MATH 215C: Differential Geometry

Lecturer: Professor Otis Chodosh

Notes by: Andrew Lin

Spring 2023

Introduction

This class will cover Riemannian geometry, using Lee's textbook (the second edition, which has more content than the first – we'll hit most of the material there). There are listed prerequisites, but we just need to review the material (and we'll do so through this class as well). But we won't spend too much time on background so that we can get through everything; instead, we can read the appendices (smooth manifolds and other things) as they come up. Chapter 1 provides some motivation for Riemannian geometry, which is a funny topic to learn for the first time because it's "boring." Inherently, we'll make some definitions and see what happens, and it'll be confusing because the topic has been modernized (and so in the days of Gauss and Riemann it was confusing what to do and getting definitions right was hard, and now we can say that we should "obviously" study something).

Organization of the course will all be done on Canvas and Gradescope; lecture notes won't be posted because we're basically following the book, skipping some parts. And office hours will be posted soon as well.

1 April 3, 2023

In this class, we'll only talk about smooth manifolds, possibly with boundary – at some point we'll want to distinguish but we don't need to worry about it for now. Recall that a **$(0, 2)$-tensor** on a manifold is a section of $T^*M \otimes T^*M \to M$, so in coordinates this is an element of the form $T_{ij} dx^i \otimes dx^j$ (so a bilinear map on the tangent space, and here we're omitting summing over i, j using **Einstein notation**, in which we sum over an index if one copy is lower and the other is upper.) We also get a pullback: if T is a $(0, 2)$ tensor on $\tilde{M}$ and $\phi : M \to \tilde{M}$ is any map, then the pullback of the tensor is defined via $\phi^* T(u, v) = T(d\phi(u), d\phi(v))$. (On the other hand, if we have a $(2, 0)$ tensor we can't generally pull it back or forward, so that's a bit awkward.)

Definition 1

A **Riemannian metric** g on M is a $(0, 2)$-tensor such that at any point $p \in M$, g_p is an inner product (meaning that it is symmetric and positive definite on $T_x M$).

Riemannian geometry is then the study of manifolds equipped with Riemannian metrics, and the point is that we can do geometry in the classical sense (we can look at two tangent vectors and look at their length, angle between them, length of curves, get a natural volume form, and so on).

Example 2
The **flat metric** $\overline{g}$ on $\mathbb{R}^n$ is such that the tangent space $T_p\mathbb{R}^n$ is canoncially identified with $\mathbb{R}^n$, and with $g_p(u, v) = u \cdot v$ under the usual dot product. More generally, if $(\tilde{M}, \tilde{g})$ is a Riemannian manifold and $F : M \to \tilde{M}$ is an immersion, then the pullback $F^*\tilde{g}$ will be a Riemannian metric on M. (Clearly it's symmetric, and because dF is nonsingular it will be an immersion.)

We'll fix some notation here: if M is already equipped with a metric g such that this pullback relation holds, we call the map a **Riemannian immersion**. And if $M \subset (\tilde{M}, g)$ is actually a submanifold, we can think of T_pM as being a subset of $T_p\tilde{M}$ (dropping the inclusion map), and then $g_p(u, v) = \tilde{g}_p(u, v)$ is just the restriction of $\tilde{g}$. So we know many Riemannian manifolds: draw a surface in $\mathbb{R}^3$, and we can equip each of those with a Riemannian metric coming from $\mathbb{R}^3$. And in the days of Gauss, this was the only theory that existed (all of his manifolds were embedded in $\mathbb{R}^3$) – Riemann was the first to study this more abstractly. But Gauss had the glimpse of the difference between "intrinsic" and "extrinsic," which we'll talk about shortly.

Example 3
If we embed the sphere $\{|x| = 1\} = S^n \subset \mathbb{R}^{n+1}$ in Euclidean space, the induced metric $\mathring{g}$ is called the **round metric**.

Definition 4
If we have a **diffeomorphism** $\phi : (M, g) \to (\tilde{M}, \tilde{g})$ with $g = \phi^*\tilde{g}$, then we say that the map is an **isometry** (we should think of isometric manifolds as being the same). A map $\phi : (M, g) \to (\tilde{M}, \tilde{g})$ is a **local isometry** if for all $p \in M$ there is some neighborhood U such that $\phi|_U$ is an isometry onto its image.

This second definition is not an equivalence relation (for example if one of the manifolds has multiple components and one of them isn't being mapped to). But note that ϕ is a local isometry if and only if the differential $d\phi_p : (T_pM, g_p) \to (T_{\phi(p)}M, \tilde{g}_{\phi(p)})$ is an isometry of inner product spaces for all p. (The inverse function theorem is used for the backwards direction.) We let $\text{Iso}(M, g)$ denote the set of isometries $\phi : M \to M$ which are isometries – there is a deep theorem (which we won't prove) that this is always a finite-dimensional Lie group (this is called the Myers-Steenrod theorem).

When we have an object in this area of math, it's often good to understand how to study it in coordinates – for example, suppose we have $g = g_{ij}dx^i \otimes dx^j$, where $g_{ij} = g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right)$ is a function on the coordinate chart. This is not symmetric, but we can define $dx^i dx^j = \frac{1}{2}(dx^i \otimes dx^j + dx^j \otimes dx^i)$ and write $g = g_{ij}dx^i dx^j$ as well.

Example 5
If we use the flat metric (dot product) in Cartesian coordinates, then the matrix g_{ij} is the identity matrix: $\overline{g} = (dx^1)^2 + \cdots + (dx^n)^2$. (So this tensor technically doesn't satisfy the rules of Einstein notation, but the point is that this isn't preserved when we change coordinates anyway, so it shouldn't.) Meanwhile, in polar coordinates in $\mathbb{R}^2$, we can check that $\overline{g} = dr^2 + r^2 d\theta^2$.

Consider a submanifold $M \subset \mathbb{R}^N$, and choose charts on the submanifold M to diffeomorphically map to subsets of $\mathbb{R}^n$. Now consider a **coordinate parameterization** X (the inverse of the chart ϕ), which is usually the map we can "actually write down." Suppose the coordinates $u^1, \cdots, u^n$ in $\mathbb{R}^n$ map to $x^1, \cdots, x^n$ in M. Then $\frac{\partial}{\partial x^i} = dX\left(\frac{\partial}{\partial u^i}\right)$, so the coefficients of the metric are

$$g_{ij} = g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = g\left(dX\left(\frac{\partial}{\partial u^i}\right), dX\left(\frac{\partial}{\partial u^j}\right)\right) = (X^*g)\left(\frac{\partial}{\partial u^i}, \frac{\partial}{\partial u^j}\right) = (X^*g)_{ij}.$$

In particular, we have a Riemanninan metric that we pull back via the diffeomorphism, and the point now is that $(X^*g)_{ij} = (X^*\overline{g})_{ij}$ (where $\overline{g}$ is the flat metric on $\mathbb{R}^N$). Indeed, the pullback pushes things only forward to the tangent space of X, so g and $\overline{g}$ are indistinguishable. Then the pullback $X^*\overline{g}$ is $(dX^1)^2 + \cdots + (dX^n)^2$, since the pullback and differential commute, and the pullback and symmetric product also commute. So

$$X^*\overline{g} = \left(\frac{\partial x^1}{\partial u^i} du^i\right)^2 + \cdots + \left(\frac{\partial x^n}{\partial u^i} du^i\right)^2 = \frac{\partial x^1}{\partial u^i}\frac{\partial x^1}{\partial u^j} du^i du^j + \cdots + \frac{\partial x^n}{\partial u^i}\frac{\partial x^1}{\partial u^j} du^i du^j,$$

and in particular we can rewrite this as $\frac{\partial X}{\partial u^i} \cdot \frac{\partial X}{\partial u^j} du^i du^j$ (where we think of X as an n-dimensional vector). And so this is how we often compute the metric in coordinates, since we actually know how to write down X.

Example 6

Let $U \subset \mathbb{R}^n$ be open, $f : U \to \mathbb{R}$ be a function, and let the graph of f be $\{(u, f(u)\} \subset \mathbb{R}^{n+1}$. Then our map is $X(u) = (u, f(u))$, so $\frac{\partial X}{\partial u^i} = (e_i, \frac{\partial f}{\partial u^i})$ and thus the metric on the graph of f is

$$g_{ij} = \delta_{ij} + \frac{\partial f}{\partial u^i} + \frac{\partial f}{\partial u^j} = (du^1)^2 + \cdots + (du^n)^2 + df^2$$

(it's a rank 1 perturbation of the identity).

Example 7

Consider a curve which is an embedded 1-dimensional manifold C in the rz-plane (always with $r > 0$), and revolve it around the z-axis to get a surface of revolution. Then $S_C = \{(x, y, z) : (\sqrt{x^2 + y^2}, z) \in C\}$ is a smooth submanifold because we avoid radius zero. Then given a parametrization $\gamma(t) = (a(t), b(t))$ of C, we get a parameterization of the surface $X(t, \theta) = (a(t)\cos\theta, a(t)\sin\theta, b(t))$. Then $\frac{\partial X}{\partial t} = (a'\cos\theta, a'\sin\theta, b')$ and $\frac{\partial X}{\partial \theta} = (-a\sin\theta, a\cos\theta, 0)$ and thus (dotting all pairs together)

$$g = (a'(t)^2 + b'(t)^2)dt^2 + a(t)^2 d\theta^2.$$

In particular, it is nice to take a unit-speed parameterization of the curve.

So given an explicit surface, computing the metric in coordinates is easy, and we'll learn how to compute very quantities related to the metric soon. (And part of what we have to learn is to actually do these computations to get a sense of the field, so we'll be forced to do some calculus.)

Example 8

Suppose $C = \{r = 1\}$, so that the surface S_C is an infinite cylinder. Parameterizing this via $\gamma(t) = (1, t)$, the metric on the cylinder is $g_{\text{cyl}} = dt^2 + d\theta^2$. In other words, S_C is locally isometric to $(\mathbb{R}^2, \overline{g})$, but this cylinder does not look like the plane globally. So what we'll try to do throughout this class is unwrap what's going on here.

Motivated by this example, we make the following definition:

Definition 9

A Riemannian manifold (M, g) is **flat** if it is locally isomorphic to $(\mathbb{R}^n, \overline{g})$; in other words, at all points p we have coordinates such that $g_{ij} = \delta_{ij}$ near p (not just at p, but finding a chart so that the metric is the identity on a neighborhood).

In particular, all 1(-Riemannian)-manifolds are flat (by choosing a unit speed parameterization; this comes down to solving an ODE), but in higher dimensions we have to solve a PDE instead. And indeed not all n-manifolds are flat (otherewise this would be a boring subject), but proving this is not so trivial. We saw that something like $dr^2 + r^2 d\theta^2$ represents the flat metric on $\mathbb{R}^2$, but it doesn't "look" flat because the matrix is $\begin{bmatrix} 1 & 0 \\ 0 & r \end{bmatrix}$. So given an explicit metric it's not so easy to show that it's not flat – doing it by hand is quite involved. So we'll develop some tools to see later on some examples of manifolds that are not flat.

We'll now mention **local frames** – this is a convenient generalization of coordinates:

Definition 10

A **local frame** is a collection of vector fields $E_1, \cdots, E_n$ near a point p such that in a neighborhood U around p, $(E_1)_q, \cdots (E_n)_q$ span $T_q M$ for all $q \in U$. We can then get a local frame of 1-forms by dualizing to get $\varepsilon^1, \cdots, \varepsilon^n$.

For example, the coordinate vector fields $\frac{\partial}{\partial x^i}$ form a local frame. And given such a local frame, we can write the metric in the form $g = g_{ij}\varepsilon^i\varepsilon^j$, where $g_{ij} = g(E_i, E_j)$. (So this is the same thing as what we just said for coordinate representations.) However, note from linear algebra (the Gram-Schmidt algorithm) that we can take $E_1, \cdots, E_n$ and produce a (still smooth) frame such that $g_{ij} = \delta_{ij}$; equivalently, we can find a local frame which is orthonormal at every point. This might be confusing with what we said previously about not all n-manfiolds being flat, and it's because in general this won't be induced by coordinates. (We can't always turn a set of vectors into coordinates, and there are obstructions in doing this.) In other words, we won't have $E_i = \frac{\partial}{\partial x^i}$ from coordinates x^i.

We'll now start doing some boring constructions we can do on Riemannian manifolds:

Definition 11

Let $(M_1, g_1), (M_2, g_2)$ be two Riemannian manifolds. Then the product $M_1 \times M_2$ is a smooth manifold, with metric $g_1 \times g_2 = \pi_1^* g_1 + \pi_2^* g_2$ (the inner products of tangent vectors in $M_1 \times M_2$ is the sum of the inner producdts in the two parts). More generally, given a positive smooth function $f > 0$ on M_1, we can define the **warped product** $M_1 \times_f M_2$ to be the manifold with metric $\pi_1^* g_1 + f \pi_2^* g_2$.

For example, we can check that surfaces of resolution are warped products, and polar coordinates $dr^2 + r^2 d\theta^2$ warp the flat metric on $\mathbb{R}_{>0}$ with the flat metric on S^1 to get a metric on $\mathbb{R}^2$.

2 April 5, 2023

Today's first topic, **Riemann submersion**, may not be too intuitive – we saw that we can pull back to get a metric on the domain of an immersion, and we can ask the same question about what it means for a submersion to be Riemannian. (Recall that this means we map from a higher-dimensional space to a lower-dimensional space.)

Definition 12

For a map $\pi : (M, g) \to (\tilde{M}, \tilde{g})$ between smooth manifolds, we have a "vertical" and "horizontal" part $V_p = \ker d\pi_p$ and $H_p = V_p^\perp$. We say that π is a **Riemannian submersion** if $d\pi : (H_p, g_p) \to (T_{\pi(p)}\tilde{M}, \tilde{g}_p)$ is an isometry.

Riemannian submersions are complicated in that we can't always find a metric to turn a submersion into a Riemannian one (and in fact it's quite rare that this can be done).

Example 13

The projection $\mathbb{R}^{n+k} \to \mathbb{R}^k$ with the flat metric is a submersion, and more generally the projection $M \times N$ onto either factor is a submersion. More generally, $\pi_M : M \times_f N \to M$ will be a Riemannian submersion but not π_N (this is good to check our understanding).

Fact 14

There's an important theorem that if G is a Lie group (smooth manifold with a group action which is also smooth, such as a matrix group), acting smoothly, isometrically, and properly on a Riemannian manifold $(\tilde{M}, \tilde{g})$ (meaning that $G \times \to M$ is smooth, the elements of G are acting as isometries, and that $G \times M \to M \times M$ via $(g, m) \mapsto (g(m), m)$ is proper), then $M = \tilde{M}/G$ has a uniquely smooth Riemannian manifold such that $\pi : (\tilde{M}, \tilde{g}) \to (M, g)$ is a Riemannian submersion. But the proof is rather boring and we won't do it here.

Example 15

If G is any discrete (in particular finite) group, we can come up with some good examples. For example, $\{\pm 1\}$ acting on S^n yields $S^n/\Gamma = \mathbb{RP}^n$, and in fact the round metric $\mathring{g}$ yields a unique round metric so that the map $S^n \to \mathbb{R}P^n$ is a Riemannian submersion. Similarly, S^1 acts on $S^{2n+1} \subset \mathbb{C}^n$ (where we think of S^1 as complex numbers of unit norm), and $S^{2n+1}/G = \mathbb{C}P^n$ yields the **Fubini-Study metric**, which is more interesting than some other examples we've described so far.

We won't talk more about submersions for now, but we can take a look at the textbook for some more details. Instead, we'll now move to some basic constructions on Riemannian manifolds. We said last time that we can basically "do all of the things an inner product allows us to do," and all we need to do is make sure nothing goes wrong when we vary the point. Recall that we have a canonical isomorphism $V \to V^*$ if we have an inner product, and this leads to a **musical isomorphism**: given a Riemannian manifold (M, g), we can construct a bundle isomorphism $\hat{g} : TM \to T^*M$ where $\hat{g}(v) : T_pM \to \mathbb{R}$ maps $w \mapsto g_p(v, w)$. In other words, if we have some $X \in \mathfrak{X}(M)$ (this notation $\mathfrak{X}(M)$ means the set of vector fields on M), then $\hat{g}(X)$ is a smooth 1-form.

To check smoothness, we'll spell this out in coordinates. Specifically, we'll do it in a local frame $E_1, \cdots, E_m$: if we want to know what $\hat{g}(X)$ looks like locally, we have

$$\hat{g}(X)(E_j) = g_{ij}X^i$$

(since we can write X as X^iE_i), or in other words $\hat{g}(X) = g_{ij}X^i\varepsilon^j$. It's common to write $g_{ij}X^i$ as $X^\flat$ ("flat" since we've "lowered the index"). We can similarly do this in reverse: if $\omega \in \Omega^1(M)$, then $\hat{g}^{-1}(\omega)$, which we denote $\omega^\sharp$ ("sharp" because we "raise the index"), will look like (with respect to a local frame) $g^{ij}\omega_iE_j$, where we haven't defined the object g^{ij} yet but it's basically the inverse matrix of the matrix g_{ij} (well, a matrix-valued function). In general, using the upper and lower g^{ij} and g_{ij} can be done on tensors – we can replace one of the vectors with a covector or vice versa by using the metric to provide an isomorphism, which corresponds to a multiplication by one of the matrices.

Example 16

The **gradient** doesn't make sense on a general smooth manifold, but if we have a function $f \in C^\infty(M)$ on a manifold (M, g) then we can form a vector out of it (from the metric) via $\nabla f = (df)^\sharp$. With respect to a local frame, this looks like

$$\nabla f = g^{ij} E_i(f) E_j,$$

or in other words in coordinates we would write this at $g^{ij} \frac{\partial f}{\partial x^i} \frac{\partial}{\partial x_j}$.

Spelling this out, the idea is that the gradient is a vector such that

$$g(\nabla f_p, v) = df(v),$$

so the "sharp" isomorphism allows us to invert this.

Example 17

We can take a $(0, 2)$-tensor $T = T_{ij} \varepsilon^i \otimes \varepsilon^j$ (which takes in two vectors) and get its trace. Specifically, we turn the second of the arguments in $T_pM \times T_pM \to R$ into T_p^*M, so that we have a map $\tilde{T}_p : T_pM \to T_pM$ that we can take the trace of.

We thus get a $(1, 1)$ tensor associated to T given by $\tilde{T} = T_{ij} g^{j\ell} \varepsilon^i \otimes E_\ell$. Then the **trace** of T is $g^{ij} T_{ij}$; alternatively, we can also say that

$$\text{tr}(T) = \sum_{i=1}^n T(E_i, E_i),$$

where E_i is a locally orthonormal frame.

Next, recall that an inner product gives us a natural volume form (since we know what a unit volume element is):

Definition 18

If (M, g) is an oriented manifold, then the volume form is defined via

$$dV_g = \varepsilon^1 \wedge \cdots \wedge \varepsilon^n,$$

where ε^i is a dual basis to the oriented orthonormal frame $E_1, \cdots, E_n$. Alternatively, in coordinates, we have $dV_g = \sqrt{\det g_{ij}} dx_1 \wedge \cdots \wedge dx^n$ (we can check that this normalization returns a volume of 1 where it should).

If (M, g) is not oriented, we don't expect such a volume form to exist – we can read about densities in the textbook, but we'll restrict ourselves to oriented manifolds. And we'll do some computations with these things on the homework.

Example 19

The **divergence** can also be defined: given a vector field $X \in \mathfrak{X}(M)$ on an oriented (not really necessary here, it's well-defined either way) manifold (M, g), we can define

$$d(X \lrcorner dV_g) = (\text{div}(X)) dV_g,$$

and this is well-defined because the n-dimensional forms form a one-dimensional vector space at each point.

We have to be a bit careful about the convention with the interior product, and specifically (on our homework) we can check that

$$\operatorname{div}\left(X^i \frac{\partial}{\partial x^i}\right) = \frac{1}{\sqrt{\det g}} \frac{\partial}{\partial x^i}(X^i \sqrt{\det g}).$$

We then get the **Laplacian** $\Delta f = \operatorname{div}(\nabla f)$, and in coordinates it looks like (also on our homework)

$$\Delta f = \frac{1}{\sqrt{\det g}} \frac{\partial}{\partial x^i}\left(g^{ij}\sqrt{\det g}\frac{\partial f}{\partial x^j}\right).$$

And on Euclidean space, we can check that this all reduces to the usual formulas because $\det g = 1$. We also get generalizations of the usual results like Green's identity, the divergence theorem, and so on (using Stokes' theorem and the definitions). And as we work with them more, we'll see that they are pretty fundamental definitions that will keep coming up.

But returning to our roots, we'll now stop working just at a point and consider more global objects like **length** and **distance**. There are some differing conventions here, and we'll use Lee's:

Definition 20

A **curve** is a parameterized arc, meaning that it is a map $\gamma : I \to M$ that is continuous (where I may be open or closed and potentially infinite). We call a curve a **curve segment** if I is compact, and we call γ **regular** if it is smooth and has nowhere vanishing derivative. Then γ is **admissible** if it is piecewise regular, meaning we have a curve (continuous everywhere) such that we can partition its domain $I = [a, b]$ into intervals $[a, a_1], [a_1, a_2], \cdots, [a_n, b]$ such that the curve on each interval is regular. (In particular, the left and right derivatives exist even at the "cusp" points.)

Definition 21

Let $\gamma : [a, b] \to (M, g)$ be an admissible curve. The **length** of the curve is given by

$$L_g(\gamma) = \int_{[a,b]} = \int_{[a,b]} |\gamma'(t)|_g dt,$$

where $|\gamma'(t)|_g = \sqrt{g_{ij}|_{\gamma(t)}(\gamma^i)'(t)(\gamma^j)'(t)}$.

The same proof as for the Euclidean metric shows that this is invariant under reparameterization, and we can always reparameterize (with the same orientation) so that $|\gamma'| = 1$ wherever the derivative exists (and we can make the right and left derivatives 1 as well).

Definition 22

For a connected Riemannian manifold (M, g), we may define

$$d_g(p, q) = \inf\{L_g(\gamma) : \gamma \text{ admissible curve between } p, q\}.$$

Clearly we need connectedness, and we can check that when the manifold is connected we do always have an admissible curve (connected implies path-connected, then replace in each chart with a segment). And the fundamental notion here is the following:

Fact 23

The definition above makes d_g a metric (in the sense of metric spaces), and the metric topology on M agrees with the topological manifold structure on M. The only thing we really must prove is that $d_g(p, q) > 0$ unless $p = q$, and the idea is to choose some chart around p not containing q, inside which $g \geq c\overline{g}$ by compactness, so any curve takes some bounded-below time to exit the chart.

So the volume form and distance are somehow fundamental things we can define on a Riemannian manifold – it's reasonable to ask about things like the surface area or distance when constrained along a surface. In general, it's often quite hard to calculate $d_g(p, q)$ unless there are a lot of symmetries – we've defined something pretty nontrivial here. And we'll be able to talk about whether the infimum is actually attained – this is not so obvious, because for example $\mathbb{R}^2 \setminus \{0\}$ does not have the distance between $(-1, 0)$ and $(1, 0)$ actually attained by any curve.

We'll close with a bit about model Riemannian metrics (with a lot of symmetry) as examples, and next time we'll talk about constructions. The point is that we can often actually do a lot of hands-on calculation which helps us understand things more concretely. Recall that $\text{Iso}(M, g)$ is the set of isometries $\phi : M \to M$.

Definition 24

For any point $p \in M$, define $\text{Iso}_p(M, g) = \{\phi \in \text{Iso}(M, g) : \phi(p) = p\}$.

Definition 25

A manifold (M, g) is **homogeneous** if $\text{Iso}(M, g)$ acts on M transitively (meaning that all points look the same), **isotropic at p** if Iso_p acts on $\{v \in T_pM : |v| = 1\}$ transitively (meaning that all directions look the same – this definition comes from the fact that isometries ϕ must have $d\phi_p$ map unit vectors to unit vectors), **isotropic** if isotropic at all p, and **frame homogeneous** (the most symmetric) if Iso acts on the set of frames $O(M, g) = \left\{\bigsqcup_{p \in M}\{\text{orthonormal bases of } T_pM\right\}$ transitively.

3 April 7, 2023

We'll do a few more examples today in preparation for getting to the "interesting part" of the class soon. Last time, we discussed isometries of a Riemannian manifold, and note that finding the full set requires a bit of theory. For example, even for the flat metric for Euclidean space, we know for any $A \in O(n)$ and $b \in \mathbb{R}^n$ that $x \mapsto Ax + b$ (an orthogonal matrix plus a translation) is an isometry. This defines the **Euclidean group** $E(n) = \left\{\begin{bmatrix} A & b \\ 0 & 1 \end{bmatrix} \in \text{GL}(n+1)\right\}$; the group acts frame homogeneously on $\mathbb{R}^n$ (so orthonormal frames can be moved to any other orthonormal frames), and it is indeed the whole isometry group (though that is a bit annoying to prove). Similarly, the sphere of radius R, $\{x; |x| = R\} \subset \mathbb{R}^{n+1}$ with the round metric, has $O(n)$ acting on it frame homogeneously, and these examples have the property that every point looks the same as every point in every direction.

Today, we'll do some less obvious properties which we might have seen in a complex analysis class:

Definition 26

Two metrics g_1, g_2 on M are **conformally related** if there exists some smooth positive function f on M with $g_2 = fg_1$.

In other words, at any point, the inner product under g_2 is just f times some inner product under g_1. (So we've changed lengths in all directions by the same amount, but not angles.)

Definition 27

A diffeomorphism $\phi : (M, g) \to (\tilde{M}, \tilde{g})$ is a **conformal diffeomorphism** if the pullback metric $\phi^*\tilde{g}$ is conformal to g. A manifold (M, g) is **locally conformally flat** if for all points $p \in M$, there is some neighborhood U of p and some $\phi : U \to \phi(U) \subset \mathbb{R}^n$ a conformal diffeomorphism.

Proposition 28

$(S^n, \mathring{g}_R)$ is locally conformally flat.

Proof. Let N be the north pole of S^n (of radius R). We'll show that $S^n \setminus N$ is conformally equivalent to $\mathbb{R}^n$ with the flat metric via **stereographic projection**: and then we can move the north pole around by homogeneity. Basically, for any point $p \in S^n(R) \setminus N$, consider the line through N and p and let $(\sigma(p), 0) \in \mathbb{R}^N \times 0$ be the intersection of that line with the plane $x_{n+1} = 0$. Explicitly, the calculation is that if $p = (\xi^1, \cdots, \xi^n, \tau)$, then

$$\sigma(p) = \frac{R\xi}{R - \tau} \implies \sigma^{-1}(u) = (\xi, \tau) = \left(\frac{2R^2u}{|u|^2 + R^2}, R\frac{|u|^2 - R^2}{|u|^2 + R^2}\right).$$

Then we want to show that the pullback of the round metric $(\sigma^{-1})^*\mathring{g}_R$ is a smooth function times the flat metric. But this is the same as the pullback $(\sigma^{-1})^*(\overline{g}_{\mathbb{R}^{n+1}}$, since the round metric is just the restriction from the full tangent space, and we can write this as $(d\xi^1)^2 + \cdots + (d\xi^n)^2 + d\tau^2$. We can then compute this explicitly from the formula above and find that

$$(\sigma^{-1})^*\mathring{g}_R = \frac{4R^4}{(|u|^2 + R^2)^2}\overline{g}_{\mathbb{R}^4},$$

so now indeed we have the result we want. □

(On the other hand, not all manifolds are locally conformally flat, and we'll see that in more detail later.) There's one other manifold that fits into this story, which we should think of as a "sphere of negative radius:"

Definition 29

There are various conventions for defining **hyperbolic space** in dimension $n > 1$ of radius $R > 0$, which are all isometric:

(a) the upper sheet of the hyperboloid in Minkowski space $\overline{q} = (d\xi^1)^2 + \cdots + (d\xi_n)^2 - d\tau^2$ (which does not yield a Riemannian metric because it's not positive). In other words, we are looking at

$$\{(\xi^1)^2 + \cdots + (\xi^n)^2 - \tau^2 = -R^2, \quad \tau > 0\}.$$

So this is a subset of Euclidean space and we use the inclusion map to pull back $\overline{q}$, and it turns out this actually yields a **Riemannian** metric on hyperbolic space,

(b) the **Beltrami-Klein metric** (explained below),

(c) the **Poincaré disk model** with metric $4R^2\frac{(du^1)^2+\cdots+(du^n)^2}{(R^2-|u|^2)^2}$ on $\{|u| < R\}$ (in particular, this blows up as $u \to R$),

(d) the **half-space** metric $R^2\frac{(dx^1)^2+\cdots+(dx^{n-1})^2+dy^2}{y^2}$ on the half-space $\{y > 0\}$.

In particular, hyperbolic space is locally conformally flat. The idea for the Beltrami-Klein metric (which is the least popular) is not conformally flat, but it's nice because the geodesics are straight lines so we might talk about it later. And the idea is that the Poincaré disk is what we get from stereographic projection from the hyperboloid down to $\tau = 0$, and the Beltrami-Klein metric comes from stereographic projection down to a different plane (tangent to the hyperboloid).

We could define less symmetric but still symmetric types of metrics – we can read the textbook for more on that – but for example if we know about Thurston's geometrization conjecture, the idea is that 3-manifolds can be chopped up into pieces which admit finite-volume relatively symmetric metrics. So continuing on in this theory is important but not for the basic theory. Instead, we'll briefly mention a few facts about Lie groups (and we should read more on our own for a fuller understanding) – note that matrix groups are the main example to keep in mind.

Definition 30

A **Lie group** is a smooth manifold G where left multiplication $L_\phi : G \to G$ sending $\phi' \mapsto \phi\phi'$ and right multiplication $R_\phi : G \to G$ sending $\phi' \mapsto \phi'\phi$ are smooth maps. The **Lie algebra** $\mathfrak{g}$ is the set of left-invariant vector fields X (meaning that $(L_\phi)_* X = X$); this is also isomorphic to the tangent space $T_e G$ (since choosing a vector gives a vector field on all of G by left multiplication). A Riemannian metric g on G is **left-invariant** if $(L_\phi)^* g = g$, which is equivalent to saying that $g(X, Y)$ is a constant function for all vector fields $X, Y \in \mathfrak{g}$.

So the left-invariant metrics are exactly the inner products on $T_e G$, since we can also pull back by the inverse. One basic thing we may want to know is what the **right** action by the group does – in particular, we may wonder when we have **bi-invariant** metrics. (For example, because addition on $\mathbb{R}^n$ is abelian, $(\mathbb{R}^n, \overline{g})$ is a bi-invariant metric.) More generally, the existence of bi-invariant metrics is usually a good thing to have for geometric purposes: we can define the conjugation map $C_\phi : G \to G$ sending $\psi \mapsto \phi\psi\phi^{-1}$, and this leads us to the **adjoint** $\text{Ad}(\phi) = (C_\phi)_*$, which is a map $\mathfrak{g} \to \mathfrak{g}$ (we must check that this preserves left-invariant vector fields and that it is linear). In particular, this yields a representation $\text{Ad} : G \to \text{GL}(\mathfrak{g})$.

Proposition 31

An inner product $\langle \cdot, \cdot \rangle$ on $T_e G$ is the same as a left-invariant metric g on G. This metric is then bi-invariant if and only if $\langle \cdot, \cdot \rangle$ is Ad-invariant, meaning that $\langle \text{Ad}(\phi)u, \text{Ad}(\phi)v \rangle = \langle u, v \rangle$.

Proof. If $X \in \mathfrak{g}$ is a left-invariant vector field, and we pushforward by $R_{\phi^{-1}}$, then

$$(R_{\phi^{-1}})^* X = (R_{\phi^{-1}})^* (L_\phi)^* X = (C_\phi)^* X = \text{Ad}(\phi)(X),$$

where the first equality comes from left-invariance of X. So

$$(R_{\phi^{-1}})^* g(X, Y) = g(\text{Ad}(\phi)X, \text{Ad}(\phi)Y) = \langle \text{Ad}(\phi)X, \text{Ad}(\phi)Y \rangle.$$

So invariance under Ad shows that this is just $g(X, Y)$ and we have the desired bi-invariance. □

Fact 32

G admits a bi-invariant metric if and only if the image $\text{Ad}(G)$ has compact closure in $\text{GL}(\mathfrak{g})$. What we're basically doing here is taking a random left-invariant metric and averaging over right actions (finding a measure requires compactness).

This may all seem very abstract, but it's actually rather concrete in the case of a matrix group: consider $\text{SL}(2, \mathbb{R}) = \{A \in M^{2\times 2} : \det(A) = 1\}$. Then the Lie algebra $\mathfrak{sl}(2, \mathbb{R})$ is the tangent space at the identity matrix, and the derivative of the determinant is the trace so we just have $\mathfrak{sl}(2, \mathbb{R}) = \{\text{tr}(A) = 0\}$. Then $\text{Ad}(A)(X) = AXA^{-1}$, and we claim that this means $\text{SL}(2, \mathbb{R})$ does not have a bi-invariant metric. Indeed, take $X_0 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ and $A_c = \begin{bmatrix} c & 0 \\ 0 & 1/c \end{bmatrix}$. Then we can check that $\text{Ad}(A_c)(X_0) = \begin{bmatrix} 0 & c^2 \\ 0 & 0 \end{bmatrix}$, which blows up as $c \to \infty$, so there is no bi-invariant metric because we do not have compact closure.

Similarly, S^3 sits in $\mathbb{C}^2$, and if we map $(\omega, z) \mapsto \begin{bmatrix} \omega & z \\ -\overline{z} & \overline{w} \end{bmatrix}$ this is in fact a diffeomorphism. The round metric is then corresponding to a bi-invariant metric on $SU(2)$. (And for another example, we can consider Berger's sphere.)

We'll finish by starting the "meat" of the course: we'll talk about **connections** on a Riemannian manifold, which are perhaps the most confusing in Riemannian geometry (and came much later than curvature). We'll use connections to define curvature (as the modern treatment goes), but it's not really the natural way to discover things. Here's how the story goes: on a Riemannian manifold, we want to define what it means to have the "shortest path between two points," but it's useful to talk about their properties in a different way, namely having "zero acceleration." But this can be quite hairy on a Riemannian manifold:

Example 33

Given a curve $\gamma(t) : I \to \mathbb{R}^n$, we get its derivative $\gamma' = \dot{\gamma}^1 \frac{\partial}{\partial x^1} + \cdots + \dot{\gamma}^n \frac{\partial}{\partial x^n}$ and $\gamma'' = \ddot{\gamma}^1 \frac{\partial}{\partial x^1} + \cdots + \ddot{\gamma}^n \frac{\partial}{\partial x^n}$. In particular, this makes sense in any coordinates – if $\gamma' \neq 0$, we can't change coordinates so that it is equal to zero.

