left.\frac{\mathrm{d}^2}{\mathrm{d}s^2}\ell(\gamma_s)\right|_{s=0} = \int_0^T \left[\frac{1}{2}\left.\frac{\partial^2}{\partial s^2}g(\dot\gamma_s, \dot\gamma_s)\right|_{s=0} - \frac{1}{4}\left.\left(\frac{\partial}{\partial s}g(\dot\gamma_s, \dot\gamma_s)\right)^2\right|_{s=0}\right]\,\mathrm{d}t,$$

where we used the fact that $g(\dot\gamma, \dot\gamma) = 1$.

We notice that the first term can be identified with the derivative of the energy function. So we have

$$\left.\frac{\mathrm{d}^2}{\mathrm{d}s^2}\ell(\gamma_s)\right|_{s=0} = \left.\frac{\mathrm{d}^2}{\mathrm{d}s^2}E(\gamma_s)\right|_{s=0} - \int_0^T \left(\left.g\left(\dot\gamma_s, \frac{\nabla}{\mathrm{d}s}\dot\gamma_s\right)\right|_{s=0}\right)^2\,\mathrm{d}t.$$

So the second part follows from the first. $\qquad\square$

3.6 Applications

This finally puts us in a position to prove something more interesting.

Synge's theorem

We are first going to prove the following remarkable result relating curvature and topology:

Theorem (Synge's theorem)**.** Every compact orientable Riemannian manifold (M, g) such that $\dim M$ is even and has $K(g) > 0$ for all planes at $p \in M$ is simply connected.

We can see that these conditions are indeed necessary. For example, we can consider $\mathbb{RP}^2 = S^2/\pm 1$ with the induced metric from S^2. Then this is compact with positive sectional curvature, but it is not orientable. Indeed it is not simply connected.

Similarly, if we take $\mathbb{RP}^3$, then this has odd dimension, and the theorem breaks.

Finally, we do need strict inequality, e.g. the flat torus is not simply connected.

We first prove a technical lemma.

Lemma. Let M be a compact manifold, and $[\alpha]$ a non-trivial homotopy class of closed curves in M. Then there is a closed minimal geodesic in $[\alpha]$.

Proof. Since M is compact, we can pick some $\varepsilon > 0$ such that for all $p \in M$, the map $\exp_p |_{B(0,p)}$ is a diffeomorphism.

Let $\ell = \inf_{\gamma \in [\alpha]} \ell(\gamma)$. We know that $\ell > 0$, otherwise, there exists a γ with $\ell(\gamma) < \varepsilon$. So γ is contained in some geodesic coordinate neighbourhood, but then α is contractible. So ℓ must be positive.

Then we can find a sequence $\gamma_n \in [\alpha]$ with $\gamma_n : [0,1] \to M$, $|\dot\gamma|$ constant, such that

$$\lim_{n \to \infty} \ell(\gamma_n) = \ell.$$

Choose

$$0 = t_0 < t_1 < \cdots < t_k = 1$$

such that

$$t_{i+1} - t_i < \frac{\varepsilon}{2\ell}.$$

So it follows that

$$d(\gamma_n(t_i), \gamma_n(t_{i+1})) < \varepsilon$$

for all n sufficiently large and all i. Then again, we can replace $\gamma_n|_{[t_i, t_{i+1}]}$ by a radial geodesic without affecting the limit $\lim \ell(\gamma_n)$.

Then we exploit the compactness of M (and the unit sphere) again, and pass to a subsequence of $\{\gamma_n\}$ so that $\gamma_n(t_i), \dot\gamma_n(t_i)$ are all convergent for every fixed i as $n \to \infty$. Then the curves converges to some

$$\gamma_n \to \hat\gamma \in [\alpha],$$

given by joining the limits $\lim_{n \to \infty} \gamma_n(t_i)$. Then we know that the length converges as well, and so we know $\hat\gamma$ is minimal among curves in $[\alpha]$. So $\hat\gamma$ is locally minimal, hence a geodesic. So we can take $\gamma = \hat\gamma$, and we are done. $\square$

Proof of Synge's theorem. Suppose M satisfies the hypothesis, but $\pi_1(M) \neq \{1\}$. So there is a path α with $[\alpha] \neq 1$, i.e. it cannot be contracted to a point. By the lemma, we pick a representative γ of $[\alpha]$ that is a closed, minimal geodesic.

We now prove the theorem. We may wlog assume $|\dot\gamma| = 1$, and t ranges in $[0, T]$. Consider a vector field $X(t)$ for $0 \leq t \leq T$ along $\gamma(t)$ such that

$$\frac{\nabla X}{\mathrm{d}t} = 0, \quad g(X(0), \dot\gamma(0)) = 0.$$

Note that since g is a geodesic, we know

$$g(X(t), \dot\gamma(t)) = 0,$$

for all $t \in [0, T]$ as parallel transport preserves the inner product. So $X(T) \perp \dot\gamma(T) = \dot\gamma(0)$ since we have a closed curve.

We consider the map P that sends $X(0) \mapsto X(T)$. This is a linear isometry of $(\dot\gamma(0))^\perp$ with itself that preserves orientation. So we can think of P as a map

$$P \in \mathrm{SO}(2n - 1),$$

where $\dim M = 2n$. It is an easy linear algebra exercise to show that every element of $\mathrm{SO}(2n - 1)$ must have an eigenvector of eigenvalue 1. So we can find

$v \in T_p M$ such that $v \perp \dot\gamma(0)$ and $P(v) = v$. We take $X(0) = v$. Then we have $X(T) = v$.

Consider now a variation $H(t, s)$ inducing this $X(t)$. We may assume $|\dot\gamma_s|$ is constant. Then

$$\frac{\mathrm{d}}{\mathrm{d}s}\ell(\gamma_s)|_{s=0} = 0$$

as γ is minimal. Moreover, since it is a minimum, the second derivative must be positive, or at least non-negative. Is this actually the case?

We look at the second variation formula of length. Using the fact that the loop is closed, the formula reduces to

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2}\ell(\gamma_s)\bigg|_{s=0} = -\int_0^T R(X, \dot\gamma, X, \dot\gamma)\,\mathrm{d}t.$$

But we assumed the sectional curvature is positive. So the second variation is negative! This is a contradiction. $\square$

Conjugate points

Recall that when a geodesic starts moving, for a short period of time, it is length-minimizing. However, in general, if we keep on moving for a long time, then we cease to be minimizing. The point where this transition happens is interesting.

As before, for a vector field J along a curve $\gamma(t)$, we will write

$$J' = \frac{\nabla J}{\mathrm{d}t}.$$

Definition (Conjugate points)**.** Let $\gamma(t)$ be a geodesic. Then

$$p = \gamma(\alpha), \quad q = \gamma(\beta)$$

are *conjugate points* if there exists some non-trivial J such that $J(\alpha) = 0 = J(\beta)$.

It is easy to see that this does not depend on parametrization of the curve, because Jacobi fields do not.

Proposition.

(i) If $\gamma(t) = \exp_p(ta)$ and $q = \exp_p(\beta a)$ is conjugate to p, then q is a singular value of exp.

(ii) Let J be as in the definition. Then J must be pointwise normal to $\dot\gamma$.

Proof.

(i) We wlog $[\alpha, \beta] = [0, 1]$. So $J(0) = 0 = J(1)$. We $a = \dot\gamma(0)$ and $w = J'(0)$. Note that a, w are both non-zero, as Jacobi fields are determined by initial conditions. Then $q = \exp_p(a)$.

We have shown earlier that if $J(0) = 0$, then

$$J(t) = (\mathrm{d}\exp_p)_{ta}(tw)$$

for all $0 \le t \le 1$. So it follows $(\mathrm{d}\exp_p)_a(w) = J(1) = 0$. So $(\mathrm{d}\exp_p)_a$ has non-trivial kernel, and hence isn't surjective.

(ii) We claim that any Jacobi field J along a geodesic γ satisfies

$$g(J(t), \dot{\gamma}(t)) = g(J'(0), \dot{\gamma}(0))t + g(J(0), \dot{\gamma}(0)).$$

To prove this, we note that by the definition of geodesic and Jacobi fields, we have

$$\frac{\mathrm{d}}{\mathrm{d}t} g(J', \dot{\gamma}) = g(J'', \dot{\gamma}(0)) = -g(R(\dot{\gamma}, J), \dot{\gamma}, \dot{\gamma}) = 0$$

by symmetries of R. So we have

$$\frac{\mathrm{d}}{\mathrm{d}t}(J, \dot{\gamma}) = g(J'(t), \dot{\gamma}(t)) = g(J'(0), \dot{\gamma}(0)).$$

Now integrating gives the desired result.

This result tells us $g(J(t), \dot{\gamma}(t))$ is a linear function of t. But we have

$$g(J(0), \dot{\gamma}(0)) = g(J(1), \dot{\gamma}(1)) = 0.$$

So we know $g(J(t), \dot{\gamma}(t))$ is constantly zero.

$\square$

From the proof, we see that for any Jacobi field with $J(0) = 0$, we have

$$g(J'(0), \dot{\gamma}(0)) = 0 \iff g(J(t), \dot{\gamma}(t)) = \text{constant}.$$

This implies that the dimension of the normal Jacobi fields along γ satisfying $J(0) = 0$ is $\dim M - 1$.

Example. Consider $M = S^2 \subseteq \mathbb{R}^3$ with the round metric, i.e. the "obvious" metric induced from $\mathbb{R}^3$. We claim that $N = (0, 0, 1)$ and $S = (0, 1, 0)$ are conjugate points.
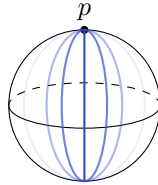
To construct a Jacobi field, instead of trying to mess with the Jacobi equation, we construct a variation by geodesics. We let

$$f(t, s) = \begin{pmatrix} \cos s \sin t \\ \sin s \sin t \\ \cos t \end{pmatrix}.$$
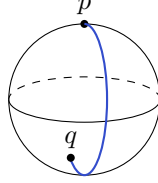
We see that when $s = 0$, this is the great-circle in the (x, z)-plane. Then we have a Jacobi field

$$J(t) = \left. \frac{\partial f}{\partial s} \right|_{s=0} = \begin{pmatrix} 0 \\ \sin t \\ 0 \end{pmatrix}.$$

This is then a Jacobi field that vanishes at N and S.

When we are *at* the conjugate point, then there are many adjacent curves whose length is equal to ours. If we extend our geodesic *beyond* the conjugate point, then it is no longer even locally minimal:

We can push the geodesic slightly over and the length will be shorter. On the other hand, we proved that up to the conjugate point, the geodesic is always locally minimal.

In turns out this phenomenon is generic:

Theorem. Let $\gamma : [0, 1] \to M$ be a geodesic with $\gamma(0) = p$, $\gamma(1) = q$ such that p is conjugate to some $\gamma(t_0)$ for some $t_0 \in (0, 1)$. Then there is a piecewise smooth variation of $f(t, s)$ with $f(t, 0) = \gamma(t)$ such that

$$f(0, s) = p, \quad f(1, s) = q$$

and $\ell(f(\,\cdot\,, s)) < \ell(\gamma)$ whenever $s \neq 0$ is small.

The proof is a generalization of the example we had above. We know that up to the conjugate point, we have a Jacobi filed that allows us to vary the geodesic without increasing the length. We can then give it a slight "kick" and then the length will decrease.

Proof. By the hypothesis, there is a $J(t)$ defined on $t \in [0, 1]$ and $t_0 \in (0, 1)$ such that

$$J(t) \perp \dot{\gamma}(t)$$

for all t, and $J(0) = J(t_0) = 0$ and $J \not\equiv 0$. Then $J'(t_0) \neq 0$.

We define a parallel vector field Z_1 along γ by $Z_1(t_0) = -J'(t_0)$. We pick $\theta \in C^\infty[0, 1]$ such that $\theta(0) = \theta(1) = 0$ and $\theta(t_0) = 1$.

Finally, we define

$$Z = \theta Z_1,$$

and for $\alpha \in \mathbb{R}$, we define

$$Y_\alpha(t) = \begin{cases} J(t) + \alpha Z(t) & 0 \leq t \leq t_0 \\ \alpha Z(t) & t_0 \leq t \leq 1 \end{cases}.$$

We notice that this is not smooth at t_0, but is just continuous. We will postpone the choice of α to a later time.

