Riemannian Geometry Notes

Simon Xiang

April 11, 2021

Notes for the Spring 2021 graduate section of Riemannian Geometry (Math 392C) at UT Austin, taught by Dr. Sadun. The course somewhat follows *Introduction to Riemannian Manifolds* (2^{nd} edition), by Lee. Source files: https://git.simonxiang.xyz/math_notes/files.html

Contents

Part I
Lee: Riemannian Manifolds

Lecture 1

Chapter 1: What is Curvature?

Introduction

These are supplementary notes, following Lee's *Introduction to Riemannian Manifolds*. To get an idea of what we're studying, Chapter 1 will start from the roots and give a high level overview of the material.

Geometry as a mathematical discipline stems from Euclidian plane geometry, the stuff you learned in middle school. Its elements are points, lines, distances, angles, and areas: the notion of equivalence comes from **congruence—** two plane figures are congruent if they can be transformed into each other by a **rigid motion of the plane**, a bijective transformation from the plane to itself preserving distance. Some theorems:

Side-Side-Side Theorem. *Two Euclidian triangles are congruent iff the lengths of their corresponding sides are equal.*

Angle-Sum Theorem. *The sum of the interior angles of a Euclidian triangle is π.*

These two seemingly simple theorems illustrate two major types of results in geometry, we call them "classification theorems" and "local-to-global theorems". The SSS theorem is a *classification theorem*. Such a theorem tells us how to determine whether two objects are equivalent. Ideal classification theorems list computable invariants and says objects are equivalent iff these invariants match. The angle-sum theorem relates a local geometric property (angle measure) to a global property (being a triangle). Most of the theorems we study are *local-to-global theorems*.

After studying points and lines, we can talk about circles. Here we state two theorems, one is a classification theorem, while the other is a local-to-global theorem (it will become clear why with time).

Circle Classification Theorem. *Two circles in the Euclidian plane are congruent iff they have the same radius.*

Circumference Theorem. *The circumference of a Euclidian circle of radius R is $2\pi R$.*

1.1 Curvature

If we want to study more stuff, we'll have to talk about curves in the plane. Arbitary curves don't vibe well with things like length and radius, so we have a new basic invariant called *curvature*, defined using calculus and is a function of position on the curve.

Formally, the **curvature** of a plane curve γ is defined as $\kappa(t) = |\gamma''(t)|$, the length of the acceleration vector, when γ is given a unit-speed parametrization. This is how we think about curvature geometrically: Given a point $p = \gamma(t)$, there are several circles tangent to γ at p, namely the circles whose velocity vector at p is the same as that of γ when both are given unit-speed parametrizations. The center of these circles lie on the line passing through p orthogonal to $\gamma'(p)$. Among these circles, there is exactly one unit-speed parametrized circle whos acceleration vector at p is the same as γ, it is called the **osculating circle**. (If acceleration is zero, replace the osculating circle by a straight line, a "circle with infinite radius"). The curvature is then $\kappa(t) = 1/R$, where R is the radius of the osculating circle. The larger the curvature , the greater the acceleration, the smaller the radius, and therefore the faster the curve is turning. A circle of radius R has constant curvature $\gamma \equiv 1/R$, while a straight line has a curvature of zero. All of this makes much more sense with a figure TODO

It is often convenient to extend the definition of curvature to allow positive and negative values, we do this by choosing a continuous unit normal vector field N along the curve, and assigning the curvature a positive sign if the curve is facing the normal vector and a negative sign if it's facing away. The resulting function κ_N along the curve is then called the **signed curvature**. We state two theorems about plane curves.

Plane Curve Classification Theorem. *Suppose γ and $\widetilde{\gamma}\colon [a,b] \to \mathbb{R}^2$ are smooth, unit-speed plane curves with unit normal vector fields N and $\widetilde{N}$, and $\kappa_N(t), \kappa_{\widetilde{N}}(t)$ represent the signed curvatures at $\gamma(t)$ and $\widetilde{\gamma}(t)$, respectively. Then γ and $\widetilde{\gamma}$ are congruent by a direction-preserving congruence iff $\kappa_N(t) = \kappa_{\widetilde{N}}(t)$ for all $t \in [a,b]$.*

Total Curvature Theorem. *If $\gamma\colon [a,b] \to \mathbb{R}^2$ is a unit-speed simple closed curve such that $\gamma'(a) = \gamma'(b)$, and N is the inward pointing normal, then*

$$\int_a^b \kappa_N(t)\, dt = 2\pi.$$

The first theorem is a classification theorem, while the second is a local-to-global theorem, relating the local property of curvature to the global (topological) property of being a simple closed curve. These generalize the circle theorems: two circles are congruent if they have the same curvature (radius), and if a circle has curvature κ and circumference C, then $\kappa C = 2\pi$.

1.2 Surfaces in Space

The natural next step is to move to three dimensions, that is, the study of general curve surfaces in space (2-dimensional embedded submanifolds of $\mathbb{R}^3$). The invariant is curvature, but it gets more complicated since a surface can curve differently in different directions.

Curvature in space is described by two numbers at each point, called the **principal curvatures**. Suppose S is a surface in $\mathbb{R}^3$, p is a point on S, and N is a unit normal vector to S at p. Here's a rough outline of how to compute principal curvature.

1. Choose a plane Π passing through p and parallel to N. The intersection of Π with a neighborhood of p in S is a place curve $\gamma \subseteq \pi$ containing p.

2. Compute the signed curvature κ_N of γ at p with respect to the chosen unit normal N.

3. Repeat *for all* normal places Π. The **principal curvatures of S at p**, denoted by minimum and maximum signed curvatures obtained.

Principle curvatures give us information about geometry, but don't answer the a paramount question in Riemannian geometry: Which properties of a surface are *intrinsic*? A property of surfaces is *intrinsic* if it is preserved by *isometries*, maps between surfaces preserving lengths of curves. To see that principle curvature isn't intrinsic, consider the embedded surface S_1, S_2 in $\mathbb{R}^3$, where S_1 is the square in the xy-plane with $0 < x < \pi, 0 < y < \pi$, and S_2 is the half-cylinder $\{(x,y,z) \mid z = \sqrt{1-y^2}, 0 < x < \pi, |y| < 1\}$. The principal curvatures of S_1 are zero, while the principal curvatures of S_2 are $\kappa_1 = 0$ and $\kappa_2 = 1$. But the map sending $(x,y,0)$ to $(x, \cos y, \sin y)$ is a diffeomorphism from S_1 to S_2, and thus an isometry.

Principle curvatures may not be intrinsic, but Guass discovered that a particular combination of them is, that is, the product $K = \kappa_1\kappa_2$ (known as the *Gaussian curvature*) is intrinsic. He named it *Theorema Egregium*, meaning "remarkable theorem". To get an idea of how Gaussian curvature works, first note that the square and half-cylinder have the same Gaussian curvature of zero (which is true by *Theorema Egregium* since they are isometric). A sphere of radius R has positive Gaussian curvature $1/R^2$, since each plane intersects the sphere in a great circle of radius R, and so the principal curvatures are $\pm 1/R \implies K = \kappa_1\kappa_2 = 1/R^2$. Similarly, "dome-shaped" objects have positive Gaussian curvature, since two principal curvatures always have the same sign, while "saddle-shaped" objects have negative Gaussian curvatures.

Model spaces of surface theory have constant Gaussian curvature. We have already seen two: Euclidian space $\mathbb{R}^2$ ($K = 0$), and the sphere of radius R ($K = 1/R^2$). The most important model surface with constant negative Gaussian curvature is the **hyperbolic plane**, whicn we'll talk about later. We state two theorems, you know the drill.

Uniformization Theorem. *Every connected 2-manifold is diffeomorphic to a quotient of one of the constant-curvature model surfaces described above by a discrete group of isometries without fixed points. So every connected 2-manifold has a complete Riemannian metric with constant Gaussian curvature.*

Gauss-Bonnet Theorem. *Suppose S is a compact Riemannian 2-manifold. Then*

$$\int_S K\,dA = 2\pi\chi(S),$$

where $\chi(X)$ is the Euler characteristic of S.

The uniformization theorem replaces the problem of classifying surfaces with classifying certain discrete groups of the models. Usually the uniformization theorem is state differently and proved with complex analysis. The Gauss-Bonnet theorem is a pure theorem of differential geometry, and arguable the most fundamental and important of them all. It relates a local geometric property (curvature) with a global topological invariant (the Euler Characteristic).

Together, these theorems place strong restrictions on the types of metrics that can occur on a given surface. For example, a consequence of Gauss-Bonnet is that the only compact, connected, orientable surface that admits a metric of strictly positive Gaussian curvature is the sphere. On the other hand, if a compact, connected orientable surface has nonpositive Gaussian curvature, Gauss-Bonnet rules out the sphere, and the uniformization theorem tells us that its universal covering space is homeomorphic to the plane.

1.3 Curvature in Higher Dimensions

Curvature becomes a lot more complicated in higher dimensions since manifolds can curve in all sorts of crazy ways. Our first issue is that in general, Riemannian manifolds don't present themselves as embedded submanifolds of Euclidian space. So we can't cut out curves by intersecting planes. However, **geodesics**—curves that are the shortest path between two points, help with our case. Examples are straight lines in Euclidian space and great circles on a sphere.

Suppose M is an n-dimensional Riemannian manifold. The most fundamental fact about geodesics is that given any $p \in M$ and any vector v tangent to M at p, there is a unique geodesic starting at p with inital velocity v. Here's a brief recipe for computing curvatures at some $p \in M$:

1. Choose a 2-dimensional subspace Π of the tangent space to M at p.
2. Look at all the geodesics through p whose initial velocities lie in Π. It turns out that near p these sweep out a certain 2-dimensional submanifold S_Π of M, which inherits a Riemannian metric from M.
3. Compute the Gaussian curvature of S_Π at p, which *Theorema Egregium* tells us can be computed from the inherited Riemannian metric. This associates a number, denoted sec(Π), called the **sectional curvature** of M at p, with the plane Π.

So the "curvature of M at p has to be interpreted as a map $\sec\colon \{\text{2-planes in } T_pM\} \to \mathbb{R}$. We again have three classes of constant (sectional) curvature model spaces: $\mathbb{R}^n$ with its Euclidian metric (for which $\sec \equiv 0$); the n-sphere of radius R, with the Riemannian metric metric inherited from $\mathbb{R}^{n+1}$ ($\sec \equiv 1/R^2$); and hyperbolic space of radius R (with $\sec \equiv -1/R^2$). Unfortunately, we have no satisfactory uniformization theorem for Riemannian manifolds in higher dimensions. In general, it is *not* true that every manifold has a metric of constant sectional curvature.

Characterization of Constant-Curvature Metrics. *The complete, connected, n-dimensional Riemannian manifolds of constant sectional curvature are, up to isometry, exactly the Riemannian quotients of the form $\widetilde{M}/\Gamma$, where $\widetilde{M}$ is a Euclidian space, sphere, or hyperbolic space with constant sectional curvature, and Γ is a discrete group of isometrics of $\widetilde{M}$ acting freely on $\widetilde{M}$.*

On the other hand, we have a number of power local-to-global theorems, which can be thought of generalizations of Gauss-Bonnet in various directions. They are consequences of the fact that positive curvature makes geodesics converge, while negative curvature makes them spread out.

Cartan-Hadamard Theorem. *Suppose M is a complete, connected Riemannian n-manifold with all sectional curvatures less than or equal to zero. Then the universal covering space of M is diffeomorphic to $\mathbb{R}^n$.*

Myer's Theorem. *Suppose M is a complete, connected Riemannian manifold with all sectional curvatures bounded below by a positive constant. Then M is compact and has a finite fundamental group.*

You can see that these theorems generalize the uniformization and Gauss-Bonnet, although not their full strength. Our goal with this course is to prove the three aforementioned theorems, among others; it is a primary goal of Riemannian geometry to improve upon and generalize the results of surface theory to higher dimensions.

Part II
Class Notes

Lecture 2

January 19, 2021

What is Riemannian geometry?? Consider $\{(x, y) \mid x^2 + y^2 < 1\}$, this is a coordinate chart. This doesn't tell us anything about the geometry of the surface, since $z = \sqrt{1-x^2-y^2}$ and $z = x^2 - y^2$ have the same coordinates. Geometry is extra data on a surface, besides the topological data. Our basic extra ingredient is the inner product, which allows us to talk about length, which tells us all sorts of things about the manifold.

2.1 Introduction to curvature

"In fact, with the inherited inner product on $\mathbb{R}^2$, this curved string is *not* curved!"

The map sending a curved string to a straight line is an isometry. So there are two different notions of curvature: extrinsic curvature talks about to what extent is something bent relative to something else, and intrinsic talks about what happens if you look locally. Locally, all 1-manifolds are isometric to $\mathbb{R}$.

Now to 2-manifolds. We have the circle, half-sphere, saddle, and half-cylinder discussed earlier. We can't really tell the difference between a cylinder and the plane, they're isometric. Extrinsically they're different, but intrinsically they're the same. The sphere is different— a circle has circumference $2\pi R$, that is, the circumference is the set of points of distance R away. What is the circumference of a sphere? Use spherical coordinates: a circle of radius R is a latitude line, and the length is 2π times the distance of the great circle cut out at R, or $2\pi\sin(R)$ if you draw out the angles. Approximately, $2\pi\sin(R) \approx 2\pi(R - \frac{R^3}{6})$. So circles are too small!

This means that spheres are in some sense, sphere shaped. Another thing is the area: the area of the cap is $A = \int 2\pi\sin(R)dR = 2\pi(1-\cos(R)) = 2\pi(\frac{R^2}{2} - \frac{R^4}{4!} + \cdots) = \pi R^2 - \frac{\pi R^4}{12} + \cdots$. So circumferences and areas are a little bit too small. If we worked in $z = x^2 + y^2$, we would find that circumferences and areas would be a little bit too big. This is why you can't flatten and orange peel, or an accurate scaled map of the world preserving angles and area.

2.2 Dual space

Suppose V is an n-dimensional vector space, with basis $\mathscr{E} = \{e_1, \cdots, e_n\}$. Then we can write any $v \in V$ as the sum $\sum v^i e_i$. So we have a natural correspondence between vectors v and coordinates $\{v^i\}$ where $v^i \in \mathbb{R}^n$. From now on, we use $v^i e_i$ to denote $\sum v^i e_i$, this is called Einstein summation notation.

The dual space $V^* = \mathrm{Hom}(V, \mathbb{R})$ is the space of linear functionals from V into the base field, the reals. One element of V^* is the function that assigns each vector to its ith coordinate $\phi^i(v) = v^i$. So we have a nice set of transformations $\{\phi^1, \phi^2, \cdots, \phi^n\}$.

Claim. The set $\{\phi^1, \phi^2, \cdots, \phi^n\}$ forms an n-dimensional basis for V^*, called the *dual basis* to $\mathscr{E}$.

Proof. Let $\alpha = \alpha_i\phi^i$. Suppose $\alpha = 0$, then for all $v, \alpha(v) = 0$. So $\alpha(e_j) = 0$, and $\alpha_i\phi^i(e_j) = 0$. Now $\phi^i(e_j) = \delta^i_j$, so $\alpha_i\delta^i_j = 0$ and therefore $\alpha_j = 0$. Now define $\alpha_j = \alpha(e_j)$. Applying this to a vector v gives $(\alpha_j\phi^j)(v) = \alpha_j(v)(\phi^j(v)) = \alpha_j(v)v^j$. Then $\alpha(v) = \alpha(v^j e_j) = v^j\alpha(e_j) = v^j\alpha_j(v)$, and these are equal. We conclude that $\alpha = \alpha_j(v)v^j$. ⊠

Summary: write arbitrary elements of V as $v^i e_i$, and the dual space as $\alpha = \alpha_j\phi^j$. We have $\phi^j(v) = v^j$, $\alpha(e_i) = \alpha_i$, and $\alpha(v) = \alpha_i v^i$ (short for $\alpha_i(v)v^i$). This is why we call V^* the dual space: pairing elements together in either order gives a number. On this vein, $V^{**} = V$, where the basis $\{e_1, \cdots, e_n\}$ is dual to $\{\phi^1, \cdots, \phi^n\}$. V is the space of contravariant vectors, while V^* is the space of covectors, which are covariant in a categorical sense.

You can visualize this by thinking of covectors as rows and vectors as columns. Then e_i is a column with a 1 in the ith slot, while ϕ^j is a row with a 1 in the jth slot. Then $\alpha(e_i) = \alpha_i$, $\phi^j(v) = v^j$, as can be seen below.

$$\alpha = \begin{pmatrix} & \alpha & \end{pmatrix}, \quad v = \begin{pmatrix} \\ v \\ \\ \end{pmatrix}, \qquad \phi^j = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \quad e_i = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

$$\begin{pmatrix} & \alpha & \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} = \alpha_i, \qquad \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \\ v \\ \\ \end{pmatrix} = v^j.$$

Suppose we have a new basis $\widetilde{\mathscr{E}} = \{\widetilde{e}_1, \cdots, \widetilde{e}_n\}$. There must be some change of basis matrix, that is, $\widetilde{e}_i = A_i^j e_j$, similarly we have a new dual basis with $\widetilde{\phi}^i = \phi^j B_j^i$[1]. If we have a vector $v = v^i e_i = \widetilde{v}^i \widetilde{e}_i$, we have $\widetilde{v}^i = \widetilde{\phi}^i(v) = B_j^i \phi^j(v) = B_j^i v^j$. When your basis vectors get big, your coordinates get small, so they transform oppositely. This is the origin of the term contravariant. We claim that $A_i^j B_j^k = \delta_i^k = B_i^j A_j^k$. To see this, note that $\delta_i^k = \widetilde{\phi}^k(\widetilde{e}_i) = \widetilde{\phi}^k(A_i^j e_j) = \phi^\ell B_\ell^k (A_i^j e_j) = B_\ell^k A_i^j \delta_j^\ell = B_j^k A_i^j = A_i^j B_j^k$.

We can kinda visualize vectors as columns and covectors as rows. Then e_i is a column with 1 in the ith while ϕ^j is a row in the jth slot. Applying α to e_i gives α_i, while applying ϕ^j to v gives v^i.

2.3 Tensors

A (k, ℓ) tensor eats k-covectors and ℓ-vectors, and pops out a number. These should be multilinear in each slot. In terms of coordinates, $T(v, w) = T(v^i e_i, w^j e_j) = v^i w^i T(e_i, e_j)$. We define $T_{ij} = T(e_i, e_j)$, so $T(v, w)$ becomes $T_{ij} v^i w^j$. This is a $(2, 0)$-tensor. What is a $(1, 1)$-tensor? This is essentially a matrix, where $S(v, \alpha) = S_i^j v^i \alpha_j$. So S is the matrix, α is a row, and v is a column. If you don't give a tensor enough information, it turns into a tensor of lower rank, which works the same in matrix multiplication. If you have a doubly covariant tensor, it turns vectors into covectors. If you have a doubly contravariant tensor, it turns covectors into vectors. Next time, we'll talk about the most important doubly covariant tensor, the inner product.

Note. The easiest way to remember why a (k, ℓ)-tensor eats k covectors and ℓ tensors is by first thinking of a tensor as an element of $\bigotimes_k V \otimes \bigotimes_\ell V^*$, then taking the dual to realize it as a multilinear map on $(V^*)^{\times k} \times V^{\times \ell}$. Also, a covariant tensor eats vectors because a covector does.

Lecture 3

January 21, 2021

3.1 A basis for tensors

Let's continue the algebra from yesterday. Recall a tensor takes two two vectors as input, denoted $T(v, w) = T(v^i e_i, w^j e_j) = v^i w^j T(e_i, e_j) = \left(\sum_{i,j}\right) T_{ij} v^i w^j$. If $V = \mathbb{R}^2$, what's an example of a covariant 2-tensor? The standard inner product $\langle v \mid w \rangle = v^1 w^1 + v^2 w^2$ works. Using the notation $g(v, w)$, we have $g_{11} = g(e_1, e_1) = 1, g_{12} = 0, g_{21} = 0, g_{22} = 1$. In general, for an inner product $g_{ij} = g_{ij}$, and if you think of it as a matrix, this will be a positive definite

[1] In this convention, A_i^j is the entry ith row and jth column of a matrix A (sometimes you see a_i^j) instead. Usually i denotes rows and j denotes columns, hence why the change of basis matrix for the dual basis is written as B_j^i.

matrix, or $g_{ij}v^iv^j \geq 0$ if $v \neq 0$ (a symmetric matrix with all eigenvectors are positive). In general, you can build an inner product out of a symmetric positive definite matrix. For example, the matrix

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{with positive eigenvectors} \quad \begin{bmatrix} \frac{1\pm\sqrt{5}}{2} \\ 1 \end{bmatrix}$$

works. Another interesting tensor is the determinant $A(v,w) = v^1w^2 - v^2w^1 = -A(w,v)$[2]. In this case, $A_{12} = -A_{21} = 1$ and $A_{11} = A_{22} = 0$.