But there's something funny happening already: if $\gamma(t) = (\cos t, \sin t) \in \mathbb{R}^2$, then $\gamma'' = -\cos t \frac{\partial}{\partial x} - \sin t \frac{\partial}{\partial y}$, but if $\gamma(t) = (1, t)$ in **polar** coordinates, it seems like we have $\gamma'(t) = \frac{\partial}{\partial \theta}$ and $\gamma''(t) = 0$. So if we want to work on a manifold, we need some notion which is invariant under change of coordinates, and we'll resolve this next time.

4 April 10, 2023

Last week, we rapidly toured through some examples and basics – we're now going to settle in and talk about connections, which are probably the most confusing (though not too complicated) objects we'll be studying. The issue is that they aren't that motivated, and our eventual destination will be to define curvature (in which connections will appear). But really, we don't need connections to define curvature – it was just discovered that it fits into the story well this way, and we've just repackaged things into a "modern" perspective.

Our motivating point was that when we changed coordinates along a curve, acceleration is zero in one case and nonzero in another. But it turns out we won't actually define acceleration yet – instead we'll look at **directional derivative of a vector field**, which is vaguely related and better for technical reasons.

Definition 34

Let $E \to M$ be a vector bundle, and let $\Gamma(E)$ denote the space of smooth sections of the bundle. A **connection** ∇ on E is a map $\nabla : \mathfrak{X}(M) \times \Gamma(E) \to \Gamma(E)$, where $(X, Y) \mapsto \nabla_X Y$, such that the following properties hold:

(i) The map $X \mapsto \nabla_X Y$ is C^∞-linear (in other words, we get something tensorial if we fix Y),

(ii) The map $Y \to \nabla_X Y$ is $\mathbb{R}$-linear,

(iii) We have the product rule (for any $f \in C^\infty(M)$)

$$\nabla_X(fY) = f\nabla_X Y + X(f)Y.$$

In other words, we are "differentiating a section in the direction of a vector field." And we know that the derivative is not C^∞-linear, so it makes sense that we just have $\mathbb{R}$-linearity in the second argument. Here $X(f)$ is the function where at each point we take the derivative of f with respect to X.

At this level of generality, this is something we can put on a vector bundle over any smooth manifold – there's no geometry in this, and we'll have to rectify that.

Example 35

On Euclidean space $\mathbb{R}^n$ with its tangent space $E = T\mathbb{R}^n$ (which is just naturally itself), an example of a connection is $\overline{\nabla}_X Y = X(Y^1)\frac{\partial}{\partial x^1} + \cdots + X(Y^n)\frac{\partial}{\partial x^n}$ (a vector-valued directional derivative). Now if we have an embedded submanifold $M^n \subset \mathbb{R}^N$, we get an **induced connection** on $E = TM$ as follows: for any section $Y \in \Gamma(TM) = \mathfrak{X}(M)$, we can extend it to a vector space $\tilde{Y} \in \mathfrak{X}(\mathbb{R}^N)$ (on charts, then using a partition of unity), and then we define the "tangential connection"

$$\nabla^T_X Y = (\overline{\nabla}_X \tilde{Y})^{\text{proj. to } TM}.$$

In other words, extend globally, compute the directional derivative on $\mathbb{R}^N$ for each coefficient, and this is a $\mathbb{R}^N$-valued function along the manifold, and then using the Euclidean inner product we orthogonally project back. We will check that this is independent of our extension $\tilde{Y}$ and that it actually defines a connection on our homework. In other words, we should think of connections either abstractly with the definition or in this concrete way. And now it turns out that on M we have a metric induced by $\mathbb{R}^N$ and this leads to a connection; however, we get the **same** connection if we choose a different embedding under the same metric. (But actually "finding" the connection is often the hard part.)

Remark 36. *In other areas of differential geometry, we may look at various connections, and Yang-Mills theory is somehow the study of connections. But we won't go much into that for now.*

Lemma 37

Let ∇ be a connection on a bundle $E \to M$. Then $(\nabla_X Y)_p$ depends only on Y **near** p and on the value of X **at** p.

Proof. For Y, suppose that Y is zero near p. We want to show that $(\nabla_X Y)_p = 0$. Choose a coordinate neighborhood U so that $Y = 0$ in U, and choose $f = 1$ near p but with the support of f contained in U. Then

$$0 = (\nabla_X fY)_p = f(p)(\nabla_X Y)_p + X(f)_p Y_p,$$

first step because fY is actually the zero section everywhere and thus we can use $\mathbb{R}$-linearity, and second step by the product rule. Then $f(p) = 1$ and $X(f)_p = 0$ because f is identically 1 near p, so $0 = (\nabla_X Y)_p$ as desired.

The proof for X is similar; we just need to also mention that (as on our homework) C^∞-linearity means we define a tensor so things only depend on the value at p. $\square$

Proposition 38

Let ∇ be a connection on $E \to M$, and let $U \subset M$ be an open set. Then there is a unique "restricted connection" ∇^U on $E|_U \to U$ such that for all $X \in \mathfrak{X}(M)$ and $Y \in \Gamma(E)$, then

$$\nabla^U_{X|_U} Y|_U = (\nabla_X Y)|_U.$$

Existence and uniqueness basically follow from the locality that we just proved; the proof is in the textbook but it's not too interesting. So we'll only apply connections to things that are locally defined.

We'll now specialize to **affine connections**, which are connections on the tangent space TM. In other words, ∇ will be a map $\mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$.

Remark 39. *A map $\mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ is often a $(1,2)$-tensor, but* **a connection is not a** $(1,2)$ **tensor** *in our case because we don't have C^∞-linearity in the second variable.*

We can now compute with respect to a local frame E_i – as we said, we're not worried about computing with respect to only locally defined things (we could always cut off E_i and extend globally, but we are fine no matter how we do that). We know that $\nabla E_i E_j$ will be some linear combination of E_ks, and we will write this as

$$\nabla_{E_i} E_j = \Gamma^k_{ij} E_k$$

for some **connection coefficients** Γ^k_{ij} (and here Einstein notation says that we only sum over k, because it appears once as a lower and once as an upper index). So now to understand the recipe for "computing the derivative of Y with respect to X at a local frame," suppose $X = X^i E_i$ and $Y = Y^j E_j$. Then

$$\nabla_X Y = \nabla_{X^i E_i}(Y^j E_j) = X^i \nabla_{E_i}(Y^j E_j)$$

by C^∞-linearity, but now we're not C^∞-linear in the inner slot so we must use the product rule (writing in the opposite order):

$$= X^i (E_i(Y^j) E_j + Y_j \nabla_{E_i} E_j).$$

We'll change the notation in the first term (just a change of dummy indices) from j to k,

$$= X^i (E_i(Y^k) E_k + Y_j \nabla_{E_i} E_j) = \left[X^i E_i(Y^k) + X^i Y_j \Gamma^k_{ij}\right] E_k,$$

where now we could also write $X^i E_i(Y^k)$ as $X(Y^k)$. Notice that in $\overline{\nabla}$, the rule was just that we "hit the coefficients with X," and for the exact same reason that acceleration was not coordinate-independent, trying to differentiate each coefficient of the vector field will go poorly. Instead, we should think of the connections as the Γ^k_{ij} "correction terms" that make the derivative into a reasonable rule. (So if we use the "parallel frame" $E_i = \frac{\partial}{\partial x^i}$ for $\overline{\nabla}$, we have $\Gamma^k_{ij} = 0$. But then if we change coordinates, because ∇ is not a tensor, those values will change.)

We'll now turn this around: suppose we have U and we've chosen a local frame on U. Then **an affine connection ∇^U on U is exactly the same thing as choosing n^3 smooth functions Γ^k_{ij} on** U; if we are told the connection coefficients, then our formula for $\nabla_X Y$ satisfies the product rule and the other properties. So as a corollary, all coordinate charts on U admit affine connections, for example by taking all Γ^k_{ij} to be zero, and thus we have the following:

Corollary 40
All manifolds M admit connections.

Proof. Choose a cover $\{U_\alpha\}$ by coordinate charts, and let ϕ_α be a partition of unity subordinate to it, and choose any connection ∇_α on U_α. Then we want to say $\nabla = \sum \phi_\alpha \nabla_\alpha$; in other words,

$$\nabla_X Y = \sum \phi_\alpha (\nabla)\alpha)_X Y.$$

(The point is that ∇_α is being multiplied by zero wherever it's not defined.) However, it's not clear the product rule is satisfied, so we do need to check that:

$$\nabla_X(fY) = \sum \phi_\alpha (\nabla_\alpha)_X (fY) = \sum \phi_\alpha f (\nabla_\alpha)_X Y + \sum \phi_\alpha X(f) Y = f \nabla_X Y + X(f) Y,$$

where **importantly** we've used that $\sum \phi_\alpha = 1$. In particular, **the sum of two connections is not a connection**, since we'd get $2X(f)Y$ instead of $X(f)Y$ in the second term. □

Proposition 41
Let ∇^0, ∇^1 be two affine connections. Then $D(X, Y) = \nabla^0_X Y - \nabla^1_X Y$ is a $(1, 2)$-tensor.

Proof. We just need to show C^∞-linearity by what we said on our homework. (Basically, we need to check linearity at X, Y and at the result; the latter can be checked by evaluating at a 1-form.). The action on 1-forms is C^∞-linear and C^∞-linearity is also true at X, and at Y the point is that the extra terms in the product rule cancel out:

$$D(X, fY) = f \nabla^0_X Y + X(f) Y - f \nabla^1_X Y - X(f) Y = f D(X, Y).$$

This "cancelling out" is going to be a common occurence later in the class as well. □

Corollary 42
The space of connections is **affine** (though this is not why it's called an affine connection): we have

$$\{\text{affine connections}\} = \{\nabla^0 + D : D \in \Gamma(T^{(1,2)}(M)\}.$$

(We do need to check that the other inclusion holds, but it's just about checking the product rule.) We'll now move on to **covariant derivatives** and connections on tensor fields: the basic idea is that if ∇ is an affine connection (meaning we can differentiate vector fields), we get a connection on tensors as well.

Proposition 43

Given an affine connection, there is a **unique** connection on each tensor bundle $T^{(k,\ell)}M$ such that:

1. the definition on $T^{(1,0)}M = TM$ agrees with the original construction,
2. on $T^{(0,0)}M = \mathbb{R} \times M$, we have $\nabla_X f = X(f)$,
3. for any two tensors F, G, $\nabla_X(F \otimes G) = \nabla_X F \otimes G + F \otimes \nabla_X G$,
4. if $F \in \Gamma(T^{k,\ell}M)$, then $\text{tr}(F) \in \Gamma(T^{k-1,\ell-1}M)$ (tracing along an upper index and a lower index by looking at the map $TM \to TM$; this is metric-independent) satisfies $\nabla_X(\text{tr}(F)) = \text{tr}(\nabla_X F)$.

This connection then also satisfies the following:

5. For any $\omega \in \Omega^1$ and $Y \in \mathfrak{X}$, we have $\langle \omega, Y \rangle$ (also denoted $\omega(Y)$) satisfying $\nabla_X \langle \omega, Y \rangle = \langle \nabla_X \omega, Y \rangle + \langle \omega, \nabla_X Y \rangle$,
6. For a general tensor $T \in \Gamma(T^{(k,\ell)}M)$, the function $T(\omega^1, \cdots, \omega^k, Y_1, \cdots, Y_\ell)$ can be differentiated with respect to the vector field X, and we have the product rule $X(T(\omega^1, \cdots, \omega^k, Y_1, \cdots, Y_\ell)) = (\nabla_X T)(\omega^1, \cdots, \omega^k, Y_1, \cdots, Y_\ell) + \sum_i T(\cdots, \nabla_X \omega^i, \cdots) + \sum_j T(\cdots, \nabla_X Y_j, \cdots)$ (where the other $\cdots$ arguments stay unchanged).

We'll prove this next time – the point is that the four properties describe a unique connection that exists, and once we find that connection then we also get two additional properties that are actually the ones that we care about.

5 April 12, 2023

Class was canceled today – the material for this lecture is pages 95-105 of Lee. Here are some of the main points (and there will be a review of this done in class on Friday):

- The result stated above is proved as follows: the first four properties imply the last two, because $\omega(Y) = \text{tr}(\omega \otimes Y)$ (this is true in coordinates) so we can use properties (3) and (4) to get property (5) and then a similar inductive process with $(k + \ell)$ traces. Now we have uniqueness because $(\nabla_X \omega)(Y) = X(\omega(Y)) - \omega(\nabla_X Y)$ by (2) and (6) tells us that the connection is uniquely determined on 1-forms, and then (6) tells us how to compute the connection on general tensors in terms of connections on 1-forms and vector fields. Existence requires us to check $C^\infty(M)$-linearity, which is true because if we replace any ω^i or Y_j with $f\omega^i$ or fY_j and expand, then the terms with f-derivatives cancel out. Finally, the defining properties of a connection follow from computation of the product rule.
- We have formulas for $\nabla_X F$ (we call this the **covariant derivative**) in a local frame: for a vector field $X = X^i E_i$ and a 1-form $\omega_i \varepsilon^i$, we have
$$\nabla_X(\omega) = (X(\omega_k) - X^j \omega_i \Gamma^i_{jk})\varepsilon^k.$$
More generally, if a (k, ℓ) tensor is written as $F = F^{i_1, \cdots, i_k}_{j_1, \cdots, j_\ell} E_{i_1} \otimes \cdots \otimes E_{i_k} \otimes \varepsilon^{j_1} \otimes \cdots \otimes \varepsilon^{j_\ell}$, then
$$\nabla_X F = \left(X(F^{i_1, \cdots, i_k}_{j_1, \cdots, j_\ell}) + \sum_{s=1}^{k} X^m F^{i_1, \cdots, p, \cdots, i_k}_{j_1, \cdots, j_\ell} \Gamma^{i_s}_{m,p} - \sum_{s=1}^{\ell} X^m F^{i_1, \cdots, i_k}_{j_1, \cdots, p, \cdots, j_\ell} \Gamma^p_{m, j_s} \right) \times E_{i_1} \otimes \cdots \otimes E_{i_k} \otimes \varepsilon^{j_1} \otimes \cdots \otimes \varepsilon^{j_\ell}$$
where p replaces the place of i_s and j_ss, respectively.

- We can thus define a map ∇ (called the **total covariant derivative**) from (k,ℓ) tensors to $(k,\ell+1)$ tensors via $(\nabla F)(\omega^1,\cdots,\omega^k,Y_1,\cdots,Y_\ell,X)=\nabla_X F(\omega^1,\cdots,\omega^k,Y_1,\cdots,Y_\ell)$ (since we have C^∞-linearity by definition of a connection), and we get a similar formula for the total covariant derivative in coordinates.
- Taking the covariant derivative twice, we can get a $(k,\ell+2)$-tensor $\nabla^2 F=\nabla(\nabla F)$. Letting $\nabla^2_{X,Y}F(\cdots)=\nabla^2 F(\cdots,Y,X)$, we have $\nabla^2_{X,Y}F=\nabla_X(\nabla_Y F)-\nabla_{\nabla_X Y}(F)$ (via a computation). For any smooth function u, we get $\nabla^2 u=\nabla(du)$, which we call the **covariant Hessian** of u.
- So now if we have a curve $\gamma: I\to M$, we may define vector and tensor fields along the curve (not necessarily requiring them to be tangent to the curve itself), and then the connection ∇ yields a **covariant derivative along the curve** D_t which is linear, satisfies $D_t(fV)=f'V+fD_tV$, and satisfies $D_tV_t=\nabla_{\gamma'(t)}\tilde{V}$ for every extension of V (if one exists). In coordinates, we have

$$D_tV(t)=\dot{V}^j(t)\partial_j|_{\gamma(t)}+V^j(t)\nabla_{\gamma'(t)}\partial_j|_{\gamma(t)}=(\dot{V}^k(t)+\dot{\gamma}^i(t)V^j(t)\Gamma^k_{ij}(\gamma(t)))\partial_k|_{\gamma(t)}.$$

  In particular, this shows that $\nabla_v Y$ at p depends only on values of Y along curves of velocity v at p (a generalization of the locality of $\nabla_X Y$ from before).
- Finally, we can return to the original motivation and define acceleration of a curve γ to be the vector field $D_t\gamma'$ and call a curve a **geodesic** if $D_t\gamma'=0$ everywhere. In coordinates, this is the requirement that (for a vector $(x(t))$

$$\ddot{x}^k(t)+\dot{x}^i(t)\dot{x}^j(t)\Gamma^k_{ij}(x(t))=0.$$

  By ODE theory, geodesics satisfy existence and uniqueness as long as the manifold does not have any boundary, and there is always a unique maximal geodesic which cannot be further extended. The geodesic starting at p with velocity v is often denoted γ_v.

6 April 14, 2023

Today's lecture will be given by Shuli Chen. Last time, we defined affine connections on the tangent bundle, and given such a connection we stated that we can extend that connection to general tensor bundles. The idea is that for a one-form $\omega\in\Omega^1$, we can define

$$(\nabla_X\omega)(Y)=X(\omega(Y))-\omega(\nabla_X Y);$$

the way to remember this is to think of this as the "product rule," since $X(\omega(Y))$ is taking a derivative of $\omega(Y)$ so we must apply ∇_X to each of the other terms. And then for a general (k,ℓ)-tensor, we may define

$$(\nabla_X F)(\omega^1,\cdots,\omega^k,Y_1,\cdots,Y_\ell)=X(F(\omega^1,\cdots,\omega^k,Y_1,\cdots,Y_\ell)$$

$$-\sum_{s=1}^k F(\omega^1,\cdots,\nabla_X\omega^s,\cdots,\omega^k,Y_1,\cdots,Y_\ell)-\sum_{s=1}^\ell F(\omega^1,\cdots,\omega^k,Y_1,\cdots,\nabla_X Y_s,\cdots,Y_\ell),$$

where we basically take the derivative on every possible entry. We do need to check that this defines a connection, which we'll just do for 1-forms. The necessary linearity is clear from the definition, and for the product rule we have

$$\begin{aligned}\nabla_X(f\omega)(Y)&=X(f\omega(Y))-f\omega(\nabla_X Y)\\&=(Xf)\omega(Y)+fX(\omega(Y))-f\omega(\nabla_X Y)\\&=(Xf)\omega(Y)+f(\nabla_X\omega)(Y).\end{aligned}$$

Thus $\nabla_X(f\omega) = (Xf)\omega + f\nabla_X\omega$ as required. Furthermore, we can check that ∇_X commutes with trace (contracting a form with a vector field, which is a natural operation) and that $\nabla_X(F \otimes G) = (\nabla_X F) \otimes G + F \otimes (\nabla_X G)$. We can then also write down computations in a local frame $\{E_i\}$: recalling that $\nabla_{E_i} E_j = \Gamma_{ij}^k E_k <$ we can compute for a 1-form

$$\begin{aligned}
\nabla_X\omega(E_k) &= X(\omega(E_k)) - \omega(\nabla_X E_k) \\
&= X(\omega(E_k)) - \omega(\nabla_{X^i E_i} E_k) \\
&= X(\omega_k) - \omega(X^i \Gamma_{ik}^j E_j) \\
&= X(\omega_k) - X^i \omega_j \Gamma_{ik}^j
\end{aligned}$$

and thus $\nabla_X \omega = (X(\omega_k) - X^i\omega_j\Gamma_{ik}^j)\varepsilon^k$. And the more general case just looks more complicated and can be found in the book. We can now use the connection to go from a (k, ℓ)-tensor $F \in \Gamma(T^{(k,\ell)}(M))$ and get $\nabla F \in \Gamma(T^{(k,\ell+1)}M)$ via

$$\nabla F(\omega^1, \cdots, \omega^k, Y_1, \cdots, Y_\ell, X) = \nabla_X F(\omega^1, \cdots, \omega^k, Y_1, \cdots, Y_\ell).$$

For notational clarity, we often denote the components ∇Y as $Y^i_{;j} E_i \otimes \varepsilon^j$, where the components are

$$Y^i_{;j} = E_j(Y^i) + Y^k \Gamma^i_{jk}.$$

Similarly, for the total derivative of a 1-form, we have

$$\nabla \omega = \omega_{i;j} \varepsilon^i \otimes \varepsilon^j; \quad \omega_{i;j} = E_j(\omega_i) - \omega_k \Gamma^k_{ji}.$$

And we can apply ∇ twice to get the second covariant derivative, defined by

$$\nabla^2_{X,Y} = \nabla^2 F(\cdots, Y, X),$$

though the ordering of X and Y may depend on the author. Notably, there is a first-order correction when we compare this to $\nabla_X \nabla_Y$: we have $\nabla^2_{X,Y} = \nabla_X \nabla_Y - \nabla_{\nabla_X Y}$. We'll check this on 1-forms: we have

$$\nabla^2_{X,Y}(\omega) = \nabla^2 \omega(Z, Y, X) = \nabla_X(\nabla \omega)(Z, Y),$$

and now applying the product rule allows us to write this as

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

Now applying the definition of the total derivative again, this simplifies as

$$= X(\nabla_Y \omega(Z)) - (\nabla_Y \omega)(\nabla_X Z) - (\nabla_{\nabla_X Y}\omega)(Z).$$

Finally, applying the product rule on the first entry yields

$$= \nabla_X(\nabla_Y \omega)(Z) + (\nabla_Y\omega)(\nabla_X Z) - (\nabla_Y \omega)(\nabla_X Z) - (\nabla_{\nabla_X Y}\omega)(Z),$$

with the middle two terms disappearing and the result following. The same type of calculation works generally as well. Remember that $\nabla_X u$ is just the directional derivative (by definition) for a function u, and in fact we have $\nabla u = du, \nabla^2 u = \nabla(du)$, $\nabla^2_{X,Y} u = X(Y(u)) - (\nabla_X Y)(u)$, and $\nabla^2 u = u_{;ij} dx^i \otimes dx^j = (\partial_j(\partial_i u) - T^k_{ij}\partial_k u) dx^i \otimes dx^j$. (Note that we should be careful about the notation ∇u because it is also used to denote the gradient $\nabla u = (du)^\sharp$.)

Fact 44

We've seen three types of derivatives, namely the covariant derivative, Lie derivative, and the exterior differential. The Lie derivative and covariant derivative both allow for directional derivatives, but $\mathcal{L}_X T$ measures how T changes along the **flow** of X, so it depends on the vector field around X. On the other hand, $\nabla_X T$ makes sense even if we only know the value of X at a single point, though we do need to first define the connection to get it. And similarly the differential does not require any geometric structure to define, but we can only apply d to differential forms, not general tensors.

Our next topic is **vector fields along curves**: if $\gamma : I \to M$ is a smooth curve, then a (smooth) vector field along γ is a (smooth) continuous map $V : I \to TM$ such that at each $t \in I$, $V(t)$ is some element of the tangent space $T_{\gamma(t)}(M)$. For example, restricting a vector field to a curve always works, but not every curve arises this way: we call a vector field along γ **extendible** if there is some smooth vector field $\tilde{v}$ on a neighborhood of $\text{im}(\gamma)$ such that $v(t) = \tilde{V}(\gamma(t))$, and we can come up with non-extendible vector field if the curve crosses back over itself.

Theorem 45

Let ∇ be an affine connection on M, and let $\gamma : I \to M$ be a smooth curve. Then there exists a unique map $D_t : \mathfrak{X}(\gamma)$ which is $\mathbb{R}$-linear (meaning $D_t(aX + bY) = aD_t(x) + bD_t(y)$), satisfies the product rule $D_t(fV) = f'V + fD_tV$, and such that if V is extendible, then we have $D_tV(t) = \nabla_{\gamma'(t)} = \tilde{V}$.

Proof. We will use the product rule to restrict just to a local path. The value of D_tV depends only on V on $(t_0, -\varepsilon, t_0 + \varepsilon)$, so if we choose ε small enough so that this is all contained within a single coordinate chart U, we have by the product rule that

$$\begin{aligned} V(t) = V^i(t)\partial_i(\gamma(t))\} \implies D_tV &= \dot{v}^i\partial_i + V^iD_t\partial_i \\ &= \dot{V}^k\partial_k + V^j\nabla_{\gamma'}\partial_j = \dot{v}^k\partial_k + v^j\nabla_{\gamma'}\partial_j \\ &= \dot{V}^k\partial_k + V^j\nabla_{\dot{\gamma}^i\partial_i}\partial_j \\ &= \dot{V}^k\partial_k + V^j\gamma^i\Gamma^k_{ij}\partial_k, \end{aligned}$$

which is independent of everything except V and γ. So this shows uniqueness, and for existence we just patch together charts and check that we satisfy the three conditions. ☐

Corollary 46

Let $Y, \tilde{Y}$ be two vector fields in $\mathfrak{X}(M)$, and suppose that $Y(\gamma(t)) = \tilde{Y}(\gamma(t))$ for some smooth γ with $\gamma(0) = p$ and $\gamma'(0) = v$. Then $\nabla_vY = \nabla_v\tilde{Y}$ at the point p.

We're now ready to talk about **geodesics**: for a curve $\gamma : I \to M$, we can define the acceleration of the curve to be $D_t\gamma'$ and call γ a **geodesic** if $D_t\gamma' = 0$. In local coordinates, these look like second-order nonlinear ODEs: if $\gamma = (x^1, \cdots, x^n)$ and $\dot{\gamma}(t) = \dot{x}^1\partial_1 + \cdots + \dot{x}^n\partial_n$, then

$$\ddot{x}^k + \dot{x}^i\dot{x}^j\Gamma^k_{ij}(x(t)) = 0.$$

Theorem 47

Let ∇ be an affine connection and let $p \in M, \omega \in T_pM, t_0 \in \mathbb{R}$. Then there is some open set I containing t_0 in $\mathbb{R}$ and some geodesic $\gamma : I \to M$ such that $\gamma(t_0) = p$ and $\gamma'(t_0) = \omega$, meaning that any two such geodesics agree on a common domain.

The idea is basically to use ODE theory and patch solutions together on the real line. And finally, we'll discuss **parallel transport**, which allows us to move vectors around:

Definition 48

A vector (or tensor) field $V \in \mathfrak{X}(\gamma)$ is **parallel** if $D_tV = 0$.

The condition for being parallel can be written as a first-order **linear** ODE $\dot{V}^k + V^j\dot{\gamma}^i\Gamma_{ij}^k(\gamma(t)) = 0$, so by existence and uniqueness and patching we have the following:

Theorem 49

Let ∇ be an affine connection, $\gamma : I \to M$ a path, $t_0 \in I$ and $v \in T_{\gamma(t_0)}M$. Then there is a unique parallel vector field $V \in \mathfrak{X}(\gamma)$ such that $V(t_0) = v$. (A similar result holds for tensor fields too.)

So we can now parallel transport a vector along a piecewise smooth curve, yielding a map $P^{\gamma}_{t_0,t_1} : T_{\gamma(t_0)}(M) \to T_{\gamma(t_1)M}$ via the map $v \mapsto V \mapsto V(t_1)$. And because D_t is $\mathbb{R}$-linear, this parallel transport map will be linear as well, and if we go backwards we can transport from t_1 back to t_0 and thus we get an isomorphism. In particular, we can parallel transport a frame along the curve as well: if $\{E_i\}$ is a parallel frame along γ, then we may write

$$V(t) = V^i(t)E_i(t) \implies D_tV(t) = \dot{V}^i(t)E_i(t),$$

and then a vector field is parallel will be constant if and only if its component functions are constant with respect to this parallel frame. Furthermore, parallel transport determines the covariant differentiation:

Theorem 50

Let ∇ be an affine connection and $V \in \mathfrak{X}(\gamma)$. Then we have

$$\nabla_t V(t_0) = \lim_{t \to t_0} \frac{P^{\gamma}_{t,t_0}(V(t)) - V(t_0)}{t - t_0}.$$

Proof. Fix a parallel frame $V = V^iE_i$, so that $P^{\gamma}_{t,t_0}(V(t)) = V^i(t)E_i(t_0)$; substituting it back in will give us the result. □

Note that it's not always possible to extend to a **parallel** vector field on an open set, even though it's always possible to extend a single vector to a vector field. (The main example to keep in mind is that parallel transporting on the sphere along a loop may not always get us back where we started.)

7 April 17, 2023

Today's topic is the **Levi-Civita connection** – as before, the issue is that we're doing things out of historical order. We've defined affine connections and various operations with them, but the basic problem is that there are infinitely many connections we can put on a manifold and there's been no obvious choice. We did see one "pseudo-obvious choice," which is the tangential connection ∇^T on a submanifold M^n of Euclidean space $\mathbb{R}^N$; essentially the miraculous fact about connections is that this is actually the correct one on M^n, and that this doesn't actually depend on the choice of embedding. (This is not going to be the way we prove that "good connections" exist, but it's what we're after.) The strategy is the following: we'll describe two properties of connections, ∇^T will have those two properties, and it'll turn out that those two properties define a unique connection.

The first is **metric compatability**: let $X, Y \in \mathfrak{X}(M)$ be two vector fields, and we'll write $\langle X, Y\rangle \in C^\infty(M)$ to be the function which is equal to $g_p(X_p, Y_p)$ at a point p.

Definition 51

An affine connection ∇ is **metric-compatible** if

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

In other words, this is a way to couple the metric to the connection via the product rule.

Proposition 52

The tangential connection ∇^T is metric-compatible.

Proof. Let $X, Y, Z \in \mathfrak{X}(M)$ extend to vector fields $\tilde{X}, \tilde{Y}, \tilde{Z} \in \mathfrak{X}(\mathbb{R}^N)$. Then we know that

$$X\langle Y, Z\rangle_M = \tilde{X}\langle \tilde{Z}, \tilde{Z}\rangle_{\mathbb{R}^N}$$

because $\langle Y, Z\rangle_M$ is the same as $\langle \tilde{Y}, \tilde{Z}\rangle_{\mathbb{R}^N}$, and then X and $\tilde{X}$ are tangent vectors so the dependence is only pointwise. But now we can just use the regular product rule since we're taking ordinary directional derivatives on $\mathbb{R}^N$:

$$= \langle \overline{\nabla}_{\tilde{X}} \tilde{Y}, \tilde{Z}\rangle + \langle \tilde{Y}, \overline{\nabla}_{\tilde{X}} \tilde{Z}\rangle.$$

But at a point $p \in M$, $\tilde{Z}$ is the same as Z (though we can't just remove the tildes on $\overline{\nabla}_{\tilde{X}} \tilde{Y}$), and now because Z is tangential we can project the other argument:

$$= \langle (\overline{\nabla}_{\tilde{X}} \tilde{Y})^T, Z\rangle + \langle Y, (\overline{\nabla}_{\tilde{X}} \tilde{Z})^T\rangle$$

and this is equal to the right-hand side we want. □

Proposition 53

Let ∇ be an affine connection on (M, g). Then the following are equivalent:

1. ∇ is metric-compatible,
2. The total covariant $(1,3)$-tensor $\nabla g = 0$,
3. In any local frame, we have $\Gamma^\ell_{ki} g_{\ell j} + \Gamma^\ell_{kj} g_{i\ell} = E_k(G_{ij})$,
4. For $V, W \in \mathfrak{X}(\gamma)$, we have $\frac{d}{dt}\langle V, W\rangle = \langle D_t V, W\rangle + \langle V, D_t W\rangle$,
5. If vector fields V, W are parallel along γ, then $\langle V, W\rangle$ is constant,
6. The parallel transport map $P^\gamma_{t_0,t_1}$ is a linear **isometry**,
7. For any smooth curve γ and any orthonormal basis at $\gamma(t_0)$, we may extend to a parallel orthonormal frame along all of γ.

Note for property (7) that it's not true that an orthonormal basis can be extended to an n-dimensional neighborhood, only along a curve. And often doing computations involving the D_t operator can be easier by using this property.

Proof. For (1) implies (2), consider the total covariant derivative ∇g. We have by definition that

$$\nabla g(Y, Z, X) = X(g(Y, Z)) - g(\nabla_X Y, Z) - g(Y, \nabla_X Z),$$

and now the right-hand side is identically zero if and only if we are metric-compatible. For (2) implies (3) (and vice versa), let Y, Z, X be written in terms of frames $E_i, , E - j, E_k$; we then have (because $\nabla_X Y = \nabla_{E_k} E_i = \Gamma^\ell_{ki} E_\ell$)

$$\nabla g(E_i, E_j, E_k) = E_k(g_{ij}) - \Gamma^\ell_{ki} g_{\ell j} - \Gamma^\ell_{kj} g_{i\ell},$$

so the left-hand side is zero if and only if this equality holds because the E_is are a basis. For (4) implies (1), note that $\nabla_X Y$ can be computed by looking at a curve whose derivative at the point is X and then computing $D_t Y|_{t=0}$. Meanwhile for (1) implies (4), V may not be extendible so we need to use a key trick: we can do the computation at a point in a coordinate neighborhood, where we write $V = V^i \partial_i$ and $W = W^j \partial_j$. Now ∂_i is extendible (because it's the restriction of the coordinate ∂_i), so

$$\frac{d}{dt}\langle V, W\rangle = \frac{d}{dt} V^i W^j \langle \partial_i, \partial_j\rangle = \frac{d}{dt}(V^i W^j)\langle \partial_i, \partial_j\rangle + V^i W^j \frac{d}{dt}(\langle \partial_i, \partial_j\rangle).$$

Now $\langle \partial_i, \partial_j\rangle$ is a function on our neighborhood, so $\frac{d}{dt}$ of it is $\nabla_{\gamma'}\langle \partial_i, \partial_j\rangle$ by the chain rule. Strictly speaking we want metric compatibility for a vector field, but in the X-slot (that is, the γ'-slot) the value only depends on the value at our specific point. So we have

$$\frac{d}{dt}(\langle \partial_i, \partial_j\rangle) = \langle \nabla_{\gamma'}\partial_i, \partial_j\rangle + \langle \partial_i, \nabla_{\gamma'}\partial_j\rangle,$$

so if we plug everything back in we have

$$\frac{d}{dt}\langle V, W\rangle = \langle \dot{V}^i \partial_i + V^i \nabla_{\gamma'}\partial_i, W^j \partial_j\rangle + \langle V^i \partial_i, \dot{W}^j \partial_j + W^i \nabla_{\gamma'}\partial_j\rangle.$$

But now $\nabla_{\gamma'}\partial_i$ is just $D_t \partial_i$ so by the product rule for D_t we just get $\langle D_t V, W\rangle + \langle V, D_t W\rangle$, as desired. (So the main idea is that the **coordinate functions are extendible** so we can use the original connection.)