We know $Y_\alpha(t)$ arises from a piecewise C^∞ variation of γ, say $H_\alpha(t, s)$. The technical claim is that the second variation of length corresponding to $Y_\alpha(t)$ is negative for some α.

We denote by $I(X, Y)_T$ the symmetric bilinear form that gives rise to the second variation of length with fixed end points. If we make the additional

assumption that X, Y are normal along γ, then the formula simplifies, and reduces to

$$I(X,Y)_T = \int_0^T \left(g(X', Y') - R(X, \dot\gamma, Y, \dot\gamma) \right) \, \mathrm{d}t.$$

Then for $H_\alpha(t, s)$, we have

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2} \ell(\gamma_s) \Big|_{s=0} = I_1 + I_2 + I_3$$

$$I_1 = I(J, J)_{t_0}$$

$$I_2 = 2\alpha I(J, Z)_{t_0}$$

$$I_3 = \alpha^2 I(Z, Z)_1.$$

We look at each term separately.

We first claim that $I_1 = 0$. We note that

$$\frac{\mathrm{d}}{\mathrm{d}t} g(J, J') = g(J', J') + g(J, J''),$$

and $g(J, J'')$ added to the curvature vanishes by the Jacobi equation. Then by integrating by parts and applying the boundary condition, we see that I_1 vanishes.

Also, by integrating by parts, we find

$$I_2 = 2\alpha g(Z, J')\big|_0^{t_0}.$$

Whence

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2} \ell(\gamma_s) \Big|_{s=0} = -2\alpha |J'(t_0)|^2 + \alpha^2 I(Z, Z)_1.$$

Now if $\alpha > 0$ is very very small, then the linear term dominates, and this is negative. Since the first variation vanishes (γ is a geodesic), we know this is a local maximum of length. $\qquad\square$

Note that we made a compromise in the theorem by doing a piecewise C^∞ variation instead of a smooth one, but of course, we can fix this by making a smooth approximation.

Bonnet–Myers diameter theorem

We are going to see yet another application of our previous hard work, which may also be seen as an interplay between curvature topology. In case it isn't clear, all our manifolds are connected.

Definition (Diameter)**.** The *diameter* of a Riemannian manifold (M, g) is

$$\mathrm{diam}(M, g) = \sup_{p, q \in M} d(p, q).$$

Of course, this definition is valid for any metric space.

Example. Consider the sphere

$$S^{n-1}(r) = \{ x \in \mathbb{R}^n : |x| = r \},$$

with the induced "round" metric. Then

$$\operatorname{diam}(S^{n-1}(r)) = \pi r.$$

It is an exercise to check that

$$K \equiv \frac{1}{r^2}.$$

We will also need the following notation:

Notation. Let $h, \hat{h}$ be two symmetric bilinear forms on a real vector space. We say $h \geq \hat{h}$ if $h - \hat{h}$ is non-negative definite.

If $h, \hat{h} \in \Gamma(S^2 T^* M)$ are fields of symmetric bilinear forms, we write $h \geq \hat{h}$ if $h_p \geq \hat{h}_p$ for all $p \in M$.

The following will also be useful:

Definition (Riemannian covering map). Let (M, g) and $(\tilde{M}, \tilde{g})$ be two Riemannian manifolds, and $f : \tilde{M} \to M$ be a smooth covering map. We say f is a *Riemannian covering map* if it is a local isometry. Alternatively, $f^* g = \tilde{g}$. We say $\tilde{M}$ is a Riemannian cover of M.

Recall that if f is in fact a universal cover, i.e. $\tilde{M}$ is simply connected, then we can (non-canonically) identify $\pi_1(M)$ with $f^{-1}(p)$ for any point $p \in M$.

Definition (Bonnet–Myers diameter theorem). Let (M, g) be a complete n-dimensional manifold with

$$\operatorname{Ric}(g) \geq \frac{n-1}{r^2} g,$$

where $r > 0$ is some positive number. Then

$$\operatorname{diam}(M, g) \leq \operatorname{diam} S^n(r) = \pi r.$$

In particular, M is compact and $\pi_1(M)$ is finite.

Proof. Consider any $L < \operatorname{diam}(M, g)$. Then by definition (and Hopf–Rinow), we can find $p, q \in M$ such that $d(p, q) = L$, and a minimal geodesic $\gamma \in \Omega(p, q)$ with $\ell(\gamma) = d(p, q)$. We parametrize $\gamma : [0, L] \to M$ so that $|\dot{\gamma}| = 1$.

Now consider any vector field Y along γ such that $Y(p) = 0 = Y(q)$. Since Γ is a minimal geodesic, it is a critical point for ℓ, and the second variation $I(Y, Y)_{[0, L]}$ is non-negative (recall that the second variation has fixed end points).

We extend $\dot{\gamma}(0)$ to an orthonormal basis of $T_p M$, say $\dot{\gamma}(0) = e_1, e_2, \cdots, e_n$. We further let X_i be the unique vector field such that

$$X_i' = 0, \quad X_i(0) = e_i.$$

In particular, $X_1(t) = \dot{\gamma}(t)$.

For $i = 2, \cdots, n$, we put

$$Y_i(t) = \sin\left(\frac{\pi t}{L}\right) X_i(t).$$

Then after integrating by parts, we find that we have

$$I(Y_i, Y_i)_{[0, L]} = -\int_0^L g(Y_i'' + R(\dot{\gamma}, Y_i) Y_i, \dot{\gamma}) \, \mathrm{d}t$$

Using the fact that X_i is parallel, this can be written as

$$= \int_0^L \sin^2 \frac{\pi t}{L} \left(\frac{\pi^2}{L^2} - R(\dot{\gamma}, X_i, \dot{\gamma}, X_i) \right) \, \mathrm{d}t,$$

and since this is length minimizing, we know this is ≥ 0.

We note that we have $R(\dot{\gamma}, X_1, \dot{\gamma}, X_1) = 0$. So we have

$$\sum_{i=2}^n R(\dot{\gamma}, X_i, \dot{\gamma}, X_i) = \mathrm{Ric}(\dot{\gamma}, \dot{\gamma}).$$

So we know

$$\sum_{i=2}^n I(Y_i, Y_i) = \int_0^L \sin^2 \frac{\pi t}{L} \left((n-1)\frac{\pi^2}{L} - \mathrm{Ric}(\dot{\gamma}, \dot{\gamma}) \right) \, \mathrm{d}t \geq 0.$$

We also know that

$$\mathrm{Ric}(\dot{\gamma}, \dot{\gamma}) \geq \frac{n-1}{r^2}$$

by hypothesis. So this implies that

$$\frac{\pi^2}{L^2} \geq \frac{1}{r^2}.$$

This tells us that

$$L \leq \pi r.$$

Since L is any number less that $\mathrm{diam}(M, g)$, it follows that

$$\mathrm{diam}(M, g) \leq \pi r.$$

Since M is known to be complete, by Hopf-Rinow theorem, any closed bounded subset is compact. But M itself is closed and bounded! So M is compact.

To understand the fundamental, group, we simply have to consider a universal Riemannian cover $f : \tilde{M} \to M$. We know such a topological universal covering space must exist by general existence theorems. We can then pull back the differential structure and metric along f, since f is a local homeomorphism. So this gives a universal Riemannian cover of M. But they hypothesis of the theorem is local, so it is also satisfied for $\tilde{M}$. So it is also compact. Since $f^{-1}(p)$ is a closed discrete subset of a compact space, it is finite, and we are done. $\square$

It is an easy exercise to show that the hypothesis on the Ricci curvature cannot be weakened to just saying that the Ricci curvature is positive definite.

Hadamard–Cartan theorem

To prove the next result, we need to talk a bit more about coverings.

Proposition. Let (M, g) and (N, h) be Riemannian manifolds, and suppose M is complete. Suppose there is a smooth surjection $f : M \to N$ that is a local diffeomorphism. Moreover, suppose that for any $p \in M$ and $v \in T_pM$, we have $|\mathrm{d}f_p(v)|_h \geq |v|$. Then f is a covering map.

Proof. By general topology, it suffices to prove that for any smooth curve $\gamma : [0,1] \to N$, and any $q \in M$ such that $f(q) = \gamma(0)$, there exists a lift of γ starting from from q.

$$
\begin{array}{ccc}
 & & M \\
 & {}^{\tilde{\gamma}}\nearrow & \downarrow f \\
[0,1] & \xrightarrow{\ \gamma\ } & N
\end{array}
$$

From the hypothesis, we know that $\tilde{\gamma}$ exists on $[0, \varepsilon_0]$ for some "small" $\varepsilon_0 > 0$. We let
$$I = \{0 \leq \varepsilon \leq : \tilde{\gamma} \text{ exists on } [0, \varepsilon]\}.$$

We immediately see this is non-empty, since it contains ε_0. Moreover, it is not difficult to see that I is open in $[0,1]$, because f is a local diffeomorphism. So it suffices to show that I is closed.

We let $\{t_n\}_{n=1}^{\infty} \subseteq I$ be such that $t_{n+1} > t_n$ for all n, and
$$\lim_{n \to \infty} t_n = \varepsilon_1.$$

Using Hopf-Rinow, either $\{\tilde{\gamma}(t_n)\}$ is contained in some compact K, or it is unbounded. We claim that unboundedness is impossible. We have

$$
\begin{aligned}
\ell(\gamma) &\geq \ell(\gamma|_{[0,t_n]}) \\
&= \int_0^{t_n} |\dot{\gamma}| \,\mathrm{d}t \\
&= \int_0^{t_n} |\mathrm{d}f_{\tilde{\gamma}(t)} \dot{\tilde{\gamma}}(t)| \,\mathrm{d}t \\
&\geq \int_0^{t_n} |\dot{\tilde{\gamma}}| \,\mathrm{d}t \\
&= \ell(\tilde{\gamma}|_{[0,t_n]}) \\
&\geq d(\tilde{\gamma}(0), \tilde{\gamma}(t_n)).
\end{aligned}
$$

So we know this is bounded. So by compactness, we can find some x such that $\tilde{\gamma}(t_{n_\ell}) \to x$ as $\ell \to \infty$. There exists an open $x \in V \subseteq M$ such that $f|_V$ is a diffeomorphism.

Since there are extensions of $\tilde{\gamma}$ to each t_n, eventually we get an extension to within V, and then we can just lift directly, and extend it to ε_1. So $\varepsilon_1 \in I$. So we are done. $\qquad\square$

Corollary. Let $f : M \to N$ be a local isometry onto N, and M be complete. Then f is a covering map.

Note that since N is (assumed to be) connected, we know f is necessarily surjective. To see this, note that the completeness of M implies completeness of $f(M)$, hence $f(M)$ is closed in N, and since it is a local isometry, we know f is in particular open. So the image is open and closed, hence $f(M) = N$.

For a change, our next result will assume a *negative* curvature, instead of a positive one!

Theorem (Hadamard–Cartan theorem)**.** Let (M^n, g) be a complete Riemannian manifold such that the sectional curvature is always non-positive. Then for every point $p \in M$, we know $\exp_p : T_p M \to M$ is a covering map. In particular, if $\pi_1(M) = 0$, then M is diffeomorphic to $\mathbb{R}^n$.

We will need one more technical lemma.

Lemma. Let $\gamma(t)$ be a geodesic on (M, g) such that $K \leq 0$ along γ. Then γ has no conjugate points.

Proof. We write $\gamma(0) = p$. Let $I(t)$ be a Jacobi field along γ, and $J(0) = 0$. We claim that if J is not identically zero, then J does not vanish everywhere else.

We consider the function

$$f(t) = g(J(t), J(t)) = |J(t)|^2.$$

Then $f(0) = f'(0) = 0$. Consider

$$\frac{1}{2} f''(t) = g(J''(t), J(t)) + g(J'(t), J'(t)) = g(J', J') - R(\dot{\gamma}, J, \dot{\gamma}, J) \geq 0.$$

So f is a convex function, and so we are done. $\qquad\square$

We can now prove the theorem.

Proof. By the lemma, we know there are not conjugate points. So we know $\exp_p$ is regular everywhere, hence a local diffeomorphism by inverse function theorem. We can use this fact to pull back a metric from M to $T_p M$ such that $\exp_p$ is a local isometry. Since this is a local isometry, we know geodesics are preserved. So geodesics originating from the origin in $T_p M$ are straight lines, and the speed of the geodesics under the two metrics are the same. So we know $T_p M$ is complete under this metric. Also, by Hopf–Rinow, $\exp_p$ is surjective. So we are done. $\qquad\square$

4 Hodge theory on Riemannian manifolds

Throughout this chapter, we will assume our manifolds are oriented, and write n of the dimension. We will write $\varepsilon \in \Omega^n(M)$ for a non-vanishing form defining the orientation.