The natural next question is: what is the space of covariant 2-tensors on $\mathbb{R}^2$? This is a vector space, what would the basis be? A natural basis would be the tensors $\phi^i \otimes \phi^j$. The tensor symbol in the middle is the **tensor product** : if $S(v_1, \cdots, v_n)$ and $T(w_1, \cdots, w_n)$ are covariant tensors (makes life simpler), then

$$S \otimes T(v_1, \cdots, v_n, w_1, \cdots, w_n) := S(v_1, \cdots, v_n)T(w_1, \cdots, w_n).$$

With this in mind, applying our basis tensor to (e_k, e_ℓ) gives us $\phi^i \otimes \phi^j(e_k, e_\ell) = \phi^i(e_k) \cdot \phi^j(e_\ell) = \delta^i_k \delta^j_\ell$. This gives us 1 if you feed it (e_i, e_j), and 0 if you feed it anything else. Since these form a basis for the sapce of covariant 2-tensors, our claim is that you can write any tensor T in the form $T = T_{ij}\phi^i \otimes \phi^j$. Then $T(e_k, e_\ell) = T_{k\ell}$ by definition.

Let us return to the inner product. If we only feed it one vector, then $g(v, \,)$ is waiting for a vector to spit out a number. This is just a covector, which gives an alternate use for the inner product besides taking two vectors and spitting out a number. This map is injective, since if it had a kernel (that is, $g(v, \,) = 0$), then $g(v, \,)(v) \neq 0$, since it's positive. Then g induces a map $g: V \to V^*$ an isomorphism: if $\alpha = g(v, \,)$, then $\alpha_j = g_{ij}v^i$. We call this *lowering an index*: we take something with an up index, multiply it through g_{ij}, resulting in a down index. The inverse that *raises* indices is the inverse matrix g^{ij}, and $g^{ij}g_{jk} = \delta^i_k$.

Now that we have a basis for 2-tensors, what about a basis for $\mathscr{T}^{k,\ell}$ (space of tensors)? It's going to be

$$\{e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_k} \otimes \phi^{j_1} \otimes \phi^{j_2} \otimes \cdots \otimes \phi^{j_\ell}\}$$

So a tensor T can be written as $T = T^{i_1\cdots i_k}_{j_1\cdots j_\ell} e_{i_1} \otimes \cdots \otimes \phi^{j_\ell}$. Furthermore, to calculate the coefficients, apply your tensor to the basis vectors to get $T^{i_1\cdots i_k}_{j_1\cdots j_\ell} = T(\phi^{i_1}, \cdots, \phi^{i_k}, e_{j_1}, \cdots, e_{j_\ell})$.

3.2 Trace of a matrix

Now let's talk about the trace of a matrix, defined as $\mathrm{Tr}(M) = M^i_i$. This is actually independent of which basis we pick. If we change our basis by $\widetilde{e}_i = A^j_i e_j$, recall that the dual basis transforms accordingly by $\widetilde{\phi}^i = B^i_j \phi^j$, where B^i_j is an inverse for A^j_i. In our new basis, $\widetilde{M}^i_j = B^i_k A^\ell_j M^k_\ell = B^i_k M^k_\ell A^\ell_j$, this is kind of like PMP^{-1} that we did in linear algebra. But this doesn't really work with tensors, what about 2, 3, 4, 5-tensors? That's why we're doing it this way.

We have the trace $\mathrm{Tr} = \widetilde{M}^i_i = B^i_k M^k_\ell A^\ell_i$. But $B^i_k A^\ell_i = \delta^\ell_k$, since these two are inverse matrices. So $\mathrm{Tr} = \delta^k_\ell M^k_\ell = M^k_k$, which is the old definition of trace. We can apply this to higher rank tensors: suppose we have a tensor that takes in a covector and two vectors, denoted $T(\alpha, v, w)$. Define a tensor $S(w) = T(\phi^i, e_i, w)$, so $S_i = T^j_{ji}$. What happens if you change basis, that is, is $T(\phi^i, e_i, w) = T(\widetilde{\phi}^i, \widetilde{e}_i, w)$?

$$\langle \textit{breakout rooms} \rangle$$

We want to show that $T^i_{ij}w^j = \widetilde{T}^i_{ij}\widetilde{w}^j$. For fixed w, let's define a new tensor, $U(\alpha, v) = T(\alpha, v, w)$. Let's take the trace of this, we've already shown that the trace of a $(1,1)$-tensor doesn't depend on basis. So the w comes along for the ride. This is a slick solution, since $\mathrm{Tr}\, U$ is well defined. Once you know that the trace is defined for $(1,1)$-tensors, you know it's defined for any (n,n)-tensor, where you pair one up index with one down index.

3.3 Tangent vectors

Enough about about tensors. Now let's talk about manifolds. Usually the vector space at a point p we care about is the tangent space $T_p(M)$ at that point. The object of importance is not a tensor at a point, but rather vector fields,

[2] You can also call this the area form, or the symplectic form.

covector fields, tensor fields, etc. The coefficients of such tensors are going to be functions of where you are, or the coordinates you're using.

In $\mathbb{R}^n$, consider a point p: what is a tangent vector there? There are several definitions, and they're all equivalent:

1. An arrow, add them head to tail.
2. An element of $\mathbb{R}^n$, a list of n numbers. This is pretty much an arrow, just take the coefficients and impose them on the standard basis.
3. Velocity. Consider all possible parametrized curves through a point, and identify all curves with the same velocity at time zero. So we mean the equivalence class of curves $\gamma(t)$ with $\gamma(0) = p$[3]. The beauty of this third definition is it makes sense on any manifold. So we can consider the curves going through this point, and take them up to equivalence.
4. Directional Derivative. This is pretty much the same thing as velocity, since for $f : \mathbb{R}^n \to \mathbb{R}$, we can take $\frac{df\cdot\gamma(t)}{dt}\Big|_{t=0} = \frac{d\gamma^i}{dt}\Big|_{t=0}\frac{\partial f}{\partial x^i}\Big|_p$. The partial derivatives $\{\partial_1, \cdots, \partial_n\}$ gives a basis for this vector space.
5. Derivations. A derivation at p is a map $D\colon C^\infty(\mathbb{R}^n) \to \mathbb{R}$ with the following properties:

   (1) $D(af + bg) = aD(f) + bD(g)$,

   (2) $D(fg) = f(p)D(g) + g(p)D(f)$.

   The partial derivatives $\partial_i|_p$ are derivations.

This idea of thinking about tangent vectors as derivations carries over very nicely to abstract manifolds, which is why we care. Understanding the ring of smooth functions means we understand what a tangent vector is, coordinate-free (we aren't anchored to $\mathbb{R}^n$).

The directional derivative is a small step forward from velocity, which is a small step forward from the näive arrow. Our goal is to show that the scary abstract idea of a derivation is just a small step forward from the directional derivative, and to do this, we show that they both use the same basis.

Claim. We have $\{\partial_i\}$ a basis for the set of derivations at p, denoted $\mathscr{D}_p(\mathbb{R}^n)$.

We want to show a couple of things:

1. To show $D(\text{constant}) = 0$, note that $D(cf) = cD(f) = cD(f) + f(p)D(c)$, so we need $D(\text{constant}) = 0$.
2. If $f(p) = g(p) = 0$, then $D(fg) = 0$, this follows from the product rule.
3. Taylor series, we have $f(x) = f(p) + \partial_i f(p)(x^i - p^i) +$ higher order terms. Then $D(f) = 0 + \partial_1 f(p)D(x^i) + 0 = D^i\partial_1 f(p)$, where $D = D^i\partial_i$.

This works for analytic functions. There is some cheating going on, but we don't need the entire Taylor series for the most part[4]. So every derivation is a linear combination of partial derivatives. From now on, think of vectors as a combination of partial derivatives, or inducing curves along a vector field.

3.4 Vector fields

A vector field $v^i(x)\frac{\partial}{\partial x^i}$ is a bunch of coefficients in combination with the basis vectors of partial derivatives. How do we change coordinates here? We do this by the chain rule, that is,

$$\text{for } \{x\} \longleftrightarrow \{y\} \quad \text{we have} \quad \frac{\partial}{\partial y^i} = \frac{\partial x^j}{\partial y^i}\frac{\partial}{\partial x^j}.$$

[3] Since you consider the curves up to equivalence, a straight line is usually the one you visualize, since it's the simplest and they're all equivalent. Of course, you can choose a wigglier curve if you want.

[4] See my notes in `independent_reading/differentiable_manifolds` for more info (or Tu's book).

You can think of $\partial x^j/\partial y^i$ as the change of basis matrix A_i^j. The inverse matrix B_j^i is just $\partial y^i/\partial x^j$. Let's try this in $\mathbb{R}^2$ with $(r,\theta) \longleftrightarrow (x,y)$. For $e_1 = \frac{\partial}{\partial x}, e_2 = \frac{\partial}{\partial y}$ and $\widetilde{e}_1 = \frac{\partial}{\partial r}, \widetilde{e}_2 = \frac{\partial}{\partial \theta}$, recall that $r = \sqrt{x^2+y^2}$, $\theta = \tan^{-1}\left(\frac{y}{x}\right)$, and $x = r\cos\theta$, $y = r\sin\theta$. So

$$\frac{\partial y}{\partial r} = \cos\theta = \frac{x}{\sqrt{x^2+y^2}},\ \frac{\partial y}{\partial r} = \sin\theta = \frac{y}{\sqrt{x^2+y^2}},\ \frac{\partial x}{\partial \theta} = -r\sin\theta = -y,\ \frac{\partial y}{\partial \theta} = r\cos\theta = x.$$

So now we can convert between the two bases, that is, we have the change of basis matrix

$$\begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} = \begin{pmatrix} \frac{x}{\sqrt{x^2+y^2}} & -y \\ \frac{y}{\sqrt{x^2+y^2}} & x \end{pmatrix}.$$

We can write the inverse matrix in terms of (x,y) or (r,θ), taking the inverse of a two by two is pretty easy. Now we know how to convert from one basis to the other.

Lecture 4

January 26, 2021

4.1 A basis for the tangent space

Consider some manifold, then how do we find a basis for the tangent space? If each point is covered by a chart, consider the coordinates for each chart and take the partial derivatives in that direction. For a topologically interesting manifold, usually just one set of coordinates won't work for the whole manifold. So it becomes imperative that we know how to change between coordinate charts. Suppose we have two coordinates $(x^1, x^2, \cdots, x^n)$ and $(y^1, y^2, \cdots, y^n)$, these give rise to the bases $\{\partial_{x^i}\}$ and $\{\partial_{y^i}\}$. The chain rule tells us that

$$\frac{\partial}{\partial x^i} = \frac{\partial y^j}{\partial x^i}\frac{\partial}{\partial y^j}.$$

Then the $\frac{\partial y^j}{\partial x^i}$ form a change of basis matrix A_i^j, and we can take everything we know about vector spaces and tensors at a *point*, and apply it to tensor *fields*. A tensor on a *fixed* vector space is a certain object given by some coefficients. If we want a tensor *field*, we need a tensor at each point, which requires understanding what a basis is for the tangent space at each point, and what the coefficients are. Now the numbers $T_{ij}^k(x)$ are *functions* of where you are, if T_{ij}^k gives the coefficients for a $(1,2)$-tensor (1-contravariant 2-covariant).

> "*Think globally, act locally.*"

This is our slogan for working with manifolds. You always want to keep a global picture of what's going on in your head, but when doing computations and such, always work in a neighborhood with coordinate charts, where everything looks like $\mathbb{R}^n$.

4.2 A basis for the dual space

What's the basis for the dual space of tangent vectors? Suppose we have a function $f : M \to \mathbb{R}$. Then we can take the derivative $df(V) := V(f)$, which is a covector field. In $\mathbb{R}^n$, we have $(df)\left(\frac{\partial}{\partial x^1}\right) = \frac{\partial f}{\partial x^1}$, and similarly $(df)\frac{\partial}{\partial x^j} = \frac{\partial f}{\partial x^j} = \partial_j f$. So $(df) = \frac{\partial f}{\partial x^j}\phi^j$, since for a covector α, we have $\alpha_i = \alpha(e_i)$. What happens when we take the derivative of the coordinate function x^i? Note that we have

$$(dx^i)\left(\frac{\partial}{\partial x^i}\right) = \frac{\partial x^i}{\partial x^j} = \delta_j^i.$$

So on a manifold with coordinates $x^1, \cdots, x^n$, the basis for T_pM is $\left\{\frac{\partial}{\partial x^i}\right\}$, and the dual basis for T_p^*M is $\{dx^i\}$.

4.3 Applications of the dual basis

Now that we've characterized the bases for the tangent and cotangent spaces, what does a tensor actually look like? Consider a $(2,3)$-tensor $T(x) = T_{ijk}^{\ell m}(x)\frac{\partial}{\partial x^\ell} \otimes \frac{\partial}{\partial x^m} \otimes dx^i \otimes dx^j \otimes dx^k$, where $T_{ijk}^{\ell m} = T(dx^\ell, dx^m, \partial_i, \partial_j, \partial_k)$. Once more: covariant is how many down indices you have, contravariant is how many up indices you have[5]. Fortunately we don't work with 5-tensors very often, but we do deal with a particular 2-covariant tensor all the time, which is the metric.

We are interested in doing geometry, so we need some way to measure lengths on a vector space. If we have the notion of an inner product at every point, then we have a metric $g = g_{ij}dx^i \otimes dx^j$, where $g_{ij}^{(x)} = g(\partial_i, \partial_j)_x = \langle \partial_i, \partial_j \rangle_x$. We will spend a ridiculous amount of time talking about this tensor at a point x. In $\mathbb{R}^n$ this inner product is just δ_j^i, but for a manifold it isn't! Let's play around with the manifold of the upper hemisphere, given by $\{(x,y,z) \mid x^2 + y^2 + z^2 = 1,\, z > 0\}$. Some possibilities for coordinates:

1. x, y, where $z = \sqrt{1-x^2-y^2}$,
2. θ, ϕ, where θ measures the angle from the north pole, and ϕ measures the longitude. So $x = \sin\theta\cos\phi$, $y = \sin\theta\sin\phi$, $z = \cos\theta$.

How do we find a metric? We have $\partial_x = (1, 0, -x/z)$, $\partial_y = (0, 1, -y/z)$. So

$$\begin{aligned}
g_{11} &= 1 + \frac{x^2}{z^2} = 1 + \frac{x^2}{1-x^2-y^2} = \frac{1-y^2}{1-x^2-y^2},\\
\underset{=g_{21}}{g_{12}} &= \frac{xy}{z^2} = \frac{xy}{1-x^2-y^2},\\
g_{22} &= 1 + \frac{y^2}{z^2} = \frac{1-x^2}{1-x^2-y^2}.
\end{aligned}$$

These are, of course, functions of x (as a point). A perfectly reasonable question to ask now would be "what is $\partial g_{ij}/\partial x^k$"? Since we now have the metric as a function of the coordinates, we can ask at what rate does it change. Now lets move onto spherical coordinates. We have

$$\begin{aligned}
\frac{dx}{dt} &= \cos\theta\cos\phi\frac{d\theta}{dt} - \sin\theta\sin\phi\frac{d\phi}{dt},\\
\frac{dy}{dt} &= \cos\theta\sin\phi\frac{d\theta}{dt} + \sin\theta\cos\phi\frac{d\phi}{dt},\\
\frac{dz}{dt} &= -\sin\theta\frac{d\theta}{dt}.
\end{aligned}$$

$$\frac{\partial}{\partial\theta} = (\cos\theta\cos\phi, \cos\theta\sin\phi, -\sin\theta), \quad \frac{\partial}{\partial\phi} = (-\sin\theta\sin\phi, \sin\theta\cos\phi, 0).$$

So

$$\begin{aligned}
g_{11} &= \cos^2\theta\cos^2\phi + \cos^2\theta\sin^2\phi + \sin^2\phi = \cos^2\theta + \sin^2\theta = 1,\\
\underset{=g_{21}}{g_{12}} &= 0,\\
g_{22} &= \sin^2\theta\sin^2\phi + \sin^2\theta\cos^2\phi = \sin^2\theta.
\end{aligned}$$

This makes sense because as you vary θ, the rate of change down the sphere is 1, or you're moving with speed 1. This basis is nice and orthogonal, but it's not orthonormal. If we consider the basis $\{\partial_\theta, (1/\sin^2(\theta)\partial_\phi\}$, then we get an orthonormal basis. The problem is, there is no set of coordinates that satisfies that the expression is exactly the derivative with respect to the coordinates. It's kind of magic that our basis in $\mathbb{R}^n$ is both orthonormal, and consists of the derivatives of the coordinates.

[5] Dr. Sadun has been doing things the other way around (he says he can never keep it straight), but I feel like the covectors and vectors should be in this order.

4.4 Vector fields and vector flows

When you see a new vector field, you should ask yourself "what does the flow look like?"

Every time we have a vector field, it induces a **flow**. Suppose we have a vector field on a manifold. Given a point, we can find a curve that follows the vector field: that is, we can find a curve $\{\gamma(t) \mid \gamma(0) = p,\ d\gamma/dt = V(\gamma(t))\}$. Moving along these flow lines gives us a *flow*, a way to turn a point at time $t = 0$ into a point at a later time. We denote this $\Phi_t^V(p)$, which is what we get when we start at a point p and flow along a vector field v for time t. Let's look at some simple vector fields and figure out what their flows are in $\mathbb{R}^2$.

(i) If $V(x) = \partial_{x^1}$, then $\Phi_t^v(x, y) = (x + t, y)$. Similarly, for $W = \partial_{x^2}$, $\Phi_t^W(x, y) = (x, y + t)$.

(ii) For $X(x) = x^1\partial_1 + x^2\partial_2$, our flow is given by $\Phi_t^X(x, y) = e^t(x, y)$. Note that the definition of flow sets up a differential equation, and the one we are solving right now is $dx^1/dt = x^1,\ dx^2/dt = x^2$.

(iii) Say $Y(x) = x^1\partial_2 - x^2\partial_1$. This is rotation at a constant speed, since we're solving the differential equations $dx^1/dt = -x^2$, $dx^2/dt = x^1$.

Recall that a vector field is a derivation. Suppose V and W are vector fields. If we have a function $f : M \to \mathbb{R}$, what kind of object is $W(f)$? It's also a function $M \to \mathbb{R}$, since the derivative of a function is a function. So is $V(W(f))$, and so on. What is $V \circ W$? It's not a vector field, but a second order differential operator, since we're taking second derivatives.

We claim that $V(W(f)) - W(V(f))$ is a *first order* differential operation. Note that $W(f) = w^i\partial_i f$, so

$$\begin{aligned}
V(W(f)) &= v^j\partial_j(w^i\partial_i f) \\
&= v^j(\partial_j w^i)\partial_i f + v^j w^i\partial_j\partial_i f. \\
W(V(f)) &= w^i\partial_i(v^j\partial_j f) \\
&= w^i(\partial_i v^j)\partial_j f + w^i v^j\partial_i\partial_j f.
\end{aligned}$$

Taking the different of the two eliminate the second order term, since mixed partial derivatives commute. So

$$V(W(f)) - W(V(f)) = v^i\partial_j w^i(\partial_i f) - w^i(\partial_i v^j)\partial_j f = (v^j\partial_j w^i - w^i\partial_j v^i)\partial_i f.$$

This is called the **Lie bracket**, and is denoted $[V, W]$. It also goes by the *commutator*, or the *Lie derivative* $\mathscr{L}_V(W)$. What do these brackets mean? If we have *any* coordinates, $[\partial_i, \partial_j] = 0$. This tells us that if we have two vector fields with a nonzero Lie bracket, then they cannot be derivatives with respect to coordinates. It's not that you haven't thought up of the right coordinates, but that there *are no coordinates* that they're derivatives of. If you recall the flows Φ_t^V, Φ_t^W above, they actually commute!