Next,(4) implies (5), (5) implies (6), and (6) implies (7) follow quickly from the definitions – for (5) implies (6), parallel translating v, w at time t_0 to Pv, Pw at time t_1 tells us that $\langle v, w\rangle = \langle Pv, Pw\rangle$, and for (6) implies (7), choosing an orthonormal basis allows us to extend along a curve and then isometry preserves orthonormality. And (7) implies (4) by choosing such a frame and then doing the computation by writing in terms of the frame so that everything works like a straight line along Euclidean space. □

Corollary 54

Let ∇ be metric-compatible with (M, g), and let γ be a smooth curve. Then we have the following:

1. $|\gamma'|$ is constant if and only if $\gamma' \perp D_t\gamma'$.
2. In particular, if γ is a geodesic (meaning $D_t\gamma' = 0$), then $|\gamma'|$ is constant.

Proof. For (1), note by condition (4) of the previous result that

$$\frac{d}{dt}|\gamma'|^2 = 2\langle \gamma', D_t\gamma' \rangle,$$

and for (2) condition (1) holds because anything is perpendicular to zero. □

If metric compatibility told us everything we need to choose a connection, then the story of connections would not be that confusing. But on any Riemannian manifold of dimension two, there's infinitely many such metric-compatible connections and thus we need to do something better. Instead we need something **torsion-free** (also called **symmetric** in Lee) – we're going to make a certain choice of property, but it's going to be a bit hard for us to see why this is the correct canonical choice.

Recall that for any two vector fields $X, Y \in \mathfrak{X}$, we can define a commutator $[X, Y] \in \mathfrak{X}$ via

$$[X, Y](f) = X(Y(f)) - Y(X(f)).$$

We can check that this is indeed a vector field (the second derivatives cancel out).

Definition 55

An affine connection ∇ is **torsion-free** (also **symmetric**) if $\nabla_X Y - \nabla_Y X = [X, Y]$.

For example, for the flat connection on Euclidean space we have

$$\overline{\nabla}_X Y = X(Y^i)\partial_i \implies \overline{\nabla}_X Y - \overline{\nabla}_Y X = (X(Y^i) - Y(X^i))\partial_i$$

and that is exactly $[X, Y]$ because either X hits f or X hits the coefficients of Y; when X hits the coefficients of Y that gives us the $X(Y^i)$ term, and hitting f gives us the previously mentioned second derivative terms. So the flat connection is symmetric. And interestingly this has nothing to do with the metric, so it's a bit odd that we need an extra property of this type to pin down the correct connection.

Note that for any coordinate vector fields we have $[\partial_i, \partial_j] = 0$, so torsion-free is equivalent to $\Gamma_{ij}^k = \Gamma_{ji}^k$ (this is the reason for the name "symmetric"). However, if we choose a local frame, the connection coefficients with respect to those frames may not be symmetric, and not all frames come from coordinate frames. (For example, choose any vector fields which locally span the tangent space and don't commute.)

Proposition 56

The affine connection ∇^T is torsion-free.

Proof. Suppose $X, Y \in \mathfrak{X}(M)$ extend to $\tilde{X}, \tilde{Y}$ on $\mathfrak{X}(\mathbb{R}^N)$. It is a fact that $[\tilde{X}, \tilde{Y}]$ extends $[X, Y]$ as well (the easiest way is to use slice coordinates for the slice manifold, since locally a manifold is a linear subspace), so

$$\nabla^T_X Y - \nabla^T_Y X = (\overline{\nabla}_{\tilde{X}} \tilde{Y} - \overline{\nabla}_{\tilde{Y}} \tilde{X})^T = ([\tilde{X}, \tilde{Y}]_{\mathbb{R}^N})^T$$

since the projection is linear and the flat connection is torsion-free, and then $[\tilde{X}, \tilde{Y}]$ is already $[X, Y]$ without any projection required. □

So we're ready now to define the Levi-Civita connection (this is one person) – like in many other fields, this is a "fundamental theorem" which is not actually that important:

Theorem 57 (Fundamental theorem of Riemannian geometry)
For any Riemannian metric (M, g), there exists a unique affine connection ∇, called the **Levi-Civita connection**, which is metric-compatible and torsion-free.

In particular, we've seen that if (M, g) arises as an embedded submanifold of $\mathbb{R}^N$, then the Levi-Civita connection is the tangential connection. So this gives us the (nontrivial) observation that indeed the embedding does not matter.

Proof. Like in many other proofs, we do uniqueness first to find a formula and see that this also gives existence. For any three vector fields, we have $X\langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle$ by metric-compatibility, and then we can rewrite this as

$$X\langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_Z X\rangle + \langle Y, [X, Z]\rangle.$$

by being torsion-free. But now cyclically permuting X, Y, Z, we also get

$$Y\langle Z, X\rangle = \langle \nabla_Y Z, X\rangle + \langle Z, \nabla_X Y\rangle + \langle Z, [Y, X]\rangle,$$

and also

$$Z\langle X, Y\rangle = \langle \nabla_Z X, Y\rangle + \langle X, \nabla_Y Z\rangle + \langle X, [Z, Y]\rangle.$$

Now adding the first two and subtracting the last, and using symmetry of the metric, we find that (because red and blue terms cancel)

$$\boxed{X\langle Y, Z\rangle + Y\langle Z, X\rangle - Z\langle X, Y\rangle = 2\langle \nabla_X Y, Z\rangle + \langle Y, [X, Z]\rangle + \langle Z, [Y, X]\rangle - \langle X, [Z, Y]\rangle}.$$

But now we've proven uniqueness, since we can compute the value of all other terms besides the connection one. And additionally this also shows existence, because we get a formula for $\langle \nabla_X Y, Z\rangle$, and we can check the product rule to see that it is actually a connection. □

8 April 19, 2023

Last time, we calculated a formula for the Levi-Civita connection, which does indeed define a "canonical" connection on any Riemannian manifold (M, g). We can now compute the coefficients of this connection and how they relate to the metric: letting coordinates for X, Y, Z be $\partial_i, \partial_j, \partial_\ell$ yields

$$2\Gamma^m_{ij} g_{m\ell} = \partial_i g_{j\ell} + \partial_j g_{\ell i} - \partial_\ell g_{ij}$$

(the Lie bracket terms disappear because we're in **coordinates**), and now if we divide by $\frac{1}{2}$ and multiply by the inverse matrix $g^{\ell k}$ we get formulas for the connection coefficients

$$\boxed{\Gamma_{ij}^k = \frac{1}{2} g^{\ell k}(\partial_i g_{j\ell} + \partial_j g_{\ell i} - \partial_\ell g_{ij})},$$

and these are called the **Christoffel symbols**. We can check metric compatibility and torsion-freeness for these as well.

From now on, we'll "never speak of another connection" – there's basically no use for any connection besides the Levi-Civita connection for the purposes of this class, and if not specified that's what ∇ will be. (Other connections are studied, such as the Chern connection on the complex manifold or arbitrary connections in Yang-Mills theory, though.)

Proposition 58

Let (M, g) be an oriented manifold. Then the volume form dV_g (a section of a subbundle of $(0, n)$-tensors) is parallel.

Proof. This computation can be very annoying if it's done in the wrong way, and here there's a useful trick to perform. We have

$$\nabla_v dV_g = D_t dV_g,$$

where we compute D_t via a curve γ with $\gamma(0) = p$ and $\gamma'(0) = v$. Remembering that we can take an orthonormal oriented basis of T_pM and parallel transport it along the **curve** γ (though not necessarily in a neighborhood of p), we have

$$1 = dV_g(E_1, \cdots, E_n) \implies 0 = \frac{d}{dt}(dV_g(E_1, \cdots, E_n)).$$

Then $\frac{d}{dt}$ satisfies the product rule even when extended to tensors, so

$$0 = (D_t dV_g)(E_1, \cdots, E_n)$$

with all other terms dropping out because the E_is are parallel. Then $D_t dV_g$ is an n-form along γ as well. □

Proposition 59

Musical isomorphisms commute with the Levi-Civita connection.

Proof. We'll just do the proof for vector fields – recall that using the metric, we can turn $Y \in \mathfrak{X}(M)$ into $Y^\flat \in \Omega^1(M)$. We may compute for any $Z \in \mathfrak{X}(M)$ using the product rule that

$$\begin{aligned}(\nabla_X(Y^\flat))(Z) &= X(Y^\flat(Z)) - Y^\flat(\nabla_X(Z)) \\ &= X\langle Y, Z\rangle - \langle \nabla_X Z, Y\rangle\end{aligned}$$

by definition of the musical isomorphism, and then this last expression is $\langle \nabla_X Y, Z\rangle = (\nabla_X Y)^\flat(Z)$. Because Z is arbitary, this means that $(\nabla_X Y)^\flat = \nabla_X(Y^\flat)$. □

(We can also try proving this with the semicolon notation in physics; the proof is basically the same but will look very different.) We'll next talk about a canonical construction we can do on a Riemannian manifold: recall that we can construct a maximally extended geodesic $\gamma_v(t)$ such that $\gamma_v(0) = p$ and $\gamma_v'(0) = v$.

Lemma 60

For all c such that the following equation holds, we have

$$\gamma_{cv}(t) = \gamma_v(ct).$$

The simplest example where geodesics don't always exist for all time is when we work on $\mathbb{R}^2 \setminus \{0\}$ and have the geodesics move radially inward. (And there's much worse things than deleting a point that could go on, too.)

Proof. Define $\tilde{\gamma}(t) = \gamma_v(ct)$, so that $\tilde{\gamma}(0) = p$ and $\tilde{\gamma}'(0) = cv$ by the chain rule. So now we just need to check that $\tilde{\gamma}$ is actually a geodesic. We know that the D_t operator depends on the curve, so we can't use the same one as we did for γ, and our goal is to show that $\tilde{D}_t\tilde{\gamma}' = 0$. We'll check this in a coordinate chart: we have

$$\tilde{D}_t\tilde{\gamma}' = \left(\frac{d}{dt}\dot{\tilde{\gamma}}^k + \Gamma_{ij}^k(\tilde{\gamma}(t))\dot{\tilde{\gamma}}^i\dot{\tilde{\gamma}}^j\right)\partial_k,$$

and now because $\dot{\tilde{\gamma}}(t) = c\dot{\gamma}_v(ct)$, we can substitute in and find that this is equal to

$$= c^2\left(\ddot{\tilde{\gamma}}^k(ct) + \Gamma_{ij}^k(\tilde{\gamma}(ct))\dot{\tilde{\gamma}}^i\dot{\tilde{\gamma}}^j\right)\partial_k,$$

and now everything here except the c^2 prefactor is just $D_t\gamma'$ at the point ct, which is zero. (by extending ∂_k and then using the product rule) $\square$

So the idea is that **scaling the velocity is just a reparameterization** of the geodesic.

Definition 61

Define the subset of the tangent space

$$\mathcal{E} = \{v \in TM : \text{the maximal geodesic } \gamma_v \text{ through } p \text{ with velocity } v \text{ exists at least for } t \in [0, 1]\},$$

and define for any $v \in \mathcal{E}$ the **exponential map** $\exp(v) = \exp_p(v) = \gamma_v(1)$.

This exponential map turns out to be important, even though it's just "flowing for time 1" (whenever that flow is possible).

Proposition 62

We have the following:

1. The set $\mathcal{E}$ is open in TM, and $\mathcal{E}_p = \mathcal{E} \cap T_pM$ is a star-shaped region around the origin $0 \in \mathcal{E}_p$ (meaning the whole line from the origin to any point is contained in the set).
2. For all $v \in T_pM$, we have $\gamma_v(t) = \exp(tv)$ if one side is defined.
3. The map $\exp : \mathcal{E} \to M$ is smooth.
4. For all $p \in M$, we have $d(\exp_p)_0 : T_0(T_pM) = T_pM \to T_pM$ is the identity map (where we're thinking of the tangent space to the tangent space as T_pM itself, since $\mathcal{E}_p$ is an open subset of T_pM).

Proof. Star-shapedness and part (2) follow from the rescaling lemma we just proved. The proofs for smoothness and openness take a bit of work, but we'll just sketch the main ideas: we can reformulate the ODE equation in coordinates as

$$\ddot{x}^k + \Gamma^k_{ij}(x(t))\dot{x}^i\dot{x}^j = 0 \implies \begin{cases} \dot{x}^i = v^i, \\ \dot{v}^k = -\Gamma^k_{ij}(x(t))v^i v^j. \end{cases}$$

(We've basically turned a nonlinear equation into a system of linear ones.) But this new variable v is the flow of a vector field on $U \times \mathbb{R}^n$, where U was the original chart. Thus we get a vector field on TM (independent of coordinates) – it's called the **geodesic flow**, and so the geodesic equation is nicer if we think of lifting the curve $(\gamma(t), \gamma'(t))$ from M to TM, under which we have flow of a nice vector field. Since this came from a coordinate-independent statement, it is reasonable that this is also coordinate-independent. Now the flow of a vector field is always defined on an open set and depends smoothly on initial conditions (details in the textbook), so in fact smoothness and openness are just ODE theory from here.

We'll check (4) more carefully: for any $v \in T_0T_pM = T_pM$, we can compute the differential by choosing a cleverly parameterized curve: we have

$$d(\exp_p)_0(v) = \left.\frac{d}{dt}\right|_{t=0} \exp_p(tv)$$

by the chain rule, and now $\exp_p(tv) = \gamma_v(t)$ by property (2) and this is $\frac{d}{dt}|_{t=0}\gamma_v(t) = v$ by definition of the curve γ_v. Thus we do have the identity map. □

Definition 63

A manifold (M, g) is **geodesically complete** if all maximal geodesics are defined on all of $\mathbb{R}$ (or equivalently, $\mathcal{E} = TM$).

Most interesting theorems in Riemannian geometry will require completeness as an assumption to avoid situations where we delete points. Remembering that we also defined a metric space structure on the set (M, g), we may ask whether the two definitions of completeness are equivalent – that will turn out to be true. But we won't assume completeness until we actually state it.

On a Riemannian manifold (M, g), since we have $d(\exp)_0 = \text{Id}$, the inverse function theorem tells us that we locally have a diffeomorphism near zero. So the other direction of this map is a coordinate chart map: we have U, V with $0 \in V \subseteq T_pM$ and $p \in U \subset M$ so that $\exp_p : V \to U$ is a diffeomorphism of open sets. In particular, if we shrink V so that it is star-shaped (this is our convention), we call U a **normal neighborhood**. Choosing an orthonormal basis for T_pM, we can identify it with $\mathbb{R}^n$ (and in particular the metric becomes the dot product). Thus $\exp^{-1}$ yields **normal coordinates** on U. (In words, we choose a basis for the tangent space T_pM, map it onto the manifold via geodesics, and then this identifies a small ball of the manifold with a small ball of coordinates.)

Normal coordinates are often used to do calculations – it's common that if we want to know some quantity that is coordinate-independent, we will be choosing normal coordinates near p.

Proposition 64

Let x^i be normal coordinates at p. Then the following hold:

1. $g_{ij}(p) = \delta_{ij}$,
2. For all $v^i\partial_i \in T_pM$, we have $\gamma_v(t) = (tv^1, \cdots, tv^n)$ as long as $\gamma_v(t) \in U$.
3. $\Gamma^k_{ij}(p) = 0$,
4. $\partial_k g_{ij}(p) = 0$.

Proof. The first property is clear because we chose an orthnormal basis to start and the differential $d(\exp)_0$ is the identity map. Property (2) follows because $\exp_p(tv) = \gamma_v(t)$ by the property (2) of Proposition 62, and the coordinates of this point are exactly $(tv^1, \cdots, tv^n)$ in normal coordinates. For (3), we know that in general coordinates the geodesic equation is $\ddot{x}^k + \Gamma_{ij}^k \dot{x}^i(t)\dot{x}^j(t) = 0$, but taking $x^i = tv^i$ (which is a geodesic) and evaluating at $t = 0$; we then have $\dot{\Gamma}_{ij}^k(0)v^i v^j = 0$, and since this is a symmetric bilinear form which vanishes the coefficients must vanish. (However, note that straight lines not through the origin may not be geodesics and in general are not.) Finally, for (4), the derivatives of the metric at p vanish, since

$$\partial_k(g_{ij}) = \partial_k\langle \partial_i, \partial_j\rangle = \langle \nabla_{\partial_k}\partial_i, \partial_j\rangle + \langle \partial_i, \nabla_{\partial_k}\partial_j\rangle$$

by metric compatibility, but now $\nabla_{\partial_k}\partial_i = 0$ at p because the connection coefficients are zero there. □

So the fact that $\Gamma_{ij}^k(p)$ is not zero but that we can choose coordinates to make it zero at a certain point shows that it's not tensorial in nature, but it explains why normal coordinates are nice. (However, the second derivatives may not vanish in general just because the first derivatives do.)

9 April 21, 2023

Now that we've developed a bit of theory, we can now determine the geodesics on model spaces. The idea is that any geodesic can be formed by taking a point and a starting velocity and maximally extending to a curve. For example, straight lines are all of the geodesics in $\mathbb{R}^n$ (specifically parameterized lines with constant speed), and on a sphere of radius R the geodesics are intersections with any 2-planes in $\mathbb{R}^{n+1}$; those circles are geodesics if traversed at constant speed. But instead of checking that, we can use the following useful fact: if $\gamma \to M \subset \mathbb{R}^N$ is a curve and $V \in \mathfrak{X}(\gamma)$ is a vector field, we can think of V as a map from the interval I to $\mathbb{R}^N$ (via the parameterization). So if V extends to M (that is, if it's a restriction of a vector field) then $D_t V = \nabla_{\dot{\gamma}} V$, and the Levi-Civita connection is the same as the tangential connection. So this is $(\overline{\nabla}_{\dot{\gamma}} V)^T$, which is the directional derivative in the direction $\dot{\gamma}$ or equivalently $\left(\frac{d}{dt}V\right)^T$, where we view V as a map in $\mathbb{R}^N$. And even if we don't extend to M we can use the usual trick that local coordinates extend, so it is true in general that $D_t V = \left(\frac{d}{dt}V\right)^T$ for an embedded submanifold. And for any such circle the tangential acceleration is indeed zero, so these are the correct geodesics.

Fact 65

For the hyperbolic plane, one fact that we may have heard is that in the Poincaré disk and half-plane models we have circular arcs or lines that intersect the boundary perpendicularly. But in the Beltrami-Klein model in fact the geodesics are straight ines, which is why it's an interesting metric.

We'll now move on to talking about how geodesics relate to length – we've mentioned previously that geodesics should generalize straight lines, which have the important property that they are the shortest path between two points. Basically everything we've done so far work so far if we have a Lorentzian manifold (where we have a nondegenerate bilinear form with one negative eigenvalue), but when we have length things become more confusing and complicated in the Lorentzian setting. So most of what we'll from now on will be limited to the Riemannian case.

Recall that admissible curves are continuous, piecewise smooth, and have nonzero derivative everywhere (including the one-sided limits to the "rough" points).

Definition 66

An admissible curve γ is **length-minimizing** if $L_g(\gamma) = \int_I |\gamma'| \le L_g(\tilde{\gamma})$ for any other curve $\tilde{\gamma}$ with the same endpoints.

It's common to study the properties of minimizers before we prove existence, and this makes use of the **calculus of variations** (a fancy way of looking at a functional). The idea is to "take a derivative and set it equal to zero" but in a slightly different way, by wiggling our curve and seeing how that changes the result. Let $I, J \subset \mathbb{R}$ be intervals, and suppose $\Gamma : J \times I \to M$ be a continuous one-parameter family of curves (so for any fixed value of J we have a curve in M, and those curves vary smoothly) – we often label the variables (s, t), so that we have curves $\Gamma_s(t) = \Gamma(s, t)$ (called the **main curves**) as well as $\Gamma_t(s) = \Gamma(s, t)$ (the **transverse curves**, explaining how the curve changes). If Γ is smooth, then we can define $\partial_t \Gamma(s, t) = (\Gamma_s)'(t)$ and similarly $\partial_s \Gamma(s, t) = (\Gamma_t)'(s)$, which are vector fields along Γ.

Definition 67

A family of curves $\Gamma : J \times I \to M$ is **admissible** if (1) it has domain $J \times [a, b]$ for some open interval J, (2) there is some partition of $[a, b]$ so that Γ is smooth on $J \times [a_{i-1}, a_i]$, and (3) $\Gamma_s(t)$ is admissible for all $s \in J$ (so the t-derivative does not vanish).

Note that $\partial_t \Gamma$ does not even need to be continuous (the derivative of an admissible curve can have jumps).

Definition 68

Let γ be a fixed curve. An admissible family Γ is a **variation** of γ if $\Gamma_0 = \gamma$, and it is a **proper variation** if Γ_s has fixed endpoints across all s.

So if we want to minimize length between two points, we think of all the ways we can wiggle away from γ while fixing the endpoints. We can then define the **variation field** $V(t) = \partial_s \Gamma(0, t)$, which will be piecewise smooth and thus in $\mathfrak{X}(\gamma)$. We then call a vector field in $\mathfrak{X}(\gamma)$ **proper** if $V(a) = V(b) = 0$. The reason we go into all of this detail is the following:

Lemma 69

Let γ be an admissible curve and $V \in \mathfrak{X}(\gamma)$. Then there exists Γ a variation of γ with variation field V, and Γ is proper if V is.

Proof. We may define $\gamma(s, t) = \exp_{\gamma(t)}(sV(t))$. In other words, we flow by the geodesic in the direction of V for time s. The exponential will not be defined for all s, but by compactness we can take s small enough so that this is defined. We must check that this induces V as the variation: for a fixed t, $\exp_{\gamma(t)}(sV(t))$ is a geodesic, specifically the geodesic starting at $\gamma(t)$ with starting velocity $V(t)$. Thus $\partial_s T|_{s=0} = V(t)$. ☐

Lemma 70 (Symmetry lemma)

Let $\Gamma : J \times [a, b] \to M$ be admissible. Then $D_s \partial_t \Gamma = D_t \gamma_s \Gamma$ on $J \times [a_{i-1}, a_i]$ (where Γ is smooth).

There's a few different ways to set up all of this – we kind of pulled D_t out of thin air, but another way to think of it is that $D_s \partial_t \Gamma$ is the connection induced on the pullback bundle on the curve. But we'll stick with our approach for now. (It's common to write $\partial_t \Gamma$ as T and $\partial_s \Gamma$ as S.)

Proof. Fix local coordinates, so that $\Gamma(s,t) = (x^1(s,t), \cdots, x^n(s,t))$. Then $\partial_t \Gamma = \partial_t x^i \partial_i$ and $\partial_s \Gamma = \partial_s x^j \partial_j$, so that (we're using the product rule, since $\partial_t x^i$ is the coefficient)

$$D_s \partial_t \Gamma = \partial_s \partial_t x^i \partial_i + \partial_t x^i \partial_s x^j \Gamma_{ij}^k \partial_k = (\partial_s \partial_t x^k + \partial_t x^i \partial_s x^j \Gamma_{ij}^k) \partial_k.$$

Similarly,

$$D_t \partial_s \Gamma = (\partial_t \partial_s x^k + \partial_j x^i \partial_t x^j \Gamma_{ij}^k) \partial_k.$$

Now the first terms agree because mixed partials of coordinates commute, and then when we swap the role of i and j in the second term we can use that the (Levi-Civita) connection is symmetric, so the two expressions are the same. ☐

We're now ready to prove an "interesting theorem:"

Theorem 71 (First variation)
Let $\gamma : [a,b] \to M$ be an admissible curve parameterized at unit speed (length is parameterization-independent). Let Γ be an admissible variation of γ with variation field V. Then the map $s \mapsto L_g(\Gamma_s)$ is smooth, and

$$\left.\frac{d}{ds}\right|_{s=0} L_g(\Gamma_s) = -\int_a^b \langle V, D_t \gamma \rangle - \sum_{i=1}^{k-1} \langle V(a_i), \Delta_i \gamma'(a_i) \rangle + \langle V(b), \gamma'(b) \rangle - \langle V(a), \gamma'(a) \rangle,$$

where $\Delta_i \gamma'(a_i) = \gamma'(a_i^+) - \gamma'(a_i^-)$.

We're basically computing the differential of the length $L_g(\Gamma_s)$ by taking a curve and then taking the derivative of the functional along that curve. We know that the derivative does not depend on the curve we took (by the chain rule), so it makes sense that the formula just involves V and not Γ.

Proof. We have Γ smooth on each $[a_{i-1}, a_i]$, so $L_g(\Gamma_s|_{[a_{i-1},a_i]})$ is smooth. We then have

$$\frac{d}{ds} L_g(\Gamma_s|_{[a_{i-1},a_i]}) = \int_{a_{i-1}}^{a_i} \frac{\partial}{\partial s} \langle T, T \rangle^{1/2} dt,$$

and now by compatibility with the metric (Proposition 53), we can use the product rule:

$$= \int_{a_{i-1}}^{a_i} \frac{1}{2} |T|^{-1/2} 2 \langle D_s T, T \rangle dt.$$

By the symmetry lemma, we may rewrite this as

$$= \int_{a_{i-1}}^{a_i} |T|^{-1/2} \langle D_t S, T \rangle dt.$$

Setting $s = 0$ and substituting in some of our definitions, we have

$$\left.\frac{d}{ds}\right|_{s=0} L_g(\Gamma_s|_{[a_{i-1},a_i]}) = \int_{a_{i-1}}^{a_i} \langle D_t V, \gamma' \rangle dt,$$

and now we want to use integration by parts. We know that $\frac{d}{dt}\langle V, \gamma' \rangle = \langle D_t, V\gamma' \rangle + \langle V, D_t \gamma' \rangle$, and now we can actually integrate the left-hand side: this simplifies to $\langle V(a_i), \gamma'(a_i) \rangle - \langle V(a_{i-1}), \gamma'(a_{i-1}) \rangle - \int_{a_{i-1}}^{a_i} \langle V, D_t \gamma' \rangle$. So then summing this up over all i yields the result. ☐

Theorem 72
Any length minimizing admissible curve is a geodesic if parameterized at unit speed.

(The other direction is false in general: on the sphere, we could go the long way around the equator which would not be length-minimizing.)

Proof. First we can show that our curve is a broken geodesic (that is, we can break it up into pieces where it's a geodesic on each part). Indeed, on an interval we can take a bump function ϕ supported away from the endpoints, and use $V = \phi D_t\gamma'$ in the first variation. (We need to be careful with this, but we've proven that for any vector field we can find an admissible variation, and that variation is proper because V vanishes at the endpoints.) Then all of the boundary terms vanish, and we have

$$\left.\frac{d}{ds}\right|_{s=0} L(\Gamma_s) = -\int_{a_{i-1}}^{a_i} \phi|D_t\gamma'|^2,$$

but now the left-hand side must be zero if Γ_s is length-minimizing. So the right-hand side is zero for all ϕ supported in the interval, which means $D_t\gamma'$ vanishes identically. Next, to show that the curve is actually unbroken, we can let V be supported near a break point a_i and apply the first variation again (now using that we know $D_t\gamma' = 0$ almost everywhere) yields $0 = -\langle V(a_i), \Delta\gamma'(a_i)\rangle$. Since we can choose $V(a_i)$ to be anything, this means $\Delta\gamma'(a_i) = 0$. Now the piecewise derivatives match up, which only tells us C^1, but because the curve solves an ODE we actually get smoothness. □

There's a geometric way of thinking about this as well – if we are not a geodesic, we can move in the direction of the acceleration to shorten the length. So rephrased, this proof shows that a unit speed curve is a **critical point** if and only if it is an unbroken geodesic (and that's why we don't necessarily need to minimize the length), and $D_t\gamma' = 0$ is an Euler-Lagrange equation. This is common in geometry – we often have a functional where we look at critical points, and we can read more about this in the textbook.

10 April 24, 2023

We've seen previously that great circles are geodesics on a sphere, and going the "long way" around a great circle can sometimes not result in minimal distance. But there's a result that shows that sufficiently short geodesics are shortest paths:

Definition 73

Suppose the exponential map $\exp_p : B_\varepsilon(0) \to M$ is a diffeomorphism onto the image (here $B_\varepsilon(0)$ makes sense because we have a metric on T_pM). Then we call $\exp_p(B_\varepsilon(0))$ the **open geodesic ball at p** of radius ε. Similarly $\exp_p(\overline{B_\varepsilon(0)})$ is the **closed geodesic ball at p** and $\exp_p(\partial B_\varepsilon(0))$ is the **geodesic sphere at p** if we have a diffeomorphism onto for some slightly larger ε.

We can always choose normal coordinates in a small neighborhood of p and define the **radial function** $r = \sqrt{(x^1)^2 + \cdots + (x^n)^2}$ and the **radial vector field** $\partial_r = \frac{x^i}{r}\partial_{x^i}$.

Theorem 74 (Gauss lemma)

In normal coordinates, the radial vector field ∂_r is a unit vector orthogonal to the geodesic spheres, also denoted $S_\varepsilon(0)$, at any point in $U \setminus \{p\}$.

This may seem obvious, but note that in normal coordinates we don't know anything about the metric of g_{ij} except at the origin. So this does require some work to show. The point is that normal coordinates do actually come from the exponential map, so that tells us something in a whole neighborhood.

Proof. Consider the curve $\gamma(t) = (tv^1, \cdots, tv^n)$ in normal coordinates, where $v \in T_pM$ is a unit vector. (This is defined for t sufficiently close to 0). Recall that this curve is a geodesic. Then

$$\partial_r|_{\gamma(t)} = \frac{tv^i}{r}\partial_{x^i} = \frac{tv^i}{t}\partial_{x^i} = v.$$

So in other words, we've trivialized the tangent bundle and basically gotten v back again at the origin, and this is equal to $\gamma'(t)$. But this is a unit vector, so ∂_r is a unit vector at each t. To show that this is perpendicular to geodesic spheres, look at the sphere of radius ρ, which we intersect at time ρ. We wish to show that if $w \in T_{\gamma(\rho)}\Sigma_\rho$ is any tangent vector, then $\langle \gamma'(\rho), w\rangle = 0$. Choose a curve σ contained in the sphere of radius ρ with $\sigma(0) = \gamma(\rho), \sigma'(0) = w$. This then yields a smooth family of curves via

$$\Gamma(s, t) = \left(\frac{t\sigma^1(s)}{\rho}, \quad , \frac{t\sigma^n(s)}{\rho}\right),$$

so that $\Gamma(0, t)$ is just the original curve γ. Let $S = \partial_s\Gamma$ and $T = \partial_t\Gamma$; notice that $S(0, 0) = 0$, $T(0, 0) = v$, $S(0, \rho) = \sigma'(0) = w$, and $T(0, \rho) = \partial_r = \gamma'(\rho)$. So now we can view $\langle S, T\rangle$ as a function fixing $s = 0$ and letting t vary along the curve γ:

$$\langle S, T\rangle = \begin{cases} 0 & t = 0, \\ \langle \partial_r, w\rangle & t = \rho. \end{cases}$$

Now by the product rule, we have

$$\frac{d}{dt}\langle S, T\rangle = \langle D_tS, T\rangle + \langle S, D_tT\rangle,$$

and the second term is zero because we have the derivative to the tangent of a geodesic. Now by the symmetry lemma, the first term is just $\langle D_sT, T\rangle = \frac{1}{2}\frac{d}{ds}|T|^2$, where this quantity means we go along the transverse curve $\Gamma^{(t)}$ and take the s-derivative. But each Γ_s is a unit-speed geodesic, so their tangent vector is a unit vector and thus $|T|^2$ is independent of s. Thus $\langle S, T\rangle$ is constant. (So notice that we do in fact need that we have the Levi-Civita connection here.) □

Corollary 75

With the definitions above, we have $\nabla r = \partial_r$ on $U \setminus \{p\}$.

Proof. The geodesic sphere is the same as the set $\{r = \text{constant}\}$, and $\partial_r \perp \{r = \text{constant}\}$. Thus ∂_r is parallel to the gradient because the gradient is orthogonal to the level sets (since $\langle \nabla f, X\rangle = X(f)$ but if X is in the direction where f is constant then this is zero). Now $\partial_r(r) = 1$ (from computing by taking the x^i derivatives and summing them up from the definition), so

$$1 = \partial_r(r) = dr(\partial r) = \langle \partial_r, \nabla r\rangle.$$

Now by the Gauss lemma ∂_r is a unit vector, and since ∇r is parallel to it this inner product can only be 1 if $\nabla r = \partial_r$. □

Proposition 76

Let $q \in B_\varepsilon(p)$. Then the radial geodesic is the unique length-minimizing curve from p to q. (Here the geodesic is the curve $\exp(tv)$ for $t \in [0, c]$, and then the length of the curve is c.)

Proof. Suppose σ is an admissible curve parameterized from p to q. Let t_0 be the last time $b(t_0) = p$ (we should throw away every thing because that was wasted length beforehand), and let b_0 be the next time after t_0 that the curve hits the geodesic sphere of radius ε. Now $r \circ \sigma$ is piecewise smooth on $[t_0, b_0]$, so we have

$$c = \int_{t_0}^{b_0} \sigma \circ r = \int_{t_0}^{b_0} \langle \nabla r, \sigma' \rangle \le \int_{t_0}^{b_0} |\sigma'| \le |b_0 - t_0| \le L(\sigma).$$

So in particular the minimum of all possible σs is indeed coming from the radial line because that's the only way all inequalities can be equalities. □

Corollary 77

On the open geodesic ball at p, we have $r(x) = d_g(x, p)$. (Note however that we cannot measure the distance between two abitrary points in the ball.) Also, open and closed geodesic balls are metric balls.

Proof sketch. We can write (and wish to prove)

$$\overline{B}_\varepsilon(p) = \{x \in M : d(x, p) < \varepsilon\}.$$

The inclusion $\subset$ is clear, and for the reverse inclusion, suppose we had some $x \in M$ with $d(x, p) \le \varepsilon$. But since ε is the distance required to get to the boundary, we cannot go any further if we're outside the ball $B_\varepsilon(p)$. □

Definition 78

The **injectivity radius** is defined via

$$\text{inj}(p) = \sup\{a : \exp_p : B_p(0) \to M \text{ is a diffeomorphism onto the image}\}.$$

It's useful to have a lower bound on the injectivity radius so that we can define coordinates in a large enough set.

Lemma 79

For any $p \in (M, g)$, there exists some $\delta > 0$ and some neighborhood U of p so that for every $q \in U$, we have $\text{inj}(q) > \delta$.