Given a coordinate patch $U \subseteq M$, we can use Gram-Schmidt to obtain a positively-oriented orthonormal frame $e_1, \cdots, e_n$. This allows us to dualize and obtain a basis $\omega_1, \cdots, \omega_n \in \Omega^1(n)$, defined by

$$\omega_i(e_i) = \delta_{ij}.$$

Since these are linearly independent, we can multiply all of them together to obtain a non-zero n-form

$$\omega_1 \wedge \cdots \wedge \omega_n = a\varepsilon,$$

for some $a \in C^\infty(U)$, $a > 0$. We can do this for any coordinate patches, and the resulting n-form agrees on intersections. Indeed, given any other choice $\omega'_1, \cdots, \omega'_n$, they must be related to the original $\omega_1, \cdots, \omega_n$ by an element $\Phi \in \mathrm{SO}(n)$. Then by linear algebra, we have

$$\omega'_1 \wedge \cdots \wedge \omega'_n = \det(\Phi)\omega_1 \wedge \cdots \wedge \omega_n = \omega_1 \wedge \cdots \wedge \omega_n.$$

So we can patch all these forms together to get a global n-form $\omega_g \in \Omega^n(M)$ that gives the same orientation. This is a canonical sch n-form, depending only on g and the orientation chosen. This is called the (Riemannian) *volume form* of (M, g).

Recall that the ω_i are orthonormal with respect to the natural dual inner product on T^*M. In general, g induces an inner product on $\Lambda^p T^*M$ for all $p = 0, 1, \cdots, n$, which is still denoted g. One way to describe this is to give an orthonormal basis on each fiber, given by

$$\{\omega_{i_1} \wedge \cdots \wedge \omega_{i_p} : 1 \leq i_1 < \cdots < i_p \leq n\}.$$

From this point of view, the volume form becomes a form of unit length.

We now come to the central definition of Hodge theory.

Definition (Hodge star). The *Hodge star operator* on (M^n, g) is a linear map

$$* : \Lambda^p T^*_x M \to \Lambda^{n-p} T^*_x M$$

satisfying the property that for all $\alpha, \beta \in \Lambda^p T^*_x M$, we have

$$\alpha \wedge *\beta = \langle \alpha, \beta \rangle_g \theta_g.$$

Since g is non-degenerate, it follows that this is well-defined.

How do we actually compute this? Since we have vector spaces, it is natural to consider what happens in a basis. Suppose $\omega_1, \cdots, \omega_n$ is an orthonormal basis of $T^*_x M$. Then we claim that

$$*(\omega_1 \wedge \cdots \wedge \omega_p) = \omega_{p+1} \wedge \cdots \wedge \omega_n.$$

We can check this by checking all basis vectors of $\Lambda^p M$, and the result drops out immediately. Since we can always relabel the numbers, this already tells us how to compute the Hodge star of all other basis elements.

We can apply the Hodge star twice, which gives us a linear endomorphism $** : \Lambda^p T^*_x M \to \Lambda^p T^*_x M$.

Proposition. The double Hodge star $** : \Lambda^p T_x^* M \to \Lambda^p T_x^* M$ is equal to $(-1)^{p(n-p)}$.

In particular,
$$*1 = \omega_g, \quad *\omega_g = 1.$$

Using the Hodge star, we can define a differential operator:

Definition (δ)**.** We define $\delta : \Omega^p(M) \to \Omega^{p-1}(M)$ for $0 \leq p \leq \dim M$ by

$$\delta = \begin{cases} (-1)^{n(p+1)+1} * \mathrm{d}* & p \neq 0 \\ 0 & p = 0 \end{cases}.$$

This is (sometimes) called the *co-differential*.

The funny powers of (-1) are chosen so that our future results work well. We further define

Definition (Laplace–Beltrami operator Δ)**.** The *Laplace–Beltrami operator* is

$$\Delta = \mathrm{d}\delta + \delta\mathrm{d} : \Omega^p(M) \to \Omega^p(M).$$

This is also known as the (Hodge) *Laplacian*.

Consider the spacial case of $(M, g) = (\mathbb{R}^n, \mathrm{eucl})$, and $p = 0$. Then a straightforward calculation shows that

$$\Delta f = -\frac{\partial^2 f}{\partial x_1^2} - \cdots - \frac{\partial^2 f}{\partial x_n^2}$$

for each $f \in C^\infty(\mathbb{R}^n) = \Omega^0(\mathbb{R}^n)$.

The negative sign is there for a good reason. For example, having a negative sign there ensures the eigenvalue of Δ are non-negative.

For a general metric $g = g_{ij} \, \mathrm{d}x^i \, \mathrm{d}x^j$ on $\mathbb{R}^n$ (or alternatively a coordinate patch on any Riemannian manifold), we have

$$\omega_g = \sqrt{\det g} \, \mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n,$$

and then we have

$$\Delta_g f = -\frac{1}{\sqrt{g}} \partial_j (\sqrt{\det g} \, g^{ij} \partial_i f) = -g^{ij} \partial_i \partial_j f + \text{lower order terms}.$$

As before, g^{ij}, as a matrix, is the inverse of g. Algebraic, it is the dual inner product on the cotangent space.

Proposition. δ is the formal adjoint of d. Explicitly, for any compactly supported $\alpha \in \Omega^{p-1}$ and $\beta \in \Omega^p$, then

$$\int_M \langle \mathrm{d}\alpha, \beta \rangle_g \omega_g = \int_M \langle \alpha, \delta\beta \rangle_g.$$

We just say it is a formal adjoint, rather than a genuine adjoint, because there is no obvious Banach space structure on $\Omega^p(M)$, and we don't want to go into that.

We can also call this the L^2 adjoint, where we define

$$\langle\langle \xi, \eta \rangle\rangle_g = \int_M \langle \xi, \eta \rangle_g \omega_g,$$

where $\xi, \eta \in \Omega^p(M)$. Then the proposition says

$$\langle\langle \mathrm{d}\alpha, \beta \rangle\rangle_g = \langle\langle \alpha, \delta\beta \rangle\rangle_g.$$

To prove this, we need to recall Stokes' theorem. Since we don't care about manifolds with boundary in this course, we just have

$$\int_M \mathrm{d}\omega = 0$$

for all forms ω.

Proof. We have

$$\begin{aligned}
0 &= \int_M \mathrm{d}(\alpha \wedge *\beta) \\
&= \int_M \mathrm{d}\alpha \wedge *\beta + \int_M (-1)^{p-1}\alpha \wedge \mathrm{d}*\beta \\
&= \int_M \langle \mathrm{d}\alpha, \beta \rangle_g \omega_g + (-1)^{p-1}(-1)^{(n-p+1)(p-1)} \int_M \alpha \wedge **\mathrm{d}*\beta \\
&= \int_M \langle \mathrm{d}\alpha, \beta \rangle_g \omega_g + (-1)^{(n-p)(p-1)} \int_M \alpha \wedge **\mathrm{d}*\beta \\
&= \int_M \langle \mathrm{d}\alpha, \beta \rangle_g \omega_g - \int_M \alpha \wedge *\delta\beta \\
&= \int_M \langle \mathrm{d}\alpha, \beta \rangle_g \omega_g - \int_M \langle \alpha, \delta\beta \rangle_g \omega_g.
\end{aligned}$$

$\square$

Corollary. Δ is formally self-adjoint.

Note that we have
$$*\Delta = \Delta *.$$

Definition (Harmonic forms). A *harmonic form* is a p-form ω such that $\Delta\omega = 0$. We write

$$\mathcal{H}^p = \{\alpha \in \Omega^p(M) : \Delta\alpha = 0\}.$$

One nice result about these forms follows if we assume that M is compact. We have a further corollary of the proposition.

Corollary. Let M be compact. Then

$$\Delta\alpha = 0 \Leftrightarrow \mathrm{d}\alpha = 0 \text{ and } \delta\alpha = 0.$$

We say α is closed and *co-closed*.

Proof. $\Leftarrow$ is clear. For $\Rightarrow$, suppose $\Delta\alpha = 0$. Then we have

$$0 = \langle\langle \alpha, \Delta\alpha \rangle\rangle = \langle\langle \mathrm{d}\delta\alpha + \delta\mathrm{d}\alpha \rangle\rangle = \|\delta\alpha\|_g^2 + \|\mathrm{d}\alpha\|_g^2.$$

Since the L^2 norm is non-degenerate, it follows that $\delta\alpha = \mathrm{d}\alpha = 0$. $\qquad\square$

In particular, in degree 0, co-closed is automatic. Then for all $f \in C^\infty(M)$, we have

$$\Delta f = 0 \Leftrightarrow \mathrm{d}f = 0.$$

In other words, harmonic forms on a compact manifold must be constant. This is a good way to demonstrate that the compactness hypothesis is required, as there are many non-trivial harmonic functions on $\mathbb{R}^n$, e.g. x.

Some of these things simplify if we know about the parity of our manifold. If $\dim M = n = 2m$, then $** = (-1)^p$, and

$$\delta = -*\mathrm{d}*$$

whenever $p \neq 0$. In particular, this applies to complex manifolds, say $\mathbb{C}^n \cong \mathbb{R}^{2n}$, with the Hermitian metric. This is to be continued in sheet 3.

4.1 Hodge decomposition theorem

We now work towards proving the Hodge decomposition theorem. This is a very important and far-reaching result.

Theorem (Hodge decomposition theorem)**.** Let (M, g) be a compact oriented Riemannian manifold. Then

– For all $p = 0, \cdots, \dim M$, we have $\dim \mathcal{H}^p < \infty$.

– We have
$$\Omega^p(M) = \mathcal{H}^p \oplus \Delta\Omega^p(M).$$

It follows that

$$\Omega^p(M) = \mathcal{H}^p \oplus \mathrm{d}\delta\Omega^p(M) \oplus \delta\mathrm{d}\Omega^p(M)$$
$$= \mathcal{H}^p \oplus \mathrm{d}\Omega^{p-1}(M) \oplus \delta\Omega^{p+1}(M).$$

Moreover, the direct sums are orthogonal with respect to the L^2 inner product. We also formally set $\Omega^{-1}(M) = 0$.

As before, the compactness of M is essential, and cannot be dropped.

This has a rather remarkable corollary. This requires recalling the notion of de Rham cohomology.

Corollary. Let (M, g) be a compact oriented Riemannian manifold. Then for all $\alpha \in H^p_{\mathrm{dR}}(M)$, there is a unique $\alpha \in \mathcal{H}^p$ such that $[\alpha] = a$. In other words, the obvious map

$$\mathcal{H}^p \to H^p_{\mathrm{dR}}(M)$$

is an isomorphism.

This is remarkable. On the left hand side, we have $\mathcal{H}^p$, which is a completely analytic thing, defined by the Laplacian. On the other hand, the right hand sides involves the de Rham cohomology, which is just a topological, and in fact homotopy invariant. It is remarkable that these can identified.

Proof. To see uniqueness, suppose $\alpha_1, \alpha_2 \in \mathcal{H}^p$ are such that $[\alpha_1] = [\alpha_2] \in H^p_{\mathrm{dR}}(M)$. Then
$$\alpha_1 - \alpha_2 = \mathrm{d}\beta$$
for some β. But the left hand side and right hand side live in different parts of the Hodge decomposition. So they must be individually zero. Alternatively, we can compute
$$\|\mathrm{d}\beta\|^2_g = \langle\langle\mathrm{d}\beta, \alpha_1 - \alpha_2\rangle\rangle_g = \langle\langle\beta, \delta\alpha_1 - \delta\alpha_2\rangle\rangle = 0$$
since harmonic forms are closed.

To prove existence, let $\alpha \in \Omega^p(M)$ be such that $\mathrm{d}\delta a = 0$. Then as before, we find that $\|\delta\alpha\| = 0$. Then if $\tilde{\alpha} \in \Omega^p(M)$ is closed, i.e. $\mathrm{d}\tilde{\alpha} = 0$, then by Hodge decomposition, we can write
$$\alpha \in \mathcal{H}^p \oplus \mathrm{d}\Omega^{p-1}(M).$$

$\square$

Why is this true? Suppose we are given some de Rham cohomology class $a \in H^p_{\mathrm{dR}}(M)$. We consider
$$B_a = \{\xi \in \Omega^p(M) : \mathrm{d}\xi = 0, [\xi] = a\}.$$
This is an infinite dimensional affine space.