A basic question when you have two vector fields is what happens if you start at a point p and flow for time v according to V, then you flow for time W along time s? Or better yet, you flow backwards along V with time $-t$, and backwards along W with time $-s$.

$$\begin{tikzcd}
{} \arrow[d, "\Phi^W_{-s}"'] & & {} \arrow[ll, "\Phi^V_{-t}"'] \\
{} & p \arrow[l, dotted, "\Phi^{[V,W]}_{st}"] \arrow[r, "\Phi^V_t"'] & {} \arrow[u, "\Phi^W_s"']
\end{tikzcd}$$

If the bracket is zero, you come back to where you started, and if it's nonzero, you don't. This bracket measures how far off you are! Showing that this is true is the homework.

4.5 The gradient

Suppose we have some $f : M \to \mathbb{R}$, then $df = \partial_1 f\, dx^1 + \partial_2 f\, dx^2 + \cdots$. This isn't what a gradient is, it's the vector that has the property that $\nabla f \cdot v = D_v f = v(f)$, while $df(v) = v(f)$. The different is the *inner product*. We have $g(\nabla f, v) = df(v)$, or $g(\nabla f, \cdot) = df$. The differential is what you get when you take the gradient and lower it an index. If you're working in $\mathbb{R}^n$ this doesn't do anything, the metric is just the identity. However in more interesting coordinates, it does do something. While df_j is a *covector*, the **gradient** ∇f^i is a *vector*. So

$$df_j = g_{ij}(\nabla f)^i, \quad \text{or} \quad (\nabla f)^i = g^{ij}(df)_j \text{ (note the inverse metric.)}$$

The differential df is defined without reference to any inner product. It's a perfectly good covector field (or 1-form) on its own. The gradient is what you get when you raise the index of df.

Lecture 5

January 28, 2021

5.1 Inheriting metrics

A **Riemannian manifold** is a manifold with a nice, smoothly varying inner product at every point. Given a topological space, how do you come up with an interesting inner product on such space? One way to do this is to sit inside a space that already has one.

Example 5.1. The inner product of two vectors on the surface of a torus is their inner product in $\mathbb{R}^3$: but how do we actually write this down? We use coordinate charts. Say we have u, v coordinates and a map $(u, v) \to \Sigma(u, v)$. A basis for the domain (in $\mathbb{R}^2$) is just ∂_u, ∂_v, so this gives rise to a basis $\partial_u \Sigma = d\Sigma(\partial_u), \partial_v \Sigma = d\Sigma(\partial v)$. These are vectors in $\mathbb{R}^3$. Then

$$g_{11} = \partial_u \Sigma \cdot \partial_u \Sigma, \quad g_{12} = g_{21} = \partial_u \Sigma \cdot \partial_v \Sigma, \quad g_{22} = \partial_v \Sigma \cdot \partial_v \Sigma.$$

In general, for a submanifold, take the metric of the big space and restrict it to tangent vectors of the submanifold. When you parametrize surfaces, we need $d\Sigma$ injective (or rank two), or else something else will map to 0, violating the rules of a metric.

Example 5.2. Consider the graph $z = f(x, y)$. This gives rise to a simple parametrization, where $\Sigma(u, v) = (u, v, f(u, v))$. Then ∂_u corresponds to $(1, 0, \partial_u f)$ and ∂_v corresponds to $(0, 1, \partial_v f)$. So

$$g_{11} = 1 + (\partial_u f)^2, \quad g_{12} = g_{21} = (\partial_u f)(\partial_v f), \quad g_{22} = 1 + (\partial_v f)^2.$$

For the upper hemisphere where $z = \sqrt{1 - x^2 - y^2} = f(x, y)$, recall that $\partial_u f = -u/\sqrt{1 - x^2 - y^2}$, $\partial_v f = -v/\sqrt{1 - x^2 - y^2}$. So

$$g_{11} = \frac{1 - v^2}{1 - x^2 - y^2}, \quad g_{12} = g_{21} = \frac{uv}{\sqrt{1 - x^2 - y^2}}, \quad g_{22} = \frac{1 - u^2}{1 - x^2 - y^2}.$$

Work out the metric for the sphere on your own, where $(u, v) \simeq (\theta, \phi) \to (\sin u \cos v, \sin u \sin v, \cos u)$. For a parametrized curve $\gamma(t)$ we only have one component, so $g_{11} = \dot{\gamma} \cdot \dot{\gamma} = \|\dot{\gamma}\|^2$. So $g = \|\dot{\gamma}\|^2 dt^2$.

5.2 Product spaces

Given two manifolds (M_1, g_1) and (M_2, g_2) with coordinates $(x^1, \cdots, x^n)$, $(y^1, \cdots, y^m)$, we have the metric $g_1 = (g_1)_{ij} dx^i \otimes dx^j$, and the metric $g_2 = (g_2)_{k\ell} dy^k \otimes dy^\ell$. Then we can define the *product manifold* $M_1 \times M_2$, with coordinates $(x^1, \cdots, x^n, y^1, \cdots, y^m)$ Recall from a G&P question that $T_{(p,q)}(M_1 \times M_2) = T_p M_1 \times T_q M_2$. So we write a vector V in $T_{(p,q)}(M_1 \times M_2)$ in the form $V = (V_1, V_2)$, where V_1 is tangent to M_1 and V_2 is tangent to M_2. Then

$$g = ((v_1, v_2), (w_1, w_2)) = g(v_1, w_1) + g_2(v_2, w_2) \quad \text{or}$$
$$g = g(g_1)_{ij} dx^i \otimes dx^j + (g_2)_{k\ell} dy^k \otimes dy^\ell.$$

The *simplest* example is $\mathbb{R}^2$, where each component is a vector in $\mathbb{R}^1$. Then $g((v_1, v_2), (w_1, w_2)) = g_1(v_1, w_1) + g_2(v_2, w_2) = v_1 w_1 + v_2 w_2$, which is exactly how the inner product works in $\mathbb{R}^2$. This generalizes to $\mathbb{R}^{n+m}$, of course.

Example 5.3. The torus is the product of two circles $\mathbb{T}^2 = S^1_{R_1} \times S^1_{R_2}$. If the metric on $S^1_{R_i}$ is given by $R_i^2 d\theta^2$, the metric for the torus just $g_1 = R_1^2 d\theta^2 + R_2^2 d\theta^2$.

5.3 The warped product

Suppose we have two manifolds M_1, M_2 and a function $f : M_1 \to \mathbb{R}^+$ (excluding 0!!). Then define $M_1 \times_{f_2} M_2 = (M_1 \times M_2, g)$, where g is *not* the product metric. The metric g at a point (p, q) is given by $g_{(p,q)} = (g_1)_{ij} dx^i \otimes dx^j + (f(p))^2 (g_2)_{k\ell} dy^k \otimes dy^\ell$. This is a **warped product**. The warped product is indeed a metric, since it's positive definite, given $f(p)$ never maps to zero.

Warped products are familiar: consider surfaces of revolution. If we parametrize $\gamma(t)$ by arc length (so it's unit-speed), then $\gamma(t) = (x(t), y(t))$, a curve in the upper half plane. Then

$$\Sigma(t, \theta) = (x(t), y(t)\cos(\theta), y(t)\sin\theta).$$

We claim that this surface Σ is equal to a product $\gamma \times_f S^1$, where $f(t) = y$. To see this, note that $\partial_t \Sigma(\dot{x}, \dot{y}\cos\theta, \dot{y}\sin\theta)$, $\partial_\theta \Sigma = (0, -y\sin\theta, y\cos\theta)$. So $g_{11} = \dot{x}^2 + \dot{y}^2 = 1$, $g_{12} = g_{21} = 0$, $g_{22} = y^2$. Then our metric is given by $g = dt^2 + (y(t))^2 d\theta^2$. The reason why $\dot{x}^2 + \dot{y}^2 = 1$ is because we used a unit speed parametrization.

Example 5.4. Concrete example: consider the half circle $\gamma(t) = (\cos t, \sin t)$ for $0 < t < \pi$. Then rotating around the x-axis gives a sphere minus antipodal poles[6], and $g = dt^2 + \sin^2(t) d\theta^2$. This looks very familiar, namely, the metric on this warped product is the exact same as the standard one for the upper half sphere.

This demonstrates the idea of the warped product giving an interesting structure. Here, $M_1 = (0, \pi)$, $M_2 = S^1$ are both flat, but the warped product gives something round.

5.4 Quotients by discrete actions

Consider the quotient space $\mathbb{R}/L\mathbb{Z}$, or the set $\{x \in \mathbb{R}\}/\sim$ where $x \sim (x + nL)$. Does the quotient space inherit a metric? Since $\mathbb{R}/L\mathbb{Z}$ is a circle, a tangent vector v_i comes from a representative w_i in $\mathbb{R}$, that is $v_i = \pi_*(w_i)$, where π_* is projection onto the quotient space. Then define $g(v_1, v_2) = g(w_1, w_2)$. This fails when two vectors upstairs (say w_i and $\widetilde{w}_i$) have different values under the inner product. For w_i and $\widetilde{w}_i$ to have the same length, the identification needs to come from a group action by isometries. Here, the group action is "translate by L", it has no fixed points, and it acts by isometries.

Proposition 5.1. *If a group G acts freely on a space (X, g) by isometries, then X/G inherits a metric from X.*

In our example above, $G = \mathbb{Z}$, and the action is $(n, x) \to x + nL$. We are inclined to use the coordinates of X for X/G, since locally (pre-identification) X/G looks like X. This isn't a new concept, we have done this for years on the circle: we have always used θ for coordinates by identifying the integers $2\pi n$. When would this fail?

Example 5.5 (Non-examples)**.** Suppose $X = (\mathbb{R}^+, dx^2)$, and $G = \mathbb{Z}$. Let the action be given by $(n, x) \to 2^n x$. This is a perfectly nice free group action (by way of our domain). However, G doesn't act by isometries. If we ask what the metric is at 1, this is the same as the metric at 2, 4, 1/2, and so on since they're all identified. This group action *stretches* things, so we don't have a well defined notion of length. The quotient is a perfectly good topological space, a circle: it just doesn't have a metric.

Now suppose $X = \mathbb{R}^2$ and $G = \mathbb{Z}_2$, so the action is rotation by π. This identifies the lower half of the plane with the upper half, and the positive real axis with the negative real axis. Besides the origin, this is a perfectly good group action acting isometrically and freely. However, the origin is a fixed point. A disk around the origin post quotient becomes a disk $\mathbb{Z}/2$, so this is not a manifold. This is an *orbifold*, a manifold with a singular point. In general, for nonfree group actions, big problems happen in neighborhoods of fixed points.

Example 5.6 (Non-non-examples)**.** Here we give some actual examples.

[6] This also omits a line (the IDL?), but we can take care of that by identifying the lines and making the quotient space.

- We have already seen $\mathbb{T}^2 = \mathbb{R}^2/\text{lattice}$, where a lattice is $\mathbb{Z}^2$ with two basis vectors. In general, we can form $\mathbb{T}^n = \mathbb{R}^n/\text{lattice}$.
- Another important example is $\mathbb{RP}^n = S^n/\pm 1$.
- The **lens spaces** $L(p,q)$ are formed by taking $S^3 \subseteq \mathbb{C}^2/\sim$, where $(z_0, z_1) \sim (\lambda z_0, \lambda^q z_1)$ and $\lambda = e^{2\pi i/p}$. The group here is $\mathbb{Z}_p$, and the identification is the group action, acting by isometries with no fixed points. Note that we want p, q to be relatively prime. Also note that $L(2,1)$, since $e^{2\pi i/2} = -1$. These lens spaces look locally like S^3, but globally have some interesting topological properties: namely $\pi_1(L(p,q)) = \mathbb{Z}_p$.
- We will also eventually see some Riemann surfaces, given by $\mathbb{H}^2/G$. Hyperbolic space has constant negative curvature, so you can realize a structure of Riemann surfaces with genus greather than 1, and give them all metrics of constant negative curvature. This is the same thing as talking about the *canonical* space of constant curvature $\mathbb{H}$, and modding it by a certain group action.

5.5 The complex projective plane

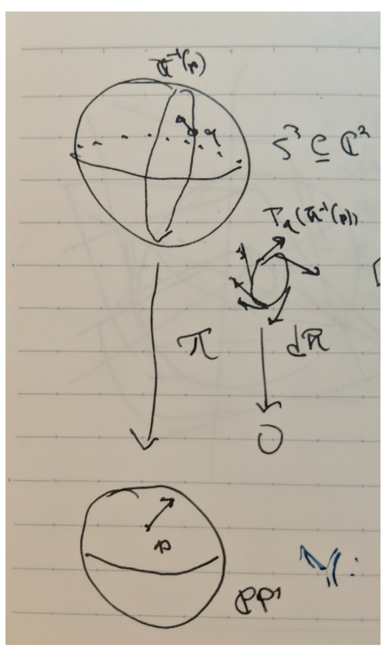
We can also quotient by continuous groups. Consider the complex projective plane $\mathbb{CP}^1$, defined by

$$\mathbb{CP}^1 := \left\{(z_1, z_2) \in \mathbb{C}^2 \setminus \{0\}\right\}/\sim, \quad (z_1, z_2) \sim (\lambda z_1, \lambda z_2),\ \lambda \in \mathbb{C} \setminus \{0\} = \mathbb{C}^*.$$

Multiplication by a scalar λ (say $3+i$) is *not* an isometry, since it scales lengths by $\sqrt{10}$. So the way we've written it as of now doesn't fit into our framework. However, we can break up $\mathbb{C}^*$ into $\mathbb{R}^+ \times U(1)$[7], an element of the unit circle times a length. So we can rewrite the definition as

$$\mathbb{CP}^1 := \left\{(z_1, z_2) \in \mathbb{C}^2 \mid (z_1)^2 + (z_2)^2 = 1\right\}/\sim, \quad (z_1, z_2) \sim (e^{i\theta} z_1, e^{i\theta} z_2).$$

Behold, *this* is an isometry. For now we restrict our discussion to lengths of vectors, since if you know what the length of a vector is, you know what the inner product is. This is because $\langle v, w\rangle = (\langle v+w, v-w\rangle - \langle v-w, v-w\rangle)/4$ which is the sum of lengths squared over four. (This is called the *polarization identity*.) Consider $S^3 \subseteq \mathbb{C}^2$ projecting onto $\mathbb{CP}^1$ (topologically a sphere) by π, then a point $p \in \mathbb{CP}^1$ corresponds to a circle in S^3. Consider a point q in this circle in S^3 corresponding to p.



[7]Recall that $U(1)$ denotes the unit complex numbers.

However, $d\pi$ is *not* injective, since it maps any vector tangent to the sphere to zero. We can write T_qS^3 as $T_q(\pi^{-1}(p)) + T_q(\pi^{-1}(p))^\perp$, so $d\pi\colon T_q(\pi^{-1}(p)) \to 0$. But $d\pi\colon T_q(\pi^{-1}(p))^\perp \to T_p\mathbb{C}\mathrm{P}^1$ is an isomorphism by the rank nullity theorem, since $T_q(\pi^{-1}(p)) = \ker d\pi$. Let us name this map α. Then $\langle v, w\rangle = g(\alpha^{-1}(v), \alpha^{-1}(w))$. We say the vectors $T_q(\pi^{-1}(p))$ are **vertical**, or they live in the fiber above a point. A vector that lives in $T_q(\pi^{-1}(p))^\perp$ is **horizontal**, since it's orthogonal to a vertical vector. So the idea is to lift vectors in the base space uniquely to *horizontal* vectors, since they don't generally have a unique vertical left (since $d\pi$ fails to be injective). What ingredients did we need to make this work?

(1) The group acts by isometries. (We need two distances in the total space to be the same.)

(2) The group needs to act freely.

This map π from $S^3 \to S^2$ is called the **Hopf fibration**. This map has some nice properties, that is, the preimage of two points will always be linked circles (if you think of S^3 as the one-point compactification of $\mathbb{R}^3$). For $\mathbb{C}\mathrm{P}^n$, we consider it as the quotient S^{2n+1}/S^1, or $S^{2n+1} = \{(z_0, \cdots, z_n) \mid |z_0|^2 + \cdots + |z_n|^2 = 1\}/S^1$, where $(z_0, \cdots, z_n) \sim (e^{i\theta}z_1, \cdots, e^{i\theta}z_n)$. What does the metric on $\mathbb{C}\mathrm{P}^1$ look like? This is the **Fubini-Study metric** todo:read about this in the book

Lecture 6

February 2, 2021

Here we talk about model spaces, and when we get around to discussing curvature and such, we will compare them to these spaces, as in "how much does a space differ from S^n?" for example.

6.1 Different notions of symmetry

Definition 6.1. A space X is **homogeneous** if all points look the same, that is, there exists an isometry $X \to X$ that takes p to q for all $p, q \in X$. For spaces (X, g) and $(Y, \widetilde{g})$, we say X, Y are **isometric** if we have a map $f : X \to Y$ such that f is a diffeomorphism, and g is the pullback of $\widetilde{g}$ (that is, $g = f^*\widetilde{g}$).

Note. Some differential topology. Recall that the derivative of a map $F : M \to N$ is also called the **push-forward**, since it induces a map $F_*\colon T_pM \to T_{F(p)}N$. Then this gives rise to a dual map on the cotangent spaces $(F_*)^*\colon T^*_{F(p)}N \to T^*_pM$, which *pulls back* tangent covectors in $T^*_{F(p)}N$ to tangent covectors in T^*_pM. Rewrite $(F_*)^*$ as F^* and call this the **pullback** of F.

Example 6.1. Some examples of homogeneous spaces:

1. Sphere
2. Plane
3. Torus

If you think of the torus as $\mathbb{R}^2/\mathbb{Z}^2$, then translations form a nice symmetry of the torus. However, even though every point the same, this doesn't hold for directions. This motivates the following definition:

Definition 6.2. A space X is **isotropic** if for all $p, q \in X$, where v, w are unit tangent vectors to p, q respectively, there exists an isometry taking $(p, v) \to (q, w)$.

Here, the sphere and plane are isotropic, but the torus is not. However, there is an even stronger condition than being isotropic. Rather than taking a single vector, you take an **orthonormal frame** at p. An orthonormal frame is just an ordered orthonormal (with respect to g) basis for the tangent space at p.

Definition 6.3. A space X is **frame isotropic** if for all $p, q \in X$, $\{v_i\}, \{w_i\}$ orthonormal frames of X at p, q respectively, for all i there exists an isometry taking $(p, v_i) \to (q, w_i)$.

An example of a space that's isotropic but not frame isotropic is $\mathbb{C}\mathrm{P}^n$ for $n > 1$. In $\mathbb{C}^n$, all directions are the same: given a vector, you can send it to $(1, 0, 0, \cdots)$ through a change of basis. Given an orthogonal *pair* of vectors, not all pairs of the same. For example, the pair $(1, 0, \cdots), (0, 1, \cdots)$ is related in a different way than $(1, 0, \cdots)$ and $(i, 0, \cdots)$, this is a complex geometric notion. What this means is that if we do something that really uses the complex structure, you can get something where the geometry tells the difference. This is what we do in $\mathbb{C}\mathrm{P}^n$: we start off with $\mathbb{C}^{n+1} \setminus \{0\}$, then we mod out by multiplication by $\lambda \in \mathbb{C} \setminus \{0\}$. Then pairs of vectors with certain relations will be preserved or not preserved by this action. That is, all points in $\mathbb{C}\mathrm{P}^n$ look the same, but given a vector and i times that vector, it behaves differently than a vector and some other vector. This is a handwavy idea, but it illustrates the point. A recap:

$$\text{frame isotropic} \implies \text{isotropic} \implies \text{homogeneous}$$

We study three different frame isotropic spaces. One is S^n, the other will be $\mathbb{R}^n$, and the third will be $\mathbb{H}^n$.

6.2 The action of the group of isometries on a space

Let a group G act transitively on a space X, that is, for all $x, y \in X$ there exists a g such that $gx = y$. Define $H_x = \{h \in G \mid h \cdot x = x\}$. This is the subgroup of G that fixes x, or the **stabilizer** of x. Let us show that

(1) H_x and H_y are conjugate.
(2) $X \simeq G/H$.