Proof sketch. Choose some normal coordinates based at p. The tangent space is then trivial, so $TU_0 = U_0 \times \mathbb{R}^n$ iin some neighborhood, and let $\mathcal{E} \subset U_0 \times \mathbb{R}^n$ be the tangent vectors (x, v) so that $\exp_x v \in U_0$. Now define a map

$$E : \mathcal{E} \to U_0 \times U_0, (x, v) \mapsto (x, \exp_x v).$$

(So this time we also keep track of how x varies, not just v.) Now the differential of E at p is (this is a map between Euclidean spaces)

$$(DE)_{(p,0)} = \begin{bmatrix} \text{Id} & * \\ 0 & \text{Id} \end{bmatrix}$$

where the blue Id comes from what we proved about the exponential map. So now we can use the inverse function theorem so that E is a diffeomorphism onto the image on a small enough neighborhood, and now we can shrink to a uniform ball. □

Corollary 80

If M is a compact Riemannian manifold, then the injectivity radius is bounded from below.

(Notice that we never proved anything about continuity, and that is in fact a subtle point.) We basically hit the injectivity radius for two reasons, either if two geodesics intersect (like going around the sphere) or if we start having weird "wiggling." And so it's a bit difficult to work with it – we can say more in a better way, and we will do that later.

We'll now think about the notions of completeness, showing that two of those definitions are actually equivalent. From now on, we'll have (M, g) be connected so that we can compute distances.

Lemma 81

Let p, q be two points and let $0 < \delta < \text{inj}(p), d(p, q)$ so that the sphere $S_\delta(p)$ separates p from q. Now let $y \in S_\delta(p)$ be some minimizer $d(\cdot, q)$ out of all points on the sphere (this exists because the distance function is minimum and the sphere is compact). Then $d(p, q) = d(p, y) + d(y, q)$.

Proof. One direction is obvious from the triangle inequality; the other asks us to show that y actually lies between p and q. Choose an admissible unit-speed curve σ; we know that it takes at least distance δ to get from p to $S_\delta(p)$, and then after that the remaining distance is at least $d(y, q)$, which is the reverse inequality. □

Lemma 82

Suppose we have some $p \in M$ such that the geodesics can be extended infinitely, meaning that $\exp_p : T_pM \to M$ is defined for all vectors in T_pM. Then the following hold:

1. For all $q \in M$, there is some minimizing geodesic from p to q (meaning that some curve attains the distance).
2. M is complete as a metric space.

Start of proof. We can only attain the minimum if we actually have a geodesic. Consider the sphere of radius ε around p such that q is outside the ball and such that we are within the injectivity radius. Choose x which is closest to q; now there is a unique radial geodesic γ from p to x because we're within the injectivity radius (by choosing normal coordinates on a bigger set). Now by assumption, we may extend this geodesic forever, and we claim that γ will run into q. Let $T = d(p, q)$ and define

$$\mathcal{A} = \left\{ b \in [0, T] : \gamma|_{[0,b]} \text{ minimizing}, d(p, q) = d(p, \gamma(b)), d(\gamma(b), q) \right\}$$

where $\gamma|_{[0,b]}$ minimizing means that any other curve from $\gamma(0)$ to $\gamma(b)$ has at least that length b. Clearly this set is closed because the distance is continuous, and we can restate the condition as $d(p, \gamma(b)) = L(\gamma|_{[0,b]})$. We know that $\varepsilon \in \mathcal{A}$ by construction, and let $A = \sup \mathcal{A}$. If $A = T$ then we are done, since $d(p, q) = d(p, \gamma(T)) + d(\gamma(T), q)$. But $d(p, q) = T$ and $d(p, \gamma(T)) = T$ so $\gamma(T)$ and q must be the same point. Otherwise, $A < T$, and now we can repeat the same game starting from A: choose a small geodesic ball centered at $y = \gamma(A)$, let z be the point on the sphere closest to q, and form a curve by connecting p to y along γ, then y to z along a radial geodesic. We'll finish the rest of this argument and part (2) next time. □

11 April 26, 2023

We'll first finish up the proof from last time:

Proof continued. Recall that for (1), we take a small ball around p (smaller than the injectivity radius), choose x closest to q, and start a geodesic with the ray in that direction. We then use a continuity method: letting $\mathcal{A}$ be the set of b such that the geodesic is minimizing and such that $\gamma(b)$ lies in between p and q, we know that $\sup A \geq \varepsilon$ and we're done if $A = T$. The next step is to take a small ball starting from $\gamma(A)$ and let z be the closest point to q. Now

$$d(p,q) = d(p,y) + d(y,q) = d(p,y) + d(y,z) + d(z,q) \geq d(p,z) + d(z,q) \geq d(p,q)$$

first step because $y = \gamma(A)$ and $\mathcal{A}$ is closed, second step by Lemma 81, and the remaining steps by the triangle inequality. So we have equalities and thus $d(p,z) = d(p,y) + d(y,z)$, so the "broken curve" from p to y and then y to z minimizes length, so in fact this is an unbroken curve by the theorem of first variation. So in fact the shortest point z is actually $\gamma(A+\delta)$ and that contradicts A being maximal if $A < T$.

For (2), suppose we have a Cauchy sequence $\{x_i\}$ of points in M. In particular, for any $p \in M$ we have $d(x_i, p)$ bounded, so $x_i = \exp_p(v_i)$, where the norms $|v_i| = d(x_i, p)$ are also bounded (here we do use part (1)). Now by taking a subsequence we can assume $v_i \to v$, so because the exponential map is continuous we have $\exp_p(v_i) \to \exp_p(v)$. So some subsequence of the Cauchy sequence converges, and that's enough to show convergence overall. □

This leads us to a nice result that we may not expect:

Theorem 83 (Hopf-Rinow)
Let (M,g) be a connected Riemannian manifold. Then being metrically complete is equivalent to being geodesically complete (meaning that maximally geodesics exist for all time).

Proof. The previous result proves the reverse direction, so we just need to do the forward one. Suppose we have a unit-speed maximal geodesic defined on (a,b) for some finite $b < \infty$. Let t_i increase to b. Then $d(\gamma(t_i), \gamma(t_j))$ is bounded below using this curve by $\int_{t_i}^{t_j} |\gamma'| = |t_i - t_j|$, and in particular $\{\gamma(t_i)\}$ forms a Cauchy sequence, so by completeness it converges to some point $p \in M$. In fact this shows that $\gamma(t) \to p$ for the whole curve (not just along the sequence). But now we know there is a neighborhood around p such that $\text{inj}(q) \geq \delta$ for all $q \in U$ (by regarding the exponential map on a whole neighborhood of the zero section), so we can extend by δ. This glues together to the previous geodesic because we have an ODE, so we have contradicted maximality of b. The same works for a. (Note that we can't just immediately take an exponential neighborhood of p, since we need to say something about how the tangent vectors converge as $t \to b$.) □

There are some important facts about covers of Riemmanian manifolds (and pulling back the metric and completeness by this map), as well as free homotopy classes and minimizing length within that homotopy class. We'll skip those discussions for now but we should read about them in the textbook. Instead, we'll move on to **curvature**, which is the main object studied in Riemannian geometry (it's used to study some fundamental quantities like length). The definition is a bit hard to motivate, but we can read a discussion in the textbook and also see how things will fit together as time goes on. (As always, we use the Levi-Civita connection.)

Definition 84
Let (M,g) be a Riemannian manifold. The **curvature** is a map $R : \mathfrak{X}(M) \times \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ given by

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

Recall that the second covariant derivatives look like

$$\nabla^2_{X,Y} Z = \nabla_X(\nabla_Y Z) - \nabla_{\nabla_X Y} Z, \quad \nabla^2_{Y,X} Z = \nabla_Y(\nabla_X Z) - \nabla_{\nabla_Y X}(Z).$$

Since the connection is linear (in fact C^∞-linear) in the subscript argument, we can thus alternatively write

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

because our connection is torsion-free. We can check that the curvature vanishes on $\mathbb{R}^n$ (since covariant derivatives are basically directional derivatives and those should commute).

Lemma 85

The curvature R is a $(1, 3)$-tensor.

Proof. Looking at the initial formula, the expression is already tensorial in X and Y: indeed,

$$\begin{aligned} \nabla^2_{X,fY} Z &= \nabla_X(\nabla_{fY} Z) - \nabla_{\nabla_X(fY)} Z \\ &= \nabla_X(f \nabla_Y Z) - \nabla_{\nabla_X(fY)} Z \\ &= X(f)\nabla_Y Z + f \nabla_X \nabla_Y Z - \nabla_{(X(f)Y + f\nabla_X Y)} Z \end{aligned}$$

and now using C^∞-linearity in the last term gives us $f\nabla^2_{X,Y} Z$; basically the same thing works for X. So now we just need to check C^∞-linearity in Z. We have

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

and now using the product rule on the first term yields $X(Y(f))Z + Y(f)\nabla_X Z + X(f)\nabla_Y Z + f\nabla_X\nabla_Y Z$; the middle two terms cancel out with the version for the second term, the first term (along with $-Y(X(f))Z$) cancels out with the $[X, Y](f)Z$ term, and then the remaining terms left indeed simplify to our calculation for $fR(X, Y)Z$ above. $\square$

We call the map $Z \mapsto R(X, Y)Z$ the **curvature endomorphism**, and in local coordinates we write

$$R = R_{ijk}{}^{\ell} dx^i \otimes dx^j \otimes dx^k \otimes \partial_\ell.$$

It's left as an exercise to us to check that

$$R_{ijk}{}^{\ell} = \partial_i \Gamma^\ell_{jk} - \partial_j \Gamma^\ell_{ik} + \Gamma^m_{jk}\Gamma^\ell_{im} - \Gamma^m_{ik}\Gamma^\ell_{jm}.$$

This means that given a metric, we can compute the Christoffel symbols, and then we can calculate the curvature from that. Note that Γ involves the derivative of the metric, so this curvature has nonlinearity depending on second derivatives of the metric. So if we're solving PDEs depending on the metric this is important, but that's only if we're an analyst.

Proposition 86

Let Γ be a smooth 1-parameter family of curves (not the same as the Christoffel symbols), and V is a smooth vector field along Γ. Then $D_s D_t V - D_t D_s V = R(\partial_s \Gamma, \partial_t \Gamma)V$.

Proof. We write $V = V^j \partial_j$ (we could have claimed that we could have reduced to just the case where $V = \partial_j$, but we actually need to write it out because there are various derivatives). We have

$$D_t V = \partial_t V^j \partial_j + V^j D_t \partial_j \implies D_s D_t V = \partial_s \partial_t V^j \partial_j + \partial_t V^j D_s \partial_j + \partial_s V^j D_t \partial_j + V^j D_s D_t \partial_j.$$

Everything except the last term together is symmetric in s and t, so $D_sD_tV - D_tD_sV$ is just the difference of the last terms, which are

$$D_sD_tV - D_tD_sV = V^j(D_sD_t\partial_j - D_tD_s\partial_j),$$

and so the tensoriality required does hold. Now if $\Gamma(s,t)$ is written as $(x^1(s,t), \cdots, x^n(s,t))$ in coordinates, we have

$$D_t\partial_j = \nabla_{\partial_t\Gamma}\partial_j = \partial_t x^i \nabla_{\partial_i}\partial_j$$

because $\partial_t\Gamma = \partial_t x^i \partial_i$ and then we use C^∞-linearity. So

$$D_sD_t\partial_j = \partial_s\partial_t x^i \nabla_{\partial_i}\partial_j + \partial_t x^i \partial_s x^k \nabla_{\partial_k}\nabla_{\partial i}\partial_j,$$

so subtracting this from the other term after relabeling i and k in one case yields

$$\begin{aligned}(D_sD_t\partial_j - D_tD_s\partial_j) &= \partial_t x^i \partial_s x^k \left(\nabla_{\partial_k}\nabla_{\partial_i}\partial_j - \nabla_{\partial_i}\nabla_{\partial_k}\partial_j\right)\\ &= \partial_t x^i \partial_s x^k R(\partial_k, \partial_i)\partial_j,\end{aligned}$$

where we use that the commutator term $\nabla_{[\partial_k,\partial_i]}\partial_j$ is zero because mixed partials commute for coordinates. So by C^∞-linearity we can now bring ∂x^i, $\partial_s x^k$, and the V^j term from before in and this yields the right-hand side. ☐

We can also lower the final index and define the **curvature tensor**

$$\text{Rm}(X, Y, Z, W) = \langle R(X, Y)Z, W\rangle.$$

Fact 87

Note that the signs of R and Rm are not "geometric" – there's a choice of sign freely in both $R(X, Y)Z$ and also separately in $\text{Rm}(X, Y, Z, W)$, so we always have to be careful with the convention.

Theorem 88

A manifold is flat (meaning it is locally isometric to Euclidean space) if and only if the curvature tensor Rm vanishes.

Proof. The forward direction can just be shown in coordinates in which the Christoffel symbols all vanish; we'll do the reverse direction. Suppose $\text{Rm} = 0$. We want to find coordinates so that $g_{ij} = \delta_{ij}$ **everywhere** in a neighborhood, and those are the coordinates that give us the isometry. We do this by starting from a point $p \in M$ and letting $\{b_1, \cdots, b_n\}$ be an orthonormal basis of T_pM. Choose some coordinates x^i near p. For each b_i, extend it along the x^1 axis by parallel transport, then extend it along the x^2 axis by parallel transport to get something in the x^1x^2-plane, and so on. This defines a vector field V on the whole box, and smoothness is not a problem by smooth dependence on initial conditions. We claim that $\nabla V = 0$. This is done by induction; we'll just check it on the x^1x^2-plane. We have $\nabla_{\partial_1}V = 0$ on the x^1-axis and $\nabla_{\partial_2}V = 0$ on the entire plane, and we know that (again coordinate vectors have commutator vanishing)

$$\nabla_{\partial_2}\nabla_{\partial_1}V - \nabla_{\partial_1}\nabla_{\partial_2}V = R(\partial_2, \partial_1 V) = 0,$$

but the second term is zero so $\nabla_{\partial_2}\nabla_{\partial_1}V = 0$ everywhere and thus $\nabla_{\partial_1 V}$ is parallel along x^2-lines. But then the parallel transport of the zero vector is the zero vector, so $\nabla_{\partial_1}V = 0$ everywhere.

Now we have parallel vector fields $E_1, \cdots, E_n$ on U, and at p they are orthonormal so by metric compatibility we have orthonormality everywhere. Then $[E_i, E_j] = \nabla_{E_i} E_j - \nabla_{E_j} E_i = 0$ because we are torsion-free. Then commuting vector fields that span everywhere are coordinate vector fields, because we can just integrate them. $\square$

Our definition of the curvature tensor is rather complicated because it has four arguments, but there will be quite a bit of symmetry that comes up from the algebraic definition. And soon we'll be able to actually say "what curvature is."

12 April 30, 2023

I was not able to attend this lecture in person, so instead these notes are transcribed from the course lecture notes.

Theorem 89

The curvature tensor satisfies the following symmetries:

1. $\operatorname{Rm}(W, X, Y, Z) = -\operatorname{Rm}(X, W, Y, Z)$,
2. $\operatorname{Rm}(W, X, Y, Z) = -\operatorname{Rm}(W, X, Z, Y)$,
3. $\operatorname{Rm}(W, X, Y, Z) = \operatorname{Rm}(Y, Z, W, X)$,
4. (First Bianchi identity) $\operatorname{Rm}(W, X, Y, Z) + \operatorname{Rm}(X, Y, W, Z) + \operatorname{Rm}(Y, W, X, Z) = 0$.

Proof. (1) is clear from the definition (since $R(X, Y)$ is antisymmetric in X and Y). For (2), it suffices to show that $\operatorname{Rm}(W, X, Y, Y) = 0$, because then we can expand $\operatorname{Rm}(W, X, Y + Z, Y + Z) = 0$ by linearity to get the result. Indeed, compatibility with the metric tells us that (since $|Y|^2 = \langle Y, Y \rangle$)

$$\begin{aligned} WX|Y|^2 &= 2\langle \nabla_W \nabla_X Y, Y \rangle + 2\langle \nabla_X Y, \nabla_W Y \rangle, \\ XW|Y|^2 &= 2\langle \nabla_X \nabla_W Y, Y \rangle + 2\langle \nabla_X Y, \nabla_W Y \rangle, \\ [W, X]|Y|^2 &= 2\langle \nabla_{[W,X]} Y, Y \rangle, \end{aligned}$$

and now if we subtract the last two equations from the first we get the desired result (by the writing out of $R(W, X)$ from the definition).

Next, we prove (4) by showing that $R(W, X)Y + R(X, Y)W + R(Y, W)X = 0$ (because then applied to any Z yields the result). For this, we can reduce to the case where X, Y, Z are parallel vector fields **at** p, because R is a tensor and thus only depends on the values of X, Y, Z at p; this means we can use normal coordinates at p and write $X = X^i \partial_i$ in those coordinates (so that $\nabla \partial_i = 0$ at p as well). However this does not allow us to make terms like $\nabla_W \nabla_X Y$ go away. So now

$$R(W, X)Y + R(X, Y)W + R(Y, W)X = \nabla_W \nabla_X Y - \nabla_X \nabla_W Y + \nabla_X \nabla_Y W - \nabla_Y \nabla_X W + \nabla_Y \nabla_W X - \nabla_W \nabla_Y X,$$

and now factoring based on the first symbol in each term yields $\nabla_W [X, Y] + \nabla_X [Y, W] + \nabla_Y [W, X]$. But now because W is parallel, $\nabla_W [X, Y] = \nabla_{[X,Y]} W + [W, [X, Y]] = [W, [X, Y]]$ (the first term goes away), and we're left with $[W, [X, Y]] + [X, [Y, W]] + [Y, [W, X]] = 0$ by the Jacobi identity. Thus identity (4) holds. Finally, if we take the first Bianchi identity and cyclically permute the entries through (W, Z, Y, X) and add them up, most terms will cancel using (1) and (2), and we are left with $2\operatorname{Rm}(Y, W, X, Z) + 2\operatorname{Rm}(X, Z, W, Y) = 0$. Finally using (1) on the first term yields identity (3). $\square$

Proposition 90 (Differential Bianchi identity)
We have $\nabla \operatorname{Rm}(X, Y, Z, V, W) + \nabla \operatorname{Rm}(X, Y, V, W, Z) + \nabla \operatorname{Rm}(X, Y, W, Z, V) = 0$.

In coordinates, this can also be written as $R_{ijlk;m} + R_{ijmk;l} + R_{ijlm;k} = 0$.

Proof. Since this is a tensor identity, again this only depends on the values of the vector fields at a point p. Again using normal coordinates and letting each vector field have constant coefficients, they are all parallel at p and pairwise commute everywhere in a neighborhood, so

$$\nabla(X, Y, Z, V, W) = W \operatorname{Rm}(X, Y, Z, V) = W \operatorname{Rm}(Z, V, X, Y) = \nabla \operatorname{Rm}(Z, V, X, Y, W)$$

using the previous result. Thus we just need to show that

$$\nabla \operatorname{Rm}(Z, V, X, Y, W) + \operatorname{Rm}(V, W, X, Y, Z) + \nabla \operatorname{Rm}(W, Z, X, Y, V) = 0$$

by rearranging all three terms using similar methods. The first term simplifies as $W \operatorname{Rm}(Z, V, X, Y) = \langle \nabla_W \nabla_Z \nabla_V X - \nabla_W \nabla_V \nabla_Z X, Y \rangle$ because W is parallel, and the other terms are this but with W, Z, V cyclically permuted. So the thing we are pairing up with Y in the total inner product is the cyclic sum of $\nabla_W \nabla_Z \nabla_V X$, minus the cyclic sum of $\nabla_Z \nabla_W \nabla_V X$, which is

$$R(W, Z)\nabla_V X + R(V, W)\nabla_Z X + R(Z, V)\nabla_W X = 0$$

again because the vector fields are parallel and because the curvature endomorphism (that is, applying $R(W, Z)$ to a vector field) is a tensor and thus sends zero to zero. □

We may also do this computation without using normal coordinates, which will also work but be more tedious.

Lemma 91 (Ricci identities)
Let X, Y, Z be vector fields and ω be a 1-form. Then we have

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

and if we define $R(X, Y)^*\omega(Z) = \omega(R(X, Y)Z)$, then we have

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

In coordinates, these formulas read ${Z^i}_{;pq} - {Z^i}_{;qp} = {R_{pqm}}^i Z^m$ and $\omega_{j;pq} - \omega_{j;qp} = -{R_{pqj}}^m \omega_m$.

We should check the textbook to see analogs of this equations for higher-order tensors in place of Z or ω (we basically get a sum of $(k + \ell)$ terms in a (k, ℓ) tensor where we apply $R(X, Y)$ to each vector field and $R(X, Y)^*$ to each one-form).

Proof. The first identity is just the definition (and written here just to show similarity with the second identity). For the second identity, recall that $\nabla^2_{X,Y} T = \nabla_X(\nabla_Y T) - \nabla_{\nabla_X Y} T$ for any tensor T, so setting T to be ω and assuming that X, Y, Z are parallel vector fields at p (so that the second term disappears), we have

$$\nabla^2_{X,Y}\omega(Z) = \nabla_X(\nabla_Y \omega)(Z) = X(\nabla_Y \omega(Z)) = X(Y(\omega(Z)) - X(\omega(\nabla_Y Z))$$

where we use parallelness and the product rule for ∇_Y. And now this last expression is $X(Y(\omega(Z)) - \omega(\nabla^2_{X,Y}Z)$; if we subtract from it the corresponding term for $\nabla^2_{Y,X}\omega$, we end up with $[X,Y]\omega(Z) - \omega(R(X,Y)Z) = -R(X,Y)^*\omega$ (since $[X,Y] = 0$), as desired. □

13 May 1, 2023

We've now defined the curvature tensor and proven a few properties, but we still don't know anything about what it means. One way to study the curvature tensor is to try to study simpler things that pop out of it, and the following is the most famous one:

Definition 92

The **Ricci curvature** is defined via

$$\mathrm{Ric}(X,Y) = \mathrm{tr}(Z \mapsto R(Z,X)Y).$$

In coordinates, $\mathrm{Ric}_{ij} = R_{kij}{}^k$, and we can also use the metric to write (where E_i is an orthonormal basis)

$$\mathrm{Ric}(X,Y) = \sum_{i=1}^n \mathrm{Rm}(E_i, X, Y, E_i).$$

We can use our symmetry relations to rewrite this as $\sum \mathrm{Rm}(Y, E_i, E_i, X) = \sum \mathrm{Rm}(E_i, Y, X, E_i)$, so $\mathrm{Ric}(X,Y) = \mathrm{Ric}(Y,X)$. So it's at the least a symmetric bilinear form, which is more manageable than a $(1,3)$-tensor. And we can also trace this using the metric to get the **scalar curvature**

$$S = \mathrm{tr}_g \mathrm{Ric} = \sum_{j=1}^n \mathrm{Ric}(E_j, E_j) = \sum_{j=1}^n \mathrm{Rm}(E_i, E_j, E_j, E_i).$$

So going from Riemann to Ricci to scalar is losing information, but it's much simpler to understand a function than a more complicated tensor. The point is that Ricci curvature is actually a geometric quantity, so even if we don't agree with the definition of Rm (in terms of which sign we use) we should end up with the same definition at that level.

Example 93

The **Einstein equations** are of the form $\mathrm{Ric} = \lambda g$ for some constant $\lambda \in \mathbb{R}$, and this comes up in physics. This is in fact the Euler-Lagrange equation for the total integral of the scalar curvature

$$\mathcal{E}(g) = \int_M S_g dV_g.$$

This functional is actually hard to work with analytically – we get an infinite saddle at critical points, so there's no minimization that we can do.

We proved the differential Bianchi identity last time, and now we can contract it to get another one:

Proposition 94 (Contracted second Bianchi identity)

We have

$$\mathrm{tr}_g(\nabla \mathrm{Ric}) = \frac{1}{2} dS,$$

where the trace is of the first and third indices: that is, $R_{ij}{}^{;i} = \frac{1}{2} S_{;j}$.

Proof. Start with the equation from Proposition 90, and first contract the first and fourth entries to get from Rm to Ric and then contract the first and third index. Using that the trace commutes with the covariant derivative (by metric compatibility), and plugging in $X, V = E_i$ and $Y, W = E_j$, the first term simplifies to $\text{tr}_g \nabla \text{Ric}(Z)$. Remembering from the proof last time that we can use the same symmetries of the curvature tensor for Rm, the second term (after swapping the third and fourth term) then becomes $-dS(Z)$, and the last term (after swapping the first and second and also third and fourth term) becomes $\text{tr}_g \nabla \text{Ric}(Z)$ as well. Since this all sums to zero, we get the desired equation. □

Proposition 95 (Schur lemma)
Let M be a connected n-dimensional manifold for $n \geq 3$. Then if we satisfy the "generalized Einstein equation" $\text{Ric} = fg$ for some $f \in C^\infty(M)$, then in fact f must be a constant.

Proof. We have by the previous result that

$$\frac{1}{2} dS = \text{tr}_g(\nabla \text{Ric}) = \text{tr}_g(\nabla(fg)),$$

and because the trace of fg is just nf the left-hand side is $\frac{n}{2} df$. Then for the right-hand side, applying it to Z yields $\sum_{i=1}^n \nabla(fg)(E_i, Z, E_i) = df(Z)$ ("metric-compatibility means we can only take derivatives on f"), but the two sides can only be equal if $df = 0$. □

Unfortunately, scalar curvature is a bit difficult to talk about in a first course, so we'll return to the Ricci curvature. There are some algebraic complications coming from symmetry: if V is a real vector space, let $\mathcal{R}$ be the vector space of "algebraic curvature tensors" in $(V^*)^{\otimes 4}$ (that is, maps where $T(x, y, z, w)$ is antisymmetric in the first two and last two slots, symmetric when we swap (x, y) with (z, w), and such that cyclically permuting y, z, w yields zero). What we've proven previously is that Rm_p lives in $\mathcal{R}(T_pM)$ for each point p.

Proposition 96
If V is an n-dimensional space, then we have $\dim \mathcal{R} = \frac{n^2(n^2-1)}{12}$.

In particular, if $n = 1$ then $\dim \mathcal{R} = 0$, so there are no nonzero curvature tensors. (And indeed, all Riemannian 1-manifolds are flat, parameterized by arclength.) For $n = 2$ we have $\dim \mathcal{R} = 1$, and in this case we can actually explicitly write down an element of $\mathcal{R}$: for any inner product g on V, we can write down the **Kulkarni-Nomizu product**

$$g \owedge g(W, X, Y, Z) = 2(g(W, Z)g(X, Y) - g(W, Y)g(X, Z)).$$

and we can check that the properties do hold. (And this fact holds in all dimensions.) But the point is that $\mathcal{R}$ is then the span of $g \owedge g$.

Corollary 97
For any 2-dimensional manifold, the Riemannian curvature tensor can be written as $\text{Rm} = \frac{1}{4} S g \owedge g$ (since this is the constant required for both sides to have the same trace), and we also have $\text{Ric} = \frac{1}{2} S g$.

We'll now begin talking about **Riemannian submanifolds**, specifically moving towards sectional curvatures. Suppose we have a Riemannian submanifold $(M, g) \subset (\tilde{M}, \tilde{g})$ (immersion is often enough because we just need conditions locally, but for simplicity we'll assume we have a submanifold). For $X, Y \in \mathfrak{X}(M)$, we can extend them to vector fields

$\tilde{X}, \tilde{Y} \in \mathfrak{X}(\tilde{M})$, and we know that we can decompose the Levi-Civita connection as

$$\tilde{\nabla}_{\tilde{X}} \tilde{Y}\big|_{p \in M} = (\tilde{\nabla}_{\tilde{X}} \tilde{Y})^T + (\tilde{\nabla}_{\tilde{X}} \tilde{Y})^\perp.$$

The first term on the right-hand side is just $\nabla_X Y$ (we only proved this for a submanifold of $\mathbb{R}^n$, but in general it's a similar idea), and the second term (the perpendicular part) is called the **second fundamental form** $\mathrm{II}(X, Y)$ (the metric is the "first fundamental form").

Proposition 98

The second fundamental form is well-defined independent of extension, and it is C^∞-bilinear and symmetric.

Proof. For symmetry, notice that

$$(\tilde{\nabla}_{\tilde{X}} \tilde{Y})^\perp - (\tilde{\nabla}_{\tilde{Y}} \tilde{X})^\perp = [\tilde{X}, \tilde{Y}]^\perp,$$

and now if we have two vector fields tangent to M then their commutator is tangent as well (we can choose slice coordinates such that there are no derivatives in the normal direction), so the right-hand side is zero. But now this proves the other two properties, because C^∞-bilinearity and well-definedness are obvious for X from the definition (since we only care about the value at the point p), and we can swap the roles of X and Y and argue the same way for Y. □

So rewriting the definition, we have the "Gauss formula"

$$\tilde{\nabla}_{\tilde{X}} \tilde{Y} = \nabla_X Y + \mathrm{II}(X, Y).$$

Here $\mathrm{II}(X, Y)$ is like a tensor, but it's not valued in the tangent space (instead only taking values in the perpendicular part). Instead, we can always define the normal bundle NM of a submanifold, and then if we have a section $N \in \Gamma(NM)$, we can define an actual $(0, 2)$ tensor

$$\mathrm{II}_N(X, Y) = \langle \mathrm{II}(X, Y), N \rangle,$$

which is a symmetric bilinear form at each point in the usual sense. We then get an associated map $W_N : \mathfrak{X}(M) \to \mathfrak{X}(M)$ (called the **Weingarten map** in the direction of N) such that

$$\langle W_N(Y), X \rangle = \mathrm{II}_N(X, Y),$$

and this tensor is self-adjoint because it comes from a symmetric tensor and thus this is also equal to $\langle Y, W_N(X) \rangle$. We can also write similarly that if $X \in \mathfrak{X}(\gamma)$, then $\tilde{D}_t X = D_t X + \mathrm{II}(\gamma, X)$ (this is the same proof with using coordinate vector fields as usual).

Proposition 99 (Weingarten equation)

Suppose N extends to some vector field on $\tilde{M}$ (we will stop writing $\tilde{N}$ for this). Then (in particular not depending on the extension)

$$(\tilde{\nabla}_X N)^T = -W_N(X).$$

Proof. We have by the product rule that

$$\langle \tilde{\nabla}_X N, Y \rangle = X \langle N, Y \rangle - \langle N, \tilde{\nabla}_X Y \rangle.$$

Again we're extending all of our vector fields to $\tilde{M}$, and the point is that because N is not a tangent vector to M we cannot apply this all with ∇ instead of $\tilde{\nabla}$. On M we know that $\langle N, Y \rangle = 0$, and at the point p we care about, X is tangent so the first term disappears. And now the right-hand side has zero tangential part, and the right-hand side is $-\text{II}_N(X, Y) = -\langle W_N(X), Y\rangle$. $\square$

These formulas are actually relatively simple compared to the ones with intrinsic calculations (involving Christoffel symbols and so on) – we should do them with a sphere and Euclidean space if we're not so sure what is going on.

Proposition 100 (Gauss equation)
Let $W, X, Y, Z \in \mathfrak{X}(M)$ (and extend these vector fields to all of $\tilde{M}$). Then

$$\tilde{\text{Rm}}(W, X, Y, Z) = \text{Rm}(W, X, Y, Z) - \langle \text{II}(W, Z), \text{II}(X, Y)\rangle + \langle \text{II}(W, Y), \text{II}(X, Z)\rangle.$$

This formula is quite nice because for a hypersurface in Euclidean space the left-hand side is zero and thus we just need to compute "derivatives of the normal vector" to find $Rm(W, X, Y, Z)$, which is just a first-order calculation. We'll do the proof of this next time!

14 May 3, 2023

We'll start by proving the Gauss equation – the idea is to just compute and not think so hard.

Proof. Since this is a tensorial fact, it's really about vectors and not vector fields: we may extend W, X, Y, Z to pairwise commuting vector fields on $\tilde{M}$ that are always tangent to M. (Note that we cannot do this and **also** be $\tilde{g}$-parallel at a fixed point). Then we have

$$\tilde{\text{Rm}}(W, X, Y, Z) = \langle \tilde{\nabla}_W \tilde{\nabla}_X Y - \tilde{\nabla}_X \tilde{\nabla}_W Y, Z\rangle$$

because the commutator is zero (in fact both on M and on $\tilde{M}$), and now remembering that we can decompose $\tilde{\nabla}_X Y$ into the tangential part $\nabla_X Y$ and the other part $\text{II}(X, Y)$, this simplifies to

$$= \langle \tilde{\nabla}_W \nabla_X Y - \tilde{\nabla}_X \nabla_W Y, Z\rangle + \langle \tilde{\nabla}_W (\text{II}(X, Y)) - \tilde{\nabla}_X (\text{II}(W, Y)), Z\rangle.$$

Now looking at the first term that we have, $\nabla_X Y$ and $\nabla_W Y$ are tangent vector fields, so in fact this simplifies to $\langle \nabla_W \nabla_X Y - \nabla_X \nabla_W Y, Z\rangle = \text{Rm}(W, X, Y, Z)$ (because the normal part goes away when we take the inner product with Z). And now for the second term, $N = \text{II}(X, Y)$ is a section of the normal bundle $\Gamma(NM)$, so we want to compute

$$\langle \tilde{\nabla}_W N, Z\rangle = W\langle N, Z\rangle - \langle N, \tilde{\nabla}_W Z\rangle$$

(note that we can't use compatibility of the g-metric because N is not a tangent vector – what we're saying here is that $\tilde{g}$ and $\tilde{\nabla}$ are metric-compatible, so we should be putting $\tilde{\nabla}$ instead of ∇). But $\langle N, Z\rangle$ is identically zero, so the first term on the right vanishes, and now $\tilde{\nabla}_W Z = \nabla_W Z + \text{II}(W, Z)$ and pairing it with N only makes the second fundamental form part survive. Thus we are left with $-\langle N, \text{II}(W, Z)\rangle$, and so this part simplifies to $-\langle \text{II}(X, Y), \text{II}(W, Z)\rangle$. The other part $-\tilde{\nabla}_X(\text{II}(W, Y)), Z$ then similarly becomes $+\langle \text{II}(W, Y), \text{II}(X, Z)\rangle$, as desired. $\square$

So the moral is that if we know everything about $\tilde{g}$ and about the second fundamental form on $\tilde{M}$, we can deduce the Riemann curvature in all tangential directions on a submanifold. But we can't quite go in reverse, only in the directions along M.