We now ask ourselves — which $\alpha \in B_a$ minimizes the L^2 norm? We consider the function $F : B_a \to \mathbb{R}$ given by $F(\alpha) = \|\alpha\|^2$. Any minimizing α is an extremum. So for any $\beta \in \Omega^{p-1}(M)$, we have
$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} F(\alpha + t\mathrm{d}\beta) = 0.$$
In other words, we have
$$0 = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} (\|\alpha\|^2 + 2t\langle\langle\alpha, \mathrm{d}\beta\rangle\rangle + t^2\|\mathrm{d}\beta\|^2) = 2\langle\langle\alpha, \mathrm{d}\beta\rangle\rangle.$$
This is the same as saying
$$\langle\langle\delta\alpha, \beta\rangle\rangle = 0.$$
So this implies $\delta\alpha = 0$. But $\mathrm{d}\alpha = 0$ by assumption. So we find that $\alpha \in \mathcal{H}^p$. So the result is at least reasonable.

Proof of Hodge decomposition theorem. The key statement is to prove that
$$\Omega^p(M) = \mathcal{H}^p \oplus \Delta\Omega^p(M).$$
Once we have this proved, the other decompositions follow from standard results such as Stokes' theorem. For example, we know that
$$\langle\langle\mathrm{d}\delta\alpha, \delta\mathrm{d}\alpha\rangle\rangle = \langle\langle\mathrm{d}\mathrm{d}\delta\alpha, \mathrm{d}\alpha\rangle\rangle = 0.$$

So we have a decomposition

$$\Delta\Omega^p(M) = \mathrm{d}\delta(\Omega^p) \oplus \delta\mathrm{d}(\Omega^p).$$

To get the other expressions, we note that

$$\mathrm{d}\Omega^{p-1}(M) = \mathrm{d}(\mathcal{H}^{p-1} \oplus \Delta\Omega^{p-1}(M)) = \mathrm{d}(\delta\mathrm{d}\Omega^{p-1}(M)) \subseteq \mathrm{d}\delta\Omega^p(M).$$

On the other hand, we certainly have the other inclusion. So the two terms are equal.

The claim that $\dim\mathcal{H}^p < \infty$ will follow as a by-product of the key argument. So we start proving it.

We want to solve

$$\Delta\omega = \alpha \tag{$\dagger$}$$

for $\omega \in \Omega^p(M)$ given $\alpha \in \Omega^p(M)$. We shall require and assume 2 (non-trivial) from analysis of PDE's. We need a bit of set up before we can state the theorems we need to quote.

Anywhere in the proof, we use always $\langle\,\cdot\,,\,\cdot\,\rangle$ for the L^2 inner product, and $\|\cdot\|$ is the L^2 norm.

Note that if (and only if) ω is a solution of $(*)$, then we trivially have

$$\langle\Delta\omega,\varphi\rangle = \langle\alpha,\varphi\rangle \tag{$*$}$$

for all $\varphi \in \Omega^p(M)$.

Note that Δ is self-adjoint, i.e. $\Delta^* = \Delta$, but we will sometimes write Δ^* to keep track of how many adjoints we have taken.

We define a linear form $\ell : \beta \in \Omega^p(M) \to \langle\omega,\beta\rangle \in \mathbb{R}$. Then we have

$$\langle\Delta\omega,\varphi\rangle = \langle\omega,\Delta^*\varphi\rangle.$$

So $(*)$ becomes

$$\ell(\Delta^*\varphi) = \langle\alpha,\varphi\rangle \tag{$**$}$$

for all $\varphi \in \Omega^p(M)$.

Definition (Weak solution)**.** Call a linear functional $\ell : \Omega^p(M) \to \mathbb{R}$ a *weak solution* of $(\dagger)$ if

(i) ℓ satisfies $(**)$

(ii) ℓ is "bounded", i.e. there is some C such that $|\ell(\beta)| < C\|\beta\|$ for all β.

We now quote the two results from PDE theory.

Theorem (Regularity theorem)**.** Every weak solution of $(\dagger)$ is of the form

$$\ell(\beta) = \langle\omega,\beta\rangle$$

for $\omega \in \Omega^p(M)$.

Note that if this is the case, then we have

$$\langle\Delta\omega,\beta\rangle = \langle\omega,\Delta^*\beta\rangle = \ell(\Delta^*\beta) = \langle\alpha,\beta\rangle.$$

for all $\beta \in \Omega^p(M)$. So it follows that $\Delta\omega = \alpha$. So what this theorem says is that all weak solutions are genuine solutions.

Note that in analysis, we had the Riesz representation theorem, which says that if we have a Hilbert space, then any linear functional is of the form $\langle \omega, \cdot \rangle$. However, C^∞ does not have a natural Hilbert space structure. Riesz representation theorem will give us some not-necessarily smooth form that represents ℓ, and the regularity theorem says this is in fact C^∞.

Theorem (Compactness theorem)**.** If a sequence $\alpha_n \in \Omega^n(M)$ satisfies $\|\alpha_n\| < C$ and $\|\Delta \alpha_n\| < C$ for all n, Then α_n contains a Cauchy subsequence.

We are not saying that $\Omega^n(M)$ is compact, because it is not complete. So the best thing we can say is that the subsequence is Cauchy.

We now know that $\dim \mathcal{H}^p < \infty$. Otherwise, by Gram-Schmidt, we obtain an infinite orthonormal sequence e_n, satisfying $\|e_n\| = 1$ and $\|\Delta e_n\| = 0$, which certainly doesn't have a Cauchy subsequence.

Now suppose we have picked an orthonormal basis $e_1, \cdots, e_m$. Then any $\xi \in \Omega^p(M)$ can be written as

$$\xi = \left(\sum_{i=1}^m \langle \xi, e_i \rangle e_i \right) + \left(\xi - \sum_{i=1}^m \langle \xi, e_i \rangle e_i \right).$$

So we can write

$$\Omega^p(M) = \mathcal{H}^p \oplus (\mathcal{H}^p)^\perp.$$

Crucially, we know $(\mathcal{H}^p)^\perp$ is a *closed* subspace. What we want to show is that

$$(\mathcal{H}^p)^\perp = \Delta \Omega^p(M).$$

Suppose $\alpha \in \mathcal{H}^p$. Then we have

$$\langle \alpha, \Delta \beta \rangle = \langle \Delta \alpha, \beta \rangle = 0.$$

So we know that

$$\Delta \Omega^p(M) \subseteq (\mathcal{H}^p)^\perp.$$

The other direction is the hard part. Suppose $\alpha \in (\mathcal{H}^p)^\perp$. We may assume α is non-zero. Since our PDE is a linear one, we may wlog $\|\alpha\| = 1$.

By the regularity theorem, it suffices to prove that $\Delta \omega = \alpha$ has a *weak* solution. We define $\ell : \Delta \Omega^p(M) \to \mathbb{R}$ as follows: for each $\eta \in \Omega^p(M)$, we put

$$\ell(\Delta \eta) = \langle \eta, \alpha \rangle.$$

Suppose that $\Delta \eta = \Delta \xi$, then we have

$$\ell(\Delta \eta - \Delta \xi) = \langle \eta - \xi, \alpha \rangle = 0$$

as $\eta - \xi \in \mathcal{H}^p$. So we have $\ell(\Delta \eta) = \ell(\Delta \xi)$. Thus, ℓ is well-defined on $\Delta \Omega^p(M) \subseteq \Omega^p(M)$.

Then property (i) of the weak solution holds by construction. We next want to show the boundedness property. We now claim that there exists a positive $C > 0$ such that

$$\ell(\Delta \eta) \leq C \|\Delta \eta\|$$

for all $\eta \in \Omega^p(M)$. To see this, we first note that we always have

$$|\langle \alpha, \eta \rangle| \leq \|\alpha\| \cdot \|\eta\| = \|\eta\|.$$

So it suffices to show that there is a $C > 0$ such that

$$\|\eta\| \leq C\|\Delta\eta\|$$

for every $\eta \in \Omega^p(M)$. Suppose not. Then we can find a sequence $\eta_n \in (\mathcal{H}^p)^\perp$ such that $\|\eta_n\| = 1$ and $\|\nabla\eta_n\| \to 0$.

But then $\|\nabla\eta_n\|$ is certainly bounded. So by the compactness theorem, we have a Cauchy subsequence η_{n_k}. Rename $\eta_{n_k} = \eta_k$. Then for any $\psi \in \Omega^p(M)$, the limit $\lim\langle\psi, \eta_k\rangle$ exists.

We define $a : \Omega^p(M) \to \mathbb{R}$ by

$$a(\psi) = \lim_{k\to\infty} \langle\psi, \eta_k\rangle.$$

Then we have

$$a(\Delta\psi) = \lim_{k\to\infty} \langle\eta_k, \Delta\psi\rangle = \lim_{k\to\infty} \langle\Delta\eta_k, \psi\rangle = 0.$$

So we know that a is a weak solution of $\Delta\xi = 0$. By the regularity theorem again, we have

$$a(\psi) = \langle\xi, \psi\rangle$$

for some $\xi \in \Omega^p(M)$. Then this forces $\Delta\xi = 0$.

Then we know $\eta_k \to \xi$ in $\|\cdot\|$. Then since $\eta_k \in (\mathcal{H}^p)^\perp$, and $(\mathcal{H}^p)^\infty$ is closed, we know $\xi \in (\mathcal{H}^p)^\perp$. But also by assumption, we have $\xi \in \mathcal{H}^p$. So $\xi = 0$. But we know $\|\xi\| = \lim\|\eta_k\| = 1$, which is bad. So ℓ is bounded.

So we now have a linear map ℓ that is well-defined and bounded on the subspace $\Delta\Omega^p(M)$. The last step requires another result from analysis.

Theorem (Hahn–Banach theorem)**.** Let L be a normed vector space, and L_0 be a subspace. We let $f : L_0 \to \mathbb{R}$ be such that $|f(x_0)| \leq \|x_0\|$ for all $x_0 \in L_0$. Then f extends to $L \to \mathbb{R}$ linear and with $|f(x)| \leq \|x\|$ for all $x \in L$.

We then apply this to $f = \frac{1}{C}\ell$, with $L == \Omega^p(M)$ and $L_0 = \Delta\Omega^p(M)$. Then f extends to a weak solution of $\Delta\omega = \alpha$. So we are done by the regularity theorem. □

Since the Hodge decomposition theorem is important, we look at one explict example, which requires easier (and more believable) analysis.

Example. Consider $M = T^n = \mathbb{R}^n/(2\pi\mathbb{Z})^n$, the n-torus with flat metric. This has local coordinates $(x_1, \cdots, x_n)$, induced from the Euclidean space. This is convenient because $\Lambda^p T^*M$ is trivialized by $\{\mathrm{d}x_{i_1} \wedge \cdots \mathrm{d}x_{i_p}\}$. Moreover, the Laplacian is just given by

$$\Delta(\alpha\mathrm{d}x_{i_1} \wedge \cdots \wedge \mathrm{d}x_{i_p}) = -\sum_{i=1}^n \frac{\partial^2\alpha}{\partial x_i^2}\mathrm{d}x_{i_1} \wedge \cdots \wedge \mathrm{d}x_{i_p}.$$

So to do Hodge decomposition, it suffices to consider the case $p = 0$, and we are just looking at functions $C^\infty(T^n)$, namely the 2π-periodic functions on $\mathbb{R}$.

Here we will quote the fact that Fourier series work.

Fact. Let $\varphi \in C^\infty(T^n)$. Then it can be (uniquely) represented by a convergenct Fourier series (to φ)

$$\varphi(x) = \sum_{k \in \mathbb{Z}^n} \varphi_k e^{ik \cdot x},$$

where k and x are vectors, and $k \cdot x$ is the standard inner product. Moreover,

$$\varphi_k = \frac{1}{(2\pi)^n} \int_{T^n} \varphi(x) e^{-ik \cdot x} \, dx.$$

Note that $\varphi_{-k} = \bar{\varphi}_k \in \mathbb{C}$, and this ensures the resulting sum is real.

This series converges, and convergence is uniform in T^n with all derivatives (term-by-term). Further, for every $m \in \mathbb{Z}$, we have $\varphi_k = o(|k|^m)$ as $k \to \infty$.