Proof. For (1), we want to show there exists a $g \in G$ such that $gH_xg^{-1} = H_y$. Since the action is transitive, we have $gx = y$ for some $g \in G$. Then for $h_x \in H_x$ (where $h_x x = x$),

$$gh_xg^{-1}(y) = gh_xx = gx = y.$$

So this fixes y, and is therefore an element of H_y. Similarly, consider $g^{-1}h_yg \in g^{-1}H_yg$, then

$$g^{-1}h_yg(x) = g^{-1}h_yy = g^{-1}y = x,$$

and therefore is an element of H_x. For (2), choose $x \in X$. Suppose we have a coset aH_x corresponding to ax, since $a(H_xx) = ax$. Since the action is transitive, we get all the points in the space. ⊠

In general, for a homogeneous space, we say $G = \{\text{isometries of } X\}$. These isometries act transitively, and we can apply our result above. For nonhomogeneous spaces this set may be small and doing this may be useless, but within the category of homogeneous spaces, we always want to look at the group of isometries. We also have $H_x = \{\text{isometries that sent } x \text{ to itself}\}$, the **isotropy group at x** (which we proved is the same for all x). This means we can just speak of the **isotropy group** of homogeneous spaces[8], rather than the isotropy group of the north pole or Austin, TX on S^2. So when we study homogeneous spaces, we always try to write it as G/H.

When can we tell when groups are isotropic? An isotropy means that H acts transitively on unit tangent vectors at a point. This means that it's big enough to send every vector to another vector, we simply look at H applied to an arbitrary vector and ask whether it hits any other vector. Frame isotropy in this context means that it takes every frame to every other frame. The group that does this is $\mathrm{O}(n)$, so if $H = \mathrm{O}(n)$ then we are frame isotropic.

Example 6.2. What are the symmetries of $\mathbb{C}\mathrm{P}^n$? In this case, $H = \widetilde{\mathrm{U}}(n)$, and since $\mathrm{U}(n)$ sends every vector to every other vector it is indeed isotropic. But it doesn't send every frame to every other frame, so $\widetilde{\mathrm{U}}(n) \neq \mathrm{O}(2n)$. However, $\mathrm{U}(1) = \mathrm{SO}(2)$, so $\mathbb{C}\mathrm{P}^1$ is frame isotropic.

6.3 Nice spaces (Part I): The plane $\mathbb{R}^n$

What's the nicest space there is? Consider $\mathbb{R}^n$ with the standard metric. The group of isometries G is generated by translations and rotations, where translations correspond to $\mathbb{R}^n$ and rotations correspond to $\mathrm{O}(n)$. There is a

[8] This group is sometimes called the **little group**.

natural action of $\mathrm{O}(n)$ on $\mathbb{R}^n$, but the actual thing we get is $\mathbb{R}^n \rtimes \mathrm{O}(n)$. The simplest way to describe this is by matrices; consider

$$H = \left(\begin{array}{c|c} A & v \\ \hline 0 & 1 \end{array}\right), \quad A \in \mathrm{O}(n),\ v \in \mathbb{R}^n.$$

Our composition rule is that

$$\left(\begin{array}{c|c} A & v \\ \hline 0 & 1 \end{array}\right)\left(\begin{array}{c|c} B & w \\ \hline 0 & 1 \end{array}\right) = \left(\begin{array}{c|c} AB & Aw+v \\ \hline 0 & 1 \end{array}\right).$$

You compose the rotations, and the subgroup where A is the identity turns the operation $Aw+v$ into simply adding vectors. If we take the quotient by $\mathrm{SO}(n)$, we get the translations. There is a subgroup where A is the identity, if we take the quotient by that subgroup we wind up with $\mathrm{O}(n)$. It isn't just the product with $\mathrm{O}(n)$ and $\mathbb{R}^n$ because the law mixes the two. So we don't call it the product, we call it the *semidirect* product. On the $\mathrm{O}(n)$ factor it's just usual multiplication, while on the $\mathbb{R}^n$ factor we have this funny business where $\mathrm{O}(n)$ interferes.

We want to figure out the isotropy group at a point, say the origin. This acts on a vector x by

$$\left(\begin{array}{c|c} A & v \\ \hline 0 & 1 \end{array}\right)\left(\begin{array}{c} x \\ \hline 1 \end{array}\right) = Ax + v.$$

The way to think about this is first rotating by A, then doing a translation by v. We want to see what happens when you act on the origin, and we want the isotropy group acting on the origin to give us back the origin. But it actually gives us back $(v \mid 1)$, so H is the set of rotations with no translational piece:

$$\left(\begin{array}{c|c} A & v \\ \hline 0 & 1 \end{array}\right)\left(\begin{array}{c} 0 \\ \hline 1 \end{array}\right) = \underbrace{\left(\begin{array}{c} 0 \\ \hline 1 \end{array}\right)}_{\text{(expected)}} = \underbrace{\left(\begin{array}{c} v \\ \hline 1 \end{array}\right)}_{\text{(actual)}} \quad \Longrightarrow \quad H_0 = \left(\begin{array}{c|c} A & 0 \\ \hline 0 & 1 \end{array}\right).$$

So $H_0 = \mathrm{O}(n)$, the rotation group. Hooray! This shows that the little group of $\mathbb{R}^n$ is the rotation group, which means $\mathbb{R}^n$ is frame isotropic. Recall $g_{ij} = \delta_{ij}$, and $g(v,w) = \sum v^i w^i$.

6.4 Nice spaces (Part II): The sphere S^n

The book works with spheres of radius R, but let's work with the standard unit sphere. This is defined as a subspace of $\mathbb{R}^{n+1}$, where $S^n := \{x \in \mathbb{R}^{n+1} \mid \|x\| = 1\}$. The set of isometries of the sphere is just the rotations in $(n+1)$-space $G = \mathrm{O}(n+1)$. Out of these, the rotations that fix the north pole just consist of the lower dimensional rotations, that is, $H_{(0,0,\cdots,1)} = \mathrm{O}(n)$.

If we wanted to restrict to things that preserve orientation, we would talk about *oriented* frames and whether a space is *oriented* frame isotropic, and ask whether $H = \mathrm{SO}(n)$, taking the quotient $\mathrm{SO}(n+1)/\mathrm{SO}(n)$. We have shown above that

$$\begin{aligned} S^n &= \mathrm{O}(n+1)/\mathrm{O}(n) \\ &= \mathrm{SO}(n+1)/\mathrm{SO}(n). \end{aligned}$$

This isn't how we usually think of a sphere, but if we really want to get into the symmetry, the quotient by (orientation preserving) rotations is the right way to think about a sphere. We can give an explicit map $\mathrm{O}(n+1) \to S^n$: an element of $\mathrm{O}(n+1)$ is an orthogonal matrix of the form $(\vec{b}_1, \cdots, \vec{b}_{n+1})$, where the columns are orthonormal. So $\mathrm{O}(n+1)$ is the set of all orthonormal frames of $n+1$ dimensions. The map simply sends $(\vec{b}_1, \cdots, \vec{b}_{n+1}) \to \vec{b}_{n+1}$.

Right multiplying by an element of $H_0 = \mathrm{O}(n)$ (as in the equation in the last section) does exactly this; it pulls out the last element.

Now we want to get coordinates on S^n and see what they look like. We do this by stereographic projection, measuring the length of a ray starting at the north pole going through a point, up to where it intersects the equatorial plane. Working out the steps is the homework, but it turns out that

$$g_{ij} = \begin{cases} 0 & \text{if } i \neq j, \\ \frac{4}{(1+\|u\|^2)^2} & \text{if } i = j. \end{cases}$$

Notice at each point, $g_{ij} = 4/(1+\|u\|^2)^2$ times the usual metric on $\mathbb{R}^n$. In other words, up to an overall scale, what happens at every point is just like what you would have on $\mathbb{R}^n$. If you take the metric on S^n and do a **conformal transformation** (rescale the metric at each point by a certain amount), we can turn $S^n \setminus \{\text{point}\}$ into $\mathbb{R}^n$. So the sphere and $\mathbb{R}^n$ are **conformal** to each other. This map doesn't preserve length, but preserves angles, which is the definition of conformal.

6.5 Nice spaces (Part III): Hyperbolic space $\mathbb{H}^n$

How about a space that curves the other way? We work in $\mathbb{R}^{n+1}$ with a **pseudometric**, since it's not positive definite. Define a pseudo inner product by

$$\langle v, w\rangle = v^1 w^1 + v^2 w^2 + \cdots + v^n w^n - v^{n+1} w^{n+1}.$$

This is **Minkowski spacetime**, with n space coordinates and one time coordinate. We usually write a vector in $\mathbb{R}^{n+1}$ as (x, t) for $x \in \mathbb{R}^n$ and $t \in \mathbb{R}$. So $\langle (v,t),(w,s)\rangle = v \cdot w - ts$. What would the unit sphere be? We actually have *two* interesting unit spheres now, since one is where you set $\langle x, x\rangle = 1$, and the other where you set $\langle x, x\rangle = -1$. Let's use $\langle x, x\rangle = -1$. This is no longer the familiar sphere you know, since it's defined by the equation

$$(x^1)^2 + (x^2)^2 + \cdots + (x^n)^2 - t^2 = -1, \quad \text{or}$$
$$t^2 = 1 + (x^1)^2 + (x^2)^2 + \cdots + (x^n)^2.$$

If $n = 2$, then $t^2 = 1 + (x^1)^2 + (x^2)^2$ defines the two sheeted hyperboloid. If we let $\mathbb{H}^n$ just be the upper half of the plane, this is *sort of* like talking about a sphere, since we *sort of* have a metric (it's indefinite), and we only consider the upper connected component. The next question to ask is "what is the metric?" Our metric is as follows: given a point p on the upper half plane and two tangent vectors, take the indefinite inner product of $\mathbb{R}^{n+1}$ and restrict it to the tangent space at a particular point. Our claim is that this doesn't give us a *pseudo*metric, it gives us an *actual* metric.

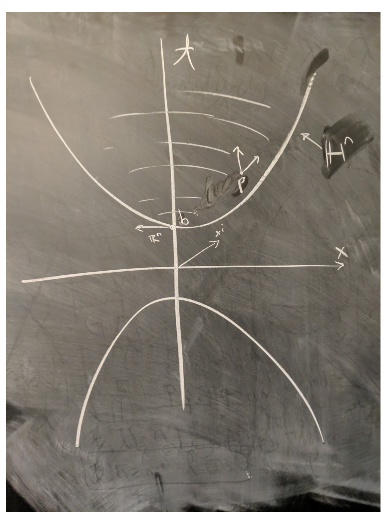


Figure 1: A model of hyperbolic space in $\mathbb{R}^{n+1}$.

For example, at the base b the tangent vectors have no t component, so the inner product is just the one on $\mathbb{R}^n$. The set of isometries are going to be the set of linear isometries of $\mathbb{R}^{n+1}$ with this funny metric $\mathrm{O}(n,1)$, and it's not hard to cook up one that sends $b \to p$. So the metric on $T_p\mathbb{H}^n$ behaves like the metric on $T_b\mathbb{H}^n$, because the map that sends $b \to q$ preserves the $(n,1)$ inner product, and once you end up at the base b the time coordinate doesn't affect you at all.

So this defines a space where the points all look alike, since $\mathrm{O}(n,1)$ sends every point to every other point. We need to explicitly construct an element of $\mathrm{O}(n,1)$ that sends $b \to q$. Let us do this in two dimensions. Say p is of the form $(p,t) = (p^1, 0, 0, \cdots, t)$. Consider

$$\left(\begin{array}{cccccc|c} \cosh(\alpha) & & & & & & \sinh(\alpha) \\ & 1 & & & & & \\ & & 1 & & & & \\ & & & 1 & & & \\ & & & & 1 & & \\ & & & & & 1 & \\ \hline \sinh(\alpha) & & & & & & \cosh(\alpha) \end{array}\right).$$

The claim is that this is an element of $\mathrm{O}(n,1)$, sending $(0,0,\cdots,1)$ to $(\sinh\alpha, 0, 0, \cdots, \cosh\alpha)$. Since $\cosh^2(\alpha) - \sinh^2(\alpha) = 1$, there exists an α such that for $(p^1, 0, 0, \cdots, t)$, $p^1 = \sinh(\alpha)$ and $t = \cosh(\alpha)$ (since $t^2 - (x^1)^2 = 1$ by the definition of our pseudometric). More explicitly, this is the map

$$\left(\begin{array}{cccccc|c} t & & & & & & x \\ & 1 & & & & & \\ & & 1 & & & & \\ & & & 1 & & & \\ & & & & 1 & & \\ & & & & & 1 & \\ \hline x & & & & & & t \end{array}\right), \quad \text{since } t^2 - x^2 = 1.$$

This is an explicit map in $\mathrm{O}(n,1)$ sending $(0,0,\cdots,1) \to (x,0,\cdots,t)$. In n dimensions, replace the component $(x,0,0,\cdots)$ with any vector of length n. This gives us our isometries, since if we have an isometry of the underlying vector space $\mathbb{R}^{n+1}$, this induces an isometry on the unit sphere. This unit sphere is actually a **Riemannian** metric since we have an isometry taking a point with a positive definite metric to another point. So for hyperbolic space,

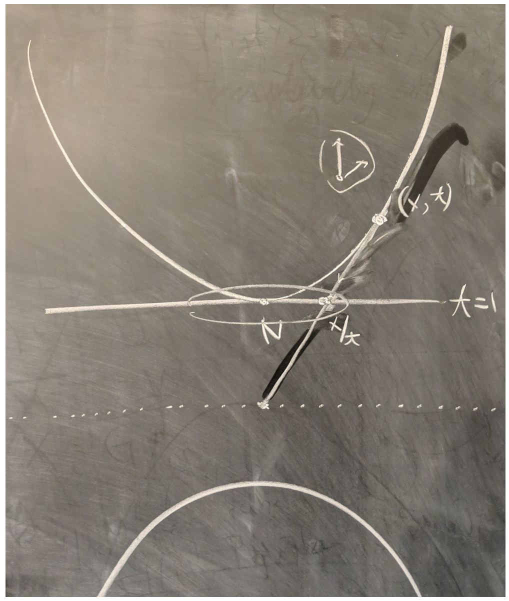
$$G = \mathrm{O}(n,1) = \left\{ A \in M(n+1) \;\middle|\; A^T \begin{pmatrix} 1 & & & & 0 \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ 0 & & & & -1 \end{pmatrix} A = \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & -1 \end{pmatrix} \right\}$$

If we replaced the -1 factor with 1, this would become the set of matrices A such that $A^T A = I$, which is just $O(n+1)$. Similarly,

$$H = O(n) = \left\{ \left(\begin{array}{c|c} B & 0 \\ \hline 0 & 1 \end{array}\right) \;\middle|\; B^T B = I \right\} < O(n+1,1).$$

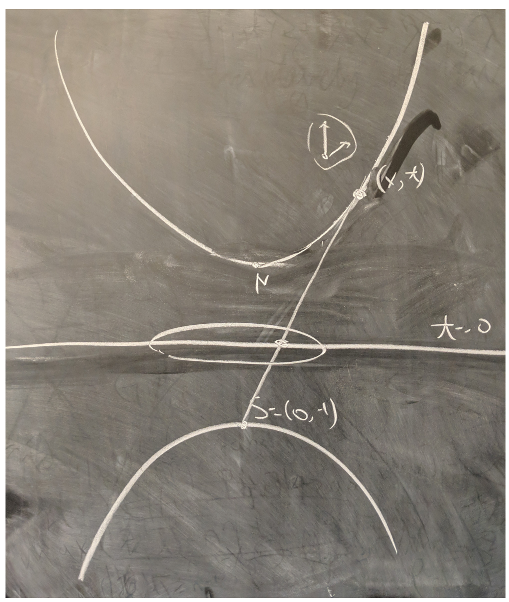
So hyperbolic space is frame isotropic! This is the abstract definition of hyperbolic space, but what does the metric look like? We need to get coordinates on $\mathbb{H}^n$, for which there will be several different models. There are three ways:

(1) Consider the plane $t = 1$. For each point (x,t), project down to the origin through that plane. This intersects $t = 1$ at a point x/t contained in the unit disk, since $t^2 - |x^2| = 1$. So $x/t < 1$ since t is bigger than x.



Parametrizing by points in the unit disk by projection gives us the **Beltrami-Klein** model.

(2) The second way is by the analogue of stereographic projection. Rather than consider a line through the north pole, we consider a line through (x, t) and the *south pole* $S = (0, -1)$.



The point ends up in a unit disk here too, but a different one: in the previous model, we had the Beltrami-Klein disk, while this is called the **Poincare disk model**. Both will give you metrics (but it turns out this metric will be conformal).

(3) The final model will give us a metric on the **upper half space**, where we take $\mathbb{R}^{n-1} \times [0, \infty)$ and map the disk onto it.

For all of these, first we work it out in two dimensions, then n dimensions. However, we actually understand n dimensional hyperbolic space abstractly: we have seen that $\mathbb{H}^n = \mathrm{O}(n, 1)/\mathrm{O}(n)$ and is frame isotropic. What is this mysterious frame isotropic space that is not a sphere nor a plane? We will see next time.

Lecture 7

February 4, 2021

todo:hyperbolic spce

Lecture 8

February 9, 2021

8.1 Recap of hyperbolic space

Last time we talked about how to view a hyperbolic space a couple different ways, listed below. (For now we limit our attention to $\mathbb{H}^2$, but in n dimensions arguments work pretty much the same way.)

(1) The hyperboloid model. Here we consider the upper half plane $z^2 = x^2 + y^2 + 1$ (for $z \geq 1$) sitting in $\mathbb{R}^3$, with the metric $ds^2 = dx^2 + dy^2 - dz^2$.

(2) The Klein-Beltrami model. Consider the disk $u^2 + v^2 < 1$, with the metric $d^2 =$ something messy.

(3) The Poincare disk model. Once again consider $u^2 + v^2 < 1$, with the metric $ds^2 = \frac{4(du^2+dv^2)}{(1-u^2-v^2)^2}$.

(4) The upper half plane. Consider all pairs (x, y) for $y > 0$, with the metric $d^2 = \frac{dx^2+dy^2}{y^2}$.

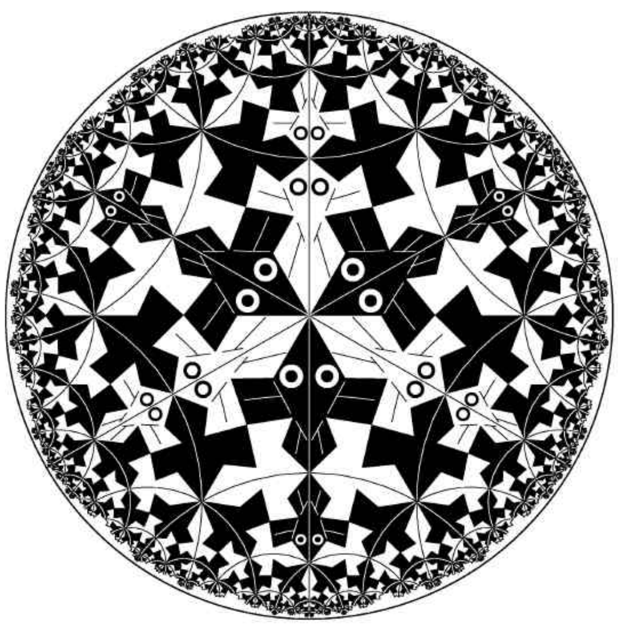


Figure 2: Obligatory Escher print.

This tells us a couple of things: all the little dudes are the same size. You can also track straight lines on the disk, the diameter lines definitely are but so are the arcs outlined in Section 8.1. It looks a little funny because this is a disk of infinite volume; the metric gets closer to zero as you approach the boundary, and if you try to integrate from the center out you'll see that it diverges.