Next, we can consider $\nabla^\perp : \mathfrak{X}(M) \times \Gamma(NM) \to \Gamma(NM)$, given by

$$\nabla^\perp_X N = (\tilde{\nabla}_X N)^\perp.$$

This will turn out to be a connection on the normal bundle (we just need to check the product rule but it follows from the product rule for $\tilde{\nabla}$). And the normal bundle has an inner product, and this connection is compatible with it in the sense that

$$\begin{aligned} X\langle N_1, N_2\rangle &= \langle \tilde{\nabla}_X N_1, N_2\rangle + \langle N_1, \tilde{\nabla}_X N_2\rangle \\ &= \langle \nabla^\perp_X N_1, N_2\rangle + \langle N_1, \nabla^\perp_X N_2\rangle. \end{aligned}$$

The next result tells us something about the normal component (in contrast to the Gauss equation):

Proposition 101 (Codazzi equations)
Let $W, X, Y \in \mathfrak{X}(M)$. Then

$$\begin{aligned} (\tilde{R}(W,X)Y)^\perp = \nabla^\perp_W(\mathrm{II}(X,Y)) - \mathrm{II}(\nabla_W X, Y) - \mathrm{II}(X, \nabla_W Y) \\ - \nabla^\perp_X(\mathrm{II}(W,Y)) + \mathrm{II}(\nabla_X W, Y) + \mathrm{II}(W, \nabla_X Y). \end{aligned}$$

Here, each row on the right-hand side is an expression for the "induced connection on the second fundamental form," which we can call $\nabla^\perp \mathrm{II}$. For example, this is saying that in Euclidean space, the second fundamental form cannot just be a general $(0,2)$ tensor – there are some restrictions. And in the hypersurface case, the Gauss and Codazzi equations are meant to be conditions for when we can make these expressions compatible.

Proof. As before, extend W, X, Y to $\tilde{M}$ and assume they commute everywhere. Fix a section N of the normal bundle $\Gamma(NM)$, and write down

$$\tilde{R}(W,X,Y,N) = \langle \tilde{\nabla}_W \tilde{\nabla}_X Y - \tilde{\nabla}_X \tilde{\nabla}_W Y, N\rangle + \langle \tilde{\nabla}_W \mathrm{II}(X,Y) - \tilde{\nabla}_X \mathrm{II}(W,Y), N\rangle$$

in the same way as the previous proof. But now we're done, because we're taking a normal component of the corresponding terms in the inner product:

$$= \langle \mathrm{II}(W, \nabla_X Y) - \mathrm{II}(X, \nabla_W Y), N\rangle + \langle \nabla^\perp_W(\mathrm{II}(X,Y)) - \nabla^\perp_X(\mathrm{II}(W,Y)), N\rangle.$$

And now because we assume W and X are parallel, this is exactly the expression that we wanted because the second and fifth terms cancel out when W and X commute. (But in general they are in the expression when we don't assume that things commute and they are necessary to have a tensorial relation.) □

We can now relate this back to some things that are more intuitive:

Definition 102
Consider a unit-speed curve $\gamma : I \to (M, g)$. Then define the **intrinsic curvature** $k = |D_t\gamma'|$ and the **extrinsic curvature** $\tilde{k} = |\tilde{D}_t\gamma'|$. (We call $D_t\gamma'$ the **acceleration** of the curve.)

So in Euclidean space, the curvature is the absolute value of the second derivative, which is the "osculating circle" definition. And if we're on a sphere, for a geodesic the intrinsic curvature vanishes (this is always true) but the extrinsic curvature will be something nonzero. And in fact we have a way of relating these in general:

Lemma 103
For any unit vector $v \in T_pM$, the second fundamental form $\mathrm{II}(v,v)$ is the acceleration of the g-geodesic through p, and so $|\mathrm{II}(v,v)|$ is the extrinsic curvature of the geodesic.

And because we have a bilinear form, understanding it only requires us to understand expressions of the form $\text{II}(v, v)$.

Proof. We have $\tilde{D}_t\gamma' = D_t\gamma' + \text{II}(\gamma', \gamma')$ by the Gauss formula (working in coordinates, the coordinate vector fields extend and we can use the analogous formula for the connection). □

Definition 104

A submanifold (M, g) is **totally geodesic** in $(\tilde{M}, \tilde{g})$ if $\tilde{g}$-geodesics with $\gamma'(t_0) \in T_{\gamma(t_0)}M$ remain in the submanifold for a short time (meaning $\gamma((t_0 - \varepsilon, t_0 + \varepsilon)) \subset M$ for some ε depending on the curve).

For example, the sphere is not totally geodesic in Euclidean space, because geodesics are straight lines tangent to the sphere. And a continuity / open-closed argument shows that if we're inside M for some short time, it'll be so for the maximal interval of the geodesic – the definition we gave is just easier to check.

Lemma 105

The following are equivalent:

1. M is totally geodesic in $\tilde{M}$,
2. g-geodesics are $\tilde{g}$-geodesics,
3. the second fundamental form II is identically zero.

Proof. For (1) implies (2), choose a $\tilde{g}$-geodesic tangent at our given point; we want to show that this is a g-geodesic. But by part (1), $\tilde{\gamma}$ must be in M for a short time, and then we can compute

$$\tilde{D}_t\tilde{\gamma}' = D_t\tilde{\gamma} + \text{II}(\tilde{\gamma}', \tilde{\gamma}'),$$

and now the left-hand side is zero because we are a $\tilde{g}$-geodesic. But the two terms on the right are tangential and normal along the curve, respectively, so they must be zero everywhere. So $D_t\tilde{\gamma}'$ vanishing means we have a g-geodesic, and since $\tilde{\gamma} = \gamma$ for a short time since they have the same initial conditions this shows our result.

Next, for (2) implies (3), we know that if γ is a g-geodesic, and thus a $\tilde{g}$-geodesic by assumption of (2), then $\tilde{D}_t\gamma' = D_t\gamma' + \text{II}(\gamma', \gamma')$; the first two terms in this equation are zero so the last one is as well. But we can always find a geodesic in any direction, so II must be zero identically.

Finally, for (3) implies (1), if we have a g-geodesic γ, then $\tilde{D}_t\gamma' = D_t\gamma' = 0$ so γ is also a $\tilde{g}$-geodesic. And now if we take a $\tilde{g}$-geodesic and a g-geodesic tangent at some point, them having the same initial conditions again means they are equal for a short time. □

Example 106

Intersecting S^3 with a plane through the origin yields a copy of S^2, and we can check that this S^2 is totally geodesic in S^3.

More generally, hypersurfaces (where M has codimension 1 in $\tilde{M}$) are the easiest case to study. We'll **assume** that there exists a unit normal vector field N (this is true locally, though globally not all normal bundles are trivial because we may have something like $\mathbb{RP}^2$).

Definition 107
The **scalar second fundamental form** is defined via

$$h(X, Y) = \mathrm{II}_N(X, Y) = \langle \mathrm{II}(X, Y), N \rangle.$$

(We should be a bit cautious here since there is sometimes a negative sign in the definition, so we have to look for conventions as usual.) Writing $s : \mathfrak{X}(M) \to \mathfrak{X}(M)$ for the Weingarten map, we can then write $h(X, Y) = \langle sX, Y \rangle$, and we can write the Gauss and Codazzi equations in terms of s and h, upon which they become easier to understand. (The idea is that knowing s, h, and N gives us everything, and we're generally working with simpler quantities this way.)

Remark 108. *If we now take a g-geodesic through some point p with velocity v, we have*

$$\tilde{D}_t \gamma' = D_t \gamma' + h(\gamma', \gamma') N,$$

and because $D_t \gamma' = 0$ the sign of $h(v, v)$ tells us about the direction of $\tilde{D}_t \gamma'$. The way to think about this is that "geodesics curve in the direction of the unit normal," but then that requires us to choose the inward-pointing normal on the sphere.

Definition 109
Since s is a self-adjoint linear map, it can be diagonalized, and the eigenvalues of s are called the **principal curvatures**. And from the principal curvatures we can define the **Gauss curvature** $K = k_1 \cdots k_n$ and the **mean curvature** $H = \frac{k_1 + \cdots + k_n}{n} = \frac{\mathrm{tr}(s)}{n}$. (However, note that most modern conventions don't use the factor of n.)

We'll now take the ambient space to be Euclidean space $\tilde{M} = |\mathbb{R}^{n+1}|$ to get a better handle of what's going on:

Proposition 110
Let $X : U \to \mathbb{R}^{n+1}$ be a local parameterization of a hypersurface (where U is an open subset of $\mathbb{R}^n$) with coordinates $(u_1, \cdots, u_n)$. Choose a unit normal $N \in \Gamma(NM)$, and let $X_i = \partial_i X$. Then

$$h(X_i, X_j) = \left\langle \frac{\partial^2 X}{\partial u^i \partial u^j}, N \right\rangle.$$

So the second fundamental form is the Hessian of our parameterization in this case.

Proof. Extend X_i, X_j locally. We have $0 = \langle X_j \langle X_i, N \rangle$ because we're differentiating the zero function by a tangent vector, which means that

$$0 = \langle \overline{\nabla}_{X_i} X_j, N \rangle + \langle X_j, \overline{\nabla}_{X_i} N \rangle$$

and now the second term is $-h(X_j, X_i)$ because $\overline{\nabla}_{X_i} N = -sX_i$. To compute the first quantity, note that $\overline{\nabla}_{X_i} X_j$ uses the Euclidean convention, and we can choose a curve so that $\gamma(t) = X(p + te_i)$. We have

$$X_j \circ \gamma = \frac{\partial X}{\partial u^j}(p + te_i) \implies (X_j \circ \gamma)'(0) = \frac{\partial^2 X}{\partial u^i \partial u^j}(p),$$

and plugging these back in yields the formula that we want. □

15 May 5, 2023

Last time, we showed a formula for how the scalar fundamental form relates back to the metric (yielding the Hessian of our parameterization). There's a follow-up result that helps us get a better understanding of principal curvatures (they're the "quadratic parts" near a point):

Corollary 111

Recall that the shape operator s has eigenvalues $k_1, \cdots, k_n$. Consider the origin in a hypersurface as a subset of $\mathbb{R}^{n+1}$. Then locally M is a graph of the equation $x^{n+1} = k_1(x^1)^2 + \cdots + k_n(x^n)^2 + O(|x|^3)$.

Proof. We have $X(u) = (u, f(u))$ locally near the origin. Taking the unit normal to be $N(p) = e_{n+1}$ and applying Proposition 110, we have $\frac{\partial^2 f}{\partial u^i \partial u^j} = \left\langle \frac{\partial^2 X}{\partial u^i \partial u^j}, N \right\rangle$, and then $h(X_i, X_j) = h(e_i, e_j)$ in these coordinates. Rotating so that our basis consists of eigenvectors, we get $h(X_i, X_j) = \delta_{ij} k_i$. □

Theorem 112 (Gauss's Theorem Egregium)

Let $M^2 \subset \mathbb{R}^3$ be a hypersurface. Then $K = \frac{1}{2}S$, so K is unchanged by local isometries (in other words, K is intrinsic to the surface).

Proof. Choose an orthonormal frame $b_1, b_2 \in T_pM$ that diagonalizes s, so that $h_{ij} = h(b_i, b_j) = k_i \delta_{ij}$. Then

$$\frac{1}{2}S = \mathrm{Rm}(b_1, b_2, b_2, b_1)$$

(because if we put b_1 in both spaces we get zero, and if we swap b_1 and b_2 we get the same thing), and by the Gauss equations this is equal to $h_{11}h_{22}$ (since we've chosen so that the remaining terms vanish), which is exactly $k_1 k_2$. □

Definition 113

The **Gaussian curvature** for a two-dimensional manifold (M, g) is $K = \frac{1}{2}S$. (We can also define the Gaussian curvature of a general hypersurface, but it will no longer be intrinsic.)

Generally the three main notions that are useful for studying are Ricci, scalar, and sectional curvature – we'll now define the last of these, which throws away the least information.

Definition 114

Let $p \in (M, g)$, and let $\Pi \subset T_pM$ be a linear subspace. Then for some small neighborhood V around zero, we have

$$S_\pi = \exp_p(\pi \cap V).$$

The **sectional curvature** $\sec(\Pi)$ is the Gaussian curvature K of S_π at p.

The restatement below is the actually "useful" version of sectional curvature, though:

Proposition 115

Let v, w form a basis for Π. Then

$$\sec(\Pi) = \frac{\text{Rm}_p(v, w, w, v)}{|v \wedge w|^2},$$

where $|v \cap w|^2 = |v|^2|w|^2 - \langle v, w\rangle^2$ is the norm on the wedge vector space.

Proof. We will assume v, w is an orthonormal basis for Π; we can check that the Gram-Schmidt process does indeed preserve the quantity above so the formula holds in general. By the Gauss equations, we have for any g-geodesic γ at p in direction of Π (which is also a geodesic for S_Π)

$$0 = D_t\gamma' = D_t^{S_\pi}\gamma' + \text{II}^{S_\pi}(\gamma', \gamma')|_p = 0,$$

and since $D_t^{S_\pi}\gamma' = 0$ the second term is also zero. And since $\text{Rm}(v, w, w, v) = \text{Rm}^{S_\pi}(v, w, w, v)$ **at** p (the second fundamental form terms are zero), this is equal to $\frac{1}{2}S$ by the same argument as before, and this is exactly the Gaussian curvature K. □

Fact 116

It turns out knowledge of $\sec(\Pi)$ for all p determines the Riemman curvature tensor uniquely, so we can recover anything from the sectional curvature. But it does involve a bit of permuting indices around, and the tensor itself is a sum of several terms of sectional curvature so may not be that enlightening.

Proposition 117

Let v_1 be a unit vector, completed to a basis $v_1, \cdots, v_n$. Then

$$\text{Ric}(v_1, v_1) = \sum_{i=2}^{n} \sec(v_1, v_i),$$

where the right-hand side means we consider the span of v_1 and v_i.

Proof. By definition, we have

$$\text{Ric}(v_1, v_1) = \text{Rm}(v_k, v, v, v_k),$$

but after tracing that exactly gives us the sectional curvature. Similarly, this also tells us that $S = \sum_{i \neq j} \sec(e_i, e_j)$. □

Theorem 118

$\mathbb{R}^n, S^n(R), \mathbb{H}^n(R)$ have sectional curvature 0, $\frac{1}{R^2}$, and $-\frac{1}{R^2}$ (constant everywhere), respectively.

Proof. We first compute the sectional curvature for S^2: a unit normal is $R^{-1}x$, so by the Weingarten equation

$$sX = -\overline{\nabla}_X N = -R^{-1}X,$$

so $s = -R^{-1}\text{Id}$ and thus $k_1, k_2 = -\frac{1}{R}$ and $K = \frac{1}{R^2}$. But now we're done because the intersection of a 2-plane with $S^n(R)$ is just $S^2(R)$, which we can count as S_π. And the same kind of argument works for $\mathbb{H}^n(R)$. □

Proposition 119

Having constant sectional curvature means that (in fact we determine the Riemann tensor)

$$R(v,w)x = c(\langle w,x\rangle v - \langle v,x\rangle w), \quad \text{Rm} = \frac{1}{2}cg \owedge g, \quad \text{Ric} = (n-1)cg, \quad S = n(n-1)c.$$

We'll now try to say something about curvature as we progress into the class, but we'll quickly mention some content from a chapter we're skipping over:

Theorem 120 (Gauss-Bonnet)

Let (M^2, g) be a compact oriented Riemannian manifold with no boundary. Then $\int K dV_g = 2\pi\chi(M)$, where $\chi(M)$ is the Euler characteristic of the manifold.

For example, if $M^2 \subset \mathbb{R}^3$ is a surface, then we can give a quick proof of this, because the left-hand side is basically the integral of $\det(dN)$, which is the signed area of $N(M)$. But for an abstract surface it takes quite a bit of work to prove, and we don't have much time to talk about it. The proof is not that enjoyable to read, but we can skim through it and see how the argument generally goes (chop up the manifold into triangles, derive a local version with boundary, and add it up). And we can see some foreshadowing of what's to come: if a surface has positive curvature $K > 0$ everywhere, then $\chi(M) > 0$ means we must have the sphere S^2, and if $K < 0$, then we must have a higher-genus surface (which is somehow "large"). We'll in fact prove some results about sectional and Ricci curvature in this direction later on.

We can now move on towards some more tools, specifically talking about **Jacobi fields**. We'll always assume that (M, g) is a **complete** Riemannian manifold (this is not always necessary, and the textbook is careful about this). If γ is a geodesic in (M, g), then we can think about variations of the curve, and we'll let Γ be a variation through **geodesics** (so that each Γ_s is still a geodesic). Since geodesics don't have break points, Γ will in fact be completely smooth. Then we can define the vector field induced by the variation

$$J = \partial_s \Gamma(0, t),$$

which is a vector field along γ.

Theorem 121 (Jacobi equation)

In the setting above, we have $D_t^2 J + R(J, \gamma')\gamma' = 0$.

Proof. Let $S = \partial_s\Gamma$ and $T = \partial_t\Gamma$. We have $D_s D_t T - D_t D_s T = R(\partial_s T, \partial_t \Gamma)T$ (we proved this by expanding in coordinates), and $D_t T = 0$ for all s because we have a family of varying geodesics. So then

$$0 = D_t D_s T + R(\partial_s\Gamma, \partial_t\Gamma)T = D_t D_t S + R(\partial_s\Gamma, \partial_t\Gamma)T = 0,$$

and now evaluating at $s = 0$ yields the result. □

The Jacobi equation is a **linear** second-order ODE in coordinates, unlike the geodesic equation, and since we can solve linear ODEs for all times (though we do have to argue as usual that the solutions overlap and so on) we get the following:

Proposition 122
Let $\gamma : I \to (M, g)$ be a geodesic and $v, w \in T_{\gamma(a)}M$. Then there exists a unique Jacobi field $J \in \mathfrak{X}(\gamma)$ (that is, something that solves the Jacob equations) such that $J(a) = v$ and $D_t J(a) = w$. In particular, letting $J(\gamma)$ be the set of Jacobi fields along γ, we have $\dim(J(\gamma)) = 2n$.

So morally, we've basically linearized the geodesic equation in the direction of deviation, and linearization is basically the derivative in some given direction.

Proposition 123
If J is a Jacobi field, then there exists a variation through the geodesic such that the variation vector field is J.

However, this is rather rare – as an exercise, we can construct an example of a map $\gamma : S^1 \to (M, g)$ (this is basically a "closed geodesic"), in which it makes sense to define a Jacobi field on this curve γ, but it will not come from an S^1-family of geodesics.

16 May 8, 2023

Last time, we stated that Jacobi fields on a geodesic arises from a variation **through geodesics**. We'll prove this fact now:

Proof. Let J be our Jacobi field. Choose any curve σ through $\gamma(0) = p$ such that $\sigma(0) = p$ and $\sigma'(0) = J(0)$ (the point is that the first-order properties of Γ are only dependent on this), and choose $V \in \mathfrak{X}(\sigma)$ (a vector field along σ) so that $V(0) = \gamma'(0)$ and $D_s V(0) = D_t J(0)$. Then we can define

$$\Gamma(s, t) = \Gamma(s, t) = \exp_{\sigma(s)}(tV(s))$$

to be our variation. (In other words, we go to s on the curve σ and then consider the geodesic through it.) So this is a variation with $\exp_{\sigma(0)}(t\gamma'(0)) = \gamma(t)$, and for any fixed s the map $t \mapsto \Gamma_s(t)$ is a geodesic. So we do have a variation and we want to check that this actually induces the correct J. Define the variation vector $W = \partial_s \Gamma(0, t)$, which is a Jacobi field; we now want to check that $W(0) = J(0)$ and $D_t W(0) = D_t J(0)$. (Indeed, since W and J both solve the same linear ODE, uniqueness then gives us the result.) But

$$W(0) = \partial_s \Gamma(0, 0) = \left.\frac{d}{ds}\right|_{s=0} \Gamma(s, 0) = \left.\frac{d}{ds}\right|_{s=0} \exp_{\sigma(s)}(0) = \left.\frac{d}{ds}\right|_{s=0} \sigma(s) = \sigma'(0) = J(0)$$

by our construction of σ. Meanwhile, by the symmetry lemma we have

$$D_t W(0) = D_t \partial_s \Gamma(0, 0) = D_s \partial_t \Gamma(0, 0) = D_s V(0) = D_t J(0),$$

where the second-to-last equality is because $\partial_t \Gamma(s, 0) = V(s)$. Thus we do induce a Jacobi field. □

Example 124

We always get some automatic Jacobi fields given any geodesic γ', which are (1) $\gamma'(t)$ and (2) $t\gamma'(t)$. Notice that $D_t\gamma' = 0$ because γ is a geodesic, and $R(\gamma',\gamma')\gamma' = 0$, and similarly $D_t(t\gamma') = \gamma'$ so $D_t^2(t\gamma') = 0$ and $R(t\gamma',\gamma')\gamma' = tR(\gamma',\gamma')\gamma' = 0$. So in both cases the Jacobi equation $D_t^2 J + R(J,\gamma')\gamma' = 0$ are indeed satisfied. Since we should think of Jacobi fields as coming from variations through geodesics, the idea is that these are actually **coming from trivial reparameterizations** (and reflecting the linearization): specifically, (1) comes from $\Gamma(s,t) = \gamma(s+t)$ and (2) comes from $\Gamma(s,t) = \gamma((1+s)t)$.

These two examples have nothing to do with geometry, so we want to work "away" from these geodesics and only look at the "normal" fields:

Definition 125

Let $\mathfrak{X}^\perp$ denote the set of vector fields on γ such that $\langle V,\gamma'\rangle = 0$. Then we define $\mathcal{J}^\perp(\gamma) = \mathcal{J}(\gamma)\cap\mathfrak{X}^\perp$.

Now if $J \in J(\gamma)$, we have

$$\frac{d}{dt}\langle J,\gamma'\rangle = \langle D_t J,\gamma'\rangle, \quad \frac{d^2}{dt^2}\langle J,\gamma'\rangle = \langle D_t^2 J,\gamma'\rangle = -\langle R(J,\gamma')\gamma',\gamma'\rangle = 0$$

by metric compatibility and then antisymmetry. This leads to the following result:

Proposition 126

Let $J \in \mathcal{J}(\gamma)$. Then the following are equivalent:

1. $J \in \mathcal{J}^\perp$,
2. $J \perp \gamma'$ at two different times,
3. J and $D_t J$ are both orthogonal to γ' somewhere,
4. J and $D_t J$ are both orthogonal to γ' everywhere.

(To go backwards, the second derivative of $\langle J,\gamma'\rangle$ being zero means that the inner product is linear in t and thus being zero at two points is enough.)

Corollary 127

We have $\dim(\mathcal{J}^\perp) = 2(n-1)$ and $\dim\mathcal{J}^T = 2$.

We'll now consider Jacobi fields that are zero at a point – recall that we're assuming we're on a complete manifold (M,g) here (though if we're only defining things on compact intervals it's not so necessary).

Lemma 128

Suppose $\gamma : I \to (M,g)$ is a geodesic with $0 \in I$ and $\gamma(0) = p$, and $J \in J(\gamma)$ is a Jacobi field with $J(0) = 0$. Then J is a variation vector field for $\Gamma(s,t) = \exp_p(t(v+sw))$, where $w = D_t J(0)$.

Notice that this result holds past normal coordinates, but the idea in those normal coordinates is that the variation comes from linearly shifting the starting vector v and then considering geodesics from that. So in normal coordinates we have the Jacobi field $tw^i\partial_i$, and we will use this later to understand how normal coordinates relate to curvature.

Proof. For s fixed this is a geodesic, so like before the variation vector field $V(t) = \partial_s \Gamma(0, t)$ is a Jacobi field and we just need to check it has the right initial conditions. We have

$$V(0) = \frac{d}{ds}\bigg|_{s=0} \Gamma(s, 0) = 0 = J(0)$$

by assumption (since $\Gamma(s, 0) = p$ for all s), and by the symmetry lemma

$$D_t V(0) = D_s \partial_t \Gamma(0, 0),$$

and now $\partial_t \Gamma(s, 0) = v + sw$ is a bit weird because it is a vector field along a constant path. So now $D_s(v + sw)$ is just w (though this might be a bit confusing if we try to check the definition – the idea is that a vector field along the constant path is the same as thinking about $V(s) \in T_p M$ and then we do have $D_s V = V'$), and this is indeed the same as $D_t J(0)$ as desired. □

Theorem 129

Let $\gamma : I \to (M, g)$ again be a geodesic, let $J \in \mathcal{J}(\gamma)$ with $J(0) = 0$, and let $w = D_t J(0)$. Then the Jacobi field tells us the derivative of the exponential map:

$$J(t) = (d \exp_p)_{tv}(tw).$$

(This does make sense to write down, since the right-hand side has us landing in the tangent space $T_{\gamma(t)}(M)$.)

Proof. From our previous result, we know the variation that induces this Jacobi field, so we can plug in and find that

$$J(t) = \partial_s \Gamma(0, t) = \frac{d}{ds}\bigg|_{s=0} (s \mapsto \Gamma(s, t)),$$

and this is the derivative of the map $s \mapsto \exp_p(tv + stw)$, which is exactly what we wrote on the right-hand side. □

Definition 130

Define (this is meant to be an extension of the first case below)

$$s_c(t) = \begin{cases} R \sin\left(\frac{t}{R}\right) & c = \frac{1}{R^2} > 0, \\ t & c = 0, \\ R \sinh\left(\frac{t}{R}\right) & c = -\frac{1}{R^2} < 0. \end{cases}$$

Proposition 131

If (M, g) has constant sectional curvature c. Then for a unit speed geodesic γ and an orthogonal Jacobi field $J \in \mathcal{J}^\perp(\gamma)$ that vanishes at zero, we have

$$J(t) = k s_c(t) E(t),$$

where $E(t)$ is a parallel perpendicular vector field.

Proof. Constant sectional curvature says that if X, Y are perpendicular unit vectors, then $R(X, Y, Y, X) = c$. So we can work out the curvature endomorphism from this, and we get the formula

$$R(J, \gamma')\gamma' = c\left(\langle \gamma', \gamma' \rangle J - \langle J, \gamma' \rangle \gamma'\right).$$

Because γ is unit-speed and by choice of J this simplifies to cJ. So in constant curvature, the Jacobi equation becomes $D_t^2 J + cJ = 0$, and we can make the ansatz $s_c(t)E(t)$ for any parallel vector field $E(t) \in \mathfrak{X}^\perp$ (since s_c solves $s_c'' + cs_c = 0$ given the conditions $S_c(0) = 0, s_c'(0) = 1$). So now $D_t(s_c E)|_{t=0} = E(0)$, and we choose $E(0) = \frac{D_t J(0)}{||D_t J(0)||}$ and $k = ||D_t J(0)||$ to make this the derivative of the Jacobi field at time zero. (Then $E(t)$ is extended by parallel transport.) □

Lemma 132

Suppose (M, g) has constant sectional curvature c, and let x^i be normal coordinates at p. Then if we define $\pi : U \to S^{N-1}$ sending $x \mapsto \frac{x}{|x|}$ and $\hat{g} = \pi^* \mathring{g}$ (this is not a metric anymore – it has some degeneracy but is still a symmetric $(0, 2)$ tensor), then in those normal coordinates we have $g = dr^2 + s_c(r)^2 \hat{g}$ on $U \setminus \{p\}$. (This is the analog of polar coordinates.)

Start of proof. When $c = 0$, we get $\overline{g} = dr^2 + r^2 \hat{g}$, which is almost what we proved – we previously showed that $\Phi : \mathbb{R}_+ \times S^{n-1} \to \mathbb{R}^n \setminus \{0\}$ mapping (ρ, ω) to $\rho\omega$ satisfies $\Phi^* \overline{g} = \mathbb{R}_+ \times_\rho S^{n-1}$, and this is $d\rho^2 + \rho^2 \mathring{g}$. Then because $\Phi^{-1}(x) = (|x|, \frac{x}{|x|})$, applying $(\Phi^{-1})^*$ to both terms yields the result.

For general c, let $g_c = dr^2 + s_c(r)^2 \hat{g}$; we wish to show that $g_c = g$. We know that

$$g(\partial_r, \partial_r) = 1 = g_c(\partial_r, \partial_r)$$

by Gauss's lemma and by plugging in the formula for ∂_r, and both metrics make ∂_i perpendicular to the constant r-spheres (by Gauss's lemma in one case, and obviously in the other because there are no cross terms). So it just remains to check that $g(w, w) = g_c(w, w)$ for any w tangent to $\{r = b\}$ and then use the parallelogram identity to show that $g = g_c$ even off the diagonal. We'll finish this next time! □

17 May 10, 2023

Last time, we checked the previous lemma about constant sectional curvature in the case $c = 0$ (we basically have an extension of polar coordinates). We'll continue the proof now:

Proof of Lemma 132, continued. As mentioned last time, it suffices to check that $g(w, w) = g_c(w, w)$ for all w tangent to the slice $\{r = b\}$, since Gauss's lemma tells us the radial direction has length 1 (which is the correct dr^2 term), and the constant r-spheres are orthogonal to the radial direction. Note that if $\overline{g} = dr^2 + r^2 \hat{g}$, then we explicitly have

$$\boxed{g_c(w, w)} = s_c(b)^2 \hat{g}(w, w) = \boxed{\frac{s_c(b)^2}{b^2} \overline{g}(w, w)},$$

and now if $\gamma = b^{-1}(tq^1, \cdots, tq^n)$ is a radial unit speed geodesic from p to q (where q is the location of our tangent vector w), we can use the explicit description of Jacobi fields in normal coordinates and the exact expression for the Jacobi field at constant curvature:

$$J(t) = \frac{t}{b} w^i \partial_i \Big|_{\gamma(t)}$$

corresponds to the variation of curves by "tilting the line" from p to q in the w direction. (Then we have $J(0) = 0$ and $D_t J(0) = b^{-1} w$.) Furthermore, $J(t)$ is actually an element of $\mathcal{J}^\perp(\gamma)$ because it's perpendicular at two points (namely p and q), so it must be of the form $J(t) = k s_c(t) E(t)$ for some parallel unit vector $E(t)$ because we have constant sectional curvature. So in this form we have $D_t J(0) = kE(0)$, and thus $k^2 = |D_t J(0)|_g^2$. Normal coordinates are parallel at p, so we can also write

$$k^2 = |D_t J(0)|_g^2 = \left| \frac{1}{b} w^i \partial_i \Big|_p \right| g = \frac{1}{b^2} |w|_{\overline{g}}^2,$$

where we're using that g is δ_{ij} at p, and the metric $\overline{g}$ is taking the sum of the squares of the coefficients. But now plugging in the boxed equality above, this is also equal to $\frac{1}{s_c(b)^2} |w|_{g_c}^2$. Plugging this into the fact that $J(b) = w$, we have

$$|w|_g^2 = |J(b)|_g^2 = k^2 s_c(b)^2 = |w|_{g_c}^2,$$

and this completes the proof. □

One way to think about this proof is that if we know this theorem, then knowing the Jacobi fields along γ can be thought about in two ways: on the one hand they are of the form $\frac{t}{b} w^i \partial_i$, but we don't know much about ∂_i away from p, but it is tangent to the constant-r sphere and under the projection π we just get a unit vector. So we're just going in the other direction of that for this proof, and we're saying that we have a warped product $I \times_{S_c} S^{n-1}$.

This result is nice because it gives us a way to see that sectional curvature is not unreasonable:

Corollary 133
Suppose $(M^n, g), (\tilde{M}^n, \tilde{g})$ both have constant sectional curvature. Then the two manifolds are locally isometric.

(Indeed, pick normal coordinates around the two points; the metrics are the same in a small enough neighborhood by our previous result so we can use the obvious map.)

Corollary 134
Suppose U is a geodesic ball (open or closed) of radius b in a manifold (M, g) of constant sectional curvature. Then we get an explicit formula in "polar coordinates"

$$\int_U f \, dV_g = \int_0^b \int_{S^{n-1}} f(\rho, w) s_c(\rho)^{n-1} d\rho dV_{\mathring{g}}.$$

We'll now move towards an interesting part of the theory: we've talked about the exponential map and how long geodesics stay minimizing, and it turns out there's a nice structure surrounding how geodesics fail to minimize. So we'll forget about curvature for a bit and see it come up again, and our first discussion point is **conjugate points**. Assume for simplicity that (M, g) is complete – we want to know when the map $d(\exp_p)_v$ not invertible. (We know that $d(\exp_p)_0 = \text{id}$, and this was important for setting up normal coordinates.) Recall that the Jacobi field starting at 0 with initial derivative $D_t J(0) = w$ along the geodesic γ with initial velocity v is $J(t) = d(\exp_p)_{tv}(tw)$, and this is singular if there exists a Jacobi field with $J(0) = 0$ and $J(t) = 0$. That motivates the following definition:

Definition 135
Let $\gamma : [a, b] \to (M, g)$ be a geodesic segment with $\gamma(a) = p$ and $\gamma(b) = q$. Then p and q are **conjugate** along γ if there exists some nontrivial $J \in \mathcal{J}(\gamma)$ with $J(a) = J(b) = 0$. The **order** of conjugacy is the dimension of the set of all $J \in \mathcal{J}(\gamma)$ for which this holds.

Proposition 136

The point $g = \exp_p v$ is conjugate to p along the curve $t \mapsto \exp_p(tv)$ if and only if v is a critical point for $\exp_p$ (that is, $d(\exp_p)_v$ is not invertible). (Note that this is a property of the geodesics, not just the points.)

Proposition 137

Suppose $\gamma : I \to M$ is a geodesic with $a \neq b \in I$. Then there is a solvable solution to the **boundary** value problem $J \in \mathcal{J}(\gamma), J(a) = v \in T_{\gamma(a)}(M), J(b) = w$ for all v, w if and only if $\gamma(a), \gamma(b)$ are **not** conjugate along $\gamma|_{[a,b]}$.

Proof. Map $J(\gamma) \to T_{\gamma(a)}M \times T_{\gamma(b)}M$ via $J \mapsto (J(a), J(b))$; this is a map between $(2n)$-dimensional vector spaces and the kernel is exactly the points where $J(a) = J(b) = 0$. □

We're now going to describe a way of thinking about Jacobi fields other than coming from variation of vector fields:

Theorem 138 (Second variation of length)

Let $\gamma : [a, b] \to (M, g)$ be a unit speed geodesic segment, and suppose $T : J \times [a, b] \to M$ is a proper variation (meaning the endpoints are left fixed). Assume for simplicity that the variation is completely smooth (the answer is the same). Then

$$\frac{d^2}{ds^2} L_g \bigg|_{s=0} = \int_a^b |D_t V^\perp|^2 - \mathrm{Rm}(V^\perp, \dot\gamma, \dot\gamma, V^\perp),$$

where $V^T = \langle V, \dot\gamma \rangle \dot\gamma$ is the tangential component and $V^\perp = V - V^T$.