Now consider

$$\ell^2 = \left\{ (\varphi_k)_{k \in \mathbb{Z}^n} : \varphi_{-k} = \bar{\varphi}_k \text{ for all } k \text{ and } \sum_k |\varphi_k|^2 < \infty \right\}.$$

Then the Fourier series gives us an $\mathbb{R}$-linear map $\mathcal{F} : C^\infty(T^n) \to \ell^2$.

This is injective, but not surjective, because the Fourier coefficients obey strong decay properties. In fact, the image is

$$H_\infty = \left\{ (\varphi_k) \in \ell^2 : \varphi_k = o(|k|^m) \text{ for all } m \in \mathbb{Z} \right\}.$$

We have an inner product on ℓ^2, given by

$$\langle \varphi, \psi \rangle = (2\pi)^n \sum_k \bar{\varphi}_k \psi_k.$$

Parseval's theorem says

$$\int_{T^n} |\varphi(x)|^2 \, dx = (2\pi)^n \sum_k |\varphi|^2.$$

So this shows that our linear map $\mathcal{F}$ is an isometry $C^\infty(T^n) \to H_\infty \subseteq \ell^2$.

So we have reduced our problem of working with functions on a torus to working with these infinite series. This makes our calculations rather more explicit.

We can now write down how the Laplacian works. By direct computation, we find

$$\mathcal{F}(\Delta\varphi) = (-|k|^2 \varphi_k).$$

In this case, we know that $\alpha \in (\mathcal{H}^0)^\perp$ iff $\alpha_0 = 0$.

Now ℓ is a weak solution of $\Delta\omega = \alpha$ if $\ell(\Delta\varphi) = \langle \alpha, \varphi \rangle$ for all $\varphi \in C^\infty(T^n)$. In terms of the coefficients, this means

$$\ell((-|k|^2 \varphi_k)) = (2\pi)^n \sum_k \bar{\alpha}_k \varphi_k.$$

Now consider the $\mathbb{R}$-linear map $\lambda : H_\infty \to H_\infty$ by

$$(\alpha_k) \mapsto \left(\lambda_k(\alpha) = \begin{cases} \frac{\alpha_k}{-|k|^2} & k \neq 0 \\ 0 & k = 0 \end{cases} \right).$$

This defines the *Green's operator* $G : C^\infty(T^n) \to C^\infty(T^n)$ satisfying

$$
\begin{array}{ccc}
C^\infty(T^n) & \xrightarrow{\ G\ } & C^\infty(T^n) \\
\downarrow & & \downarrow \\
H_\infty & \xrightarrow{\ \lambda\ } & H_\infty
\end{array} \ .
$$

It is immediate that we have $G(\alpha) \in (\mathcal{H}^0)^\perp$ for all α. We set

$$
\ell(\beta) = \langle G(\alpha), \beta \rangle,
$$

which is bounded by Cauchy–Schwartz. In fact, the strong solution is $\omega = G(\alpha) \in C^\infty(T^n)$. This is what was promised by the regularity theorem.

For the compactness part, we restrict to $n = 1$, so that $T^n = S^1$. We efine the "*Hilbert cube*" by

$$
K = \left\{ \varphi_k) \in \ell^2 : \varphi_{-k} = \bar\varphi_k \in \mathbb{C} \text{ and } |\varphi_k| \frac{1}{|k} \text{ for all } k \neq 0 \right\} .
$$

It is an exercise to show that K is sequentially compact, i.e. Bolzano–Weierstrass holds.

The key property in all of the above is that Δ corresponds in ℓ^2 to the diagonal map

$$
\varphi \mapsto -|k|^2 \varphi_k.
$$

Since $|k|^2$ is non-zero for all but one k. So we were able to invert it on the complement of a finite-dimensional subspace. Such operators are particularly nice in analysis, and this is related to them being *ellipticelliptic*.

More can be found in Well's "Differentiable analysis on complex manifolds".

We now start a new topic which is more geometric in nature, but continues certain thoughts in Hodge decomposition.

4.2   Introduction to Bochner's method

Let $E \to M$ be a vector bundle over a Riemannian manifold with inner product on fibers, varying smoothly in the usual way. For a covariant derivative

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

we may define the formal adjoint $(\nabla^E)^* : \Gamma(E \otimes T^*M) \to \Gamma(E)$ by

$$
\int_M \langle \nabla\alpha, \beta \rangle_{E,g}\, \omega_g = \int_M \langle \alpha, \nabla^*\beta \rangle_E\, \omega_g
$$

for any $\alpha \in \Gamma(E)$ and $\beta \in \Gamma(E \otimes TM)$. Then

$$
\nabla^*\nabla : \Gamma(E) \to \Gamma(E)
$$

is the *covariant Laplacian*.

This is something we can do on any vector bundle, but in reality, we are only going to use it on one particular bundle — $E = T^*M$, and we just take ∇ to be

the (induced) Levi-Civita of g. We can check that for each local orthonormal frame field $e_1, \cdots, e_n$ on $U \subseteq M$, we have

$$\nabla^* \nabla \alpha|_U = -\sum_{i=1}^n \nabla_{e_i} \nabla_{e_i} \alpha.$$

To see this, we recall that

$$\nabla \alpha|_U = \sum_{i=1}^n (\nabla_{e_i} \alpha) \otimes \varepsilon_i,$$

where ε_i is the dual orthonormal co-frame field. It is an exercise to show that

$$\nabla^* \sum \beta_i \otimes \varepsilon_i = -\sum_{i=1}^n \nabla_{e_i} \beta_i.$$

This will become clearer as we progress, when we prove more general theory.

We now have a very obvious question — is this $\nabla^* \nabla$ the same as the Laplace–Beltrami operator Δ for 1-forms?

It turns out they are different, and Bochner's method is about exploiting the difference between the two operators.

It turns out the difference is related to the Ricci curvature. We now define the Ricci curvature again, which is now a linear endomorphism $\mathrm{Ric} : \Omega^1(M) \to \Omega^1(M)$.

To reconcile this with the other notion of Ricci curvature we had, given an $\alpha \in \Omega^1(M)$, we can find a unique $Y_\alpha \in \mathrm{Vect}(M)$ such that

$$\alpha(z) = g(Y_\alpha, Z)$$

for all $Z \in \mathrm{Vect}(M)$. Then given $\alpha \in \Omega^1(M)$, we define $\mathrm{Ric}(\alpha) \in \Omega^1(M)$ by

$$\mathrm{Ric}(\alpha)(X) = \mathrm{Ric}(X, Y_\alpha),$$

where the second Ric is the Ricci curvature we are used to.

Theorem (Bochner–Weitzenböck formula)**.** On an oriented Riemannian manifold, we have
$$(\Delta \alpha)(X) = (\nabla^* \nabla \alpha)(X) + \mathrm{Ric}(\alpha)(X).$$
for $\alpha \in \Omega^1(M)$ and $X \in \mathrm{Vect}(M)$. In other words, we have

$$\Delta \alpha = \nabla^* \nabla \alpha + \mathrm{Ric}(\alpha),$$

or, even more concisely,

$$\Delta = \nabla^* \nabla + \mathrm{Ric}.$$

To prove this, we start with a rather technical lemma.

Lemma (Ricci identity)**.** Let M be any Riemannian manifold, and $X, Y, Z \in \mathrm{Vect}(M)$ and $\alpha \in \Omega^1(M)$. Then

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

This is not a difficult thing to prove.

Proof. We have

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

Then we have

$$\langle \alpha, \nabla_{[X,Y]} Z - [\nabla_X, \nabla_Y] Z \rangle = [X,Y]\langle \alpha, Z \rangle - \langle \nabla_{[X,Y]} \alpha, Z \rangle - \langle \alpha, [\nabla_X, \nabla_Y] Z \rangle.$$

Also, we can compute

$$(XY - YX)\langle \alpha, Z \rangle = \langle [\nabla_X, \nabla_Y]\alpha, Z \rangle + \langle \alpha, [\nabla_X, \nabla_Y] Z \rangle.$$

And so we are done.   □

In ordinary multi-variable calculus, we had the notion of the divergence. This makes sense in general as well. Given any $X \in \text{Vect}(M)$, we have

$$\nabla X \in \Gamma(TM \otimes T^*M) = \Gamma(\text{End}\, TM).$$

Now we can take the trace of this, because the trace doesn't depend on the choice of the basis.

Definition (Divergence)**.** The *divergence* of a vector field $X \in \text{Vect}(M)$ is

$$\text{div}X = \text{tr}(\nabla X),.$$

It is not hard to see that this extends the familiar definition of the divergence. Indeed, we pick $e_1, \cdots, e_n$ a local orthonormal frame field. Then locally, we have

$$\text{div}X = \sum_{i=1}^{n} g(\nabla_{e_i} X, e_i).$$

We have, for any $f \in C^\infty(M)$,

$$\text{div}(fX) = \text{tr}(\nabla(fX)) = f\text{div}X + \langle \mathrm{d}f, X \rangle.$$

This is the product rule for divergence.

Now let $\theta \in \Omega^1(M)$ and $X_\theta \in \text{Vect}(M)$ be the dual via g. Then we have

$$\langle \theta, V \rangle = g(X_\theta, V)$$

for all $V \in TM$. Equivalently, we have

$$\langle \alpha, X_\theta \rangle = g(\alpha, \theta)$$

for all $\alpha \in T^*M$

Proposition. Let θ, X_θ be as above. Then

$$\delta\theta = -\text{div}X_\theta.$$

So the divergence we defined is not really a new operator. It is just, up to a sign, the L^2 adjoint of d.

This is a rather ambitious thing to prove. To prove this, we will need a lemma.

Lemma. We have

$$(\operatorname{div} X)\omega_g = \mathrm{d}(i(X)\omega_g),$$

for all $X \in \operatorname{Vect}(M)$, where $i(X) : \Omega^p(M) \to \Omega^{p-1}(M)$ is the *interior product* with X that sends $\psi(\,\cdot\,, \cdot\,, \cdots\,, \cdot\,)$ to $\psi(X, \cdot\,, \cdots\,, \cdot\,)$.

Some people also write the operator as

$$i(X)\psi = X \lrcorner \psi.$$

We now prove our proposition assuming the lemma

Proof of proposition. By the formal adjoint property of δ, we know that for any $f \in C^\infty(M)$, we have

$$\int_M g(\mathrm{d}f, \theta)\omega_g = \int_M f \cdot \delta\theta \cdot \omega_g.$$

We want to show that

$$\int_M g(\mathrm{d}f, \theta)\omega_g = - \int_M f \cdot \operatorname{div} X_\omega \cdot \omega_g.$$

We have

$$\int_M \operatorname{div}(f X_\theta)\omega_g = \int_M g(\mathrm{d}f, \theta)\omega_g + \int_M f \operatorname{div} X_\theta \cdot \omega_g.$$

By the lemma, we have

$$\operatorname{div}(f X_\theta)\omega_g = \mathrm{d}(i(f X_\theta)\omega_g).$$

So the LHS vanishes. So we are done. □

We then prove the lemma

Proof of lemma. We will assume two facts from sheet 3:

– $\nabla \omega_g = 0$.

– In local coordinates, for any p-form ψ, we have

$$\mathrm{d}\psi = \sum_{k=1}^{n} \mathrm{d}x_k \wedge \nabla_{\partial_k}\psi.$$

Now as in sheet 1, we have

$$\nabla_Y(i(X)\psi) = i(\nabla_Y X)\psi + i(X)\nabla_Y \psi/$$

Thus

$$\nabla_Y(i(X)\omega_g) = i(\nabla_Y X)\omega_g.$$

We also have

$$d(i(X)\omega_g)$$
$$= \sum_{k=1}^{n} \mathrm{d}x_k \wedge \nabla_k(i(X)\omega_g)$$
$$= \sum_{k=1}^{n} \mathrm{d}x - k \wedge i(\nabla_k X)\omega_g$$
$$= \sum_{k=1}^{n} \mathrm{d}x_k \wedge i(\nabla_k X)(\sqrt{\det g}\ \mathrm{d}x_k \wedge \cdots \mathrm{d}x_n \wedge \mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_{k-1}(-1)^{(k-1)(n-1)})$$
$$= \langle \mathrm{d}x_k, \nabla_k X \rangle \omega_g$$
$$= (\mathrm{div}X)\omega_g$$

$\square$

Corollary. If θ is a 1-form, then

$$\delta\theta = -\sum_{k=1}^{n} i(e_k)\nabla_{e_i}\theta = -\sum_{k=1}^{n} \langle \nabla_{e_k}\theta, e_k \rangle$$

where $\{e_k\}$ as before denotes a local orthonormal frame field.