8.2 The upper half plane

Let x,y for $y>0$. It's easier to package them in a complex number $z=x+iy$. So our metric is given by

$$ds^2=\frac{dx^2+dy^2}{y^2}=\frac{dz\,d\overline{z}}{(\operatorname{Im}z)^2}=\frac{|dz|^2}{(\operatorname{Im}z)^2}.$$

We claim there is a natural group acting on this space, namely $\mathrm{SL}(2,\mathbb{R}):=\left\{\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)\,\middle|\,ad-bc=1\right\}$. This acts on $\mathbb{C}$ by the rule $\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)\cdot z=\frac{az+b}{cz+d}$. You can think of this action as $\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)\left(\begin{smallmatrix}z\\1\end{smallmatrix}\right)=\left(\begin{smallmatrix}az+b\\cz+d\end{smallmatrix}\right)$, so when the matrix acts on z itself, it transforms the ratio of $z:1$ to $az+b:cz+d$. If you think about the linear action of matrices on $\mathbb{C}^2$ and the induced action it has on quotients, this is really an action on $\mathbb{CP}^1$. Call this ratio $\frac{az+b}{cz+d}=w$, and this transformation is called a **Möbius transformation**. Does this really take the upper half plane to the upper half plane? To do this, let's see if $\operatorname{Im}w=(w-\overline{w})$[9] is positive. We have

$$\begin{aligned}w-\overline{w}&=\frac{az+b}{cz+d}-\frac{a\overline{z}+b}{c\overline{z}+d}\\&=\frac{z-\overline{z}}{|cz+d|^2}.\end{aligned}$$

[10]So $\operatorname{Im}w=\frac{\operatorname{Im}z}{|cz+d|^2}>0$, and this action is really a transformation of the upper half plane. Further more, this is a local isometry. Since $w=\frac{az+b}{cz+d}$, $dw=\frac{a(cz+d)-c(az+b)}{(cz+d)^2}dz=\frac{dz}{(cz+d)^2}$. Similarly, $d\overline{w}=\frac{d\overline{z}}{(c\overline{z}+d)^2}$, and $dw\,d\overline{w}=\frac{dz\,d\overline{z}}{|cz+d|^4}$. But the imaginary part transforms by $\frac{1}{|cz+d|^2}$, so

$$\frac{|dw|^2}{(\operatorname{Im}w)^2}=\frac{|dz|^2}{(\operatorname{Im}z)^2},$$

and our metric is preserved. We have just shown that we have a group acting on the upper half plane by isometries; the next question to ask is "is the action transitive?" Here are some particularly interesting elements of $\mathrm{SL}(2,\mathbb{R})$.

- $\begin{pmatrix}1&a\\0&1\end{pmatrix}$: this sends $z\mapsto z+a$, which translates z by a in the *real* direction.
- $\begin{pmatrix}\sqrt{b}&0\\0&\frac{1}{\sqrt{b}}\end{pmatrix}$: this sends $z\mapsto bz$, which scales z by b.

If you consider i as a "center" point for your model, we have $\left(\begin{smallmatrix}1&a\\0&1\end{smallmatrix}\right)\left(\begin{smallmatrix}\sqrt{b}&0\\0&1/\sqrt{b}\end{smallmatrix}\right)\cdot i=a+bi$, and composing the matrices gives $\left(\begin{smallmatrix}\sqrt{b}&a/\sqrt{b}\\0&1/\sqrt{b}\end{smallmatrix}\right)$ as the map sending $i\mapsto a+bi$. Our action is transitive, since i goes to any point, and the inverse matrix sends any point to i. Then any point gets sent to any point by first sending it to i, then the other point.

The next question to ask is "what is the isotropy group of a point?" We have

$$\begin{pmatrix}a&b\\c&d\end{pmatrix}i=i,\quad\frac{ai+b}{ci+d}=i\implies ai+b=di-c.$$

Then $a=d$ and $c=-b$, so we are looking at matrices of the form $\left(\begin{smallmatrix}a&b\\-b&a\end{smallmatrix}\right)$; furthermore these matrices have determinant 1, so $a^2+b^2=1$. This becomes all matrices of the form $\left(\begin{smallmatrix}\cos\theta&\sin\theta\\-\sin\theta&\cos\theta\end{smallmatrix}\right)$, which is just $\mathrm{SO}(2)$. The point is applying one of these matrices to i rotates the plane by 2θ. Note that

$$i\mapsto\frac{\cos\theta i+\sin\theta}{-\sin\theta i+\cos\theta}=\frac{ie^{-i\theta}}{e^{-i\theta}}.$$

The derivative map, how much you rotate tangent vectors, is $\frac{1}{(cz+d)^2}=\frac{1}{e^{-2i\theta}}=e^{2i\theta}$. In particular, if $\theta=\pi$, then we rotate back to 2π. This is not intrinsic to i, in general, $\left(\begin{smallmatrix}-1&0\\0&-1\end{smallmatrix}\right)z=z$. So $\mathrm{Isom}(\mathbb{H}^2)=\mathrm{SL}(2,\mathbb{R})/\pm I=\mathrm{PSL}(2,\mathbb{R})$, and

$$\begin{aligned}\mathbb{H}&=\mathrm{SL}(2,\mathbb{R})/\,\mathrm{SO}(2)\\&=\mathrm{PSL}(2,\mathbb{R})/\,\mathrm{PSO}(2).\end{aligned}$$

[9]It's actually $\frac{w-\overline{w}}{2i}$, but we're just trying to check whether it's positive or not.
[10]Some algebra was skipped.

8.3 Geodesics

Let (M, g) be a Riemannian manifold, and $p, q \in M$. We want to define the distance from p to q. The first thing we do is define distance along a path: think of a path $\gamma(t)$ such that $\gamma(t_0) = p$ and $\gamma(t_f) = q$. Then we define the **length** of γ by

$$L(\gamma) = \int_{t_0}^{t_f} \sqrt{g(\dot{\gamma}, \dot{\gamma})}\, dt.$$

This doesn't depend on parametrization, if you reparametrize you get the same integral. Define the **distance** from p to q as

$$d(p, q) = \inf_{\gamma} L(\gamma),$$

where γ is a path from p to q. Unfortunately, a **geodesic** has two meanings, and we use both. One is that a geodesic is a locally length minizing path. The second definition of the geodesic is something that minimizes a different functional called the **energy** of the path. Things that minimize energy are length minimizing paths parametrized at constant speed. So for every geodesic, there's a *best* parametriation where you move along at constant speed, and we say "geodesic" we speak of that parametrization. For now we're just talking about geometric paths, and want something that minimizes length.

The geodesic joining two points is not always unique. Consider $\mathbb{R}^2 \setminus \{0\}$, and the line connecting $(-1, 0)$ and $(1, 0)$. There is no path of length 2 connecting them, since it would have to go through the origin (not in the space). You can find a path that takes a small detour, as small as you want, so the inf(all paths) still equals 2. But there is no geodesic from $(-1, 0)$ to $(1, 0)$. So geodesics don't always exist globally[11]: they do always exist locally however, which we will soon see.

8.4 Geodesics in nice spaces

Our three nice spaces are the plane, the sphere, and hyperbolic space.

(1) Consider $p, q \in \mathbb{R}^2$ with $\gamma(0) = p$ and $\gamma(1) = q$. Can you show the shortest path is a straight line? (This was homework for diff geo.) Assume $p = (0, 0)$ and $q = (1, 0)$ for simplicity (since length is invariant under isometries, and the argument works for arbitrary lengths). If our path is given by $y = \gamma(x)$, then

$$\begin{aligned} L(\gamma) &= \int_0^1 \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx \\ &\geq \int_1^0 \sqrt{1}\, dx = 1. \end{aligned}$$

We get equality iff y is constant, since wiggling in the y direction only makes the $(dy/dx)^2$ factor bigger, increasing the length. (Wow this was a lot simpler than the diff geo proof.)

(2) Now we move to S^2. If (φ, θ) parametrize the sphere, then $ds^2 = d\varphi^2 + \sin^2\varphi (d\theta)^2$. (Figure out which coordinate corresponds to the azimuthal or whatever angle from that expression.) What's the shortest path from the north pole to a point (φ, θ)? We have

$$\begin{aligned} &\int \sqrt{\left(\frac{d\varphi}{dt}\right)^2 + \sin^2\varphi \left(\frac{d\theta}{dt}\right)^2}\, dt \\ &\geq \int \left|\frac{d\varphi}{dt}\right| dt \geq \varphi_{\text{final}}. \end{aligned}$$

This is the same argument as before: any wiggles in the θ direction only contribute to $(d\theta/dt)^2$, increasing the length. You also always want $d\varphi/dt = |d\varphi/dt|$, i.e. no doubling back. So the length minimizing path is when θ is constant and φ is monotonic, i.e. you're going straight down the longitudinal line. This continues

[11] They always exist for *compact* manifolds.

all the way down to the south pole. So we have just shown that a great circle is a geodesic. To parametrize this, you can send $t = \cos\theta(0,0,1) + \sin\theta\vec{v}$, where $\vec{v}$ lives on a great circle.

Suppose we start at some other point p with tangent vector v, where $p \in \mathbb{R}^3$ and $p \perp v \in \mathbb{R}^3$. Can you find a geodesic that goes through p along v? Geodesics are invariant under isometry, so send p to the north pole. Now find a geodesic/great circle going through v, and rotate back. We have shown that $\gamma(t) = \cos(t)p + \sin(t)v$ is a geodesic.

Spherical triangles have area of the sum of the angles minus π. In a plane this sum minus π is zero, but in the sphere it's not. (A neat proof is in the diff geo textbook/my diff geo notes.)

(Digression!). We never discussed what area means for a Riemannian manifold. The idea of area is that you chop things up into little pieces, and find the area of each piece. This is a parallelogram spanned by two vectors v, w. This is going to be some sort of determinant. We want

$$\begin{aligned} |v||w|\sin\theta &= \sqrt{|v|^2|w|^2 - |v|^2|w|^2\cos^2\theta} \\ &= \sqrt{(v\cdot v)(w\cdot w) - (v\cdot w)^2} \\ &= \sqrt{g_{11}g_{22} - g_{12}^2} \\ &= \sqrt{\det g}. \end{aligned}$$

We gave an argument for two dimensions, but the same thing works for arbitary dimensions. In general

$$\int \sqrt{\det g}\, d^n x = \text{volume}.$$

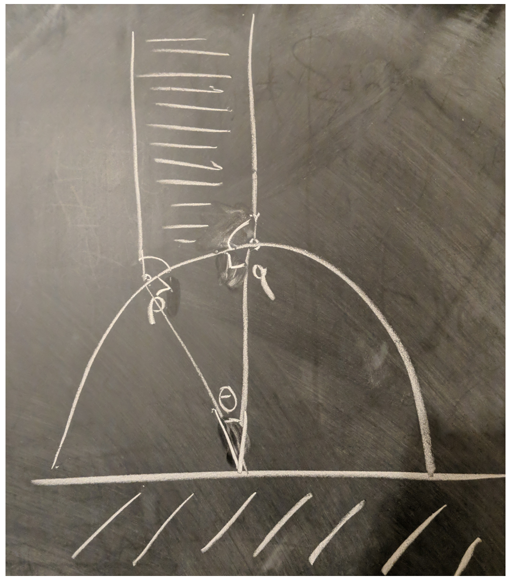
In the case of S^2, $g_{11} = 1, g_{12} = g_{21} = 0, g_{22}\sin^2\varphi$. So $dA = \sin\varphi d\varphi d\theta$, and then simply integrate. For stereographic coordinates, $g_{11} = g_{22} = \frac{4}{(1+u^2+v^2)^2}, g_{12} = 0$, so $\sqrt{\det} = \frac{4}{(1+u^2+v^2)}$. So $dA = (4\,du\,dv)/(1+u^2+v^2)^2$. In the Poincare disk, just replace u^2, v^2 with minus signs. In the upper half plane, $g_{11} = g_{22} = \frac{1}{y^2}, g_{12} = 0$, $\det = \frac{1}{y^r}, \sqrt{\det} = \frac{1}{y^2},\ dA = \frac{1}{y^2}dx\,dy$.

(3) Now we turn our attention to hyperbolic space. Suppose we have two point i, iy_1. Then the geodesic is also a straight line, since

$$d = \int \frac{\sqrt{dx^2 + dy^2}}{y} \geq \int_1^{y_1} \left|\frac{dy}{y}\right| = |\ln|y_1||.$$

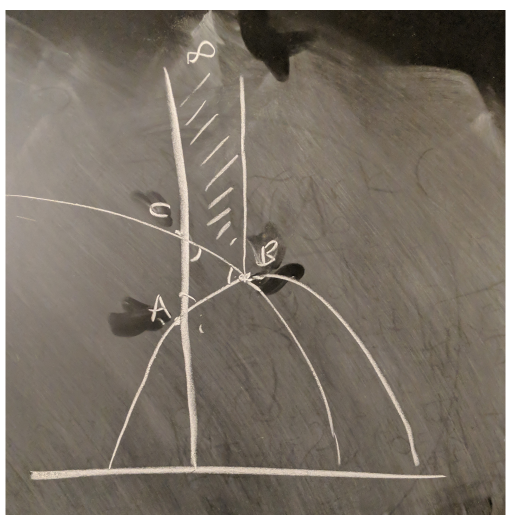
This diverges as you approach the outer edge, and also as you approach the center. Now for arbitrary p, q, send $p \to i$ and rotate until q winds up above i by our aforementioned transformations. What happens when you send a line through an isometry? The translation $\left(\begin{smallmatrix}1 & a\\ 0 & 1\end{smallmatrix}\right)$ shifts a vertical line to the right by a factor of a, while dilation $\left(\begin{smallmatrix}\sqrt{b} & 0\\ 0 & 1/(\sqrt{b})\end{smallmatrix}\right)$ does nothing to a vertical line. Now for the Mobius transformation $\left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right)$, $z \mapsto \frac{az+b}{cz+d}$, as z gets large, this is approximately a/c. As z gets small, this is approximately b/d. So the image of a vertical line is going to be something that starts on the real axis at b/d and goes to a/d (you know b/d and a/c are distinct since $ad - bc = 1$). The remarkable thing is that this turns out to be a circle centered at the midpoint $(b/d + a/c)/2$, with radius $(a/c - b/d)/2 = (ad - bc)/2cd = 1/2cd$. So our geodesics are either vertical lines or semicircles.

Consider an **ideal triangle**, which is the limit of a triangle as one of the vertices goes off to infinity.



The sum of the angles at p and q will be less than π. How much less? The claim is that it will be less than π by exactly θ, since summing up $\angle p, \angle q$, the right angles, and θ gives exactly 2π. The right angles plus θ is $\pi + \theta$, so $\angle p + \angle q = \pi - \theta$.

How does this relate to areas? Recall that we want to integrate $\int \frac{dx\,dy}{y^2} = \int \frac{dx}{y}$, since $\int_{y_0}^{\infty} \frac{dy}{y^2} = \frac{1}{y_0}$. On a circle centered at the x-axis, $\int \frac{dx}{y} = \int d\theta = \theta$. So in hyperbolic space, the angle deficit is the area. One last thing: usually you don't care about ideal triangles, you care about real triangles. Send it through an isometry takes one vertex to i, the other above i, and the third somewhere else.



The claim is that the sum of the angles of ΔABC is the sum of the angles of $\Delta AB\infty$ (ideal triangle) *minus* the sum of the angles of $\Delta BC\infty$ plus π. So the deficit for the real triangle is the deficit of the big ideal triangle minus the deficit of the little ideal triangle. Skipping some steps, we have shown that the sum of the angles in hyperbolic space is $\pi -$ Area Δ.

Corollary 8.1. *The area of a triangle is less than π.*

Hyperbolic geometry is wack. See you next time.

Lecture 9

February 11, 2021

9.1 More on geodesics in hyperbolic space

Last time we were talking about geodesics in hyperbolic space. We have seen that vertical lines are geodesics, and an element $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ acting on the line sends it to an arc starting at a/c and ending at b/d WLOG. Then shifting by isometries, we can assume that $a/c = -b/d$, which implies $c/d = -a/b$. So

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} : iy \to \frac{b + aiy}{d + ciy} = \frac{b\left(1 + \frac{a}{b}iy\right)}{d\left(1 + \frac{c}{d}iy\right)}.$$

Since the expresssions in the numerator and denominator are conjugate, the image of iy is a point of constant magnitude b/d, which implies that geodesics are indeed parts of a circle centered at the x-axis. Now that we know what geodesics are, this tells us a lot of things. Consider a point $(a, 0)$, with our geodesics a circle of radius R. Parametrize by θ where $x = a + R\cos\theta$, $y = R\sin\theta$, $dx = -R\sin\theta\, d\theta$, $dy = R\cos\theta\, d\theta$, with metric

$$ds^2 = \frac{dx^2 + dy^2}{y^2} = \frac{R^2 d\theta^2}{R^2\sin^2\theta} = \frac{d\theta^2}{\sin^2\theta} \implies ds = \pm\frac{d\theta}{\sin\theta}.$$

The $\pm$ sign stems from the fact that when calculating distance we can go around a circle either clockwise or counterclockwise. Integrating gives us $S = \pm\ln|\csc\theta + \cot\theta| = \pm\ln|R/y + (x-a)/y| = \pm\ln|(R + x - a)/y|$. This diverges logarithmically as $x \to a + R$, $y \to 0$, which is expecting since this is an isometry of the vertical line. As $x \to a - R$ and $y \to 0$, both factors approach zero so it's a little trickier. However you can rewrite this as $\ln|y/(R - x - a)|$, and so the denominator approaches zero as expected and this also diverges logarithmically. So because of our metric, this arc is infinitely long on both sides.

Given two points $p, q \in \mathbb{H}^2$, we want to find the semicircle going through p and q. To do this, draw a line through p and q, then the perpendicular bisector will hit the x axis at a point an equal distance R from both p and q. This forms a triangle, and now you can just compare values of $\csc\theta, \cot\theta$, etc. So we have just figured out the *entire* metric in hyperbolic space, since given any two points we can calculate the distance between them.

9.2 Triangles in hyperbolic space

Recall from last time that the sum of angles α, β of an ideal triangle is $\pi - \theta$, and Area $= \theta$. How do we know this? We have

$$\iint_A \frac{dx\,dy}{y^2} = \int dx \int_{y(x)}^{\infty} \frac{1}{\tilde{y}^2}\, d\tilde{y} = \int \frac{dx}{y} = \int \frac{dx}{y} = \int \frac{R\sin\theta\, d\theta}{R\sin\theta} = \int -d\theta = -\theta.$$

The negative value stems from the fact that we integrate from left to right, while θ is measured from right to left. So $\alpha + \beta = \pi - $Area, a surprisingly nice result when compared to our weird expression for length. To find a similar expression for *actual* triangles, say α, β, γ are the angles of a triangle T where β, γ lie on a semicircle. Then the complement of α is $\pi - \alpha$, and denote the angle of the ideal triangle corresponding to β by δ.[12] Let B denote the ideal triangle formed by $\pi - \alpha$ and δ. We have

$$\begin{aligned} \pi - \alpha + \delta &= \pi - (\text{Area of } B) \\ \delta + \beta + \gamma &= \pi - (\text{Area of } B) - (\text{Area of } T) \\ \alpha + \beta + \gamma - \pi &= -\text{Area}(T) \\ \boxed{\pi - (\alpha + \beta + \gamma) = \text{Area}(T).} \end{aligned}$$

Wow! We figured out the area of an arbitrary triangle just by using symmetry, no messing with Christoffel symbols or curvature or anything.

[12] This is for triangles with two vertices α, γ positioned directly on top of each other, but you can write any triangle this way up to isometry.

Corollary 9.1. *The area of a triangle is always less than π.*

To visualize this result, imagine a triangle with two base points infinitely close to the x-axis, and a third point all the way out in the y direction (say $y = 3 \times 10^{45}$). Then by our metric $ds^2 = (dx^2 + dy^2)/y^2$, as y becomes very large the two "parallel" vertical geodesics stemming from our two basepoints get very close. Likewise, the parallel lines get infinitely close to the endpoints of the semicircle connecting the basepoints. So just imagine a triangle in the Poincare disk model with infinitely tight pinched edges.

This is in contrast with the sphere, where a triangle has bloated edges. There, a triangle can even have three right angles. This completes our tour of hyperbolic space, and chapter three of the book! Now we move on to talking about geodesics.[13]

9.3 Geodesics as energy minimizing curves

For a Riemannian manifold (M, g), let γ be a path taking a point $p \in M$ to a point $q \in M$, where $\gamma(0) = p$ and $\gamma(t) = q$. Recall that $L(\gamma) = \int \sqrt{g(\dot{\gamma}, \dot{\gamma})}\, dt$. The inner product of velocity with itself is velocity squared, taking the square root gives speed and integrating WRT time gives length. Our previous notion of geodesics was thinking about them as length minimizing curves. Now consider the **energy** of a path, where

$$E(\gamma) = \int g(\dot{\gamma}, \dot{\gamma})\, dt.$$

We will study curves that minimize energy and see what they look like. In one dimension, consider $M = (\mathbb{R}, dx^2)$. We want a path from 0 to L in time T, that is, $\gamma(0) = 0$, $\gamma(T) = L$, and for this path to minimize $\int_0^T (d\gamma/dt)^2\, dt = \int_0^T (dx/dt)^2\, dt$. We expect this to be the path with constant velocity, since any other path gets traversed faster, ergo more energy.