This looks a lot like the Jacobi equation – this is the quadratic form that generates it, but we'll talk about that later. Notice that $V^\perp$ is the only thing that factors into the expression for length here.

Proof. (The case where we aren't smooth everywhere gives us boundary terms that we need to check cancel out. But there's no questions about differentiating L since it is smooth.) We saw that

$$\frac{d}{ds} L(\Gamma_s) = \int_a^b \frac{\langle D_t S, T \rangle}{\langle T, T \rangle^{1/2}} dt$$

by the product rule and using the symmetry lemma. Notice that we're now not enforcing that the Γ_s are unit-speed because that will make the expressions harder to work with. Differentiating this equation again and plugging in $s = 0$ (so using that $|T| = 1$ at $s = 0$), we're left with

$$\frac{d^2}{ds^2}\bigg|_{s=0} = \int_a^b \langle D_s D_t S, T \rangle + \langle D_t S, D_s T \rangle - \frac{1}{2} \frac{\langle D_t S, T \rangle \langle D_s T, T \rangle}{\langle T, T \rangle^{3/2}} = \int_a^b \langle D_s D_t S, T \rangle + |D_t S|^2 - \langle D_t S, T \rangle^2.$$

But now we can write $D_t V$ instead of $D_t S$ in the second term, and similarly we can rewrite the last term

$$\frac{d^2}{ds^2}\bigg|_{s=0} = \int_a^b \langle D_s D_t S, T \rangle + |D_t V|^2 - \langle D_t V, \dot\gamma \rangle^2.$$

For the first term, we want to swap the D_t and D_s because then we are differentiating a geodesic and thus the term vanishes when we integrate it, so using integration by parts it simplifies to $\langle D_t D_s S, T \rangle + \mathrm{Rm}(V, T, V, T) = \langle D_t D_s V, T \rangle - \mathrm{Rm}(V, T, T, V)$. (Here we use that if W is a vector field along Γ, $D_s D_t W - D_t D_s W = R(\partial_s \Gamma, \partial_t \Gamma) W$.) But any tangential component of V yields zero, so we can replace Vs with $V^\perp$s and get the Rm term that we want.

And now we check that everything else works out: we have

$$\int_a^b \langle D_t D_s S, T\rangle = \int_a^b \frac{d}{dt}\langle D_s V, T\rangle - \langle D_s V, D_t T\rangle.$$

The second term vanishes because we have a geodesic, and the first term becomes $\langle D_s S, T\rangle_a^b$. But now at $t = a$, $\Gamma(s, a)$ is constant and thus $\partial_s \Gamma(s, a) = S(s, a) = 0$, meaning $D_s S(s = 0, a) = 0$. (The same holds for $t = b$.) Finally,

$$D_t V^T = \langle D_t V, \dot\gamma\rangle\dot\gamma = (D_t V)^T$$

(orthogonal progression commutes with D_t; note that this requires γ to be a geodesic), and so subtracting both sides from $D_t V$ yields $D_t V^\perp = (D_t V)^\perp$. Thus

$$|D_t V^\perp|^2 = |D_t V|^2 - \langle D_t V, \dot\gamma\rangle^2,$$

and so the two boxed terms above do simplify correctly. □

18 May 12, 2023

Today's lecture will be given by Shuli Chen. We'll start by discussing the **index form**: recall that if $\Gamma_s : I \times [a, b] \to M$ is a variation field, then we computed the second variation

$$\left.\frac{d^2}{ds^2}\right|_{s=0} L_g(\Gamma_s) = \int_a^b (|D_t V^\perp|^2 - \mathrm{Rm}(V^\perp, \gamma', \gamma', V^\perp))dt.$$

(Remember this only depends on the normal part of the curve, since the tangential part is just a reparameterization of the curve which doesn't change the length.) Polarizing this expression, we can define a symmetric bilinear form as follows:

Definition 139

Let $\gamma : [a, b] \to (M, g)$ be a unit-speed geodesic. The **index form** is a

$$I(V, W) = \int_a^b (\langle D_t V, D_t W\rangle - \mathrm{Rm}(V, \gamma', \gamma', W))\, dt$$

for $V, W \in \mathfrak{X}^\perp(\gamma)$.

Since this definition comes from the second variation, we immediately get the following:

Corollary 140

If γ is a unit-speed geodesic that minimizes length, then $I(V, V) \geq 0$ for all $V \in \mathfrak{X}^\perp(\gamma)$.

Our goal is to relate the index form to the Jacobi equation:

Proposition 141
Let $V, W \in \mathfrak{X}^\perp(\gamma)$ be piecewise smooth vector fields. Then the index form may be computed via the formula

$$I(V, W) = -\int_a^b \langle D_t^2 V + R(V, \gamma')\gamma', W\rangle dt + \langle D_t V, W\rangle|_a^b - \sum_{i=1}^{k-1} \langle \Delta_i D_t V(a_i), W(a_i)\rangle,$$

where the a_is are the break points of the curve.

Proof. For simplicity we'll assume that V, W are smooth (otherwise we can just add together the contributions). Then by integration by parts, we have

$$\frac{d}{dt}\langle D_t V, W\rangle = \langle D_t^2 V, W\rangle + \langle D_t V, D_t W\rangle$$

and now integrating everything and plugging in yields the result. □

This next result is an exercise but it's similar to showing that unit-speed length-minimizing curves are geodesics:

Corollary 142
Suppose γ is a geodesic and V is a proper (vanishing at endpoints) piecewise smooth normal vector field. Then $I(V, W) = 0$ for all proper normal piecewise smooth W if and only if V is a Jacobi field.

We'll now move to discussing geodesics and their interaction with conjugate points – last time, we defined that for a geodesic $\gamma : [a, c] \to M$, then $p = \gamma(a)$ and $q = \gamma(c)$ are conjugate if there exists a nonzero Jacobi field $J \in \mathcal{J}(\gamma)$ with $J(a) = J(c) = 0$. In this setting, we say that γ **has a conjugate point** if there is some $b \in (a, c]$ with $\gamma(a), \gamma(b)$ conjugate, and we say that γ has an **interior conjugate point** if $b \neq c$.

Theorem 143
If γ is a geodesic with an interior conjugate point, then there exists a proper normal vector field $X \in \mathfrak{X}^\perp(\gamma)$ along γ such that $I(X, X) < 0$. In particular, this means that γ does not minimize length from $\gamma(a)$ to $\gamma(c)$.

Proof. The idea is to slightly manipulate the Jacobi field. By definition, there is some Jacobi field J vanishing at $\gamma(a)$ and $\gamma(b)$, and now we can define $V(t) = J(t)$ for $t < b$ and 0 for $t \geq b$. This is piecewise smooth, and computing its index form we see that $I(V, V) = 0$. But now because $D_t V$ has a "jump" at $t = b$ (otherwise it would be zero identically), we may define $W \in \mathfrak{X}^\perp(\gamma)$ to be a proper vector field such that $W(b) = -D_t J(b) = \Delta D_t V(b)$ (by multiplying by some bump function). We have $W(b) \neq 0$, and now if we define $X_\varepsilon = V + \varepsilon W$, we find that

$$I(X_\varepsilon, X_\varepsilon) = I(V, V) + 2\varepsilon I(V, W) + \varepsilon^2 I(W, W),$$

and now if we compute $I(V, W)$ we only get the contribution from the jump part because V is a Jacobi field, yielding $-2\varepsilon|D_t V|^2$. So now picking ε small enough we can make this term dominate the $\varepsilon^2 I(W, W)$ term, and X_ε does the job. □

On the other hand, if there are no interior conjugate points, γ does minimize among nearby curves (though not global curves, since Jacobi fields only deal with things locally). To prove that, though, we need an intermediate result first:

Lemma 144
Suppose $J_1, J_2 \in \mathcal{J}(\gamma)$ are two Jacobi fields. Then $\langle D_t J_1, J_2\rangle - \langle J_1, D_t J_2\rangle$ is constant along the curve.

Proof. This is the "Wronskian of a second-order ODE:" we can compute by the product rule that

$$\frac{d}{dt}\left(\langle D_t J_1, J_2\rangle - \langle J_1, D_t J_2\rangle\right) = \langle D_t^2 J_1, J_2\rangle - \langle J_1, D_t^2 J_2\rangle.$$

Now by the Jacobi equation we can turn these into curvature terms, yielding

$$= \langle R(J_1, \gamma')\gamma', J_2\rangle - \langle J_1, R(J_2, \gamma')\gamma'\rangle.$$

But this becomes zero if we use the symmetry properties of Rm. ☐

Theorem 145

Suppose $\gamma : [a, b] \to M$ is a unit speed geodesic without interior conjugate points. Then if $V \in \mathfrak{X}^{\perp}(\gamma)$ is a proper normal piecewise smooth vector field, then $I(V, V) \geq 0$, with equality if and only if V is a Jacobi field. In particular, if $\gamma(b)$ is not conjugate to $\gamma(a)$, then $I(V, V) > 0$ (so "the curve locally minimizes length").

Proof. Assume $a = 0$ and $p = \gamma(0)$ for simplicity. Fix an orthonormal basis $w_1, \cdots, w_n \in T_pM$ with $w_1 = \gamma'(0)$, and extend $w_2, \cdots, w_n$ to the unique Jacobi fields $J_2, \cdots, J_n$ satisfying $J_\alpha(0) = 0$ and $J'_\alpha(0) = w_\alpha$. We know that these Jacobi fields do not vanish on the interior of the interval, so they provide an orthonormal frame along the curve, and in the interior we can write $V(t) = V^\alpha(t)J_\alpha$. We can then in fact extend this definition to the endpoints $0, b$ – indeed, the Jacobi fields $J(t) = d(\exp_p)_{tv}(tw)$ can be written in the nice form in normal coordinates around p as $J_\alpha(t) = t\frac{\partial}{\partial x^\alpha}\big|_{\gamma(t)}$ (where t is the distance to the point). But this means J_α only vanishes to first order so $V(t) = V^\alpha(t)J_\alpha(t)$ lets us define V^α. And at the other endpoint similarly J_α either vanishes or it doesn't but in either case we can extend V^α. So now we have

$$|D_t V|^2 - \text{Rm}(V, \gamma', \gamma', V) = \frac{d}{dt}\langle D_t V, V\rangle - \langle D_t^2 V, V\rangle - \text{Rm}(V, \gamma', \gamma' a, V),$$

and since $D_t V = D_t(V^\alpha J_\alpha) = \dot{V}^\alpha J_\alpha + V^\alpha D_t J_\alpha$, meaning that (again by the product rule) $D_t^2 V = V^\alpha D_t^2 J_\alpha + 2\dot{V}^\alpha D_t J_\alpha + \ddot{V}^\alpha J_\alpha$. Replacing D_t^2 with the corresponding Jacobi term, we thus have

$$D_t^2 V = -R(V, \gamma')\gamma' + 2\dot{V}^\alpha D_t J_\alpha + \ddot{V}^\alpha J_\alpha,$$

so substituting back in yields

$$\frac{d}{dt}\langle D_t V, V\rangle - \langle \ddot{V}^\alpha J_\alpha + 2\dot{V}^\alpha D_t J_\alpha, V\rangle$$
$$= \frac{d}{dt}\langle \dot{V}^\alpha J_\alpha + V^\alpha D_t J_\alpha, V\rangle - \langle \ddot{V}^\alpha J_\alpha + 2\dot{V}^\alpha D_t J_\alpha, V\rangle,$$

and now by compatibility with the metric again yields

$$= \frac{d}{dt}\langle V^\alpha D_t J_\alpha, V\rangle + \langle \ddot{V}^\alpha J_\alpha, V\rangle + \langle \dot{V}^\alpha D_t J_\alpha, V\rangle + \langle \dot{V}^\alpha J_\alpha, D_t V\rangle - \langle \ddot{V}^\alpha J_\alpha + 2\dot{V}^\alpha D_t J_\alpha, V\rangle,$$

and now some terms here will cancel out and we are left with $\frac{d}{dt}\langle V^\alpha D_t J_\alpha, V\rangle + \langle \dot{V}^\alpha J_\alpha, D_t V\rangle - \langle \dot{V}^\alpha D_t J_\alpha, V\rangle$. Plugging in the expression for $D_t V$ we found above further cancels to $\frac{d}{dt}\langle V^\alpha D_t J_\alpha, V\rangle + |\dot{V}^\alpha J_\alpha|^2 + \langle \dot{V}^\alpha J_\alpha, V^\beta D_t J_\beta\rangle - \langle \dot{V}^\alpha D_t J_\alpha, V^\beta J_\beta\rangle$; the last two terms cancel by our Wronskian lemma above because their difference is constant and starts off as zero. Finally, the $\frac{d}{dt}$ term goes away when we integrate out to find

$$I(V, V) = \int_a^b |\dot{V}^\alpha J_\alpha|^2 \geq 0,$$

with equality only if we have a Jacobi field as desired. ☐

Example 146

Recall that great circles are geodesics on a sphere, and diametrically opposite points are conjugate points. So then great circles indeed minimize length all the way up to the diameter (the result above is inconclusive about what happens at the conjugate point), and then past that point we no longer minimize length.

Remark 147. *In the textbook, it's mentioned that the index form can also be interpreted as a Hessian of the* **Morse index***, which is the maximum dimension of a vector space such that $I(V, V)$ is strictly negative on that vector space. And this number will also equal the number of conjugate points on the curve γ, counted with multiplicity. Similar results also hold for higher-dimensional minimal surfaces.*

We'll make a final definition for our last topic of **cut points** – conjugate points tell us how things work locally, but cut points tell us more global results. The example to keep in mind is the cylinder, which is flat and has no conjugate points but for which going past antipodal points causes length minimization to fail.

Definition 148

Let $v \in T_pM$ and let γ be a geodesic starting at p with velocity v. The **cut time** of (p, v) is then defined via

$$t_{\text{cut}}(p, v) = \sup\{b > 0 : \gamma_v|_{[0,b]} \text{ is length-minimizing}\}.$$

The **cut point** is defined to be $\gamma_v(t_{\text{cut}}(p, v))$, and the **cut locus** of p $\text{Cut}(p)$ is the set of points q such that q is a cut point for some (p, v).

19 May 15, 2023

We'll discuss cut points today – we proved last time that passing a conjugate point on a geodesic means that we are no longer length-minimizing. There are also other ways where we can cease to be minimizing (for example on a cylinder, where the cylinder is always a local diffeomorphism but where going halfway around the circle means we should have taken a shorter path), but we can basically say that this is the only other method. We'll always consider complete manifolds here – recall that the cut time of (p, v) is the supremum of all times b for which the geodesic $\gamma|_v$ is length-minimizing on $[0, b]$. (We know this is positive because we can take normal coordinates, where straight lines are length-minimizing, but it could be infinite on something like $\mathbb{R}^2$.) We then care about the set of cut points $\gamma_v(t_{\text{cut}}(p, v))$ for some fixed p across all v.

Proposition 149

Let $v \in T_pM$ be a unit vector and suppose $c = t_{\text{cut}}(p, v)$. Then we have the following:

1. $\gamma_v|_{[0,b]}$ is minimizing for all $0 < b < c$, and this is the unique unit-speed minimizing geodesic from $\gamma(0)$ to $\gamma(b)$.
2. If $c < \infty$, then $\gamma_v|_{[0,c]}$ minimizes length, and at least one of the following hold: **(a)** $\gamma_v(c)$ is conjugate to p along γ_v, and/or **(b)** there exist at least two minimizing unit-speed geodesics from p to $\gamma_v(c)$.

The canonical example of a cut locus is the sphere, for which (because geodesics are great circles) the cut locus of the north pole is the south pole. But it's not a very good example, since both **(a)** and **(b)** occur and that's true in general (we can wiggle the great circle to generate a Jacobi field that shows conjugacy of the two poles, but we can also just take a completely different great circle to get a geodesic). Meanwhile, only **(b)** holds for the cylinder. (An example where **(a)** holds but not **(b)** is less trivial – take the paraboloid and consider the "meridional" geodesic straight down. As we go along that curve, for a while there will only be one minimizing geodesic, but later on there will be a **bifurcation** and that can only happen if the differential vanishes and thus that corresponds to a conjugate point.)

Remark 150. *Notice that on the cylinder and on the sphere in the examples we've described, we start traversing another geodesic in reverse, but this is not true if we form a "helix" shape on the cylinder as a geodesic instead. In general, it is true that for the "closest geodesic" things do close up, but we can look up Klingenberg's work for more.*

Proof. For (1), if there were a shorter curve up to time b, then (by completeness) we must have a unit-speed geodesic segment. But then the two curves must have different velocities (or else by uniqueness of ODEs they would be the same), and then we can do the same "rounding the corner" proof to show that this cannot occur.

For (2), choose $b_i \uparrow c$; by continuity of the distance function we know that $d(\gamma(c), p) = \lim d(\gamma(b_i), p) = \lim L(\gamma|_{[0,b_i]}) = \lim b_i = c$. Thus $\gamma_v|_{[0,c]}$ is a length minimizer. Now assume that $\gamma(c)$ is not conjugate to p along γ_v; we now need to find another geodesic of the same length. Choose $c_i \downarrow c$; by assumption we know that $\gamma_v|_{[0,c_i]}$ are all not length minimizing, so there must be a unit-speed minimizing geodesic σ_i from p to $\gamma_v(c_i)$ which is not γ.

Now $\sigma_i'(0)$ is a unit vector in T_pM, so we can pass to a subsequence so that $\sigma_i'(0)$ converges to some $w \in T_pM$. Since $L(\sigma_i) = \ell_i$ converges to c (again by continuity of the distance function and knowing that $\gamma_v|_{[0,c]}$ is a minimizer), we want to show that $\sigma : t \mapsto \exp_p(tw)$ is another minimizing geodesic between the relevant points. First notice that $\sigma(c) = \gamma_v(c)$, because $\exp(\ell_i w_i) = \gamma_v(c_i)$ by definition and then taking the limit yields $\exp(cw) = \gamma_v(c)$. Then σ is a unit-speed curve and achieves the minimum distance c, so all we need to check is that this is actually a new geodesic.

Here is where we use that $\gamma_v(c)$ and p are not conjugate along γ_v, which is the same as saying that $(d\exp_p)_{cv}$ is invertible. In particular, the inverse function theorem then tells us that there is a neighborhood W around $cv \in T_pM$ such that $(\exp_p)|_W$ is a diffeomorphism onto its image. But if $\sigma_i'(0)$ converged to v, then $\ell_i\sigma_i'(0)$ would eventually be in W, and then $\sigma(\ell_i) = \exp_p(\ell_i\sigma_i'(0))$ is the same point as $\gamma_v(c_i) = \exp_p(c_i v)$. So $c_i v = \ell_i\sigma_i'(0)$, which is a contradiction because by definition σ_i and γ_v are not the same curve. □

Definition 151

The **tangent cut locus** is the set of "vectors at the cut time"

$$\mathrm{TCL}(p) = \{v \in T_pM : |v| = t_{\mathrm{cut}}\left(p, \frac{v}{|v|}\right)\}.$$

The **injectivity domain** $\mathrm{ID}(p)$ is similarly the set of v such that $|v| < t_{\mathrm{cut}}\left(p, \frac{v}{|v|}\right)$.

Theorem 152

Define the unit tangent space $UTM = \{(p, v) \in TM : |v| = 1\}$. The map $t_{\mathrm{cut}} : UTM \to (0, \infty]$ is continuous.

Proof. Suppose $(p_i, v_i) \to (p, v)$ and $c_i = t_{\mathrm{cut}}(p_i, v_i)$. Let $b = \liminf_i c_i$ and $c = \limsup_i c_i$; we wish to show that $c \le t_{\mathrm{cut}}(p, v) \le b$ (as inequalities for the extended reals), then we get the desired result.

First to show that $c \le t_{\mathrm{cut}}(p, v)$, first assume that $c < \infty$. If $v_i \to v$ and $p_i \to p$, then because the exponential map and distance function are continuous we converge to the minimizing curve γ_v on $[0, c]$. And if this is true, then $t_{\mathrm{cut}}(p, v) \ge c$, as desired. Meanwhile, if $c = \infty$, we can just chop off the γ_vs and use that γ_v is minimizing on any $[0, t]$.

Next, to show that $t_{\text{cut}}(p, v) \leq b$, we may assume b is finite. Now assume $c_i \to b$ (we can always pass to a subsequence where this is true); by the previous result either we have **(a)** $\gamma_{v_i}(c_i)$ conjugate to p or **(b)** σ_i is another minimizing geodesic. We can pass to a further subsequence so that either **(a)** or **(b)** holds for the whole sequence. In case **(a)** we're done because $d(\exp_p)$ is continuous and thus the limit will also be conjugate, and in case **(b)** we do the same proof as before where we look at γ_v restricted to $[0, c]$, pick a different minimizing geodesic, and make the same local diffeomorphism argument. □

This leads us to some nice consequences:

Corollary 153
We have $\text{TCL}(p) = \partial\text{ID}(p)$ and $\text{Cut}(p) = \exp_p(\text{TCL}(p))$.

Theorem 154
Suppose (M, g) is complete and connected. Then $\text{Cut}(p)$ has measure zero in M, $\exp|_{\overline{\text{ID}(p)}}$ is surjective, and $\exp|_{\text{ID}(p)}$ is a diffeomorphism onto $M \setminus \text{Cut}(p)$.

(The first is an analysis fact – the graph of a continuous function has measure zero. The second fact is true because we can always connect p and q by a minimizing geodesic, and the time it takes is less than the cut time. Finally, we have $\exp(\text{ID}(p)) = M \setminus \text{Cut}(p)$ because the cut points are the only ones in the closure, injectivity holds because having two minimizing geodesics means we're not in ID, and we have a diffeomorphism because of the exponential map.)

Corollary 155
Every compact connected smooth manifold is homeomorphic to $B^n/\sim$, where $\sim$ somehow identifies points in ∂B^n.

(We do this by identifying the closure of $\text{ID}(p)$ with the closed ball; we have to argue that this kind of set is homeomorphic to a closed ball, which does take some work.)

20 May 17, 2023

We'll start off by tying up some loose ends. Recall that the injectivity radius at a point p is the biggest radius of the tangent space T_pM on which the exponential map is a diffeomorphism onto its image:

Proposition 156
If (M, g) is a complete, connected manifold, then $\text{inj}(p) = d(p, \text{Cut}(p))$ (and the cut locus is empty if and only if the injectivity radius is infinite).

Proof. Let $d = d(p, \text{Cut}(p))$. For any $a < d$, we know that $\exp_p(v)$ is not a cut point along the radial (exponential) geodesic for any $v \in B_a(0)$, so in particular $B_a(0) \subseteq \text{ID}$. On the other hand, if $a > d$, then there is some cut point $q \in \text{Cut}(p)$ with $d(p, q) = d < a$. Choosing a minimal geodesic between p and q, we see that $\gamma = \exp_p(tv)$ for some v with $|v| = d$ with $\gamma(1) = q$. But q is a cut point, so geodesics don't minimize past this point on γ and thus injectivity is broken. □

Corollary 157
If (M, g) is a complete connected manifold, then $p \mapsto \text{inj}(p)$ is continuous (since $\text{inj}(p) = \min\{t_{\text{cut}}(p, v) : |v| = 1\}$, t_{cut} is continuous on UTM, and the unit ball is compact).

We're now going to move to the topic for the rest of this class, **comparison theory** – it may seem like we prove a sequence of random theorems, but they basically tell us "how to think about curvature bounds" (and thus say something quantitatively above our manifold). Our basic tool will be the Riccati equation.

Fact 158
We've often written ∇r instead of $\text{grad}(r)$, but we'll use the latter to be consistent with Lee from here.

Definition 159
Let $U \subset (M, g)$ be a normal neighborhood of p (that is, the image of a star-shaped neighborhood of our tangent space). Recall that this defines the radial vector field $\partial_r = \text{grad}(r)$ on $U \setminus \{p\}$ and the covariant Hessian $\nabla^2 r$. Define the $(1, 1)$-tensor (the **Hessian operator**) $\mathcal{H}_r = \nabla \partial_r$ (which we often will think of as an element of $\Gamma(\text{End}(TM))$); in other words, $\mathcal{H}_r = (\nabla^2 r)^\sharp$ satisfies $\langle \mathcal{H}_r(v), w\rangle = \nabla^2 r(v, w)$.

Lemma 160
We have the following:

1. $\mathcal{H}_r$ is self-adjoint,
2. $\mathcal{H}_r(\partial_r) = 0$,
3. $\mathcal{H}_r$ restricted to the tangent of a levelset of r is the shape operator with respect to the unit normal $N = -\partial_r$.

Proof. Property (1) is clear because $\nabla^2 r$ is symmetric (that is, we have an endomorphism associated to be symmetric bilinear form). For (2), notice that $\nabla_{\partial_r} \partial_r = 0$ because radial curves are geodesics (and ∂_r is the associated velocity vector). For (3), recall that ∂_r is perpendicular to $T\{r = b\}$, so in particular $\mathcal{H}_r$ is an element of $\text{End}(T_q\{r = b\})$ (the image is always tangent to the spheres). Since the shape operator is defined by $h(X, Y) = \langle sX, Y\rangle$, and the Weingarten equation reads $-sX = \nabla_X N$, taking the inward-pointing unit normal yields $sX = \nabla_x \partial_r = \mathcal{H}_r(X)$, as desired. $\square$

Proposition 161
Let $U \subset (M, g)$ be a normal neighborhood, and let $\gamma : [0, b] \to U$ be a radial geodesic with unit speed (starting at p). Let $J \in \mathcal{J}^\perp(\gamma)$ be a normal Jacobi field with $J(0) = 0$. Then we have $D_t J(t) = \mathcal{H}_r(J(t))$ for all $t \in (0, b]$.

In other words, $\mathcal{H}_r$ tells us how much the Jacobi field stretches as it moves.

Proof. Let $v = \gamma'(0)$ and $w = D_t J(0)$; note that $v \perp w$ (since a normal Jacobi field has normal derivative). Find a curve $\sigma(s)$ on the unit sphere S_1 such that $\sigma(s) \in S_1$, $\sigma(0) = v$, and $\sigma'(0) = w$. Now we get a family of varying geodesics $\Gamma(s, t) = \exp_p(t\sigma(s))$, and also $\partial_t \Gamma(s, t) = \partial_r|_{\Gamma(s,t)}$. (So this is similar to how we've been varying geodesics previously, but we want to make sure we stay valued in the unit sphere.) Now by the chain rule,

$$\partial_s \Gamma(0, t) = d(\exp_p)_{tv}(tw) = J(t)$$

with the last step coming from what we've previously proved about Jacobi fields. Therefore,

$$D_t J = D_t\, \partial_s \Gamma|_{s=0} = D_s \partial_t \Gamma = D_s \partial_r|_{\Gamma(s,t)},$$

but now the right-hand side is extendible (since ∂_r is defined on the whole neighborhood from the start) so it can also be written as $\nabla_{\partial_s \Gamma} \partial_r|_{s=0} = \nabla_J \partial_r$, which is exactly $\mathcal{H}_r(J)$. $\square$

Note that the previous formula only makes sense away from 0, not because $J(t)$ doesn't make sense but because $\mathcal{H}_r$ is singular (we'll talk about this now).

Proposition 162

Let $U \subset (M, g)$ be a normal neighborhood. Then g has sectional curvature identically c on U if and only if $\mathcal{H}_r = \frac{s_c'(r)}{s_c(r)} \pi_r$ on $U \setminus \{p\}$, where π_r is the projection to $T\{r = b\}$.

In particular, we can write out explicit formulas

$$\frac{s_c'(t)}{s_c(t)} = \begin{cases} \frac{1}{t} & c = 0 \\ \frac{1}{R} \cot \frac{t}{R} & c = \frac{1}{R^2}, \\ \frac{1}{R} \coth \frac{t}{R} & c = -\frac{1}{R^2}, \end{cases}$$

and this shows that the eigenvalues of $\mathcal{H}_r$ blow up to infinity as $t \to 0$.

Proof. For the forward direction, suppose we have a radial unit-speed geodesic γ. Choose $E_1(t), \cdots, E_n(t)$ to be a unit orthonormal frame extended so that $E_n(t) = \gamma'(t)$. Now we know that for any normal Jacobi field J with $J(0) = 0$, we have $D_t J = \mathcal{H}_r(J(t))$ for $t \in (0, 1]$. But we also know that $J = s_c(t) E_i(t)$ for all $1 \le i \le n-1$ are all in $\mathcal{J}^\perp$. So putting these together,

$$D_t J = s_c' E_i$$

(no product rule on the first term since E_i is parallel by definition), and then we also have for all i that

$$\mathcal{H}_r(E_i) = \frac{s_c'}{s_c} E_i$$

for all $1 \le i \le r-1$, and because $\mathcal{H}_r(E_n) = 0$ this means we exactly have $\frac{s_c'(r)}{s_c(r)} \pi_r$ as desired. (Notice that if $c = \frac{1}{R^2}$, we can't have $b \ge \pi R$ because we can't have a geodesic of length at least this.)

For the other direction, if J is a Jacobi field with $J(0) = 0$, we proven that $D_t J = \mathcal{H}_r(J)$, but by assumption this is equal to $\frac{s_c'}{s_c} \pi_R J$. But Gauss's lemma tells us that $\pi_r J = J$, so we in fact have $D_t J = \frac{s_c'}{s_c} J$ and thus $D_t(sc^{-1} J) = 0$. And now recalling the proof that constant sectional curvature yields $g = dr^2 + s_c(r)^2 \hat{g}$ in a neighborhood, we only used the constant sectional curvature fact in this last part. $\square$

Theorem 163 (Riccati equation)

Let $U \subset (M, g)$ be a normal neighborhood (with set up as before). Then along a radial unit speed geodesic γ, we have the equality of endomorphisms $D_t \mathcal{H}_r + \mathcal{H}_r^2 + R_{\gamma'} = 0$ (where $\mathcal{H}_r^2(w) = \mathcal{H}_r(\mathcal{H}_r(w))$ and $R_{\gamma'}(w) = R(w, \gamma')\gamma'$).

Proof. We first check that the two sides are equal on the vector field ∂_r. We have, by the product rule

$$D_t \mathcal{H}_r(\partial_r) = \frac{d}{dt}(\mathcal{H}_r(\partial_r)) - \mathcal{H}_r(D_t \partial_r).$$

The first term is zero because $\mathcal{H}_r(\partial_r)$ is identically zero and the second term is zero because $D_t\partial_r$ is identically zero (we have a geodesic). Similarly the other two terms are zero, so this equality does hold in the radial direction. So now if $z \in T_{\gamma(t_0)}\{r = t_0\}$ is a tangential vector, we may write $z = z^i\partial_i$, and we saw that $tz^i\partial_i|_{\gamma(t)}$ is a Jacobi field (translating the straight line in the direction z); in fact, it is a normal Jacobi field because it is normal at both 0 (where it vanishes) and at t_0 (by choice). So by a previous result we have $D_tJ = \mathcal{H}_rJ$; the Jacobi equation tells us that (taking another derivative)

$$-R(J, \gamma')\gamma' = D_t^2 J = D_t\mathcal{H}_r(J) + \mathcal{H}_r(\mathcal{H}_r(J))$$

where we've used the formula again on the last term. Evaluating at t_0, the two sides simplify to the problem statement at z. □

Since $\mathcal{H}_r$ is the shape operator on the spheres, this tells us how the shape operator evolves as we change the radius (depending on curvature), so this is rather fundamental and tells us something really geometric.

21 May 19, 2023

Last time, we defined the Hessian operator $\mathcal{H}_r = \nabla\partial_r$ (in a normal neighborhood) and we proved the Ricatti equation $D_t\mathcal{H}_r + \mathcal{H}_r^2 + R_{\gamma'} = 0$, which is one of the big things we will get a lot of mileage out of (in terms of relating curvature to geometry). We need to take this first-order ODE and somehow say that "if $R_{\gamma'}$ is big, then we can say something about the other terms." Recall that we say that two self-adjoint endomorphisms A, B satisfy $A \le B$ if $\langle Av, v\rangle \le \langle Bv, v\rangle$ for all v (or equivalently the matrix $B - A$ is nonnegative definite).

Theorem 164 (Riccati comparison theorem)
Suppose $\gamma : [a, b] \to M$ is a geodesic, $\eta, \tilde{\eta}$ are self-adjoint endomorphisms along $\gamma|_{(a,b]}$, satisfying the Riccati-type equations $D_t\eta + \eta^2 + \sigma = 0$ and $D_t\tilde{\eta} + \tilde{\eta}^2 + \tilde{\sigma} = 0$, where $\sigma, \tilde{\sigma}$ are continuous self-adjoint endomorphisms with $\tilde{\sigma}(t) \ge \sigma(t)$ for all $t \in [a, b]$. Now if $\lim_{t\downarrow a}(\tilde{\eta}(t) - \eta(t)) \le 0$ (and the limit exists), then $\tilde{\eta}(t) \le \eta(t)$ for all $t \in (a, b]$.

Recall that in constant curvature, the Hessian operator looks like $\mathcal{H}_r^c = \frac{s_c'(r)}{s_c(r)}\pi_r$, which blows up as $r \to 0$. So the point is that $\tilde{\eta}$ and η do not need to be defined as $t \to a$, though the difference should be well-defined. And here when we say that η is an endomorphism along γ, we mean that $\eta(t) \in \operatorname{End}(T_{\gamma(t)}M)$ for each t.

To prove this, let $E_1(t), \cdots, E_n(t)$ be a parallel orthonormal frame along γ, and let $H, \tilde{H}, S, \tilde{S}$ be the matrices of $\eta, \tilde{\eta}, \sigma, \tilde{\sigma}$ respectively (so that $\eta(E_i) = H_i^jE_j$). The t-derivative only hits η and not E_i because the frame is parallel, so the t-derivative only hits the matrix entries: thus, we have $H' + H^2 + S = 0$ and similarly $\tilde{H}' + \tilde{H}^2 + \tilde{S} = 0$. Now the Riccati equation is equivalent to a purely Euclidean fact:

Theorem 165
Let $H, \tilde{H} : (a, b] \to S(n, \mathbb{R})$ (the space of symmetric $n \times n$ real-valued matrices) and $S, \tilde{S} : [a, b] \to S(n, \mathbb{R})$ satisfy $H' + H^2 + S = 0$ and $\tilde{H}' + \tilde{H}^2 + \tilde{S} = 0$. Assume that $\tilde{S}(t) \ge S(t)$ for all $t \in [a, b]$ and that $\lim_{t\downarrow a}(\tilde{H}(t) - H(t))$ exists and is nonpositive. Then $\tilde{H}(t) \le H(t)$ for all $t \in (a, b]$.