Proof. According to the proposition, we have

$$\delta\theta = -\mathrm{div}X_\theta = -\sum_{i=1}^{n} g(\nabla_{e_i}X_\theta, e_i) = -\sum_{k=1}^{n} \langle \nabla_{e_i}\theta, e_i \rangle,$$

where the last equality follows from the fact that

$$\langle \nabla_{e_i}\theta, e_i \rangle + \langle \theta\nabla_{e_i}e_i \rangle = e_i \langle \theta, e_i \rangle,$$

and also we know that

$$g(\nabla_{e_i}X_\theta, e_i) + g(X_\theta, \nabla_{e_i}e_i) = e_i g(X_\theta, e_i).$$

$\square$

We will assume a version for 2-forms (the general result is again on the third example sheet):

Proposition. If $\beta \in \Omega^2(M)$, then

$$(\delta\beta)(Y) = -\sum_{k=1}^{n} (\nabla_{e_k}\beta)(e_k, Y).$$

In other words,

$$\delta\beta = -\sum_{k=1}^{n} i(e_k)(\nabla_{e_k}\beta).$$

We now take a closer look at ∇^* (and $\nabla^*\nabla$). We have a vector bundle $E \to M$ with inner product h on fibers, and a "compatible" connection $\nabla = \nabla^E : \Omega_M^p(E) \to \Omega_M^{p+1}(E)$. As in the case of the Levi-Civita connection, "compatible" means

$$\mathrm{d}(h(\alpha, \beta)) = h(\nabla\alpha, \beta) + h(\alpha, \nabla\beta)$$

for $\alpha, \beta \in \Gamma(E) = \Omega_M^0(E)$.

More generally, if

$$\alpha = \sum_I \alpha_I \, \mathrm{d}x^I, \beta = \sum_J \beta_J \, \mathrm{d}x^J \in \Omega_M^p(E)$$

are local p forms, where $I, J \subseteq \{1, \cdots, n\}$ are indices, then we have a Hodge star

$$*\beta = \sum_J \beta_J * \mathrm{d}x^J.$$

We can then define

$$\nabla^* = - * \nabla *$$

when $p = 1$.

Similar to scalar forms, we can check that

$$\nabla^*\beta = -\sum_{k=1}^n i(e_k)\nabla_{e_k}\beta.$$

In particular, when $E = T^*M$ with the canonical connection induced by the Levi-Civita connection, when $\alpha \in \Omega^1(M) = \Omega_M^0(T^*M)$, then

$$\nabla\alpha = \sum_\ell \varepsilon_\ell \otimes (\nabla_{e_\ell}\alpha) \in \Gamma(T^*M \otimes T^*M),$$

where ε_ℓ are the dual orthonormal frame fields. Then we have

$$\nabla^*\nabla_\alpha = -\sum_{k=1}^n \nabla_{e_k}\nabla_{e_k}\alpha$$

Finally, to prove Bochner-Weitzenböck, we need to pick our e_k a bit more cleverly. We fix a point $p \in M$. We let $e_1, \cdots, e_k$ be a *normal* local orthonormal frame field at p, i.e. in addition to being orthonormal, we need

$$\nabla_Y e_i|_p = 0$$

for every i and any $Y \in \mathrm{Vect}(M)$. This is again on the example sheet. The word "normal" comes from "normal neighbourhood". Then for any 1-form $\alpha \in \Omega^1(M)$ and any $X \in \mathrm{Vect}(M)$, then

$$
\begin{aligned}
\langle \mathrm{d}\delta\alpha, X \rangle &= X(\delta\alpha) \\
&= -\sum_{i=1}^n X\langle \nabla_{e_i}\alpha, e_i \rangle \\
&= -\sum_{i=1}^n \langle \nabla_X \nabla_{e_i}\alpha, e_i \rangle.
\end{aligned}
$$

We also have

$$d\alpha(X, Y) = X\langle\alpha, Y\rangle - Y\langle\alpha, X\rangle - \langle\alpha, [X, Y]\rangle$$
$$= \langle\nabla_X\alpha, Y\rangle - \bar{\nabla}_Y\alpha, X\rangle.$$

Recall that our goal is to understand

$$\Delta\alpha = d\delta\alpha + \delta d\alpha$$

for $\alpha \in \Omega^1(M)$. We have previously managed to understand $d\delta\alpha$. For a fixed point $p \in M$, pick a normal orthonormal frame field at p.

$$\langle d\delta\alpha, X\rangle = -\sum_{k=1}^{n}\langle\nabla_X\nabla_{e_k}\alpha, e_k\rangle.$$

We are going to exploit this normality quite a lot. Since we know

$$\nabla_Y e_k = 0$$

for all Y at p, we know

$$\nabla_{e_k} Y = [e_k, Y]$$

at p. Then we have

$$\langle d\delta\alpha, X\rangle = -\sum_{k=1}^{n}(\nabla_{e_k}d\alpha)(e_k, X).$$

We nwo note that for any $\beta \in \Omega^2(M)$, we have

$$(\nabla_Y\beta)(E, X) = Y(\beta(E, X) - \beta(\nabla_Y E, X) - \beta(E, \nabla_Y X)).$$

So we have

$$\langle\delta(d\alpha), X\rangle = \sum_{k=1}^{n}\Big[-e_k(d\alpha(e_k, X) + d\alpha(e_k, \nabla_{e_k}X))\Big]$$
$$= \sum_{k=1}^{n}\Big[-e_k(e_k\langle\alpha, X\rangle - X\langle\alpha, e_k\rangle - \langle\alpha, [e_k, X]\rangle)\Big]$$
$$+ \langle\nabla_{e_k}\alpha, \nabla_{e_k}X\rangle - \langle\nabla_{\nabla_{e_k}X}\alpha, e_k\rangle$$
$$= -\sum_{k=1}^{n}\langle\nabla_{e_k}\nabla_{e_k}\alpha, X\rangle + \sum_{k=1}^{n}\langle\nabla_{e_k}\nabla_X\alpha, e_k\rangle - \sum_{k=1}^{n}\langle\nabla_{\nabla_{e_k}X}\alpha, e_k\rangle.$$

Thus

$$\langle\nabla\alpha, X\rangle = -\sum_{k=1}^{n}\langle\nabla_{e_k}\nabla_{e_k}\alpha, X\rangle + \sum_{k=1}^{n}\nabla_{e_k}\langle(\nabla_X - \nabla_X\nabla_{e_i} - \nabla_{e_i,[X]})\alpha, e_i\rangle$$
$$= \langle\nabla^*\nabla\alpha, X\rangle + \Big\langle\alpha, \sum_{k=1^n} R(e_i, X)e_i\Big\rangle$$
$$= \langle\nabla^*\nabla\alpha, X\rangle + g(\alpha, \text{Ric}(X, \cdot))$$
$$= \langle\nabla^*\nabla\alpha, X\rangle + \text{Ric}(\alpha)(X)$$

For the last equality, recall that $\text{Ric}(\alpha)(X)$ in coordinates is

$$\text{Ric}_{ij}\, g^{ik}a_k X^j.$$

4.3   Applications of Bochner–Weitzenböck

We are now going to understand how Riemannian geometry relates to topology. Previously, we mostly used sectional curvature, and exploited properties of geodesics. Here we are going to use Ricci curvature instead.

Theorem. Let (M, g) be a compact connected oriented manifold. Then

- If $\mathrm{Ric}(g) > 0$ at each point, then $H^1_{\mathrm{dR}}(M) = 0$.

- If $\mathrm{Ric}(g) \geq 0$ at each point, then $b^1(M) = \dim H^1_{\mathrm{dR}}(M) \leq n$.

- If $\mathrm{Ric}(g) \geq 0$ at each point, and $b^1(M) = n$, then g is flat.

Proof. We use Bochner–Weitzenböck.

- For any $\alpha \in \Omega^1(M)$, we have

$$g(\delta\alpha, \alpha) = g(\nabla^*\nabla, \alpha) + \mathrm{Ric}(\alpha, \alpha).$$

In local coordinates, $\mathrm{Ric}(\alpha, \alpha)$ means

$$\mathrm{Ric}_{ij}\, g^{ik} g^{j\ell} \alpha_k \alpha_\ell.$$

Since M is compact, we can integrate over M, and we get

$$\langle\langle \delta\alpha, \alpha \rangle\rangle = \langle\langle \nabla^*\nabla\alpha, \alpha \rangle\rangle + \int_M \mathrm{Ric}(\alpha, \alpha)\omega_g$$

$$= \|\nabla\alpha\|_2^2 + \int_M \mathrm{Ric}(\alpha, \alpha)\omega_g.$$

By assumption, we have $\mathrm{Ric}(\alpha, \alpha) \geq 0$ and vanishes iff $\alpha = 0$.

Now if $\alpha \in \mathcal{H}^1_M \cong H^1_{\mathrm{dR}}(M)$, then $\delta\alpha = 0$. So RHS $= 0$. So by looking at the curvature term, we know $\alpha = 0$.

- If $\delta\alpha = 0$, then by a similar calculation, we know that we must have $\nabla\alpha = 0$. So if we know $\alpha(x)$ for some fixed $x \in M$, then we know the value of α everywhere by parallel transport. Thus α is determined by the initial condition $\alpha(x)$, Thus there are $\leq n = \dim T^*_x M$ linearly independent such α.

- If $b^1(M) = n$, then we can pick a basis $\alpha_1, \cdots, \alpha_n$ of $\mathcal{H}^1_M$. Then we can pick a dual basis $X_1, \cdots, X_n \in \mathrm{Vect}(M)$, and they must be parallel, i.e. $\nabla X_i = 0$. Indeed, we have

$$\langle \alpha_j, \nabla X_i \rangle + \langle \nabla\alpha_j, X_i \rangle = \nabla\langle \alpha_j, X_i \rangle.$$

But $\langle \alpha_j, X_i \rangle$ is constantly 0 or 1 depending on i and j, So the RHS vanishes. Similarly, the second term on the left vanishes. Since the α_j span, we know we must have $\nabla X_i = 0$.

Now we have

$$R(X_i, X_j)X_k = (\nabla_{[X_i, X_j]} - [\nabla_{X_i}, \nabla_{X_j}])X_k = 0,$$

Since this is valid for all I, j, k, we know R vanishes at each point. So we are done.

$\square$

Bochner–Weitzenböck can be exploited in a number of similar situations.

Note that in the third part of the theorem, we haven't actually proved the optimal statement. We can deduce more than the flatness of the metric, but requires some slightly advanced topology. We will provide a sketch proof of the theorem, making certain assertions about topology.

Proposition. In the case of (iii), M is in fact isometric to a flat torus.

Proof sketch. We fix $p \in M$ and consider the map $M \to \mathbb{R}^n$ given by

$$x \mapsto \left(\int_p^x \alpha_i \right)_{i=1,\cdots,n} \in \mathbb{R}^n,$$

where the α_i are as in the previous proof. The integral is taken along any path from p to x, and this is not well-defined. But by Stokes' theorem, and the fact that $d\alpha_i = 0$, this only depends on the homotopy class of the path.

In fact, $\int_p^x$ depends only on $\gamma \in H_1(M)$, which is finitely generated. Thus, $\int_p^x \alpha_i$ is a well-defined map to $S^1 = \mathbb{R}/\lambda_i\mathbb{Z}$ for some $\lambda_i \neq 0$. Therefore we obtain a map $M \to (S^1)^n = T^n$. Moreover, a bit of inspection shows this is a local diffeomorphism. But since the spaces involved are compact, it follows by some topology arguments that it must be a covering map. But again by compactness, this is a finite covering map. So M must be a torus. So we are done. $\square$

Note that Bochner–Weitzenböck, which we proved for 1-forms, is in fact valid for forms of any degree. To do so, we consider $E = \Lambda^p T^*M$, and then we have a map

$$\nabla : \Gamma(E) = \Omega_M^0(E) \to \Omega_M^1(E),$$

and this has a formal adjoint

$$\nabla^* : \Omega_M^1(E) \to \Omega_M^0(E).$$

Now if $\alpha \in \Omega^p(M)$, then

$$\Delta\alpha = \nabla^*\nabla\alpha + \tilde{R}(\alpha),$$

where $\tilde{R}$ is a linear map $\Omega^p(M) \to \Omega^p(M)$ depending on $R^i_{j,k\ell}, \mathrm{Ric}_{ij} \, g^{ij}$. We can write

$$\mathfrak{R} = R_{ij,k\ell} \in S^2\Lambda^2 T^*M.$$

Then we can prove that if $\mathfrak{R} > 0$ at all points, then $\mathcal{H}^k(M) = \{\}$ for all $k = 1, \cdots, n-1$.