The Cauchy-Schwarz inequality for L_2 space. Recall that Cauchy-Schwarz tells us that $\|\langle u, v\rangle\| \leq \langle u, u\rangle \cdot \langle v, v\rangle$. Applying this to the norm on the space of square-integrable functions (where $\langle f, g\rangle = \int_A f(x)\overline{g(x)}\, dx$) gives

$$\left| \int_{\mathbb{R}^n} f(x)\overline{g(x)}\, dx \right|^2 \leq \int_{\mathbb{R}^n} |f(x)|^2\, dx \int_{\mathbb{R}^n} |g(x)|^2\, dx.$$

Now that we know this very useful inequality, note that

$$\begin{aligned}
TE(\gamma) &= \int_0^T 1^2\, dt \int_0^T \left(\frac{dx}{dt}\right)^2 dt \\
&\geq \left(\int_0^T \left(1 \cdot \frac{dx}{dt}\, dt\right)^2 \right) = L^2. \\
E(\gamma) &\geq \frac{L^2}{T}.
\end{aligned}$$

We have equality precisely when (dx/dt) is a constant. A more elementary argument is that

$$\begin{aligned}
&\int \left(\frac{dx}{dt}\right)^2 dt - \int \left(\frac{dx}{dt} - \frac{L}{T} + \frac{L}{T}\right)^2 dt \\
&\qquad = \int \left(\frac{dx}{dt} - \frac{L}{T}\right)^2 + \underbrace{\frac{2L}{T}\left(\frac{dx}{dt} - \frac{L}{T}\right)}_{=0} + \frac{L^2}{T^2}\, dt \\
&\qquad = \int \left(\frac{dx}{dt} - \frac{L}{T}^2\right) dt + \frac{L^2}{T}.
\end{aligned}$$

[13] The subtitles for the recorded lecture always say "genetics" instead of geodesics, and I find it hilarious that we so often abruptly switch from talking about math to talking about biology.

You minimize this expression by setting the left integral to zero, or setting $(dx/dt) = L/T$ everywhere. This is essentially the same proof, since proving Cauchy-Schwartz uses a complete the square argument like this. Let us also solve this by using an overkill technique, called calculus of variations. This generalizes and is quite useful, which is why we're demonstrating how to prove this three times over.

9.4 Calculus of variations

Given a function $f(x)$ on $\mathbb{R}^n$, how do you find the minima? We compute ∇f and set it equal to zero. This may also find maxima and saddle points, but if we're at a minima, it's locally constant. Another way to say this is that $f(\vec{x}+\delta\vec{x}) = f(\vec{x}) + L(\delta\vec{x}) + O(\delta\vec{x})$. We want $L(\delta\vec{x}) = 0$, otherwise we would have a nontrivial change in f to the first order. This linear function L is $\nabla \cdot f$, or df applied to the change in x.

Back to our original problem. We have $x(t)$, where $x(0) = 0$ and $x(T) = L$. We change $x(t)$ to $x(t) + \delta x(t)$, and compute $E(X+\delta x) - E(x) = L(\delta x) + O(\delta x^2)$. We throw away the higher order terms, and now we want $L(\delta x) = 0$. So

$$\begin{aligned} E(x+\delta x) &= \int \frac{d(x+\delta x)^2}{dt} \\ &= \int \underbrace{\left(\frac{dx}{dt}\right)^2}_{=E(x)} + 2\frac{dx}{dt}\frac{d(\delta x)}{dt} + \underbrace{\left(\frac{d\,\delta x}{dt}\right)^2}_{\substack{\text{second order,}\\ \text{ignore}}} dt. \end{aligned}$$

Then

$$\begin{aligned} \delta E &= \int 2\frac{dx}{dt}\frac{d(\delta x)}{dt}\,dt \\ &= \underbrace{2\frac{dx}{dt}\delta(x)\Big|_0^T}_{\substack{=0,\ \delta x(0)=0,\\ \delta(T)=0}} - \int 2\frac{d^2x}{dt^2}\delta x(t)\,dt, \\ \delta E &= -\int 2\ddot{x}\delta x\,dt, \quad \frac{\delta E}{\delta x(t)} = -2\ddot{x}(t). \end{aligned}$$

When will this linear function of δx be zero? This is true only if $\ddot{x} = 0$. So our solution must have $\ddot{x}(t) = 0$, $\dot{x}(t) =$ constant. Since we're going from position 0 to position L over time T, we must have $\dot{x} = L/T$.

For an arbitrary manifold, consider a path γ from p to q. How would you minimize $E(\gamma)$? Geometrically we want to take the shortest path possible: once you go along a path with total length, we basically arrive at the one dimensional problem, where we travel a distance L along time T. We do this by going at constant speed L/T. So for an energy minizing geodesic, we want to:

(1) Minimize length.
(2) Go at constant speed, given by $\sqrt{g(\dot{\gamma},\dot{\gamma})}$.

By minimizing energy, we're killing two birds with one stone: we get minimal length for free, while also achieving constant speed. Functionally, minimalizing energy is easier than minimizing length because there's no square root involved.

9.5 Motivating connections

In $\mathbb{R}^n$, you minimize length by requiring $\ddot{x} = 0$ and $\dot{x} =$ constant. In $S^2 \subseteq \mathbb{R}^3$, we have already determined geodesics to be great circles, so we go around the equator (by sending any point to a pole) at constant speed. Our acceleration points towards the center of the sphere, so $\ddot{x} \perp T_pS^2$. If you're driving on the sphere, you never turn (acceleration points toward a different T_pS^2) and never slow down or speed up (acceleration in the forwards/backwards direction).

In hyperbolic space, what is $\ddot{x}$? This is a trick question. You know what it means to take a second derivative with respect to coordinates. When we say $\ddot{x} = 0$ in $\mathbb{R}^n$, we really mean $\ddot{x}_i = 0$. Using different coordinates (such as polar), it would not be true that $\ddot{r}, \theta = 0$. This statement that $\ddot{x} = 0$ is coordinate dependent, and in Euclidian space we have these beautiful coordinates to play with. For S^2, the acceleration refers to taking the second derivative with respect to coordinates in $\mathbb{R}^3$.

We don't have these coordinates in hyperbolic space: we have models, but those are all different! To answer this, we'll need to develop a coordinate free way of taking coordinates, to answer the question "what does $\frac{d}{dt}$(vector) even mean?" We know $\frac{d}{dt}$(position) is just a tangent vector. But $\ddot{x}$ is the derivative of a *tangent vector*, and we take this derivative at different points. Given a vector field on a manifold, what is its derivative? This is what **connections** are for, along with *covariant derivatives* and *parallel transport*. Given a nice connection, we can ask what the second derivative is, then set the acceleration to zero, and nice things happen. But we need to develop this machinery first.

Lecture 10

February 25, 2020

Recall that geodesics in $\mathbb{R}^n$ between two points p and q are straight lines, explicitly given by $X(t) = p + (q-p)t/T$, moving with constant velocity $\dot{X}(t) = (q-p)/T$, with $\ddot{X} = 0$. On S^2, geodesics are great circles, parametrized by (assuming we're at the equator) $\vec{x} = (x, y, z) = (\cos t, \sin t, 0)$. So $\dot{\vec{x}} = (-\sin t, \cos t, 0)$ and $\ddot{\vec{x}} = (-\cos t, -\sin t, 0)$. The acceleration has no tangential component, so it's only enough to keep us on the surface of the sphere.

But what about hyperbolic space? What does a derivative along a function mean? Given a function $f(x)$ along a curve, we have

$$\frac{d}{dt} f(x) = \frac{d}{dt}(f \circ x(t)) = \lim_{h \to 0} \frac{f(x(t+h)) - f(x(t))}{h}.$$

It's perfectly sensible to subtract the value of a function at two points. However,

$$\frac{d}{dt}\dot{x} \stackrel{?}{=} \lim_{h \to 0} \frac{f(\dot{x}(t+h)) - f(\dot{x}(t))}{h}.$$

The issue is $\dot{x}(t+h)$ and $\dot{x}(t)$ live in two different tangent spaces. How do we take vectors in one space and move them to vectors in another space? The thing that "connects" vectors in one space and moves them to another is called a *connection*.[14]

10.1 Recap of vector bundles

Let M be a manifold. The rough definition of a vector bundle is that you assign to each point a vector space, and tie them all together somehow.

Definition 10.1. A **vector bundle** is a smooth manifold E along with a projection map $V \to E \xrightarrow{\pi} M$ (where V is a fixed vector space) such that

(1) $\pi^{-1}(p)$ is a vector space,

(2) There exists a neighborhood U of p such that $\pi^{-1}(U) \simeq U \times V$,

(3) Transition maps are linear transformations.

Essentially (1) says that each fiber has a vector space structure, (2) tells us we can locally assign coordinates to preimages of neighborhoods, and (3) says that the transition maps respect such bases and the linear structure of the upstairs vector space.

[14] You can do this using flow, and the result is a *Lie derivative*. But that isn't the current direction we're going.

Example 10.1. An example of a vector bundle is the **tangent bundle** TM, since the T_pM is a vector space at each point, and locally we have a basis $\{\partial x^i\}$, and going from one tangent space to another we have a change of basis matrix. Another example is the **cotangent bundle** T^*M, where we consider all tangent covectors rather than vectors. On this vein we can look at the space of all (k,ℓ)-tensors $T^{(k,\ell)}M$. So $T^{(0,1)}M = TM$ and $T^{(1,0)}M = T^*M$.

If $M \subseteq \mathbb{R}^n$, we can look at the **normal bundle**, the space of all vectors perpendicular to M. We could also look at the **trivial bundle** $M \times V$, where we have a global structure, so we can use the same basis for V everywhere, leading to one chart for the bundle structure.

Definition 10.2. A (smooth) **section** of a vector bundle E is a (smooth) map $s\colon M \to E$ such that $\pi \circ s = \mathrm{id}$. In other words, for $p \in m$, we have $s(p) \in \pi^{-1}(p)$.

So a section smoothly assigns to every point of the manifold an element of the fiber living above it.

10.2 Connections

Let $v \in T_pM$, and s be a section of some vector bundle $E \to M$. We want to take $\nabla_v s$, or the derivative in the v-direction of s. Suppose $b_1(x), \cdots, b_m(x)$ is a basis for the fiber of p, and $e_1, \cdots, e_n$ is a basis for T_pX. A section $s(x)$ can be written as $s^\alpha(x)b_\alpha(x)$ since it lives in the fiber of p, a vector space. Here are some things we want out of derivatives:

(i) $\nabla_{e_i} s = \nabla_{e_i}(s^\alpha b_\alpha) = \nabla e_i(s^\alpha)b_\alpha + s^\alpha(\nabla e_i b_\alpha)$,

(ii) $\nabla_{e_i} b_\alpha = (A_i)^\beta b_\beta$,

(iii) $\left(\nabla_{e_i} s\right)^\beta = \partial_{e_i} s^\beta + (A_i)^\beta_\alpha s^\alpha$.

Normally we see equations like "$\nabla_i = \partial_i + A_i$". The ∇_i is a **covariant derivative**, which is the ordinary derivative of the coefficients plus an extra matrix. We need a matrix for every direction we take a derivative in, and these matrices A_i are by choice (not given). If someone tells you what the derivative of the basis vectors are, you can figure out the derivative of the section by linearity. The ∇_i is also called a **connection**. This idea works in any vector bundle. But we *really* care about tangent bundles.

In the case of a tangent bundle TM with basis $e_1, \cdots, e_n$ for the tangent space and $e_1, \cdots, e_n$ for the fiber, the derivative in the e_i direction of e_j is some linear combination of the e_k's. We denote this

$$\nabla_{e_i} e_j = \Gamma^k_{ij} e_k, \quad \text{or} \quad \Gamma^k_{ij} = \left(\nabla_{e_i} e_j\right)^k$$

where the second expression means that you take the kth coefficient of the derivative $\nabla_{e_i} e_j$. If somebody tells you what the Γ's are, you know how to take covariant derivatives. More generally you might have vector fields $v = v^i e_i$, $w = w^j e_j$, and the natural question is "what is the covariant derivative in the vth direction of w?" Then the kth component is given by

$$\boxed{(\nabla_v(w))^k = v^i(\partial_i w^k) + \Gamma^k_{ij} v^i w^j.}$$

In the expression $\nabla_i = \partial_i + A_i$, to be more precise, the A_i refers to a matrix $(A_i)^\beta_\alpha$, where the i refers to the direction you're taking the derivative with respect to, and α, β are coordinates in the fiber. Likewise, for $\Gamma^k_{ij} = (\Gamma_i)^k_j$, i refers to the direction of the derivative and j, k tells us this is a linear transformation within the vector space.

Our first approach was based on faith, that we somehow know that $\nabla_{e_i} e_j = \Gamma^k_{ij} e_k$, and from here we know the derivative of anything. An alternate approach is more axiomatic (this is the approach the book uses). We want to have this operation where we take the derivative of any vector field in the direction of any other vector field, say $\nabla_V(W)$. Assume X, Y, Z are vector fields. Then a **connection** $\nabla\colon X \times Y \to Z$ satisfies

(1) $\nabla_X(c_1 Y + c_2 Z) = c_1(\nabla_X Y) + c_2(\nabla_x Z)$,

(2) $\nabla_{fX+gY} = f\nabla_X Z + g\nabla_Y Z$,

(3) $\nabla_X(fZ) = f\nabla_X Z + X(f)Z$.

If we have a connection and a local basis for our tangent space $\{e_i\}$, then we can ask what $\left(\nabla_{e_i} e_j\right)^k = \Gamma^k_{ij}(x)$ is.

10.3 Connection under a change of coordinates

The next question is "what happens under a change of coordinates?" Say we have coordinates $(x^1, \cdots, x^n)$ on M, giving us a basis $e_i = \partial/\partial x^i$ for the tangent space. Note that section and vector field are synonymous, since a vector field is a section of the tangent bundle. Then $(\nabla_i s)^k = \partial_i s^k + \Gamma^k_{ij} s^j$. We call $(\nabla_i s)^k$ the covariant derivative and $\partial_i s^k$ the ordinary derivative.

Now let's switch to coordinates $(y^1, \cdots, y^n)$ with an alternate basis $\widetilde{e}_i$. This gives us a change of basis matrix $\widetilde{e}_i = \partial/\partial y^j = A^j_i e_j$, and the dual basis $\widetilde{\phi}^i = B^i_j \phi^j$. We also know $A^j_i = \partial x^j/\partial y^i$, $B^i_j = \partial y^i/\partial x^j$. Then $\nabla_{\widetilde{e}_i} \widetilde{e}_j = \widetilde{\Gamma}^k_{ij} \widetilde{e}_k$. The question is: how do the $\widetilde{\Gamma}^k_{ij}$ compare to the old Γ^k_{ij}? In our words, we want to compute $\widetilde{\Gamma}^k_{ij}$ in terms of A'_s, B'_s, Γ'_s. Note that

$$\begin{aligned}
\nabla_{\widetilde{e}_i} \widetilde{e}_j &= \nabla_{A^\ell_i e_\ell} A^m_j e_m = A^\ell_i \nabla_{e_\ell} (A^m_j e_m) \\
&= A^\ell_i A^m_j \nabla_{e_\ell} e_m + A^\ell_i (\partial_\ell A^m_j) e_m \\
&= A^\ell_i A^m_j \Gamma^p_{\ell m} e_p + A^\ell_i (\partial_\ell A^m_j) B^k_m \widetilde{e}_k \\
&= (A^\ell_i A^m_j B^k_p \Gamma^p_{\ell m} + A^\ell_i \partial_\ell A^m_j B^k_m) \widetilde{e}_k.
\end{aligned}$$

So

$$\boxed{\widetilde{\Gamma}^k_{ij} = \underbrace{A^\ell_i A^m_j B^k_p \Gamma^p_{\ell m}}_{\text{Lie tensor}} + \underbrace{A^\ell_i B^k_m (\partial_\ell A^m_j)}_{\text{Exterior independent of } \Gamma!}.}$$

This tells us a little bit about the Γ's: they aren't tensors, they transform like a $(2,1)$-tensor plus an inhomogeneous bit. It is not true that connections are tensors, but if you *add* a tensor to a connection, it transforms like a tensor, and the extra bit is invariant since we already know the connection. So

(1) If ∇ is a connection and T is a $(2,1)$-tensor, then $(\nabla + T)$ is a connection.

(2) If ∇ and ∇' are connections, then $\nabla - \nabla'$ is a *tensor*.

(3) If a connection ∇_0 exists, then the set of *all* connections can be expressed in the form $\nabla_0 + \mathscr{T}^{2,1} M$.

The space of connections isn't an vector space, but it is an affine space. Does such a ∇_0 exist? Let us construct a connection on S^2. Say we have two coordinate charts (u_1, v_1) and (u_2, v_2) constructed by stereographic projection. Construct a partition of unity on both charts: say $\rho_1 := 1$ on N, 1 all the way down to the Tropic of Cancer, hit zero on the Tropic of Capricorn, and is zero at S. Define ρ_2 in a similar way for the second chart (u_2, v_2). These are compactly supported, where $\rho_1 + \rho_2 = 1$.

Define ∇_0 with $\Gamma^k_{ij} = 0$ (not defined on the south pole), and likewise $\widetilde{\nabla}_0$ with $\widetilde{\Gamma}^k_{ij} = 0$ (not defined on the north pole) corresponding to our charts. Then the sum $\nabla = \rho_1 \nabla_0 + \rho_2 \widetilde{\nabla}_0$ is defined everywhere, since the left term extends to 0 near the south pole (since $\rho_1 = 0$ there) and likewise $\rho_2 \widetilde{\nabla}_0$ extends to the north pole. Is this a connection? Yes, because you can think of this sum as $\nabla = \nabla_0 + \rho_2(\widetilde{\nabla}_0 - \nabla_0)$, a connection plus a tensor. This argument extends to arbitrary manifolds. So every manifold admits a connection, and the set of connections looks like an affine space based on the set of $(2,1)$-tensors.

Lecture 11

March 2, 2021

Recap: Let $\mathfrak{X}(M)$ be the set of vector fields on M, or the sections of TM. Recall that $\nabla \colon \mathfrak{X} \times \mathfrak{X} \to \mathfrak{X}$, $(X, Y) \to \nabla_X Y$. We want $\nabla_{fX} Y = f \nabla_X Y$, linearity WRT X, Y, and $\nabla_X(fY) = f \nabla_X Y + (X(f))Y$. Note that $X(f) = X \partial_i f$. If we had something that satisfied every condition besides the extra term, that would just be a tensor. This on the other hand, is a differential operator. If we have a basis $\{e_i\}$, then $\nabla_{e_i} e_j = \Gamma^k_{ij} e_k$. Usually we have $e_i = \partial/\partial x^i$. Furthermore, $(\nabla_x y)^k = x(y^k) + \Gamma^k_{ij} x^i y^j$, or "$\nabla_i = \partial_i + \Gamma_i$".

If $\nabla, \widetilde{\nabla}$ are connections, we also saw that $\widetilde{\nabla} - \nabla$ is a *tensor*. So $(\nabla_X - \nabla_X)(fY) = f(\widetilde{\nabla}_X - \nabla_X)(y)$, a tensor (the extra components cancel). Furthermore the set of connections is characterized by elements of the for $\nabla_0 + (1,2)$-tensors, if ∇_0 exists (it does).