Proof. Define $A = \tilde{H} - H$, which extends continuously to $[a, b]$ by assumption, and define $B = -\frac{1}{2}(\tilde{H} + H)$, which is only necessarily defined on $(a, b]$. Now

$$A' = \tilde{H}' - H' = -\tilde{H}^2 + H^2 - \tilde{S} + S = AB + BA - \tilde{S} + S \leq AB + BA,$$

(by expanding out AB and BA and using our assumption $\tilde{S} \geq S$), and similarly

$$B' = \frac{1}{2}(H^2 + \tilde{H}^2) + \frac{1}{2}(S + \tilde{S})$$

can now be bounded by throwing away the first term and then bounding the second from below: there is some $k \in \mathbb{R}$ such that $B' \geq k\text{Id}$. Integrating this from t to b yields $B(b) - B(t) \geq k(b - t)\text{Id}$, or in other words $B(t) \leq -B(b) + k(b - t)\text{Id}$ and thus we can write $B \leq K\text{Id}$ on $(a, b]$ for some constant $K > 0$ (by bounding $B(b)$ from below as well). Now we want to use a kind of maximum principle: define a function $f : [a, b] \times S^{n-1} \to \mathbb{R}$ via

$$f(t, x) = e^{-2kt}\langle Ax, x\rangle.$$

Recall that our goal is to show that $A \leq 0$. Suppose not for the sake of contradiction; then $f > 0$ somewhere and thus there exists $(t_0, x_0) \in (a, b] \times S^{n-1}$ attaining the maximum (since we extended A continuously to a and $f(a, x) \leq 0$ by assumption). Now $\langle Ax, x\rangle$ is maximized at x_0, meaning that x_0 is an eigenvector and thus $A(t)x_0 = \lambda_0 x_0$ for some $\lambda_0 > 0$ (positive because $f(t_0, x_0) = e^{-2kt_0}\lambda_0$ is assumed to be positive. We now have $\left.\frac{\partial f}{\partial t}\right|_{t_0}$ nonnegative (since it's either zero if it's an interior maximum, or it's potentially positive if it's a maximum on the right endpoint)

$$0 \leq \left.\frac{\partial f}{\partial t}\right|_{t_0} = e^{-2Kt}\left(\langle (AB + BA)x_0, x_0\rangle - 2K\langle Ax_0, x_0\right)$$

where we've used the product rule and also the inequality for A' that we derived beforehand, and thus by self-adjointness

$$\begin{aligned} 0 \leq e^{-2Kt}\left(2\langle Ax_0, Bx_0\rangle - 2K\langle Ax_0, x_0\right) & \\ &\leq e^{-2Kt}\lambda_0\left(2\langle \lambda x_0, (B - 2K\text{Id})x_0\rangle\right) \\ &\leq e^{-2kt}\lambda_0\langle -Kx_0, x_0\rangle \\ &< 0, \end{aligned}$$

which is a contradiction. Thus A must be nonpositive everywhere. ☐

We'll now prove some **sectional curvature comparison** results: we can now see how to control geometric properties more explicitly.

Theorem 166 (Hessian comparison)
Let $U \subset (M, g)$ be a normal neighborhood of p. Then we have the following:

1. If the sectional curvature is at most c everywhere, then $\mathcal{H}_r \geq \frac{s_c'(r)}{s_c(r)}\pi_r$ on $U_0 \setminus \{p\}$, where U_0 is either U if $c \leq 0$ or the subset $\{x \in U : r < \pi R\}$ if $c = \frac{1}{R^2}$.
2. If the sectional curvature is at least c everywhere, then $U = U_0$ and $\mathcal{H}_r \leq \frac{s_c'(r)}{s_c(r)}\pi_r$ on $U \setminus \{p\}$.

In other words, small sectional curvature means bigger than the model metric and vice versa. And the technical situation with U_0 is something we'll see often, but we shouldn't worry too much about it – it's just that (for example) past $t = \pi$ the expression $\frac{s_1'}{s_1} = \frac{\cos t}{\sin t}$ is blowing up to $-\infty$, and we don't want our inequality to go past where the model stops making sense.

Proof. Use normal coordinates on U, and define r as usual to be the radial variable (which agrees with the distance function from the center). Define $\mathcal{H}_r^c = \frac{s_c'(r)}{s_c(r)}\pi_r$ on $U_0 \setminus \{p\}$, where π_r is still the projection onto the constant-r-sphere tangent space. Now along a radial geodesic, we have $D_t\pi_r = 0$ (since we can write π_r in terms of a parallel frame), and $\pi_r^2 = \pi_r$ because we have a projection. Thus

$$D_t\mathcal{H}_r^c = \frac{s_c''}{s_c}\pi_r - \frac{(s_c')^2}{s_c^2}\pi_r,$$

and now by definition of s_c and using that $\pi_r^2 = \pi_r$ this simplifies to

$$= -c\pi_r - (\mathcal{H}_r^c)^2.$$

But now we have the equation $D_t\mathcal{H}_r^c + (\mathcal{H}_r^c)^2 + c\pi_r = 0$, and we also have the actual Riccati equation $D_t\mathcal{H}_r + \mathcal{H}_r^2 + R_{\gamma'} = 0$. Now $\langle R_{\gamma'}(w), w\rangle = \langle R(w, \gamma')\gamma', w\rangle$, and if w is any unit vector perpendicular to γ' then this number will be a sectional curvature, so in case (1) it's at most c and in case (2) it's at least c. Meanwhile, plugging in γ' for w yields zero for both $c\pi_r$ and $R_{\gamma'}$. Thus we've checked $R_{\gamma'} \le c\pi_r$ in case (1) and $R_{\gamma'} \ge c\pi_r$ in case (2) on both the normal and tangential parts, and so now we want to apply Riccati comparison but need to check the limit of the differences. (This is not trivial in general – for example the Hessian of the radial vector is $\frac{1}{r}\partial_r$.) Indeed, $\frac{s_c'}{s_c} = \frac{1}{r} + O(r)$, while the Hessian of a function in coordinates is given by $\nabla^2 r = (\partial_j\partial_k r - \Gamma_{jk}^m \partial_m r)dx^j \otimes dx^k$. So

$$\mathcal{H}_r = g^{ij}(\partial_j\partial_k r - \Gamma_{jk}^m\partial_m r)\partial_i \otimes dx^k,$$

and now "being lazy" the Christoffel symbols are smooth and vanish at the origin and are thus $O(r)$; meanwhile $g^{ij} = \delta^{ij} + O(r)$ because the metric is δ^{ij} at the origin and the inverse is continuous. So now we get a contribution from the Euclidean Hessian $\frac{1}{r}\pi_r$ and the rest is $O(r)$; thus the difference indeed goes to zero as $r \to 0$ (in particular the limit exists) and so the Riccati comparison theorem tells us the result. But we have to actually check that $U = U_0$ in case (2): if $c = \frac{1}{R^2}$ and $U \supsetneq U_0$, then there is a radial geodesic γ starting from the origin which goes beyond πR. We've proven that $\mathcal{H}_r \le \frac{s_c'}{s_c}\pi_r$ up until U_0, but the right-hand side goes to $-\infty$ as $r \to \pi R$. And now the idea is that if the neighborhood were bigger than U_0, then $\mathcal{H}_r$ would be defined on a smooth identity and thus can't go to negative infinity, which would contradict the comparison theorem. □

Corollary 167 (Principal curvature comparison theorem)
Let $U \subset (M, g)$ be a normal neighborhood. Then we have the following (because the shape operator restricted to level sets is the Hessian operator):

1. If the sectional curvature is at most c everywhere, then any part of an r-level sphere in U has all principal curvatures satisfying $k \ge \frac{s_c'}{s_c}$ (for $r < \pi R$ if $c = \frac{1}{R^2}$, or for all r if $c \le 0$).
2. If the sectional curvature is at least c everywhere, then any part of an r-level sphere in U has all principal curvatures satisfying $k \le \frac{s_c'}{s_c}$.

22 May 22, 2023

Last time, we bounded the Hessian in the "opposite direction" as our sectional curvature bounds. Since Jacobi fields are controlled by the Hessian, we can now bound Jacobi fields from above and below:

Theorem 168

Let $\gamma : [a, b] \to (M, g)$ be a geodesic segment and $J \in \mathcal{J}^{\perp}(\gamma)$ be a Jacobi field vanishing at the origin.

1. If the sectional curvature is at most c everywhere, then $|J(t)| \geq s_c(t)|D_t J(0)|$ for all $t \in [0, b_1]$, where $b_1 = b$ if $c \leq 0$ and $\min\{b, \pi R\}$ if $c = \frac{1}{R^2}$.
2. If the sectional curvature is at least c everywhere, then $|J(t) \leq s_c(t)|D_t J(0)|$ for all $t \in [0, b_2]$, where b_2 is either the first conjugate point or b if no such point exists.

Proof. We may assume that $D_t J(0) \neq 0$ – otherwise J is just zero and we're automatically done. Choose b_0 to be as large as possible so that $\gamma|_{[0,b_0]}$ has no interior conjugate point and so that $s_c(t) > 0$ for all $t \in (0, b_0)$. Now in case (1), if $c = \frac{1}{R^2}$, then $b_0 \leq b_1$ (since if we hit πR then $s_c(t)$ vanishes, but we may have an interior conjugate point before that) and similarly if $c \leq 0$ we have that same inequality because $b_1 = b$. Similarly in case (2), we also have $b_0 \leq b_2$. So in either case, we are working on a "shorter interval" than we mean to be, but we'll prove that in fact we have equality and have gotten all the way to b_1 or b_2, respectively.

Now **assume** that $\gamma([0, b_0))$ is contained in some normal neighborhood U of $p = \gamma(0)$. (We'll then develop a hack to extend it in general. Notice that in this theorem we have not assumed that everything is contained in such a U, since in the past we had things like the Hessian operator that required a normal neighborhood.) Define the function

$$f(t) = \log(s_c(t)^{-1}|J(t)|) = \frac{1}{2}\log(s_c(t)^{-2}|J(t)|^2);$$

taking a derivative yields

$$f'(t) = \frac{\langle D_t J, J\rangle}{|J(t)|^2} - \frac{s_c'(t)}{s_c(t)}$$

and now because we're in a normal neighborhood we have $D_t J = \mathcal{H}_r(J)$. But now by Hessian comparison, being bounded above by c says that $\mathcal{H}_r \geq \frac{s_c'}{s_c}\pi_r$ on $U_0 \setminus \{p\}$ (with U_0 "cut off" in the same way as usual), and being bounded below by c says that $\mathcal{H}_r \leq \frac{s_c'}{s_c}\pi_r$. So in particular, that means $f' \geq 0$ in case 1, and notice that we must stop at $b_0 \leq \pi R$ so we don't have to worry about leaving the range of Hessian comparison. Meanwhile, in case 2 we have $f' \leq 0$. By L'Hopital's rule twice we thus have

$$\lim_{t\downarrow 0} \frac{|J(t)|^2}{s_c(t)^2} = \lim_{t\downarrow 0} \frac{\langle D_t J(t), J(t)\rangle}{s_c'(t)s_c(t)} = \frac{|D_t J(0)|^2}{s_c'(0)^2}$$

(where we dropped the term $\langle D_t J(t)^2, J(t)\rangle$ in the numerator because that's still zero anyway), and we normalized $s_c'(0) = 1$ by definition. So now we're done, since this ratio is increasing or decreasing depending on which case we're in, and at time zero it goes to $|D_t J(0)|^2$ so it will continue to satisfy the inequality.

So now we need to fix the normal neighborhood problem – it's "dangerous" because if we have a skinny cylinder, in which the curvature is bounded by 1 (since $0 \leq 1$) but the geodesic going around the skinny way is not contained in a normal neighborhood and there are no problems with hitting πR or getting to a conjugate point. To do that, consider $\gamma(t) = \exp_p(tw)$; in the tangent space, if we consider the ray $tw \in T_pM$ for $t \in [0, b_0]$, we know that $d(\exp_p)_{tw}$ is invertible for all $t \in [0, b_0)$. Now around every point in that interval, we can choose a small ball of radius $r(t)$ so that $d(\exp_p)$ is invertible on that ball, and in fact it is invertible on all segments connecting the origin to that ball. So now we can define

$$W = \bigcup_{t\in[0,b_0)} \text{cone to origin}(B_{r(t)}(tw)) \cup \{\text{small ball around } 0\},$$

so that W is a star-shaped open neighborhood such that $\exp_p$ is a local diffeomorphism when restricted to W. If we now define the pullback

$$\tilde{q} = (\exp_p)^* g,$$

the curve $\tilde{\gamma} : t \mapsto tw$ for $[0, b_0)$ is a geodesic with respect to this metric, and we can also define a vector field along $\tilde{\gamma}$ via $\tilde{J}(t) = d(\exp_p)^{-1}(tw)J(t)$ which is a Jacobi field. (So the exponential map says that we have a **local** isometry to g, so we can basically transplant everything back – we've unwrapped the geodesic so it sits in the normal neighborhood, in which radial lines are still geodesics.) And now we can repeat the previous step because $\tilde{g}$ has the same curvature bounds as g. The key is that we're basically passing to a cover so that there are no global obstruction issues.

It now remains to just show that $b_0 = b_1$ or $b_0 = b_2$ in the respective cases. For case (1), the only way that we could have $b_0 < b_1$ is if b_0 were a conjugate point to $\gamma(0)$ along γ (since the $s_c(t) > 0$ case was accounted for by the definition of b_1). That would mean that there is a Jacobi field $J \in \mathcal{J}^\perp$ with $J(0) = J(b_0) = 0$, but J is a continuous function on the whole line and thus the inequality we just proved for $[0, b_0)$ should hold at b_0 as well. Since $s_c(t) > 0$, this is a contradiction. And in case (2), if $b_0 < b_2$ then we must have had $s_c(b_0) = 0$, but then any Jacobi field must vanish at b_0 by the inequality we proved and we would have been a conjugate point (and thus $b_0 = b_2$). □

We've previously seen that if we have constant sectional curvature c, then our metric looks like $g = g_c = dr^2 + s_c(r)^2\hat{g}$ (since Gauss's lemma tells us that the dr term is correct and what's left is tangent to the constant-r curves; then we check that both metrics give the correct length on any such w). From there, we said that $tw^i\partial_i$ restricted to $\gamma(t)$ is a Jacobi field (by wiggling in the direction of w); to compute $|w|_g$ we write the field in the form $J(t)ks_c(t)E(t)$ and check that this is equal to $|w|_{g_c}$. Looking back at this argument, if we had inequality, **the same proof gives an inequality for the metrics** as well:

Theorem 169 (Metric comparison theorem)
Let $U \subset (M, g)$ be a normal neighborhood. If the sectional curvature is bounded from above by c, then $g \geq g_c$ on $U_0 \setminus \{p\}$, and if the sectional curvature is bounded from below by c, then $g \leq g_c$. (Here we're comparing in the usual sense of bilinear forms.)

So this is a result that tells us that we can actually compare lengths in terms of sectional curvature, and thus we've arrived at something really geometric.

Theorem 170 (Laplacian comparison, theorem 1)
Let $U \subset (M, g)$ be a normal neighborhood. If the sectional curvature is bounded above by c, then $\Delta r \geq (n-1)\frac{s_c'}{s_c}$ on $U_0 \setminus \{p\}$.

The other direction of the inequality works as well, but we'll talk about Ricci curvature comparison for the other direction (which will be stronger).

Proof. The Hessian comparison theorem says that $\mathcal{H}_r \geq \frac{s_c'}{s_c}\pi_r$, and the trace of π_r is $(n-1)$ so taking traces on both sides yields the result. □

Theorem 171 (Conjugate comparison, theorem 1)
Let (M, g) have sectional curvature bounded from above by c. If $c \leq 0$, then there are no conjugate points, and if $c = \frac{1}{R^2}$, then there are no conjugate points until at least distance πR.

Jacobi fields are meant to model the infinitesimal behavior of geodesics, and conjugate points mean that geodesics are "coming back together." So this result basically says that "coming back" cannot happen too quickly and in fact cannot happen at all if c is nonpositive.

Proof. This follows from the Jacobi comparison theorem, since $|J(t)|$ has to be positive up until some point because $s_c(t)$ is positive. $\square$

23 May 24, 2023

Today, we'll finally prove a "properly geometric" result:

Theorem 172 (Günther volume comparison theorem)
Let (M, g) be connected with sectional curvature bounded from above by c, and let $p \in M$. Let $\delta_0 = \text{inj}(p)$ if $c \le 0$ and $\delta_0 = \min\{\pi R, \text{inj}(p)\}$ if $c = \frac{1}{R^2}$, and for $\delta \in (0, \delta_0)$ define $V_g(\delta)$ to be the volume of the geodesic ball of radius δ at p and $V_c(\delta)$ be the volume of the geodesic ball in the corresponding model (either $H^n(R), S^n(R)$, or $\mathbb{R}^2$). Then the function $\frac{V_g(\delta)}{V_c(\delta)}$ is nondecreasing as δ increases, and $\lim_{\delta \downarrow 0} \frac{V_g(\delta)}{V_c(\delta)} = 1$.

Basically, Riemannian manifolds are flat on small scales, so $V_g(\delta)$ is approximately $V_c(\delta)$ for small δ. And in particular, this means that $V_g(\delta) \ge V_c(\delta)$ for all $\delta < \delta_0$, with equality if and only if the geodesic ball $B_\delta(p)$ has sectional curvature c everywhere (meaning it's isometric to the ball). Note that monotonicity here is much stronger than the inequality $V_g \ge V_c$, even if it is harder to wrap our head around.

Proof. Notice that the part $V_g(\delta) \ge V_c(\delta)$ follows from the metric comparison theorem (since $g \ge g_c$, meaning that $\det(g) \ge \det(g_c)$ and the square root of the determinant is the volume form – however, we do need the "square root trick" to prove that $g \ge g_c$ implies $\det(g) \ge \det(g_c)$). Proving monotonicity will be more involved – we work in a normal neighborhood containing $B_\delta(p)$, which exists because we work below the injectivity radius. We have $g^{ij}\partial_j r = r^{-1}x^i$ since the left-hand side is $\text{grad}(r)$ and the right-hand side is how we defined ∂_r (so this is Gauss's lemma). We can then compute the Laplacian by the product rule:

$$\begin{aligned}
\Delta r &= \frac{1}{\sqrt{\det g}}\partial_i\left(g^{ij}\sqrt{\det g}\,\partial_j r\right) \\
&= \frac{1}{\sqrt{\det g}}\partial_i\left(\sqrt{\det g}\,r^{-1}x^i\right) \\
&= r^{-1}x^i\partial_i \log\sqrt{\det g} + \partial_i\left(r^{-1}x^i\right) \\
&= \partial_r\left(\log\sqrt{\det g}\right) + \frac{n-1}{r} \\
&= \partial_r \log\left(r^{n-1}\sqrt{\det g}\right)
\end{aligned}$$

where we've computed $\partial_i(r^{-1}x^i) = \frac{n-1}{r}$ by using ordinary calculus. But now by the Laplacian comparison theorem, we know that

$$\partial_r \log(r^{n-1}\sqrt{\det g}) \ge (n-1)\frac{s_c'}{s_c} = \partial_r \log(s_c(r)^{n-1}),$$

so $\log\left(\frac{r^{n-1}\sqrt{\det g}}{s_c(r)^{n-1}}\right)$ is nondecreasing. We can thus define (as a function on our normal neighborhood)

$$\lambda(r, \omega) = \frac{r^{n-1}\sqrt{\det g}}{s_c(r)^{n-1}},$$

such that if we choose a geodesic in the normal neighborhood this is increasing along a radial geodesic. We will now rewrite the volume $V_g(\delta)$ in terms of λ: notice that for any function f, on Euclidean space we have

$$\int_{B_\delta} f\, dV_g = \int_0^\delta \int_{S^{n-1}} f \rho^{n-1} d\rho dV_g$$

(basically just generalized polar coordinates), and thus if we are integrating the ball in normal coordinates we have

$$V_g(\delta) = \int_{B_\delta(0)} \sqrt{\det g} dV_g = \int_0^\delta \int_{S^{n-1}} \sqrt{\det g}\, \rho^{n-1}\, d\rho dV_{\mathring{g}}$$

Meanwhile, we have

$$V_c(\delta) = |S^{n-1}| \int_0^\delta s_c(\rho)^{n-1} d\rho,$$

so the ratio of the volumes is (bringing in the expression for $V_c(\delta)$ into the integral for $V_g(\delta)$)

$$\frac{V_g(\delta)}{V_c(\delta)} = \frac{1}{|S^{n-1}|} \int_{S^{n-1}} \frac{\int_0^\delta \lambda \cdot s_c^{n-1}(\rho) d\rho}{\int_0^\delta s_c(\rho)^{n-1} d\rho} dV_{\mathring{g}}.$$

To complete the proof, suppose that $\delta_1 \le \delta_2$. We will prove the result by "clearing denominators:" we have

$$\boxed{\int_0^{\delta_1} \lambda(\rho, \omega) s_c(\rho)^{n-1} d\rho \int_0^{\delta_2} s_c(\rho)^{n-1} d\rho} = \int_0^{\delta_1} \lambda s_c^{n-1} d\rho \cdot \left(\int_0^{\delta_1} s_c^{n-1} d\rho + \int_{\delta_1}^{\delta_2} s_c^{n-1} d\rho \right).$$

Now because λ is increasing, we can bound the first term in the parentheses on the right as (omitting the $d\rho$s for convenience, and remembering that λ is a function of ρ)

$$\begin{aligned}
&\le \int_0^{\delta_1} \lambda s_c^{n-1} \int_0^{\delta_1} s_c^{n-1} + \lambda(\delta_1, \omega) \int_0^{\delta_1} s_c^{n-1} \int_{\delta_1}^{\delta_2} s_c^{n-1} \\
&\le \int_0^{\delta_1} \lambda s_c^{n-1} \int_0^{\delta_1} s_c^{n-1} + \int_0^{\delta_1} s_c^{n-1} \int_{\delta_1}^{\delta_2} \lambda s_c^{n-1} \\
&= \boxed{\int_0^{\delta_2} \lambda s_c^{n-1} \int_0^{\delta_1} s_c^{n-1}},
\end{aligned}$$

and now rearranging the boxed terms show monotonicity in δ, as desired. □

We do have to stop at the injectivity radius – for example consider a cylinder, for which the volume only grows linearly once we get to the diameter of the circle – but we'll see a result later on that removes this type of injectivity radius condition.

Notice that we've only proven results for sectional curvature bounded from above for the last few results, and that's because we can use Ricci curvature instead for the other direction. Notice that if the sectional curvature is bounded from above by c, then $\text{Ric} \ge (n-1)c$, since $\text{Ric}(v, v)$ is the sum of all sectional curvatures coming from pairing v with the other elements of the orthonormal basis.

Remark 173. *We may ask why we don't consider inequalities that look like Ric $\le (n-1)c$ (times g, as bilinear forms) – it turns out that such results are "useless," since at a technical level AM-GM is only able to go one way. There is also a result by Lohkamp which says that if we have a compact (M^n, g) for $n \ge 3$, then there is some $\tilde{g}$ that approximates g in a C^0 sense, but Ric$_{\tilde{g}} < -1$. So we can start with the round metric on the 3-sphere and perturb in C^0 and be bounded from above by -1, and C^0 perturbations preserve length and volume up to a small error. So we shouldn't expect "true statements" to hold in this direction.*

Theorem 174 (Laplacian comparison, theorem 2)
Let $U \subset (M, g)$ be a normal neighborhood with $\text{Ric} \geq (n-1)c$. Then $\Delta_r \leq (n-1)\frac{s_c'}{s_c}$ on $U_0 \setminus \{p\}$.

We can't just trace the Hessian comparison theorem for this direction – we've already traced at the level of curvature so we shouldn't expect that to do anything for us here.

Proof. We know that $\Delta r = \text{tr}(\mathcal{H}_r)$, and trace commutes with D_t, so we can trace the Riccati equation $D_t\mathcal{H}_r + \mathcal{H}_r^2 + R_{\gamma'} = 0$ to find

$$\frac{d}{dt}\Delta r + \text{tr}(\mathcal{H}_r^2) + \text{tr}(R_{\gamma'}) = 0.$$

We'll now compute some of these terms in normal coordinates. Choosing an orthonormal frame $E_1, \cdots, E_n$, we have

$$\text{tr}(R_{\gamma'}) = \sum_{i=1}^n \langle R_{\gamma'}(E_i), E_i\rangle = \sum_{i=1}^n \langle R(E_i, \gamma')\gamma', E_i\rangle,$$

and this is exactly the Ricci curvature $\text{Ric}(\gamma', \gamma')$, which is bounded from below by $(n-1)c$ by assumption. Now if we assume the frame diagonalizes $\mathcal{H}_r$ with $E_1 = \partial_r$ (it's a null vector so it's an eigenvector) and with eigenvalues λ_i, then

$$\text{tr}(\mathcal{H}_r^2) = \sum_{i=2}^n \langle \mathcal{H}_r(\mathcal{H}_r(E_i)), E_i\rangle = \sum_{i=2}^n \lambda_i^2,$$

and now by AM-GM (or Cauchy-Schwarz) we can bound this as

$$\text{tr}(\mathcal{H}_r^2) \geq \frac{(\sum_{i=2}^n \lambda_i)^2}{n-1} = \frac{(\Delta r)^2}{n-1}.$$

So if we take the differentiated Riccati equation and divide everything through by $(n-1)$, we get

$$\frac{d}{dt}\left(\frac{\Delta r}{n-1}\right) + \left(\frac{\Delta r}{n-1}\right)^2 + \tilde{S} = 0,$$

where $\tilde{S}$ is $\frac{1}{n-1}$ times the leftover part from Cauchy-Schwarz plus the $\text{Ric}(\gamma', \gamma')$ term. Thus $\tilde{S} \geq c$ – notice that the inequalities go in the wrong direction here if we were trying to bound by Ricci in the other direction. But now this is a one-dimensional Riccati equation (scalar-valued) and we can use Riccati comparison to $\frac{s_c'}{s_c}$ (which satisfies $H' + H^2 + c = 0$) to show that $\tilde{H} = \frac{\Delta r}{n-1} \leq H$, as desired. □

24 May 26, 2023

We'll continue with Ricci curvature today, using a proof technique from last time:

Theorem 175 (Conjugate point comparison, theorem 2)
Suppose (M, g) is a manifold with $\text{Ric} \geq \frac{n-1}{R^2}$ for some R. Then geodesic segments of length at least πR have conjugate points.

Proof. Suppose γ is contained in a normal neighborhood U of $\gamma(0)$. We've proven that $\Delta r = \partial_r \log(r^{n-1}\sqrt{\det g})$, and by the Laplacian comparison theorem we know this is at most $\partial_r \log(s_c(r)^{n-1})$. Now $\frac{r^{n-1}\sqrt{\det g}}{s_c(r^{n-1})}$ goes to 1 as $r \to 0$ and the function is decreasing, so $r^{n-1}\sqrt{\det g} \leq s_c(r)^{n-1}$ along γ. But at $r = \pi R$, the right-hand side is zero and the left-hand side is positive, which is a contradiction. So conjugate points must appear before then.

Now in general, let $w = \gamma'(0)$. We have $tw \in T_pM$ and can construct a star-shaped neighborhood containing all such points $\{tw\}$ and with $d\exp_p$ nonsingular on W (we're just fattening up the curve); now define $\tilde{g} = \exp_p^*(g)$ on W. We know that $\{tw\}$ is a geodesic for $\tilde{g}$ and has a conjugate point if its length is at least πR (by the previous part), so there is some Jacobi field $\tilde{J} \in \mathcal{J}^\perp(\tilde{\gamma})$ so that $\tilde{J}(0) = \tilde{J}(r_0) = 0$ for some $r_0 \in (0, \pi R]$. Consider

$$J(t) = (d\exp_p)_{tw}(\tilde{J}(t));$$

this is a vector field along the original geodesic, and it's a normal Jacobi field because $d\exp$ is a local isometry and thus the pictures are exactly the same. So γ must have had a conjugate point. □

Corollary 176

If $\text{Ric} \geq \frac{n-1}{R^2}$, then $\text{inj}(p) \leq \pi R$ for all p.

(Indeed, large injectivity radius implies long geodesics.)

Corollary 177

If (M, g) is **complete** and satisfies $\text{Ric} \geq \frac{n-1}{R^2}$, then it is compact and has diameter at most πR.

(Indeed, if there were points further apart, the completeness would allow us to connect them with a minimizing geodesic, but minimizing geodesics have no interior conjugate points.) This latter corollary can be proved in a more typical way than how it's covered in the textbook, so we'll discuss that as well:

Alternate proof of Corollary 177. Suppose we had two points $p, q \in M$ with $d_g(p, q) = L > \pi R$. Then there is a minimizing geodesic γ connecting p to q. By the second variation formula, we have for any proper variation vector field V that

$$\int_0^L |D_t V^\perp|^2 - R(V^\perp, \gamma', \gamma', V^\perp) dt \geq 0.$$

Choose a parallel orthonormal basis $\{E_1, \cdots, E_n\}$ along γ and choose $E_n = \gamma'$. Let $V = \sin(\pi L^{-1} t) E_i$; since V is already perpendicular for $i < n$, we have $D_t V^\perp = \frac{\pi}{L}\cos(\pi L^{-1} t) E_i$ (since the time-derivative of E_i is zero). Then

$$\frac{\pi^2}{L^2} \int_0 \cos\left(\frac{\pi}{L} t\right)^2 \geq \int \sin\left(\frac{\pi}{L} t\right)^2 R(E_i, E_n, E_n, E_i),$$

and now since we're summing over n we do get control over the Ricci curvature:

$$\frac{\pi^2}{L^2}(n-1) \int_0^L \cos\left(\frac{\pi}{L} t\right)^2 \geq \int_0^L \sin\left(\frac{\pi}{L} t\right)^2 dt \cdot \text{Ric}(E_n, E_n) \geq \frac{n-1}{R^2} \int_0^L \sin\left(\frac{\pi}{L} t\right)^2 dt,$$

and now the integrals are the same (by direct computation) so $\frac{\pi^2}{L^2} \geq \frac{1}{R^2}$ and thus $L \leq \pi R$. □

Theorem 178 (Bishop-Gromov)

Suppose $\text{Ric} \geq (n-1)c$. Let $V_g(\delta)$ be the volume of the **metric** ball of radius δ at p and $V_c(\delta)$ be the volume of the metric ball in the model. Then $\frac{V_g(\delta)}{V_c(\delta)}$ is nonincreasing as δ increases (up to $\delta < \text{inj}(p)$). If (M, g) is complete, this holds for all δ. Thus $V_g(\delta) \leq V_c(\delta)$, with equality implying that the metric ball has constant curvature c everywhere.

(Metric balls are geodesic balls until we hit the injectivity radius, and then after that point geodesic balls don't make sense anymore.) This is a big improvement from Günther's result – the theory of manifolds with Ricci bounds from below is quite rich, and this is basically the most important fact.

Proof. If $\delta < \text{inj}(p)$, we make use of the Laplacian comparison theorem. Recall that the injectivity domain $\text{ID} \subset T_pM$ is the set of "vectors before the cut time," and computing in normal coordinates (by identifying (T_pM, g) with $(\mathbb{R}^n, \overline{g})$ we have

$$V_g(\delta) = \int_{\text{ID}(p)\cap B_\delta(0)\subset T_pM} \sqrt{\det g}\, dV_g.$$

(Intuitively, we need to make sure we don't go past the cut points, because then there is a shorter radial path and thus we would be counting the same contribution to the volume twice.) The cut locus has measure zero, so we don't need to think about its contributions to the integral. So if we delete the cut locus from the metric ball, the exponential map is a diffeomorphism. We previously defined the decreasing function $\lambda = \frac{\sqrt{\det g}\rho^{n-1}}{s_c(\rho)^{n-1}}$; because of the direction our bound goes, we can extend it by zero outside of $\text{ID}(p)$ to get a function $\tilde{\lambda}$. Similarly we may extend s_c by zero for $\rho \geq \pi R$ to get $\tilde{s}_c$, so that like before we have

$$V_g(\delta) = \int_0^\delta \int_{S^{n-1}} \tilde{\lambda}\tilde{s}_c^{n-1} d\rho dV_{\mathring{g}}, \quad V_c(\delta) = \int_0^\delta \int_{S^{n-1}} \tilde{s}_c^{n-1} d\rho dV_{\mathring{g}}.$$

And now repeating the proof of Günther's theorem (but with inequalities reversed) allows us to compare the Euclidean area to the g area, showing that the ratio is indeed nonincreasing.

For the equality case, note that if $V_g(\delta) = V_c(\delta)$ for $\delta < \text{inj}(p)$, then we have equality in the Laplacian comparison theorem, which means that all of the eigenvalues in the Hessian besides the one for π_r are all equal. And we know that $\Delta r = (n-1)\frac{s_c'}{s_c}$ in the equality case, so the only option is that $\mathcal{H}_r = \frac{s_c'}{s_c}\pi R$ and that implies that the sectional curvature is c on all of $B_\delta(p)$.

When M is complete, the argument here shows that $\tilde{\lambda} = 1$ everywhere for $0 < \rho < \delta$ and $\tilde{s}_c > 0$. So if $c \leq 0$ in fact $\text{ID}(p) \cap B_\delta(0) = B_\delta(0)$ and we do have the entire geodesic ball, and otherwise if $\delta > \pi R$ we have the whole manifold M by the previous corollary and thus $V_g(\pi R) = V_c(\pi R)$. ☐

Corollary 179

If (M, g) is compact with $\text{Ric} \geq \frac{n-1}{R^2}$, then the volume of (M, g) is at most the volume of $S^n(R)$ (by taking $\delta = \pi R$ in the inequality $V_g(\delta) \leq V_c(\delta)$). In particular, Ricci curvature being bounded from below controls both volume and diameter from above.

Next lecture, our goal will be to prove things in the other direction, showing that curvature bounded from above gives us "bigness."

25 May 31, 2023

Our final topic of the class is **curvature and topology** – the point is to "relate local to global," since curvature is a local property (since we need only the metric in a neighborhood) and then piecing that together gives us global information. We've proven global results about geometric quantities like diameter so far, and we now want to go for something more topological.

We'll start with the constant sectional curvature case, skipping the proofs of some facts that we had to skip over earlier in the class:

Theorem 180 (Killing-Hopf)
Let (M, g) be a complete, connected, simply connected manifold (of dimension $n \geq 2$) with sectional curvature c everywhere. Then (M, g) is isometric to either $\mathbb{R}^n$, $S^n(R)$, or $\mathbb{H}^n(R)$.