If $\mathfrak{R} \geq 0$ only, then we have

$$b^k(M) \leq \binom{n}{k} = \dim \Lambda^k T^*M,$$

and moreover $\Delta\alpha = 0$ iff $\nabla\alpha = 0$.

5 Riemannian holonomy groups

Again let M be a Riemannian manifold, which is always assumed to be connected. Let $x \in M$, and consider a path $\gamma \in \Omega(x, y)$, $\gamma : [0, 1] \to M$. At the rather beginning of the course, we know that γ gives us a parallel transport from $T_x M$ to $T_y M$. Explicitly, given any $X_0 \in T_x M$, there exists a unique vector field X along γ with

$$\frac{\nabla X}{\mathrm{d}t} = 0, \quad X(0) = X_0.$$

Definition (Holonomy transformation)**.** The *holonomy transformation* $P(\gamma)$ sends $X_0 \in T_x M$ to $X(1) \in T_y M$.

We know that this map is invertible, and preserves the inner product. In particular, if $x = y$, then $P(\gamma) \in \mathrm{O}(T_x M) \cong \mathrm{O}(n)$.

Definition (Holonomy group)**.** The *holonomy group* of M at $x \in M$ is

$$\mathrm{Hol}_x(M) = \{P(\gamma) : \gamma \in \Omega(x, x)\} \subseteq \mathrm{O}(T_x M).$$

The group operation is given by composition of linear maps, which corresponds to composition of paths.

We note that this group doesn't really depend on the point x. Given any other $y \in M$, we can pick a path $\beta \in \Omega(x, y)$. Writing P_β instead of $P('b)$, we have a map

$$\mathrm{Hol}_x(M) \longrightarrow \mathrm{Hol}_y(M)$$

$$P_\gamma \longmapsto P_\beta \circ P_\gamma \circ P_{\beta^{-1}} \in \mathrm{Hol}_y(M).$$

So we see that $\mathrm{Hol}_x(M)$ and $\mathrm{Hol}_y(M)$ are isomorphic. In fact, after picking an isomorphism $\mathrm{O}(T_x M) \cong \mathrm{O}(T_y M) \cong \mathrm{O}(N)$, these subgroups are conjugate as subgroups of $\mathrm{O}(n)$. We denote this class by $\mathrm{Hol}(M)$.

Note that depending of what we want to emphasize, we write $\mathrm{Hol}(M, g)$, or even $\mathrm{Hol}(g)$ instead.

Really, $\mathrm{Hol}(M)$ is a *representation* (up to conjugacy) induced by the standard representation of $\mathrm{O}(n)$ on $\mathbb{R}^n$.

Proposition. If M is simply connected, then $\mathrm{Hol}_x(M)$ is path connected.

Proof. $\mathrm{Hol}_x(M)$ is the image of $\Omega(x, x)$ in $\mathrm{O}(n)$ under the map P, and this map is continuous from the standard theory of ODE's. Simply connected means $\Omega(x, x)$ is path connected. So $\mathrm{Hol}_x(M)$ is path connected. $\square$

It is convenient to consider the *restricted* holonomy group.

Definition (Restricted holonomy group)**.** We define

$$\mathrm{Hol}_x^0(M) = \{P(\gamma) : \gamma \in \Omega(x, x) \text{ nullhomotopic}\}.$$

As before, this group is, up to conjugacy, independent of the choice of the point in the manifold. We write this group as $\mathrm{Hol}^0(M)$.

Of course, $\mathrm{Hol}^0(M) \subseteq \mathrm{Hol}(M)$, and if $\pi_1(M) = 0$, then they are equal.

Corollary. $\mathrm{Hol}^0(M) \subseteq \mathrm{SO}(n)$.

Proof. $\mathrm{Hol}^0(M)$ is connected, and thus lies in the connected component of the identity in $\mathrm{O}(n)$. $\qquad\qquad\square$

Note that there is a natural action of $\mathrm{Hol}_x(M)$ and $\mathrm{Hol}_x^0(M)$ on T_x^*M, $\Lambda^p T^*M$ for all p, and more generally tensor products of T_xM.

Fact.

- $\mathrm{Hol}^0(M)$ is the connected component of $\mathrm{Hol}(M)$ containing the identity element.

- $\mathrm{Hol}^0(M)$ is a *Lie subgroup* of $\mathrm{SO}(n)$, i.e. it is a subgroup and an immersed submanifold. Thus, the Lie algebra of $\mathrm{Hol}^0(M)$ is a Lie subalgebra of $\mathfrak{so}(n)$, which is just the skew-symmetric $n \times n$ matrices.

  This is a consequence of Yamabe theorem, which says that a path-connected subgroup of a Lie group is a Lie subgroup.

We will not prove these.

Proposition (Fundamental principle of Riemannian holonomy). Let (M, g) be a Riemannian manifold, and fix $p, q \in \mathbb{Z}_+$ and $x \in M$. Then the following are equivalent:

(i) There exists a (p, q)-tensor field α on M such that $\nabla \alpha = 0$.

(ii) There exists an element $\alpha_0 \in (T_xM)^{\otimes p} \otimes (T_x^*M)^{\otimes q}$ such that α_0 is invariant under the action of $\mathrm{Hol}_x(M)$.

Proof. To simplify notation, we consider only the case $p = 0$. The general case works exactly the same way, with worse notation. For $\alpha \in (T_x^*M)^q$, we have

$$(\nabla_X \alpha)(X_1, \cdots, X_q) = X(\alpha(X_1, \cdots, X_q)) - \sum_{i=1}^q \alpha(X_1, \cdots, \nabla_X X_i, \cdots, X_q).$$

Now consider a loop $\gamma : [0, 1] \to M$ be a loop at x. We choose vector fields X_i along γ for $i = 1, \cdots, q$ such that

$$\frac{\nabla X_i}{\mathrm{d}t} = 0.$$

We write

$$X_i(\gamma(0)) = X_i^0.$$

Now if $\nabla \alpha = 0$, then this tells us

$$\frac{\nabla \alpha}{\mathrm{d}t}(X_1, \cdots, X_q) = 0.$$

By our choice of X_i, we know that $\alpha(X_1, \cdots, X_q)$ is constant along γ. So we know

$$\alpha(X_1^0, \cdots, X_q^0) = \alpha(P_\gamma(X_1^0), \cdots, P_\gamma(X_q^0)).$$

So α is invariant under $\mathrm{Hol}_x(M)$. Then we can take $\alpha_0 = \alpha_x$.

Conversely, if we have such an α_0, then we can use parallel transport to transfer it to everywhere in the manifold. Given any $y \in M$, we define α_y by

$$\alpha_y(X_1, \cdots, X_q) = \alpha_0(P_\gamma(X_1), \cdots, P_\gamma(X_q)),$$

where γ is any path from y to x. This does not depend on the choice of γ precisely because α_0 is invariant under $\mathrm{Hol}_x(M)$.

It remains to check that α is C^∞ with $\nabla \alpha = 0$, which is an easy exercise. $\quad\square$

Example. Let M be oriented. Then we have a volume form ω_g. Since $\nabla \omega_g = 0$, we can take $\alpha = \omega_g$. Here $p = 0$ and $q = n$. Also, its stabilizer is $H = \mathrm{SO}(n)$. So we know $\mathrm{Hol}(M) \subseteq \mathrm{SO}(n)$ if (and only if) M is oriented.

The "only if" part is not difficult, because we can use parallel transport to transfer an orientation at a particular point to everywhere.

Example. Let $x \in M$, and suppose

$$\mathrm{Hol}_x \subseteq H\mathrm{U}(n) = \{g \in \mathrm{SO}(2n) : gJ_0g^{-1} = J_0\},$$

where

$$J_0 = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}.$$

By looking at $\alpha_0 = J_0$, we obtain $\alpha = J \in \Gamma(\mathrm{End}\,TM)$ with $\nabla J = 0$ and $J^2 = -1$. This is a well-known standard object in complex geometry, and such a J is an instance of an *almost complex structure* on M.

Example. Recall (from the theorem on applications of Bochner–Weitzenböck) that a Riemannian manifold (M, g) is flat (i.e. $R(g) \equiv 1$) iff around each point $x \in M$, there is a parallel basis of parallel vector fields. So we find that (M, g) is flat iff $\mathrm{Hol}^0(M, g) = \{id\}$.

It is essential that we use $\mathrm{Hol}^0(M, g)$ rather than the full $\mathrm{Hol}(M, g)$. For example, we can take the Klein bottle
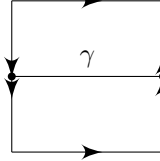
with the flat metric. Then parallel transport along the closed loop γ has

$$P_\gamma = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

In fact, we can check that $\mathrm{Hol}(K) = \mathbb{Z}_2$. Note that here K is non-orientable.

5.1 Holonomy and curvature

Definition (Holonomy algebra)**.** The *holonomy algebra* $\mathfrak{hol}(M)$ is the Lie algebra of $\mathrm{Hol}(M)$.

Thus $\mathfrak{hol}(M) \leq \mathfrak{so}(n)$ up to conjugation.

Now consider some open coordinate neighbourhood $U \subseteq M$ with coordinates $x_1, \cdots, x_n$. As before, we write

$$\partial_i = \frac{\partial}{\partial x_i}, \quad \nabla_i = \nabla_{\partial_i}.$$

The curvature may also be written in terms of coordinates $R = R^i_{j,k\ell}$, and we also have

$$R(\partial_k, \partial_\ell) = -[\nabla_k, \nabla_\ell].$$

Thus, $\mathfrak{hol}(M)$ contains

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} P(\gamma_t),$$

where γ_t is the square
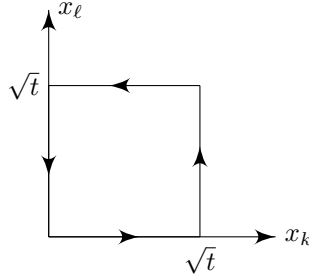
By a direct computation, we find

$$P(\gamma_t) = I + \lambda t R(\partial_k, \partial_\ell) + o(t).$$

Here $\lambda \in \mathbb{R}$ is some non-zero absolute constant that doesn't depend on anything (except convention).

Differentiating this with respect to t and taking the limit $t \to 0$, we deduce that at for $p \in U$, we have

$$R_p = (R^i_{j,k\ell})_p \in \Lambda^2 T^*_p M \otimes \mathfrak{hol}_p(M),$$

where we think $\mathfrak{hol}_p(M) \subseteq \operatorname{End} T_p M$. Recall we also had the $R_{ij,k\ell}$ version, and because of the nice symmetric properties of R, we know

$$(R_{ij,k\ell})_p \in S^2 \mathfrak{hol}_p(M) \subseteq \lambda^2 T^*_p M \otimes \Lambda^2 T^*_p M.$$

Strictly speaking, we should write

$$(R^{i\ k}_{j\ \ell})_p \in S^2 \mathfrak{hol}_p(M),$$

but we can use the metric to go between $T_p M$ and $T^*_p M$.

But there is more to these. So far, we have only been doing something rather tautological.

Consider a Lie subgroup $G \subseteq \operatorname{GL}_n(\mathbb{R})$. There is a standard representation ρ of $\operatorname{GL}_n(\mathbb{R})$ on $\mathbb{R}^n$, which restricts to a representation $(\rho, \mathbb{R}^n)$ of G. This induces a representation $(\rho^k, \Lambda^k(\mathbb{R}^*))$ of G.

This representation is in general not irreducible. We decompose it into irreducible components

$$(\rho^k_i, W^k_i),$$

where

$$\Lambda^k(\mathbb{R}^*) = \bigoplus_i W^k_i.$$

We now bring in differential geometry. Consider a manifold M with a G-structure, i.e. there is an atlas of coordinate neighbourhoods where the transition maps satisfy

$$\left(\frac{\partial x_\alpha}{\partial x'_\beta}\right)_p \in G$$

for all p. Then we can use this to split our bundle of k-forms into well-defined vector sub-bundles with typical fibers W_i^k:

$$\Lambda^k T^* M = \bigoplus \Lambda_i^k.$$

We can furthermore say that every G-equivariant linear map $\varphi : W_i^k \to W_j^\ell$ induces a morphism of vector bundles $\phi : \Lambda_i^k \to \Lambda_j^\ell$.