11.1 Connections of covectors and general tensors

How do we take derivatives of covector fields? Or $(1,1)$-tensors, $(5,7)$-tensors, whatever. Suppose α is a covector field, and v is a vector field. Then $\alpha(v)$ is just a function, and we would like for

$$\partial_i(\alpha(v)) = (\nabla_i \alpha)(v) + \alpha(\nabla_i v)$$

to be true. We know what $\nabla_i v$ means for v a vector field. We *define* $(\nabla_i \alpha)(v) = \partial_i(\alpha(v)) - \alpha(\nabla_i v)$, so our product rule works. Consider

$$\begin{aligned}
(\nabla_i \phi^k)(e_j) &= \partial_i(\phi^k(e_j)) - \phi^k(\nabla_i e_j) \\
&= \partial_i \delta_j^k - \gamma_{ij}^k \\
&= -\Gamma_{ij}^k.
\end{aligned}$$

So $\nabla_i e_j = \Gamma_{ij}^k e_k$, and $\nabla_i \phi^k = -\Gamma_{ij}^k \phi^j$. The roles are reversed: in the first you take the derivative of j and get k (with some coefficients), and in the second expression you take the derivative of k and get j. This has to be the case because of where our indices are placed. This is the same idea of taking minus the transpose of the other, where the derivative on rows is minute the transpose of the derivative of columns. Then

$$\begin{aligned}
(\nabla_i \alpha) &= \nabla_i(\alpha_k \phi^k) \\
&= (\partial_i \alpha_k)\phi^k + \alpha_k \nabla_i \phi^k \\
&= (\partial_i \alpha_k)\phi^k - \alpha_k \Gamma_{ij}^k \phi^j,
\end{aligned}$$

$$\boxed{(\nabla_i \alpha)_j = \partial_i \alpha_j - \Gamma_{ij}^k \alpha_k.}$$

So that's how we take the covariant derivative of a covector field. What about a $(1,1)$-tensor? For $T(v,\alpha)$ eating a vector and covector, ideally we would have

$$\partial_i(T(v,\alpha)) = (\nabla_i T)(v,\alpha) + T(\nabla_i v, \alpha) + T(v, \nabla_i \alpha).$$

We know what $T(\nabla_i v, \alpha)$ and $T(v, \nabla_i \alpha)$ are, so consider this as a definition for $(\nabla_i T)(v,\alpha)$. Similarly, for a (k,ℓ)-tensor we just have k covariant terms and ℓ contravariant terms instead of one.

11.2 Metric connections

Now that we know how to take the derivative of tensors, let us compute the derivative of the *metric* $\nabla_i(g)(e_j, e_k)$. Then

$$\begin{aligned}
\nabla_i(g)(e_j, e_k) &= \partial_i g_{jk} - g(\nabla_i e_j, e_k) - g(e_j, \nabla_i e_k) \\
&= \partial_i g_{jk} - g(\Gamma_{ij}^\ell e_\ell, e_k) - g(e_j, \Gamma_{ik}^m e_m), \\
(\nabla_i g)_{jk} &= \partial_i g_{jk} - \Gamma_{ij}^\ell g_{\ell k} - \Gamma_{ik}^m g_{jm} \\
&= \partial_i g_{jk} - \Gamma_{ijk} - \Gamma_{ikj}.
\end{aligned}$$

What we did was lower an index k that used to be at the top, and wrote it as Γ_{ijk}, where the *third* index is the lowered one. Ideally, the derivative of the metric should be zero. If $\nabla_i g = 0$, we say ∇ is a **metric connection**. For metric connections, we know that $\Gamma_{ijk} + \Gamma_{ikj} = \partial_i g_{jk}$. However, not every connection is metric.

Example 11.1. Consider the Euclidan connection ∇^E on $\mathbb{R}^n$. Given the standard basis, we have $\nabla_i^E e_j = 0$, or $\Gamma_{ij}^k = 0$. This depends on basis; we work in $\mathbb{R}^2$ with basis (r,θ), where $\widetilde{e}_1 = \partial/\partial r$, $\widetilde{e}_2 = \partial/\partial\theta$. Given $\partial/\partial r = x(\partial/\partial x) + y(\partial/\partial y)$, $\partial/\partial\theta = x(\partial/\partial y) - y(\partial/\partial x)$, what are the values of $\nabla_{\frac{\partial}{\partial\theta}} \frac{\partial}{\partial\theta}, \nabla_{\frac{\partial}{\partial\theta}} \frac{\partial}{\partial r}, \nabla_{\frac{\partial}{\partial r}} \frac{\partial}{\partial\theta}, \nabla_{\frac{\partial}{\partial r}} \frac{\partial}{\partial r}$?

First, recall that

$$\widetilde{e}_1 = \frac{x\partial_x + y\partial_y}{\sqrt{x^2+y^2}} = \frac{x}{\sqrt{x^2+y^2}}\partial_x + \frac{y}{\sqrt{x^2+y^2}}\partial_y, \quad \widetilde{e}_2 = -y\partial_x + x\partial_y,$$

so

$$\begin{aligned}
\nabla_{\frac{\partial}{\partial x}}\widetilde{e}_1 &= \frac{y^2\partial_x - xy\partial_y}{(x^2+y^2)^{3/2}} = -\frac{y}{(x^2+y^2)^{3/2}}\widetilde{e}_2, & \nabla_{\frac{\partial}{\partial y}}\widetilde{e}_1 &= \frac{-xy\partial_x + x^2\partial_y}{(x^2+y^2)^{3/2}} = \frac{x}{(x^2+y^2)^{3/2}}\widetilde{e}_2\\
\nabla_{\frac{\partial}{\partial x}}\widetilde{e}_2 &= \partial_y, & \nabla_{\frac{\partial}{\partial y}}\widetilde{e}_2 &= -\partial_x.
\end{aligned}$$

Then

$$\begin{aligned}
\nabla_{\widetilde{e}_1}\widetilde{e}_1 &= \frac{x}{r}\left(-\frac{y}{r^3}\right)\widetilde{e}_2 + \frac{y}{r}\left(\frac{x}{r^3}\right)\widetilde{e}_2 = 0, & \nabla_{\widetilde{e}_2}\widetilde{e}_1 &= -y\left(\frac{-y}{r^3}\right)\widetilde{e}_2 + x\left(\frac{x}{r^3}\right)\widetilde{e}_2 = \frac{\widetilde{e}_2}{r},\\
\nabla_{\widetilde{e}_1}\widetilde{e}_2 &= \frac{x}{r}\partial_y + \frac{y}{r}(-\partial_x) = \frac{\widetilde{e}_2}{r}, & \nabla_{\widetilde{e}_2}\widetilde{e}_2 &= -y\partial_y - x\partial_x = -r\widetilde{e}_1.
\end{aligned}$$

11.3 The tangential connection

Suppose we have a Riemannian manifold (M,g) sitting inside a bigger Riemannian manifold N. (The motivating example is $N=\mathbb{R}^n$.) Suppose $\overline{\nabla}_j\overline{g}$ is a connection (and $\overline{g}$ a metric) for N. Given a vector in T_pN, we have the orthogonal projection $\pi_p\colon T_pN \to T_pN$, taking normal vectors to zero and tangent vectors to themselves. Let us try to define a connection on M. Suppose X,Y are vector fields on M, define

$$\nabla_X^T Y = \pi\left(\overline{\nabla}_X Y\right).$$

Even though X,Y are only defined on M, we can always extend it to a neighborhood of p in some way. Showing that the choice of extension is not a factor is on the homework.

Example 11.2. Let $M=S^1$, $N=\mathbb{R}^2$. A tangent vector is of the form ∂_θ, so $\nabla_{\partial_\theta}\partial_\theta = \pi\left(\overline{\nabla}_{\partial_\theta}\partial_\theta\right) = \pi(-r\partial_r) = 0$.

Classic differential geometry was to figure out how surfaces sit in $\mathbb{R}^3$. The question boiled down to "what is this projection?", which amounts to understanding the normal vector is, because all π does it vanquish the normal component. So we have things like the Gauss map that gives a normal vector at ecah point. Somebody (probably Gauss/Euler) said we should work with intrinsic properties, doing things like figuring out areas and angle sums of triangles and circumferences of circles directly on the surface itself, without regards to **N**. It turns out we get the exact same results! So the way something sits in $\mathbb{R}^3$ actually turns out to be intrinsic.

Lecture 12

March 4, 2021

12.1 The tangential connection in S^2

Consider $S^2 \subseteq \mathbb{R}^3$, with the angle θ sweeping across the equator and φ measuring longitudinal distance from the north pole. So $\Sigma(\varphi,\theta) = (\sin\varphi\cos\theta, \sin\varphi\sin\theta, \cos\varphi)$. In terms of coordinates, say $x^1=\varphi$ and $x^2=\theta$. We know $\partial_\varphi\Sigma = (\cos\varphi\cos\theta, \cos\varphi\sin\theta, -\sin\varphi) = e_1$, $\partial_\theta\Sigma = (-\sin\varphi\sin\theta, \sin\varphi\cos\theta, 0) = e_2$. So $g_{11}=1, g_{12}=0, g_{22}=\sin^2\varphi$. We want to compute $\nabla_{e_1}e_1, \nabla_{e_1}e_2, \nabla_{e_2}e_1, \nabla_{e_2}e_2$. For tangential connections, we use the connection on the big space $\mathbb{R}^3$, which is trivial. We have

$$\begin{aligned}
\overline{\nabla}_{e_1}e_1 &= (-\sin\varphi\cos\theta, -\sin\varphi\cos\theta, -\cos\varphi) = -\Sigma, & \overline{\nabla}_{e_1}e_2 &= (-\cos\varphi\sin\theta, \cos\varphi\cos\theta, 0) = \cot\varphi e_2,\\
\overline{\nabla}_{e_2}e_1 &= (-\cos\varphi\sin\theta, \cos\varphi\cos\theta, 0) = \cot\varphi e_2, & \overline{\nabla}_{e_2}e_2 &= (-\sin\varphi\cos\theta, -\sin\varphi\sin\theta, 0) = -\sin\varphi\cos\varphi e_1.
\end{aligned}$$

To calculate $\overline{\nabla}_{e_1} e_2$, note that $\overline{\nabla}_{e_1} e_2 \perp e_1$. So the projection is some multiple of e_2: to figure out what the multiple is, we want to calculate

$$\pi_{e_2}(\overline{\nabla}_{e_1} e_2) = \frac{(\overline{\nabla}_{e_1} e_2)\cdot e_2}{e_2 \cdot e_2} e_2 = \frac{\sin\varphi\cos\varphi}{\sin^2\varphi} e_2 = \frac{\cos\varphi}{\sin\varphi} e_2 = \cot\varphi e_2.$$

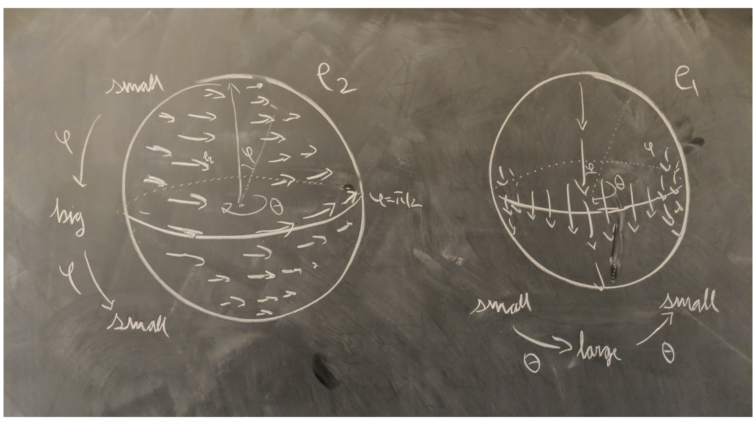
Why are we just taking derivatives with respect to φ and θ to calculate the ∇'s? In $\mathbb{R}^3$, $(\overline{\nabla}_v w)^j = v(w^j) = v^i \partial_i w^j$. Since $e_1 = \partial_\varphi, e_2 = \partial_\theta$, we have $\overline{\nabla}_{e_1} = \partial_\varphi$, $\overline{\nabla}_{e_2}\partial_\theta$. After doing these calculations, we just project down. Given an orthonormal frame $\{e_i\}$, and a vector v, then

$$\pi_{\text{plane}} v = \frac{v\cdot e_1}{e_1\cdot e_1} e_1 + \frac{v\cdot e_2}{e_2\cdot e_2} e_2 = \frac{v\cdot e_i}{e_i\cdot e_i} e_i \quad \text{(implicit sum)}.$$

The final component $\overline{\nabla}_{e_2} e_2$ is the tricky one. It looks sort of like the position, but it has no third component. It's not entirely pointing in the normal direction, it has a tangential component. Since $\overline{\nabla}_{e_2} e_2 \perp e_2$, and $\overline{\nabla}_{e_2} e_2 \cdot e_1 = -\sin\varphi\cos\varphi,\ e_1\cdot e_1 = 1$, we conclude $\overline{\nabla}_{e_2} e_2 = -\sin\varphi\cos\varphi e_1$. Finally, we conclude that

$$\begin{aligned} \nabla_{e_1} e_1 &= 0, & \nabla_{e_1} e_2 &= \cot\varphi e_2, \\ \nabla_{e_2} e_1 &= \cot\varphi e_2, & \nabla_{e_2} e_2 &= -\sin\varphi\cos\varphi e_1. \end{aligned}$$

Now that we know this, we can take the derivative of *any* vector field on S^2 with respect to *any* other vector field on S^2, since we've worked out what all the Christoffel symbols are. How do we geometrically interpret our results?



The vector field e_2 is small at the poles and grows larger as it goes south from the north pole (moving in the φ direction), then starts to decrease as it approaches the south pole. A similar thing happens for e_1, which is large at the great circle connecting the poles, but gets smaller as you go farther out east or west (in the θ direction). The precise coefficient $\cot\varphi$ is a calculation, but this is the general idea.

Now that we've developed the theory of connections, let's return to the geodesic.

Definition 12.1. A **geodesic** with respect to a covariant derivative ∇ is a curve satisfying $\nabla_{\dot{x}}\dot{x} = 0$. In terms of coordinates, we have $\ddot{x}^k + \Gamma^k_{ij}\dot{x}^i\dot{x}^j = 0$.

Is this equivalent to the other definitions of a geodesic? It turns out there is a certain connection that satisfies this, the *Levi-Civita connection*.

12.2 The fundamental theorem of Riemannian geometry

Let's recall some properties that connections can have.

- A connection is **metric** if $\nabla g = 0$, which is equivalent to $X(g(Y,Z)) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z)$. In particular, for $X = e_i, Y = e_j, Z = e_k$, we have $\partial_i g_{jk} = g(\nabla_i e_j, e_k) + g(e_j, \nabla_i e_k) = g(\Gamma^\ell_{ij} e_\ell, e_k) + g(e_j, \Gamma^m_{ik} e_m) = \Gamma^\ell_{ij} g_{\ell k} + \Gamma^m_{ik} g_{jm}$, then set $\Gamma^\ell_{ij} g_{\ell k} = \Gamma_{ijk}$ and $\Gamma^m_{ik} g_{jm} = \Gamma_{ikj}$. In short, $\boxed{\partial_i g_{jk} = \Gamma_{ijk} + \Gamma_{ikj}.}$

Note that our tangential connection $\nabla = \pi(\overline{\nabla})$ on S^2 is metric. The question is, does $X(g(Y,Z)) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z)$? We have $g(\nabla_X Y, Z) + g(Y, \nabla_X Z) = g(\pi(\overline{\nabla}_X Y), Z) + g(Y, \pi(\overline{\nabla}_X Z))$: the difference between ∇ and $\overline{\nabla}$ of something results in something in the perpendicular direction. Then taking the inner product of something in the perpendicular direction with the tangent direction gives zero, so $g(\pi(\overline{\nabla}_X Y), Z) + g(Y, \pi(\overline{\nabla}_X Z)) = g(\overline{\nabla}_X Y, Z) + g(Y, \overline{\nabla}_X Z)$. Then since $\overline{\nabla}$ is metric, ∇ must be metric as well.

- A connection is **symmetric** (or **torsion-free**) if $[X,Y] = \nabla_X Y - \nabla_Y X$.

In $\mathbb{R}^n$ this is true, since we compute the bracket of vector fields is by taking the ordinary derivatives of the coefficients, which is the same as taking covariant derivatives of the coefficients. What does this mean in terms of Christoffel symbols? If $X = \frac{\partial}{\partial x^i}, Y = \frac{\partial}{\partial x^j}$, then $[X,Y] = 0$. So $0 = \nabla_i e_j - \nabla_j e_i = \Gamma_{ij}^k e_k - \Gamma_{jk}^k e_k$, so $\Gamma_{ij}^k = \Gamma_{ji}^k$. In Euclidian space, both coefficients are zero, so the standard connection is symmetric.

Theorem 12.1. *If $\overline{\nabla}$ is metric and symmetric, so is the tangential connection ∇^T.*

Proof. We have already shown the metric part of this result. Consider $[X,Y] = \overline{\nabla}_X Y - \overline{\nabla}_Y X$ by assumption. Then $\pi([X,Y]) = \pi(\overline{\nabla}_X Y) - \pi(\overline{\nabla}_Y X)$. Since $[X,Y]$ is the bracket of two tangent vectors fields and therefore a tangent vector field, applying the *tangential* projection does nothing. Furthermore, $\pi(\overline{\nabla}) = \nabla$, so we have $[X,Y] = \nabla_X Y - \nabla_Y X$. ⊠

Fundamental Theorem of Riemannian Geometry. *For any Riemannian manifold (M,g), there exists a unique connection that is metric and symmetric.*

Corollary 12.1. *The tangential connection ∇^T is completely determined by the metric g.*

The result of Corollary 12.1 is strange, since ∇^T is an *extrinsic* property: it's all about how something is sitting in space, and is *not* preserved by isometries. However, it's determined by an *intrinsic* property, the metric. When Gauss came up with this theorem in the context of surfaces, he called it the **Theorema Egregium**. Now let's prove our big theorem.

Proof of the Fundamental Theorem of Riemannian Geometry. Recall that a metric connection implies that $\partial_i g_{jk} = \Gamma_{ijk} + \Gamma_{ikj}$, and a symmetric connection satisfies $\Gamma_{ij}^k = \Gamma_{ji}^k$, which implies $\Gamma_{ijk} = \Gamma_{jik}$. Then consider

$$\begin{aligned} \partial_i g_{jk} &= \Gamma_{ijk} + \Gamma_{ikj} \\ +(\partial_j g_{ik} &= \Gamma_{jik} + \Gamma_{jki}) \\ -(\partial_k g_{ij} &= \Gamma_{kij} + \Gamma_{kji}), \end{aligned}$$

and since $\Gamma_{kij} = \Gamma_{ikj}, \Gamma_{kji} = \Gamma_{jki}$, four out of six terms cancel and we are left with

$$\begin{gathered} \partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij} = 2\Gamma_{ijk} \implies \\ \Gamma_{ijk} = \frac{1}{2}\left(\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}\right), \\ \boxed{\Gamma_{ij}^k = \frac{1}{2} g^{km}(\partial_i g_{jm} + \partial_j g_{im} - \partial_m g_{ij}).} \end{gathered}$$

Here, g^{km} denote the coefficients of the dual metric tensor, or inverse matrix. This connection is called the **Levi-Civita** connection. We have Γ_{ij}^k symmetric since interchanging the roles of $\partial_i g_{jk}$ and $\partial_j g_{ik}$ does nothing (addition is commutative). This is also metric since

$$\Gamma_{ijk} + \Gamma_{ikj} = \frac{1}{2}\left(\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}\right) + \frac{1}{2}\left(\partial_i g_{jk} + \partial_k g_{ij} - \partial_j g_{ik}\right) = \partial_i g_{jk}.$$ ⊠

12.3 Computing Christoffel symbols of the Levi-Civita connection

Example 12.1. Consider $\mathbb{R}^2$ with polar coordinates $e_1 = \partial_r, e_2 = \partial_\theta$. Our metric is $g_{11} = 1, g_{12} = 0, g_{22} = r^2$. There are eight Γ_{ij}^k's, but only a couple are interesting. We know $\partial_1 g_{22} = 2r$, while all other derivatives are zero. First, let's compute the Γ_{ijk}'s. All are zero but $\Gamma_{122} = \frac{1}{2}(\partial_1 g_{22} + 0 - 0) = r, \Gamma_{212} = \frac{1}{2}(0 + \partial_1 g_{22} - 0) = r$, and $\Gamma_{221} = -r$. So

$$\begin{aligned}
\Gamma_{11}^1 &= 0,\\
\Gamma_{11}^2 &= 0,\\
\Gamma_{12}^1 &= 0,\\
\Gamma_{12}^2 &= \frac{1}{2}g^{2m}(\partial_1 g_{2m} + 0 - 0) = \frac{1}{2}g^{22}(\partial_1 g_{22}) = \frac{1}{2r^2}2r = \frac{1}{r},\\
\Gamma_{21}^1 &= 0,\\
\Gamma_{21}^2 &= \Gamma_{12}^2 = \frac{1}{r},\\
\Gamma_{22}^1 &= \frac{1}{2}g^{1m}(0 + 0 - \partial_m g_{22}) = -r,\\
\Gamma_{22}^2 &= 0.
\end{aligned}$$

Example 12.2. Let us return to S^2. We have $g_{11} = 1, g_{12} = 0, g_{22} = \sin^2\varphi$. The only relevant derivative is $\partial_1 g_{22} = 2\sin\varphi\cos\varphi$, all other derivatives are zero. Then $\Gamma_{122} = \Gamma_{212} = \frac{1}{2}(\partial_1 g_2 2) = \sin\varphi\cos\varphi$. Similarly, $\Gamma_{221} = -\sin\varphi\cos\varphi$. When you raise the indices, you get $\Gamma_{12}^2 = \Gamma_{21}^2 = \frac{\sin\varphi\cos\varphi}{\sin^2\varphi} = \cot\varphi$, and $\Gamma_{22}^1 = -\sin\varphi\cos\varphi$. This is exactly what we got from our extrinsic calulation at the beginning of lecture!