The idea of the proof is as follows: we've showed that all points have neighborhoods isometric to balls in the corresponding model (by choosing normal coordinates and working out what the metric is – it ends up being the same as one of the models in polar coordinates). Now picking a base point p, we get an isometry from a neighborhood of p to a ball in the model, and now we need to construct a map globally to any other point q. What we do is consider some point p' on the way from p to q and look at the neighborhood and isometry (U', ϕ') at p'; these two isometries may not agree on the overlap, but if they do then we can keep gluing together until we get to q and this doesn't depend on the path (since being simply connected means the paths are homotopic). This thus gives us a local isometry, and **local isometries from complete manifolds are covering maps**. But the map from (M, g) to the model is a covering map and the models are all simply connected for $n \geq 2$, so it must be a bijection. (So the only thing we haven't done here is explain how we "adjust" the overlaps – for this we use frame-homogeneity of the model spaces, and it's a good strategy to be aware of.)

Corollary 181
Let (M, g) be a complete, connected manifold with sectional curvature c everywhere. Then (M, g) is isometric to the quotient of a model space by Γ, where Γ is a discrete subspace of the isometries of the model acting freely.

(Indeed, if we pass to the universal cover $\tilde{M}$ and pull back the Riemannian metric via the covering map, then it is a fact that $(\tilde{M}, \tilde{g})$ is complete. Then the rest is from the theory of covering spaces, and we can look up the details ourselves.) We call (model)/Γ a **space form** – the theory of space forms is interesting, and it turns into a question in group theory (with some geometry). When the model is $\mathbb{R}^n$, we call Γ a **crystallographic group** – these are not fully classified in higher dimensions yet. On the other hand, the problem is solved for S^n: in even dimensions we just have S^n and $\mathbb{RP}^n$, while in odd dimensions there are many (for example, thinking of S^3 as a subset of $\mathbb{C}^2$, we can act via $\mathbb{Z}/p\mathbb{Z}$ via $(z, w) \mapsto (e^{im}z, e^{in}w)$ for m, n relatively prime, leading us to **lens spaces**). Finally, it is an active field of study to understand what's happening in $\mathbb{H}^n$.

The "first" local-to-global theorem we should prove in most classes is the following:

Theorem 182 (Cartan-Hadamard)
Let (M, g) be a complete connected manifold with sectional curvature nonpositive everywhere. Then for all p, $\exp_p : T_pM \to M$ is a covering map, so in particular the universal cover $\tilde{M}$ is diffeomorphic to $\mathbb{R}^n$, and if $\pi_1(M) = 0$, then $\exp_p$ is the diffeomorphism.

Proof. If sectional curvature is nonpositive everywhere, then there are no conjugate points along any geodesic, and thus $\exp_p$ is a local diffeomorphism. Let $\tilde{g} = \exp_p^* g$; then $\tilde{g}$ is complete (since we can just check that the exponential map is defined at one point on the tangent space, and $\tilde{\exp}$ is defined on $T_0(T_pM)$ (take rays from the origin, which are geodesics). So now we have $(T_pM, \tilde{g}) \overrightarrow{\exp_p} (M, g)$ a local isometry from a complete manifold and therefore we have a covering map. □

Definition 183

If (M, g) is a complete connected manifold with $\pi_1(M) = 0$ and nonpositive sectional curvature everywhere, then we call M **Cartan-Hadamard**.

From the point of view of geometry, there is still something we can say – in particular, hyperbolic space and Euclidean space are both Cartan-Hadamard manifolds but they are very different.

Definition 184

Let $\gamma : \mathbb{R} \to (M, g)$ be a nonconstant geodesic parameterized on all of $\mathbb{R}$, such that $d(\gamma(t), \gamma(s)) = L(\gamma|_{[t,s]})$ (that is, we are minimizing between any two fixed points). Then we call γ a **line**.

Proposition 185

If (M, g) is Cartan-Hadamard, then we have the following:

1. The injectivity radius of M is infinite,
2. Any nonconstant geodesic is a line,
3. Any two points are contained in a line,
4. Metric balls are the same as geodesic balls,
5. If we pick a point q and define $r(x) = d_g(q, x)$, then r is C^∞ on $M \setminus \{q\}$ and r^2 is C^∞ everywhere.

Proof. (1) follows from the Cartan-Hadamard theorem, since $\exp_p$ is a diffeomorphism. Then there is a unique geodesic from p to q, so we can't have any other nonconstant geodesic between those two points, which shows both (2) and (3). Then (4) is clear because the injectivity radius is infinite, and for (5) we note that the entire manifold M is a normal neighborhood of q. □

Proposition 186

Suppose (M, g) is Cartan-Hadamard, and $A, B, C \in M$ are tangent vectors three points that are not collinear (meaning that the third point does not lie on the line through the other two). Let a, b, c be the lengths of BC, AC, AB, and use the usual notation $\angle C$ for the angle between the vectors in the tangent space at C from the two lines AC and BC. Then we have $c^2 \ge a^2 + b^2 - 2ab \cos \angle C$ and $\angle A + \angle B + \angle C \le \pi$.

Proof. The injectivity radius is infinite, so we may use normal coordinates at C. Letting $\overline{g}$ be the Euclidean metric in those coordinates, we have $g \ge \overline{g}$ by metric comparison, and the geodesic between A and B is not necessarily a straight line (because its not radial). If $\overline{a}$ is the distance $d_{\overline{g}}(C, B)$ and so on, then $c^2 \ge \overline{c}^2$ (since $g \ge \overline{g}$, and then the length of the curve in $\overline{g}$ is just the distance in ordinary coordinates) and by ordinary Law of Cosines this is equal to $\overline{a}^2 + \overline{b}^2 - 2\overline{a}\overline{b} \cos \angle C$. But now everything in this expression is the same with g or $\overline{g}$ because everything is along radial geodesics, and the metric is δ_{ij} at the origin in normal coordinates.

To prove the angle sum inequality, consider the comparison triangle in Euclidean space with the same lengths a, b, c as our original triangle. Since cos is decreasing, the g-angles are at most the $\overline{g}$-angles in the original triangle (since the g-angle $\cos \angle C \ge \frac{a^2+b^2-c^2}{2ab}$, and the right-hand side is $\cos \angle C$ in $\overline{g}$-comparison). Since the $\overline{g}$-angles sum to π, the g-angles must sum to at most π. □

This type of result is morally similar to "detecting curvature from the metric structure" – we could prove relations on triangles which is purely in terms of that metric space structure instead of the inner product, and we can look up Toponogov's theorem for this (it gives us a way to say when a metric space satisfy triangle-type inequalities, and this turns out to be quite important for taking "limits of manifolds").

As a result of Cartan-Hadamard, we also get a few other consequences:

Corollary 187

Suppose M is a compact manifold with $\pi_1(M) = 0$. Then there does not exist a metric on M with nonpositive sectional curvature everywhere.

(So for example, the sphere does not admit a metric with nonpositive sectional curvature everywhere.) And the idea is that we ideally can use this to classify the space of allowed metrics and push towards non-existence results.

Corollary 188

Let (M, g) be a complete manifold with nonpositive sectional curvature everywhere. Then the higher homotopy groups satisfy $\pi_k(M) = 0$ for all $k \geq 2$ (we call such a manifold **aspherical**).

Next time, we'll prove some further results, saying a bit more about the fundamental group "not being too small" and then moving on to the positive-curvature case as well.

26 June 2, 2023

Our first topic today will be relating π_1 and nonpositive sectional curvature. One extension of what we discussed last time is the following:

Proposition 189

If M is compact with nonpositive sectional curvature, then $|\pi_1(M)| = \infty$.

Proof. The universal cover $\tilde{M}$ is diffeomorphic to $\mathbb{R}^n$. □

We'll prove things in another direction now, showing that groups are "not too abelian" and thus not too small:

Lemma 190

Suppose (M, g) is Cartan-Hadamard. Let $f(x) = \frac{1}{2}d(x, p)^2$ for some point $p \in M$. Then f is strictly geodesically convex, meaning that for any nonconstant geodesic segment $\gamma : [0, 1] \to M$ of constant speed, we have

$$f(\gamma(t)) < (1 - t)f(\gamma(0)) + tf(\gamma(1)).$$

Proof. We'll work in normal coordinates at p, so that $f = \frac{1}{2}r^2$. By the Hessian comparison theorem, we can bound $\mathcal{H}_r \geq \frac{1}{r}\pi_r$ on $M \setminus \{p\}$, so away from p we have

$$\nabla(\text{grad}(f)) = \nabla(r\text{grad}(r)) = \text{grad}(r) \otimes dr + r\mathcal{H}_r \geq \text{grad}(r) \otimes dr + \pi_r,$$

and now the right-hand side is the identity endomorphism because it's true on the orthogonal component and along the radial component. Thus

$$\langle \mathcal{H}_f(v), v \rangle \geq |v|^2,$$

and the left-hand side is completely smooth even at p because we have the Hessian of a smooth function. Then

$$\frac{d}{dt}f(\gamma(t)) = \langle \text{grad}(f)|_{\gamma(t)}, \gamma'(t)\rangle, \quad \frac{d^2}{dt^2}f(\gamma(t)) = \langle D_t\text{grad}(f), \gamma'\rangle.$$

where we've used compatibility with D_t for the second derivative. And because we have an extendible vector field, this simplifies to

$$= \langle \mathcal{H}_f(\gamma'), \gamma'\rangle \geq |\gamma'|^2 > 0,$$

so the function is strictly convex on $[0, 1]$, implying the result. □

Lemma 191

Let (M, g) be Cartan-Hadamard, and let $S \subset M$ be a compact subset which contains at least two points. Then there is a unique closed ball of smallest radius that contains S.

(This is not true for the set of two diametrically opposite points on a sphere, so the Cartan-Hadamard assumption is necessary here.)

Proof. Existence is clear by taking smaller and smaller balls and using compactness and closedness. For uniqueness, suppose there are two smallest balls containing S $\overline{B}_r(p_1)$ and $\overline{B}_r(p_2)$. Let γ be a geodesic from p_1 to p_2 parameterized at constant speed on $[0, 1]$, and let $m = \gamma\left(\frac{1}{2}\right)$. Then for any $q \in S$, we have

$$\frac{1}{2}d(q, m)^2 < \frac{1}{2}(d(q, p_1)^2 + d(q, p_2)^2) \leq r^2$$

by strict convexity, so by compactness again there must be some $r' < r$ such that $S \subset \overline{B}_{r'}(m)$, a contradiction. □

Theorem 192 (Cartan fixed point)

Let G be a compact Lie group acting smoothly and isometrically on a Cartan-Hadamard manifold (M, g). Then there is some fixed point $p_0 \in M$, meaning that $gp_0 = p_0$ for all $g \in G$.

Again, this requires the Cartan-Hadamard assumption, since rotations on a sphere or translations on Euclidean space do not have fixed points.

Proof. Fix some $q_0 \in M$ and let $S = Gq_0$ be its orbit under the action. This is a compact set, since G is compact. If the orbit is q_0 itself, then we have found a fixed point. Otherwise, by the previous lemma, there is some unique $\overline{B}_r(p_0)$ containing S of minimal radius r. Now for any $g \in G$, because the group acts isometrically, we have $g \cdot \overline{B}_r(p_0) = \overline{B}_r(g \cdot p_0)$. Thus

$$\overline{B}_r(g \cdot p_0) \supset g \cdot S = S$$

(since S is an orbit). But now by uniqueness of the smaller-radius ball, this means we must have $gp_0 = p_0$. Since g was arbitrary, p_0 is a fixed point. □

Corollary 193 (Cartan torsion)

Let (M, g) be complete with nonpositive sectional curvature. Then $\pi_1(M)$ is torsion-free.

Proof. Let $(\tilde{M}, \tilde{g})$ be the universal cover of M. Then $M = \tilde{M}/\Gamma$, where $\Gamma \subset \text{Iso}(\tilde{M}, \tilde{g})$ is the group of decke transformations (diffeomorphisms that interchange the pieces of the cover) and $\Gamma \cong \pi_1(M)$. But now if ϕ is a torsion element, then $\langle \phi \rangle \subseteq \Gamma$ is a finite (cyclic) group that acts on $\tilde{M}$ smoothly and isometrically. So by the Cartan fixed point theorem, ϕ has a fixed point, and this can only happen for deck transformations if we have the identity element. $\square$

This puts some restrictions on $\pi_1(M)$ – in particular, not all manifolds will admit nonpositive sectional curvature. But we can do even better if we require strictly negative sectional curvature:

Theorem 194 (Preissmann)
Let (M, g) be a compact connected manifold with everywhere negative sectional curvature. Then any abelian subgroup $\Gamma \leq \pi_1(M)$ is $\mathbb{Z}$.

In particular, this means that something like T^2 (with $\pi_1(M) = \mathbb{Z}^2$) does not admit negative sectional curvature. To prove this result, we'll need to prove various other results:

Definition 195
Let $\phi \in \text{Iso}(M, g)$ be an isometry. We call a geodesic $\gamma : \mathbb{R} \to M$ an **axis** if the isometry satisfies $\phi(\gamma(t)) = \gamma(t+c)$ (for example, translations have an axis along the direction of translation, and dilations have the y-axis as an axis in the upper-half-space model for the hyperbolic space; however, translations have no axis).

Proposition 196
Let (M, g) be compact. Then there is a closed geodesic loop of least length in each free homotopy class of loops (where homotopies are maps from $S^1 \times [0, 1] \to M$, and where there is no required fixed basepoint).

Two proofs of proposition. This is a rather important result, so we'll show two different proofs. The first is to fix p and find geodesic segments $p \to p$ of least length (among all loops in the homotopy class that pass through p). This is easy to do by lifting any representative to the universal cover $\tilde{M}$; this yields a path from some $\tilde{p}$ to a well-defined endpoint $\tilde{p}'$. We can then choose the length-minimizing curve in the universal cover and then project it back down. Now vary p to achieve the smallest length (by compactness and continuity). We now need to check that this is a geodesic loop – it cannot have corners because we can "round the corner" by the first variation formula, and this preserves the homotopy class.

For a second proof strategy, we can do things more directly: choose a sequence of constant-speed curves limiting towards the minimum. In charts, these curves are Lipschitz, so we can get C^α-convergence to a Lipschitz curve by Arzela-Ascoli, and in such convergence length only goes down. So this allows us to find a Lipschitz curve achieving the minimum, and in normal coordinates we have radial segments so we have smoothness. $\square$

Lemma 197
Let (M, g) be compact and connected, and let $\pi : (\tilde{M}, \tilde{g}) \to (M, g)$ be the universal cover. Then any deck transformation $\phi \in \text{Iso}(\tilde{M}, \tilde{g})$ has an axis that (when projected down) covers a closed geodesic shortest in the homotopy class.

Proof sketch of lemma. Pick a point $\tilde{x} \in \tilde{M}$ and consider $\phi(\tilde{x})$ (which lies over the same base point in M as $\tilde{x}$). Choose a curve $\tilde{\gamma}$ from $\tilde{x}$ to $\phi(\tilde{x})$, which quotients to a loop γ starting and ending at x in M. This is not homotopically trivial if ϕ is nontrivial, so we can pick some γ_1 of least length in the homotopy class containing some point 0. Looking at its lift, which is a path from some point $\tilde{\gamma}_1(0)$ to $\tilde{\gamma}_1(1)$, we claim that $\tilde{\gamma}_1(1) = \phi(\tilde{\gamma}_1(0))$. To prove that this is true, we can try varying the curve γ a little bit and seeing that $\tilde{x}$ and $\phi(\tilde{x})$ vary in the same way in a small neighborhood, and then use an open-closed argument. So now we can extend $\tilde{\gamma}_1$ to a curve $\mathbb{R} \to (\tilde{M}, \tilde{g})$ (it's a complete manifold so we can extend the geodesic segment), and we just need to show that $\phi(\tilde{\gamma}_1(t)) = \tilde{\gamma}_1(t+1))$, meaning that we've found an axis. We already know this holds for $t = 0$, and it suffices to show this holds for the velocity vector by uniqueness of geodesics. But this is true because the velocity vector at the beginning and end are the same in the original curve γ_1. □

Lemma 198

Let (M, g) be Cartan-Hadamard with negative sectional curvature everywhere. Then axes are unique up to reparameterization.

Proof of lemma. Assume we have two axes γ_1, γ_2 that don't intersect for some isometry ϕ. Let A, C be points on γ_1, γ_2, respectively, and $B = \phi(A), D = \phi(C)$. Then we can draw a geodesic σ from A to C and a geodesic $\phi(\sigma)$ from B to D, forming a "parallelogram" $ABDC$. The angles sum to 2π (since the corresponding angles at A and B are equal because ϕ is an isometry), but triangles ABC and BCD each of angle sum less than π, which is a contradiction because $\angle ACD \le \angle ACB + \angle BCD$ and $\angle ABD \le \angle ABC + \angle CBD$.

On the other hand, if we have two axes that intersect, suppose $\gamma_1(0) = \gamma_2(0)$. Then we have $\gamma_1(1) = \gamma_2(1)$ as well, but in a Cartan-Hadamard manifold this can only happen if $\gamma_1 = \gamma_2$. □

Next time, we'll finish the proof of Preissmann's theorem and discuss positive curvature (and then we'll discuss some more "modern state of the field" topics on the last day of class).

27 June 5, 2023

We'll start by proving Preissmann's theorem, which follows from the work we've already done. Recall that the Cartan torsion theorem already tells us that $\pi_1(M)$ cannot have $\mathbb{Z}/p\mathbb{Z}$ subgroups if M has negative (in fact nonpositive sectional curvature):

Proof of Theorem 194. Let H be an abelian subgroup of $\pi_1 M$; recall that we can always think of this as a subgroup of the Deck transformations of the universal cover $(\tilde{M}, \tilde{g})$. We've proven that for any non-identity $\phi \in M$ there is always an axis, and we used negative sectional curvature and triangle comparison to show that this is unique. So we have $\phi(\gamma(t)) = \gamma(t + c)$ under some unit-speed parameterization of the axis γ, and by abelianness we have for any $\psi \in H$

$$\psi(\phi(\gamma(t))) = \phi(\psi(\gamma(t))) = \psi(\gamma(t + c)),$$

so $\psi \circ \gamma$ is also a geodesic which is also an axis for ϕ, meaning that $\psi \circ \gamma$ **is** γ up to reparameterization $\gamma(t + a)$ (it's still unit-speed because we have an isometry, and it cannot be $\gamma(-t + a)$ because for $t = \frac{a}{2}$ we would have $\psi(\gamma(a/2)) = \gamma(a/2)$, but non-identity Deck transformations don't have fixed points). This thus defines a map $F : H \to \mathbb{R}$ (with additive structure of the reals) mapping each ψ to its corresponding a (and mapping the identity to zero), which is a homomorphism. Furthermore, if $F(\psi) = 0$, then each point of the geodesic is fixed and thus ψ is the identity. So F maps H isometrically onto $F(H)$, and we can define $b = \inf\{a \in F(H) : a > 0\}$. We know that each axis $\pi(\gamma)$ covers a closed geodesics in M, and the length of closed geodesics in (M, g) is at least $2\text{inj}(M)$ (because otherwise there are two paths between opposite points on the geodesic). Thus b is isolated from elements of $F(H)$ that are bigger than b (we cannot have points approaching the infimum), so in fact b is in $F(H)$. And now we claim that b generates $F(H)$, because otherwise we can use the division algorithm to get something strictly between 0 and b. □

Remark 199. *In the case $n = 2$, by Gauss-Bonnet having curvatures positive, zero, and negative correspond to having S^2, the torus, and the multi-holed tori, which have $\pi_1 = 0$, $\pi_1 = \mathbb{Z}^2$, and π_1 "exponentially big." There's a story in three dimensions as well, and we'll talk about it next time if we have time.*

We'll now turn to **positive** curvature – we've met our first topological result for this already. We previously proved that if (M, g) is complete and connected with $\text{Ric} \geq \frac{n-1}{R^2}$ everywhere. Then M is compact with diameter at most πR, and $|\pi_1(M)| < \infty$ (since the universal cover must also be compact because it satisfies the same assumptions, and then each point in the preimage must have a small ball disjoint from the others, so we can only have finitely many points in this preimage). This theorem is sharp in the sense that S^n satisfies this results, and it's interesting to ask whether this is the only example that attains inequality.

Theorem 200 (Cheng's maximal diameter theorem)
Let (M, g) be a complete manifold with $\text{Ric} \geq \frac{n-1}{R^2}$ and $\text{diam}(M) = \pi R$. Then M is isometric to $S^n(R)$.

(We'll skip the proof for now and maybe talk about it next time if we have time.)

Remark 201. *We may ask about the "nearly sharp" case too, where we have a bound $|diam(M) - \pi R| < \varepsilon$ instead for some $\delta > 0$. Then we may ask whether M is $\varepsilon(\delta)$-closed to $S^n(R)$ under some appropriate metric – we'll say a bit about this next time.*

If we weaken $\text{Ric} \geq \frac{n-1}{R^2}$ to $\text{Ric} \geq 0$, we still have some strong control – on the homework, we will prove the following:

Theorem 202 (Milnor)
Let (M, g) be a complete connected manifold when $\text{Ric} \geq 0$, and suppose Γ is a finitely generated subgroup of $\pi_1(M)$. Then S is a finite generating set, then we have the "polynomial growth"

$$\left|\left\{g \in \Gamma : g = \prod_i^m s_i, s_i \in S\right\}\right| \leq Cm^n.$$

Milnor conjectured that π_1 itself is always finitely generated – this is true for $n = 3$ (proved by Yau and Liu) but there turns out to be a counterexample for $n = 7$ (by Brué, Naber, and Semola, showing that $\pi_1 = \mathbb{Q}/\mathbb{Z} \subset S^1$).

Lots of problems in this area of math are still open, but we can prove the following:

Theorem 203 (Symge)
Let (M, g) be compact with positive **sectional** curvature (which is stronger than a Ricci bound). If n is even and M is orientable, then $\pi_1 = 0$, and if n is odd then M is orientable.

Proof. In the first case, suppose there is some nonzero class $[\gamma]$ (we're not being too careful with basepoints here yet) in π_1. Then there is some least-length representative in the homotopy class which is a closed geodesic loop. (Indeed for each fixed point, we can find the least-length loop via the universal cover, and if we pick the best base point it will work out.) Then the second variation of length says that

$$\int_{S^1} |D_t V|^2 - R(V, \gamma', \gamma', v) \geq 0$$

for all vector fields $V \in \mathfrak{X}^{\perp}(\gamma)$. Now if we pick some basepoint $p \in \gamma$ and consider parallel transform $\tilde{P} : T_pM \to T_pM$ along the curve, then $\tilde{P}(\gamma'(0)) = \gamma'(0)$ (because the velocity vector is parallel) and thus we can restrict $\tilde{P}$ to the normal space $P : N_p\gamma \to N_p\gamma$ because parallel transport preserves inner products. Now $\tilde{P}, P$ are isometries and preserve orientation on $T_pM, N_p\gamma$ respectively, and in particular $\det P = 1$. Since $N_p\gamma$ has odd dimension, it must have some eigenvector of 1 (they cannot all be -1), say v. Then define the vector field via $V = P_t(v)$ (it must close back up smoothly because it solves an ODE); we have $D_t V = 0$ and thus we have second variation $\int -\mathrm{Rm}(V, \gamma', \gamma', V)$, which must be negative by assumption on our sectional curvature, which is a contradiction because we are length-minimizing.

We'll just sketch the second case: if M were non-orientable, then there would be a free homotopy class of loops which reverses orientation. So now we do the same thing, where we're again in a length-minimizing loop but now with $\det P = -1$ and with $N_p\gamma$ an even-dimensional space. So again there must be a $+1$ eigenvector and we can proceed as before. □

Corollary 204

Let (M, g) be compact and connected with positive sectional curvature and n even. If M is orientable, then $\pi_1 = 0$, and if it is non-orientable, then $\pi_1 = \mathbb{Z}/2\mathbb{Z}$.

On the other hand, in odd dimensions we can have **lens spaces** and thus the behavior is indeed more complicated.

Corollary 205

There is no metric on the even-dimensional space $\mathbb{RP}^n \times \mathbb{RP}^n$ with positive sectional curvature everywhere, even though the standard metric on $\mathbb{RP}^n$ does have positive sectional curvature.

Along these lines with the distinction between positive and nonnegative sectional curvature, we have the following:

Conjecture 206 (Hopf)

$S^2 \times S^2$ does not have positive sectional curvature.

This is probably the most famous problem in all of geometry now, but it is still unsolved despite lots of attempts!

28 June 7, 2023

We'll begin by proving Theorem 200, which is a nice result using various things we've proved in this class:

Proof. The setup here is that we have a complete manifold (M, g) with $\mathrm{Ric} \geq \frac{n-1}{R^2}$, which implies that the manifold is compact (note that just being positive everywhere is not enough – we need a uniform bound). Then we want to prove that if the diameter is exactly πR, then M is isometric to $S^n(R)$.

By compactness, we can find two points p_1, p_2 attaining the diameter with $d(p_1, p_2) = \pi R$. Take two values $\delta_1 + \delta_2 = \pi R$ and consider the metric balls $B_{\delta_1}(p_1)$ and $B_{\delta_2}(p_2)$; these open balls are disjoint by the triangle inequality, so we have

$$\mathrm{vol}(M) \geq \mathrm{vol}(B_{\delta_1}(p_1)) + \mathrm{vol}(B_{\delta_2}(p_2)).$$

Thus by Bishop-Gromov (monotonicity for the volume ratios, even past the injectivity radius), the ratio $\frac{\text{vol}(B_\delta(p_i))}{V_c(\delta)}$ is nonincreasing, and where $V_c(\delta)$ is the volume of the ball in $S^n(R)$. This means that

$$\frac{\text{vol}(B_{\delta_i}(p_i))}{V_c(\delta_i)} \geq \frac{\text{vol}(B_{\pi R}(p_i))}{V_c(\pi R)} = \frac{\text{vol}(M)}{|S^n(R)|},$$

because all we're missing from M is the cut locus, which is a set of measure zero. Thus, plugging this into our first inequality, we have

$$\text{vol}(M) \geq \frac{\text{vol}(M)}{|S^n(R)|}(V_c(\delta_1) + V_c(\delta_2)) = \frac{\text{vol}(M)}{|S^n(R)|}|S^n(R)| = \text{vol}(M),$$

where the middle equality comes from considering the balls of radius δ_1, δ_2 at two antipodal points on the sphere (this covers all of $S^n(R)$ up to a set of measure zero). So we must have had equalities everywhere, and rewriting using the closed balls instead of the open balls (which is changing again by a set of measure zero), this means

$$\text{vol}(M) = \text{vol}(\overline{B}_{\delta_1}(p_1)) + \text{vol}(\overline{B}_{\delta_2}(p_2)).$$

Additionally, we also know from the Bishop-Gromov equality case that $\text{vol}(M) = \text{vol}(S^n(R))$. Now we claim that for any $p \in M$, we have $d(p_1, p) + d(p, p_2) = \pi R$. Indeed, the $\geq$ direction is true by the triangle inequality, and if we had $d(p_1, p) + d(p, p_2) > \pi R$ for some p, we could choose $\delta_1 + \delta_2 = \pi R$ such that p is not in either closed ball $\overline{B}_{\delta_i}(p_i)$ (by choosing δ_i slightly smaller than $d(p_i, p)$). And now we have a contradiction – since these balls are compact, we then have some $\varepsilon > 0$ such that $B_\varepsilon(p)$ is disjoint from those balls, and that must add more to the volume of M, a contradiction.

So now consider any unit-speed geodesic γ starting at p_1, and choose some $\varepsilon > 0$ such that $\gamma|_{[0,\varepsilon]}$ is length-minimizing. We know that $d(\gamma(\varepsilon), p_2) = \pi R - \varepsilon$ (by what we just proved), and now we can choose some minimizing geodesic σ from $\gamma(\varepsilon)$ to p_2 of this length $\pi R - \varepsilon$. The concatenation of $\gamma_{[0,\varepsilon)}$ and σ is then a curve from p_1 to p_2 of length πR, which is also the distance, so it must be a minimizing geodesic and in fact there is no kink at ε so $\sigma = \gamma|_{[\varepsilon,\pi R]}$ by uniqueness of ODEs.

Thus the injectivity radius of p_1 is at least πR (otherwise some radial geodesic would stop minimizing, which we just proved does not occur), and it is also at most πR by cut locus comparison. So $M \setminus \{p_2\}$ is a **geodesic** ball $B_{\pi R}(p_1)$, and similarly $M \setminus \{p_1\}$ is $B_{\pi R}(p_2)$. Now if we define the functions $r_i = d(\cdot, p_i)$, then $r_1 + r_2 = \pi R$, and by Laplace comparison on $M \setminus \{p_1, p_2\}$ we have

$$\Delta r_1 \leq (n-1)\frac{s_c'(r_1)}{s_c(r_1)} = -(n-1)\frac{s_c'(r_2)}{s_c(r_2)} \leq -\Delta r_2$$

with the middle step coming from the explicit definition of s_c. But because $r_1 + r_2 = \pi R$, $-\Delta r_2 = \Delta r_1$, so equalities everywhere means that $M \setminus \{p_1, p_2\}$ has sectional curvature equal to c everywhere. Since sectional curvature is continuous, it's c everywhere, and by Killing-Hopf that means we are isometric to $S^n(R)/\Gamma$. Quotienting divides the volume but $\text{vol}(M) = \text{vol}(S^n(R))$, so we must have $S^n(R)$ itself. □

Following up on Remark 201, it turns out that having diameter $\pi R - \varepsilon$ does not mean we need to have something close to $S^n(R)$. The metric on S^n is a warped product $dt^2 + \sin^2 t g_{S^{n-1}}$ with a lower-dimensional sphere (we can think of this as the "spherical suspension" of S^{n-1}), but if we consider the spherical suspension of $\mathbb{RP}^n$, locally this is isometric to S^n and thus we have the same Ricci curvature. But this won't be a smooth manifold – we get two singular points at $t = 0$ and $t = \pi$ which we cannot close up, and this is not just a coordinate charts issue. So we can glue something there, and we will get something topologically and geometrically different from $S^n(R)$ that still has approximately correct diameter. However, M is indeed (in a very weak sense, specifically Gromov–Hausdorff) close to a spherical suspension – this counterexample is essentially the only one there is.

We'll close with some thoughts on "what more we could have covered in this class if we have more time," followed by some modern developments:

Theorem 207 (Berger, Klingenberg (1960s))
Let (M, g) be a compact manifold with $\pi_1(M) = 0$ and all sectional curvatures satisfying $1 \leq \sec(\Pi) < 4$. Then M is homeomorphic to S^n.

On the other hand, $\mathbb{CP}^n$ with the Fubini-Study metric has sectional curvatures between 1 and 4, attaining both endpoints. The proof is quite involved but mostly uses tools we've developed plus a few strengthenings – there is some Morse theory required for various dimensions though.

Theorem 208 (Cheeger-Gromoll, "Soul Theorem" (1972))
Let (M, g) be a complete connected noncompact manifold with nonnegative sectional curvature. Then there exists a compact totally geodesic submanifold S, called the "soul," so that M is diffeomorphic to the normal bundle of S.

This is basically a result saying that "we understand sectional curvature topologically" – the simplest example is where S is a paraboloid and the soul is a point, or where S is a cylinder and the soul is a circle.

Theorem 209 (Cheeger-Gromoll (1971))
Let (M, g) be a complete connected manifold with $\text{Ric} \geq 0$ which contains a line. Then M is isometric to $N \times \mathbb{R}$, where N is a Riemannian $(n-1)$-manifold with $\text{Ric} \geq 0$.

Finally, we'll say a few words about **Ricci flow** – the idea is that PDEs are used in modern geometry and analysis does enter the picture. We've mentioned the equation $\text{Ric} = 0$ (the Einstein equation), which is an elliptic PDE in a certain sense (like $\Delta u = 0$), and we could also study the equation

$$\frac{\partial g}{\partial t} = -2\text{Ric}_{g(t)}$$

where g is a time-dependent metric. Since a symmetric 2-tensor at each point lives in a vector space, the time-derivative is easy to take and the left-hand side makes sense, and this is in fact a parabolic differential equation if interpreted correctly (like $\partial_t u = \Delta u$).

Lemma 210
For any $p \in (M, g)$, there exists a chart U centered at p with $\Delta g x^i = 0$. (These are called **harmonic coordinates**, and they are in some sense better than normal coordinates.)

In harmonic coordinates, the Ricci curvature (by a computation) takes the form

$$\text{Ric}_{ij} = -\frac{1}{2}\Delta g_{ij} + O(\partial g^2, g^2),$$

which explains the $-\frac{1}{2}$ factor and explains why the equation $\text{Ric} = 0$ is like requiring that g is harmonic.

Theorem 211
Let (M^3, g) be a compact, simply connected manifold with $\text{Ric} > 0$. Then M is diffeomorphic to S^3.

Proof sketch. Take g and flow by Ricci flow (we must prove that a solution to this PDE exists); a Ricci-positive metric may be "bumpy," but the flow essentially smooths out the bumps and makes g more round. However, the flow also "shrinks things:" for example if we have a "changing-radius" equation $g(t) = R(t)^2 \mathring{g}$, we have $\text{Ric}_{g(t)} = \text{Ric}_{\mathring{g}} = (n-1)\mathring{g}$, so the Ricci flow equation becomes

$$(R^2)' = -2(n-1) \implies R(t)^2 = R(0)^2 - 2(n-1)t$$

and thus the sphere will be shrunk away in finite time (which is different from what happens for the heat equation). So instead the idea is that Ricci flow will shrink us to a "round point," where zooming in gives us something very round and we have to figure out how to rephrase this. □

Hamilton later proved something similar for M^4 with positive curvature operator, and then later the following results were proven:

Theorem 212 (Brendle-Schoen (2009))
Let (M, g) be a manfiold with $\pi_1 = 0$ and sectional curvature satisfying $1 \le \sec(\Pi) < 4$. Then (M, g) is actually **diffeomorphic** to S^n.

This then leads us to the most famous result, the Poincaré conjecture:

Theorem 213 (Perelman)
Let M^3 be a compact manfiold with $\pi_1 = 0$. Then M is diffeomorphic to S^3 (so the Ricci positive assumption is not necessary).

The idea for why this is more difficult than Hamilton's theorem is that when we perform Ricci flow, the "skinny parts" will pinch off faster. So we won't get rounder and rounder and then disappear – instead, we have to stop and do "surgery" before we continue to flow. And there is in fact more with what was proved involving the geometrization of 3-manifolds, but that's something we can read on our own.