Now suppose further that $\mathrm{Hol}(M) \leq G \leq \mathrm{O}(n)$. Then by the fundamental principle of Riemannian holonomy, we deduce that parallel transport preserves this decomposition into sub-bundles. So ∇ restricts to a well-defined connection on each Λ_i^k.

Thus, if $\xi \in \Gamma(\Lambda_i^k)$, then $\nabla \xi \in \Gamma(T^* M \otimes \Lambda_i^k)$, and then we have $\nabla^* \nabla \xi \in \Gamma(\Lambda_i^k)$. But we know the covariant Laplacian is related to Laplace–Beltrami via the curvature. We only do this for 1-forms for convenience of notation. Then if $\xi \in \Omega^1(M)$, then we have

$$\Delta \xi = \nabla^* \nabla \xi + \mathrm{Ric}(\xi).$$

We can check that Ric also preserves these subbundles. Then it follows that $\Delta : \Gamma(\Lambda_j^1) \to \Gamma(\Lambda_j^1)$ is well-defined.

It follows that $\mathcal{H}^k = \bigoplus \mathcal{H}_i^k$, where $\mathcal{H}_i^k = \ker(\Delta : \Gamma(\Lambda_j^1) \to \Gamma(\Lambda_j^1))$.

Theorem. Suppose M is compact, so $\mathcal{H}^k \cong H_{\mathrm{dR}}^k(M)$. Then

$$H_{\mathrm{dR}}^k(M) = \bigoplus H_{i,\mathrm{dR}}^k(M),$$

and

$$b^k = \sum_i b_i^k.$$

These are the *refined Betti numbers*.

Theorem. Let (M, g) be a connected and oriented Riemannian manifold, and consider the decomposition of the bundle of k-forms into irreducible representations of the holonomy group,

$$\Lambda^k T^* M = \bigoplus_i \Lambda_i^k.$$

In other words, on each fiber $(\Lambda_i^k)_x \subseteq \Lambda^k T_x^* M$ is an irreducible representation of $\mathrm{Hol}_x(g)$. Then

  (i) For all $\alpha \in \Omega_i^k(M) = \Gamma(\Lambda_i^l)$, we have $\Delta \alpha \in \Omega_i^k(M)$.

  (ii) If M is compact, then we have a decomposition

$$H_{\mathrm{dR}}^k(M) = \bigoplus H_{i,\mathrm{dR}}^k(M),$$

where

$$H_{i,\mathrm{dR}}^k(M) = \{[\alpha] : \alpha \in \Omega_i^k(M), \Delta \alpha = 0\}.$$

We have only proved this for $k = 1$, but the same proof technique can be used to do it for arbitrary k.

6   Manifolds with non-negative Ricci curvature

This is an introduction to Cheeger–Gromoll splitting theorem.

Before we can state the theorem, we need to do a bit of preparation.

Rays and lines on a Riemannian manifold

Definition (Ray). Let (M, g) be a Riemannian manifold. A *ray* is a map $r(t) : [0, \infty) \to M$ if r is a geodesic, and minimizes the distance between any two points on the curve.

Definition (Line). A *line* is a map $\ell(t) : \mathbb{R} \to M$ such that $\ell(t)$ is a geodesic, and minimizes the distance between any two points on the curve.

We have seen from the first example sheet that if M is a complete unbounded manifold, then M has a ray from each point, i.e. for all $x \in M$, there exists a ray r such that $r(0) = x$.

Definition (Connected at infinity). A complete manifold is said to be *connected at infinity* if for all compact set $K \subseteq M$, there exists a compact $C \supseteq K$ such that for every two points $p, q \in M \setminus C$, there exists a path $\gamma \in \Omega(p, q)$ such that $\gamma(t) \in M \setminus K$ for all t.

We say M is *disconnected at infinity* if it is not connected at infinity.

Note that if M is disconnected at infinity, then it must be unbounded, and in particular non-compact.

Lemma. If M is disconnected at infinity, then M contains a line.

Proof. Note that M is unbounded. Since M is disconnected at infinity, we can find a compact subset $K \subseteq M$ and sequences $p_m, q_m \to \infty$ as $m \to \infty$ (to make this precise, we can pick some fixed point x, and then require $d(x, p_m), d(x, q_m) \to \infty$) such that every $\gamma_m \in \Omega(p_m, q_m)$ passes through K.

In particular, we pick γ_m to be a minimal geodesic from p_m to q_m parametrized by arc-length. Then γ_m passes through K. By reparametrization, we may assume $\gamma_m(0) \in K$.

Since K is compact, we can pass to a subsequence, and wlog $\gamma_m(0) \to x \in K$ and $\dot{\gamma}_m(0) \to a \in T_x M$ (suitably interpreted).

Then we claim the geodesic $\gamma_{x,a}(t)$ is the desired line. To see this, since solutions to ODE's depend smoothly on initial conditions, we can let

$$\ell(t) = \lim_{m \to \infty} \gamma_m(t).$$

Then we know

$$d(\ell(s), \ell(t)) = \lim_{m \to \infty} d(\gamma_m(s), \gamma_m(t)) = |s - t|.$$

So we are done. □

Let's look at some examples.

Example. The elliptic paraboloid

$$\{z = x^2 + y^2\} \subseteq \mathbb{R}^3$$

with the induced metric does not contain a line. To prove this, we can show that any geodesic that is not a meridian must intersect itself.

Example. For any complete metric g on $S^{n-1} \times \mathbb{R}$ contains a line since it is disconnected at ∞.

Theorem (Cheeger–Gromoll line-splitting theorem (1971)). If (M, g) is a complete connected Riemannian manifold containing a line, and has $\mathrm{Ric}(g) \geq 0$ at each point, then M is isometric to a Riemannian product $(N \times \mathbb{R}, g_0 + \mathrm{d}t^2)$ for some metric g_0 on N.

We will not prove this, but just see what the theorem can be good for.

First of all, we can slightly improve the statement. After applying the theorem, we can check again if N contains a line or not. We can keep on proceeding and splitting lines off. Then we get

Corollary. Let (M, g) be a complete connected Riemannian manifold with $\mathrm{Ric}(g) \geq 0$. Then it is isometric to $X \times \mathbb{R}^q$ for some $q \in \mathbb{N}$ and Riemannian manifold X, where X is complete and does not contain any lines.

Note that if X is zero-dimensional, since we assume all our manifolds are connected, then this implies M is flat. If $\dim X = 1$, then $X \cong S^1$ (it can't be a line, because a line contains a line). So again M is flat.

Now suppose that in fact $\mathrm{Ric}(g) = 0$. Then it is not difficult to see from the definition of the Ricci curvature that $\mathrm{Ric}(X) = 0$ as well. If we know $\dim X \leq 3$, then M has to be flat, since in dimensions ≤ 3, the Ricci curvature determines the full curvature tensor.

Recall (from example sheets) that a manifold is *homogeneous* if the group of isometries acts transitively. In other words, for any $p, q \in M$, there exists an isometry $\phi : M \to M$ such that $\phi(p) = q$. This in particular implies the metric is complete. We shall see that if M is homogeneous, then so is X.

Now if $p = (p_0, v)$ and $q = (q_0, v)$ in the splitting, then every minimal geodesic between p and q must lie in $X \times \{v\}$.

We now show that X is compact. Suppose not. Then X is unbounded. We will obtain a line on X. By assumption, for all $n = 1, 2, \cdots$, we can find p_n, q_n with $d(p_n, q_n) \geq 2n$. Since X is complete, we can find a minimal geodesic γ_n connecting these two points, parametrized by arc length. By homogeneity, we may assume that the midpoint $\gamma_n(0)$ is at a fixed point x_0. By passing to a subsequence, we wlog $\dot{\gamma}_n(0)$ converges to some $a \in T_{x_0}(X)$. Then we use a as an initial condition for our geodesic, and this will be a line. The proof is similar to what we have already discussed.

A similar argument gives

Lemma. Let (M, g) be a compact Riemannian manifold, and suppose its universal Riemannian cover $(\tilde{M}, \tilde{g})$ is non-compact. Then $(\tilde{M}, \tilde{g})$ contains a line.

Proof. We first find a compact $K \subseteq \tilde{M}$ such that $\pi(K) = M$. Since $\tilde{M}$ must be complete, it is unbounded. Choose p_n, q_n, γ_n like before. Then we can apply deck transformations so that the midpoint lies inside K, and then use compactness of K to find a subsequence so that the midpoint converges. $\qquad \square$

We do more applications.

Corollary. Let (M, g) be a compact, connected manifold with , $\mathrm{Ric}(g) \geq 0$. Then

– The universal Riemannian cover is isometric to the Riemannian product $X \times \mathbb{R}^N$, with X compact, $\pi_1(X) = 1$ and $\mathrm{Ric}(g_X) \geq 0$.

– Let $p \in M$ be such that $\mathrm{Ric}(g)_p > 0$. Then $\pi_1(M)$ is finite.

– Denote by $I(\tilde{M})$ the group of isometries $\tilde{M} \to \tilde{M}$. Then $I(\tilde{M}) = I(X) \times E(\mathbb{R}^q)$, where $E(\mathbb{R}^q)$ is the group of rigid Euclidean motions,

$$\mathbf{y} \mapsto A\mathbf{y} + \mathbf{b}$$

where $\mathbf{b} \in \mathbb{R}^n$ and $A \in \mathrm{O}(q)$.

– If $\tilde{M}$ is homogeneous, then so is X.

Proof.

– This is direct from Cheeger–Gromoll and the previous lemma.

– If $\mathrm{Ric}(g)_p > 0$, then we cannot have any non-trivial splitting. So $\tilde{M} \not\cong X \times \mathbb{R}^q$ with $q \geq 1$. So $\tilde{M}$ must be compact. So by standard topology, $|\pi_1(M)| = |\pi^{-1}(\{p\})|$.

– We use the fact that $E(\mathbb{R}^q) = I(\mathbb{R}^q)$.

Since $g \in I(\tilde{M})$, we know g takes lines to lines. Since all lines in $\tilde{M} \times \mathbb{R}^q$ are of the form $p \times \mathbb{R}$ with $p \in X$ and $\mathbb{R} \subseteq \mathbb{R}^q$ an affine line.

Then $g(p \times \mathbb{R}) = p' \times \mathbb{R}$, possibly for some other copy of $\mathbb{R}$. By taking unions, we deduce that $g(p \times \mathbb{R}^q) = p' \times \mathbb{R}^q$. Now for any $X \times \mathbf{a}$ with $\mathbf{a} \in \mathbb{R}^q$, we have $X \times \mathbf{a} \perp p \times \mathbb{R}^q$ for all $p \in X$.

So we know

$$g(X \times \mathbf{a}) = g(X \times \mathbf{b}$$

for some $\mathbf{b} \in \mathbb{R}^q$. We deduce that $g(p, a) = (q, b)$, with $h_X \in I(X), e \in E(\mathbb{R}^q)$ such that

$$h_X(p) = q, \quad e(\mathbf{a}) = \mathbf{b},$$

which are independent of $\mathbf{a}$ and p respectively.

$\square$

We can look at more examples.

Proposition. Consider $S^n \times \mathbb{R}$ for $n = 2$ or 3. Then this does not admit any Ricci-flat metric.

Proof. Note that $S^n \times \mathbb{R}$ is disconnected at ∞. So any metric contains a line. Then by Cheeger–Gromoll, $\mathbb{R}$ splits as a Riemannian factor. So we obtain $\mathrm{Ric} = 0$ on the S^n factor. Since we are in $n = 2, 3$, we know S^n is flat, as the Ricci curvature determines the full curvature. So S^n is the quotient of $\mathbb{R}^n$ by a discrete group, and in particular $\pi_1(S^n) \neq 1$. This is a contradiction. $\square$

Let G be a Lie group with a bi-invariant metric g. Suppose the center $Z(G)$ is finite. Then the center of $\mathfrak{g}$ is trivial (as of the Lie algebra of $G/Z(G)$, which has trivial center). From sheet 2, we find that $\mathrm{Ric}(g) > 0$ implies $\pi_1(G)$ is finite. The converse is also true, but is harder. This is done on Q11 of sheet 3 — if $\pi_1(G)$ is finite, then $Z(G)$ is finite.

Index