So we don't need to know how exactly it's sitting in $\mathbb{R}^3$, we only need to know how the metric works. There's a particular space that doesn't sit particularly well in $\mathbb{R}^3$ but for which we understand the metric quite well... the hyperbolic plane! It's not the tangential connection, but the unique symmetric connection instead. The metric on $\mathbb{H}$ is $g_{11} = g_{22} = \frac{1}{y^2}$, and $g_{12} = 0$. Once we know what the Christoffel symbols are, we know what a geodesic is!

Lecture 13

March 9, 2021

13.1 Geodesics as paths with no acceleration

Recall our three notions of a geodesic.

(1) A geometric path that "locally" minimizes distance,

(2) A parametrized path $\gamma(t)$ that has stationary energy (where $E(\gamma) = \int_0^T g(\dot\gamma, \dot\gamma)\,dt$),

(3) A path with no acceleration; $\nabla_{\dot\gamma}\dot\gamma = 0$.

We have already related (1) and (2), and now we want to show (2) and (3) are equivalent, with the Levi-Civita connection. In $\mathbb{R}^n$, $g_{ij} = \delta_j^i$, so $\partial_i g_{jk} = 0$. Consider $\gamma^{\text{new}}(t) := \gamma(t) + \delta\gamma(t)$, then we are interested in making $E(\gamma^{\text{new}} - E(\gamma) = \int(\ )\,\delta\gamma + O(\delta\gamma^2)$. This is the same process as taking derivatives: for f a multivariable function, we have $f(x + dx) = f(x) + L(dx) + O(dx^2)$, and $L(dx)$ is called the derivative df. Then taking $\lim_{dx\to 0}(f(x + dx) - f(x) - L(dx))/|dx|$ is precisely what a derivative is. If $f(x + dx)$ is stationary, then $df = \frac{\partial f}{\partial x^i}dx^i = 0$, so we set $\delta E = 0$.

$$E(\gamma) = \int_0^T g_{ij}\dot{\gamma}^i\dot{\gamma}^j\,dt, \tag{1}$$

$$\delta E = \int_0^T 2g_{ij}\dot{\gamma}^i(\delta\dot{\gamma}^j)\,dt \tag{2}$$

$$= \underbrace{2g_{ij}\dot{\gamma}^i\delta\gamma^j\Big|_0^T}_{=0} - \int 2g_{ij}\ddot{\gamma}^i\delta\gamma^j\,dt \tag{3}$$

From (1) to (2), for a general manifold the g_{ij} will change, since g_{ij} becomes a function of time. But here it stays the same. We get to (3) by integration by parts, and the first component vanishes because $\gamma(0) = p,\ \gamma(T) = q$, so $\delta\gamma(0) = 0,\ \delta\gamma(T) = 0$. In general, we have

$$0 = -2\int g_{ij}\ddot{\gamma}^i(\delta\gamma^j)\,dt.$$

The only way this integral is zero for all possible values of $\delta\gamma^j$ is if $g_{ij}\ddot{\gamma}^i$ is zero. Then $g_{ij}\ddot{\gamma}^i = 0$, so a geodesic in $\mathbb{R}^n$ is something with zero acceleration. For a general manifold, since g_{ij} is no longer constant we get a new metric and other terms, which look like

$$g_{k\ell}\ddot{\gamma}^\ell + (\text{derivatives of } g)\dot{\gamma}^i\dot{\gamma}^j = 0.$$

Working out these terms is homework, they turn out to be precisely Γ_{ij}^k. For an arbitrary Riemannian manifold, without loss of generality we can assume that we're working in a single chart, which locally is just $\mathbb{R}^n$ with a funny metric.

13.2 Parallel transport

Suppose V is a vector field and γ is a curve. What is $\nabla_{\dot{\gamma}}V$? (Here ∇ is an arbitrary connection.) Since $V = v^i(x)e_i$, then

$$\begin{aligned}\nabla_{\dot{\gamma}}V = \nabla_{\dot{\gamma}}(v^ie_i) &= \dot{\gamma}(v^i)e_i + v^i\nabla_{\dot{\gamma}}e_i\\ &= \dot{\gamma}^j(\partial_jv^i)e_i + v^i\dot{\gamma}^j\nabla_{e_j}e_i\\ &= \dot{\gamma}^j(\partial_jv^i)e_i + \Gamma_{ij}^kv^i\dot{\gamma}^je_k\\ &= \dot{v}^ie_i + \Gamma_{ij}^kv^i\dot{\gamma}^je_k.\end{aligned}$$

Definition 13.1. We say V is **parallel** along the path γ if and only if $\nabla_{\dot{\gamma}}V = 0$.

todo:figure?

Given $V(\gamma(t))$, does there exist a parallel V? We want $0 = \dot{v}^ke_k + \Gamma_{ij}^kv^i\dot{\gamma}^je_k$. In other words, we want $\dot{v}^k = -\Gamma_{ij}^kv^i\dot{\gamma}^j$. This is just a linear first order differential equation, since we want to solve for the derivative of v in terms of v! As long as the data is smooth (Γ is a smooth function of position, γ^j a smooth function of time), then by the Existence-Uniquess theorem for differential equations we have a unique solution. Since the ODE is linear, the solution gives a linear map from T_pM to $T_{\gamma(t)}M$, which we call **parallel transport**. Without a connection the tangent spaces are just vector spaces and we wouldn't know what to do, but given a connection, we can "drag" one vector to another along the curve, in particular we can send frames to frames this way.

todo:figure about transporting frames?

A natural question is "does parallel transport depend on γ?" In $\mathbb{R}^n$, it doesn't matter what path you transport by, you always get the same result, since $\dot{v}^k = 0$ which implies v^k is a constant. However on another manifold, it might depend on which path you take.

Example 13.1. In S^2, transporting a vector along two different paths to the north pole gives a different result depending on what path you take, because there's a singularity at the north pole.

todo:figure

Curvature means that things twist when we go along loops. Whenever the answer depends on path, then going along a closed loop does not give the identity. However, we do get a linear transformation $T_pM \to T_pM$. We claim this linear transformation is an *isometry*. Note that since our connection is metric,

$$\frac{d}{dt}g(v,v) = \dot{\gamma}(g(v,v)) = g(\nabla_g v, v) + g(v, \nabla_g v) = g(0,v) + g(v,0) = 0.$$

Similarly, under parallel transport the derivative of two vectors doesn't change. Going around a loop gives you a rotation of your tangent space. On a two dimensional surface, all rotations commute, since the rotation group is abelian. For a closed loop, you can "chop up" the interior of the closed curve and see how much rotation you get from each bit, then sum. So the rotation from parallel transport by γ a closed loop is $\int \kappa\, dA$ for some function K. This function κ is called the **Gauss curvature**. On S^n it's zero, in $\mathbb{H}$ it's negative 1, and on $\mathbb{R}^n$ its zero. So if you want to know the curvature at a point, just go along an infinitesmal path, and we get that κ is a function of $g, \Gamma, \partial\Gamma$. (Sneak peek of future content.)

In short, if you what a connection is, you know what parallel transport is. And if you know what parallel transport is, you know what curvature is. If the connection happens to be the Levi-Civita connection, then Γ and $\partial\Gamma$ come from the metric, so we get formulas for curvature in terms of the metric.

Example 13.2. Let's think about $S^3 \subseteq \mathbb{R}^4$, the space of quaternions. That is, $x = x^1 + x^2 i + x^3 j + x^4 k$, where $i^2 = j^2 = k^2 = ijk = -1$. Then $ij = k, jk = i, ki = j$, but $ji = -k, kj = -i, ik = -j$. This is a nonabelian extension of the complex numbers. Multiplication of the unit quaternions forms a group of order eight. Given the vector field $e_1(x) = ix, e_2(x) = jx, e_3(x) = kx$, we claim all these vectors are tangent to S^3 and form an orthonormal basis for the tangent space. This gives us a frame, but we don't have a coordinate system. We could also consider the alternate frame $\widetilde{e}_1(x) = xi, \widetilde{e}_2(x) = xj, \widetilde{e}_3(x) = xk$.

Let us define some interesting connections. One is by defining $\nabla^L_{e_i} e_j = 0$. In the homework we show this connection is metric but has torsion. Another connection says that $\nabla^K_{\widetilde{e}_i}\widetilde{e}_j = 0$. This is also metric but not symmetric. A third connection is defined by $\nabla^M = (\nabla^L + \nabla^R)/2$, which will turn out to be metric and symmetric, so it's the Levi-Civita connection. Curvature in three dimensions is a bit more subtle since rotations don't commute. Eventually we get to curvature in higher dimensions, which is a tensor with four indices, a big mess.

13.3 A coordinate free approach to the fundamental theorem

Sometimes we don't have coordinates, and just want to talk about things in terms of frames or vector fields. The point is that frames e_i may not be coordinates $\partial/\partial x^i$. What are $\nabla_{e_i} e_j, [e_i, e_j]$, and $g(e_i, e_j)$? In a coordinate frame, the brackets are zero since mixed partials commute. However, this doesn't always hold for arbitrary frames. It would be nice if we had a formula for $\nabla_{e_i} e_j$ in terms of $[e_i, e_j]$ and $g(e_i, e_j)$. In full generality, we could talk about $g(\nabla_X Y, Z)$, $[X, Y]$ etc, $g(X, Y)$ etc, for X, Y, Z vector fields. The fundamental theorem says that there exists a unique metric is, so we can figure out $g(\nabla_X Y, Z)$ in terms of the other data. But there are other formulas for the Levi-civita connection, which we'll talk more about next time. This expression is called **Koszul's formula**.

Lecture 14

March 11, 2021

14.1 General formulas for the Levi-Civita connection

Consider $g(Y, Z)$ which is just a function, so we can take its directional derivative $X(g(Y,Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z)$. (Denote $\langle X, Y\rangle = g(X,Y)$). So

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

is the equation for being metric. For a symmetric connection, we have

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

In a coordinate basis $[X,Y] = 0$, so $\Gamma_{ij}^k = \Gamma_{ji}^k$. Now just by changing the names of the vector fields we have $Y\langle X,Z\rangle = \langle \nabla_Y X, Z\rangle + \langle X, \nabla_Y Z\rangle$ and $Z\langle X,Y\rangle = \langle \nabla_Z X, Y\rangle + \langle X, \nabla_Z Y\rangle$. Then

$$\begin{aligned}
+X\langle Y,Z\rangle &= \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle \\
+Y\langle X,Z\rangle &= \langle \nabla_Y X, Z\rangle + \langle X, \nabla_Y Z\rangle \\
-Z\langle X,Y\rangle &= \langle \nabla_Z X, Y\rangle + \langle X, \nabla_Z Y\rangle \\
&\Longrightarrow \\
X\langle Y,Z\rangle + Y\langle X,Z\rangle - Z\langle X,Y\rangle &= \langle \nabla_X Y + \nabla_Y X, Z\rangle + \langle \nabla_X Z - \nabla_Z X, Y\rangle + \langle \nabla_Y Z - \nabla_Z Y, X\rangle \\
&= 2\langle \nabla_X Y, Z\rangle - \langle [X,Y],Z\rangle + \langle [X,Z],Y\rangle + \langle [Y,Z],X\rangle \\
&\Longrightarrow
\end{aligned}$$

$$\boxed{\langle \nabla_X Y, Z\rangle = \frac{1}{2}\left(X\langle Y,Z\rangle + Y\langle X,Z\rangle - Z\langle X,Y\rangle + \langle [X,Y],Z\rangle - \langle [X,Z],Y\rangle - \langle [Y,Z],X\rangle\right).}$$

This equation is called **Koszul's formula**, which is the most general formula for the Levi-Civita connection. There are some bases we care about:

(1) If $X = \partial_i, Y = \partial_j, K = \partial_k$, then $\Gamma_{ijk} = \frac{1}{2}\left(\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}\right)$ since all the bracket terms are zero. This is the form we're familiar with of the Levi-Civita connection, and the most common one as well.

(2) Say we have an orthonormal basis $\{E_i\}$, where $X = E_i, Y = E_j, Z = E_k$. Then $[E_i, E_j] = c_{ij}^k E_k$, where the coefficients c_{ij}^k tell you to what extent are the brackets nonzero. If $c_{ijk} = \langle [E_i, E_j], E_k\rangle$, then $\Gamma_{ijk} = \Gamma_{ij}^k = \frac{1}{2}(c_{ijk} - c_{ikj} - c_{kji})$.

(3) Given a general frame, we have $\Gamma_{ijk} = \frac{1}{2}(\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}) + \frac{1}{2}(c_{ijk} - c_{ikj} - c_{jki})$.

14.2 Return to geodesics

Recall that a geodesic satisfies $\nabla_{\dot{x}}\dot{x} = 0$, and $\ddot{x}^k + \Gamma_{jk}^k \dot{x}^i \dot{x}^j = 0$. Say we have a point p and a vector v_0, we want to talk about a geodesic starting at p with velocity v_0. We want to convert this second order differential equation of n variables into a first order ODE of two variables. Say $v(t) = \dot{x}(t)$, then what is $\dot{v}$? Recall $\nabla_v v = 0 \iff \dot{v}^k + \Gamma_{jk}^k(x) v^i v^j = 0$, so

$$\begin{aligned}
\dot{v}^k &= -\Gamma_{ij}^k(x) v^i v^j, \\
\dot{x}^k &= v^k.
\end{aligned}$$

These are our two desired differential equations. Then we have unique solutions since the Γ's are smooth functions of x (the particular smooth function is given by the Levi-Civita connection). In other words, there exists a unique solution $x(0) = p, v(0) = v_0$ for a short time (locally).

For an arbitrary amount of time, consider the line going through the origin in $\mathbb{R}^2 \setminus \{0\}$, which fails near the origin. Another example is the open interval $M = (a,b), g = dx^2$; after a certain amount of time you fall of the edge of the world. If somebody gives you a manifold with a starting point and velocity, we want to run a geodesic for as long as possible in the $\pm t$ direction. This motivates the following definition.

Definition 14.1. A geodesic is **maximal** if it can't be extended.

Is there a maximal geodesic? Given a starting point, take the union of the geodesics on an interval. This is unique, and we can't extend this any farther because this implies a bigger interval. So given any point and any starting vector, there is always a unique maximal geodesic.

14.3 The exponential map

Say we have a manifold M, $p \in M$, and the tangent space T_pM. Then the **exponential map** is defined as $\exp_p(v) = \gamma(1)$, where γ is a geodesic with $\gamma(0) = p, \dot{\gamma}(0) = v$. Is this defined on all of T_pM? This is defined for all sufficiently small v (for large v you may fall of the edge of the world). You can think of $\exp_p(sv) = \gamma_{sv}(1) = \gamma_v(s)$. For every p there exists a small neighborhood that looks like $\mathbb{R}^n$, so exp is well defined on the tangent space in that neighborhood. So think of the exponential map as the map

$$\exp_p : \text{Nbd of } 0 \in T_pM \to \text{Nbd of } p \in M.$$

Given a compact manifold we can extend this, so geodesics run forever and the exponential map is defined for all t. What is $d\exp_p|_{v=0}$? This is the identity, since $d\exp_p|_{v=0} : T_0(\text{Nbd of } 0 \in T_pM) \to T_pM$ which is the same as a map $T_pM \to T_pM$. At $v = 0$, this is saying "given an infinitesmally small vector, where do you wind up?" This is the same thing as saying "how fast are you moving at time zero if you have a large vector?", since how far you wind up with an infinitesmally small vector is how far you wind up with an ordinary geodesic in a short amount of time. So this asks how geodesics at small time, which is just the derivative of a geodesic at 0, which is just v!

The nice thing about the identity is that it's invertible. Suppose we have two manifolds $M, N, f : M \to N, p \in M$, $q = f(p) \in N$, $df : T_pM \to T_qN$. If df_p is invertible, then f is a local diffeomorphism, so $f|_V$ is a diffeomorphism $U \to V$ (where U, V are neighborhoods of p, q). So $\exp_p$ is a diffeomorphism, since it takes neighborhoods to neighborhoods. Let $r = \sup\{\text{radii } \mid \exp_p \text{ is a diffeomorphism on } B_p\}$. We say r is the **injectivity radius** at p.

Example 14.1. Consider M the torus, if we draw it as a rectangle, say it has width L_1 and height L_2, $L_2 > L_1$. Then the injectivity radius at p is $L_2/2$, since any points past $L_2/2$ wrap around. The exponential map is a local diffeomorphism, but it fails to be injective; this is why it's called the *injectivity* radius. On a sphere with p the north pole, our injectivity radius is π (since the circumference is 2π).

You might think the injectivity radius is about the topology, since $H_1(\mathbb{T}) = \mathbb{Z} \oplus \mathbb{Z}$ is nontrivial and we have a cycle to wrap around. It turns out the injectivity radius isn't just about the topology, since S^2 has no interesting first homology group. A handwavy way to think about the injectivity radius is the biggest radius such that a neighborhood of size r around p looks like a ball.

If we have orthonormal coordinates for T_pM, this induces coordinates on T_qN by the exponential map. Then $g_{ij}(p) = \delta_{ij}$, or even stronger, we have $g_{ij}(\exp_p(v)) = \delta_{ij} + O(v^2)$. Consider a ball of radius $\varepsilon < r$ in T_pM, for $p, q \in U_p$ (for U_p a neighborhood around p in M 0. Is there a geodesic connecting p and q? Sure there is, since exp is a diffeomorphism. The geodesic is also locally unique. If $q = \exp_p(v)$, then what is the distance from p to q? It's the inf of all lengths; not all geodesics globally minimize length, but a length minizming curve is certainly a geodesic. This is equivalent to the energy concept, that the variational equations for energy give us geodesics WRT the Levi-Civita connection. Since the shortest path gives us a geodesic, and there exists exactly one geodesic, then the geodesic must be the shortest path. So the distance from p to q is the magnitude of v.

In other words, the distance function from p is just the magnitude function in T_pM. If we use polar coordinates on T_pM, the metric in the radial direction $g_{rr} = 1$, so the metric will always be dr^2+something.

14.4 Tubular neighborhoods

Suppose we have a Riemannian manifold M and a submanifold N. The tubular neighborhood theorem states that if N is compact, the set of all points within ε of N is a tubular neighborhood, and is diffeomorphic to a neighborhood of the zero section of the normal bundle. If N is not compact, then ε is not globally chosen (it varies from point to point). What does a neighborhood of $p \in N$ look like? A neighborhood of p in the big space looks like a neighborhood of the origin in T_pM. But $T_pM = T_pN \oplus N_pN$, so we parametrize by $(p, v) \to \exp_p(v)$.

Lecture 15

March 23, 2021

15.1 Calculus of variations

Suppose we have $z(t)$, where $z(0) = 0, z(2) = 0$. We want to minimize the integral $\mathscr{L} = \int_0^2 \frac{1}{2}\dot{z}^2 - 32z\,dt$. We will do this two ways: one is the "sloppy" version with our usual notation for the calculus of variations, and one by a mathematically precise method. The usual notation has a precise meaning behind it, but it just seems sloppy.

Lecture 16

April 6, 2021

In two dimensions, parallel transport around an orientation preserving loop is a rotation, which you can parametrize by just a number θ. You can sum up the rotations around tiny little loops. Define K to be the rotation per unit block on an area A bounded by a loop γ, then define the size of rotation by $P_\gamma = \iint_A K\,d\text{Area}$.

16.1 Gauss-Bonnet Theorem