MATH 144 NOTES: RIEMANNIAN GEOMETRY

ARUN DEBRAY
MARCH 13, 2014

These notes were taken in Stanford's Math 144 class in Winter 2014, taught by Professor Rick Schoen. I live-TEXed them using `vim`, and as such there may be typos; please send questions, comments, complaints, and corrections to `a.debray@math.utexas.edu`

Contents

1. Manifolds: 1/7/14

"What is Riemannian geometry? If you want a one-line summary, it's the study of curved spaces."

...though not every space is curved. Uncurved spaces are flat, e.g. Euclidean space, $\mathbb{R}^n$. Riemannian geometry involves lots of calculation, so various notational bookkeeping tricks are common. For example, the coordinates in $\mathbb{R}^n$ are written with superscripts, as $x^1, \dots, x^n$.

Geometry in $\mathbb{R}^n$ is due to the inner product: $\langle \mathbf{v}, \mathbf{w} \rangle = \sum_{i=1}^n v_i w_i$ (where $\mathbf{v} = (v_1, \dots, v_n)$, and $\mathbf{w}$ is analogous). This is very useful for measuring lengths, angles, and even k-dimensional volume of k-submanifolds, and is considered flat Riemannian geometry. For another example, Gauss found a coordinate-invariant quantity that indicates whether a space is flat, and it happens that all flat spaces locally look like $\mathbb{R}^n$.

For a more exotic example (though maybe not if you're a physicist), one has flat Lorentz geometry, given by Minkowski space $\mathbb{R}_1^{n+1}$, in which coordinates are denoted $x^0, \dots, x^n$, and the inner product is $\langle \mathbf{v}, \mathbf{w} \rangle = -v_0 w_0 + \sum_{i=1}^n v_i w_i$. The Euclidean metric on $\mathbb{R}^n$ is positive definite, so it is called a Riemannian metric, but the Minkowski metric is indefinite, and is thus called a Lorentz metric.

This space turns out to be important for special relativity, especially when $n = 3$. It has much more intricate geometry than Euclidean space. For example, there are three kinds of vectors: *time-like*, where $\langle \mathbf{v}, \mathbf{v} \rangle < 0$, *space-like*, in which $\langle \mathbf{v}, \mathbf{v} \rangle > 0$, and *null*, for which $\langle \mathbf{v}, \mathbf{v} \rangle = 0$, so that the magnitude of the first component is equal to that of the last n. These lie on a 45° cone called a lightcone; space-like vectors lie outside of this lightcone, and time-like vectors lie on the inside.

The group of linear transformations that preserve this inner product is called the Lorentz group, and the transformations themselves are called Lorentz transformations. These transformations preserve the types of vectors, so, in some sense, this space has a distinguished direction.

Before talking about curved spaces, it's probably important to formalize exactly what a space is. Geometry ought to be a property of the structure of an object, so it should be independent of coordinates. Though in other parts of math one might use a more general definition, Riemannian geometry uses the notion of smooth n-dimensional manifolds.

Definition. A chart on M (the thing that will eventually be a manifold) is a pair (φ, U), where $U \subseteq M$, $\varphi : U \to \mathbb{R}^n$ is one-to-one, and $\varphi(U)$ is open in $\mathbb{R}^n$.

For example, the 2-sphere S^2 and 2-torus T^2 are two-dimensional manifolds. At a point, one can map a neighborhood into $\mathbb{R}^2$. Since φ is one-to-one, then its inverse φ^{-1} can be thought of as assigning coordinates to a neighborhood U in M.

Definition. A smooth manifold is a set M such that $M = \bigcup_{i \in I} U_i$, such that there exist charts (φ_i, U_i) for each $i \in I$, subject to the condition that if $U_i \cap U_j \neq \emptyset$, then $\varphi_i(U_i \cap U_j)$ is open in $\mathbb{R}^n$ and the map $\varphi_j \circ \varphi_i^{-1} : U_i \cap U_j \to \mathbb{R}^n$ is smooth.[1]

Like all definitions of manifolds, this one sounds kind of scary, but in essence, an n-dimensional manifold is a set that locally looks like $\mathbb{R}^n$: the charts give it local coordinates, and the smoothness condition ensures that these sets of coordinates are compatible with each other. Note that the rigorous definition above wasn't assembled until the 20$^{\text{th}}$ Century, considerably after Riemann did his stuff.

For convenience, "smooth" will mean C^∞ for this class. But there are lots of other interesting classes of manifolds obtained by changing this requirement to mean, for example, complex-analytic, or C^2, or so on.

Finally, note that Kühnel's definition, which is given here, is different from those given in most advanced textbooks, which require M to already have some topology, so that each φ in a chart is a local homeomorphism (which automatically implies $\varphi(U)$ is open). However, these two definitions seem to be compatible with each other.

Definition. An atlas is a collection of charts that cover M, i.e. such that for every $p \in M$ there exists a chart (φ, U) in the atlas with $p \in U$.

For a given manifold there may be many atlases, so one assumes the maximal atlas for a given manifold, which is unique. This atlas is called a differentiable structure on M. There's an interesting nuance in that there exist topologically identical spaces which have different differentiable structures, but that is far beyond the scope of this class.

Examples. Some examples of manifolds will make the definition less abstruse.

1. $\mathbb{R}^n$ and $\mathbb{R}_1^{n+1}$ as given above are manifolds.
2. If $U \subseteq \mathbb{R}^n$ is open, then it is a submanifold. There is a single chart, (U, id).
3. Submanifolds of $\mathbb{R}^n$ are an important class, with lots of examples. Special cases include curves and surfaces in space. These have extrinsic geometry, given by how the submanifold sits in the ambient space, as well as intrinsic geometry, given by making measurements on or along the submanifold. This class will focus on the intrinsic geometry, though some submanifolds, such as curves, only have interesting extrinsic geometry. As for what these things actually are:

   Definition. An n-dimensional submanifold is a set $M \subseteq \mathbb{R}^m$ where $n < m$ such that for every $p \in M$ there exists an open $U \subseteq \mathbb{R}^m$ and a $f : U \to \mathbb{R}$ such that $f(q) = 0$ for all $q \in M \cap U$ and (in the case $m = n + 1$; the full case will be given below) $\nabla f(p) \neq 0$.

   This definition is again kind of abstract, but let's see what it implies: by the Inverse and/or Implicit Function Theorems, this is the condition that M is locally the graph of a smooth function: $M = \{(x^1, \ldots, x^n), f(x^1, \ldots, x^n) : x^1, \ldots, x^n \in U\}$. Then, $x^1, \ldots, x^n$ can be taken to be local coordinates, i.e. they are the result of the chart maps: $\varphi(x^1, \ldots, x^n, f(x^1, \ldots, x^n)) = (x^1, \ldots, x^n)$. Thus, submanifolds are in fact manifolds!

   In the more general case where $n < m$ by some possibly larger amount, the function instead becomes $f : U \to \mathbb{R}^{m-n}$ and the requirement is that its differential map (i.e. the matrix of partial derivatives) $Df|_p$ is onto.[2] Thus, one gets charts in the same way.

[1] Since one also has a smooth map in the opposite direction given by switching i and j, this condition is equivalent to requiring it to be a diffeomorphism.

[2] That is, it's surjective. Since the differential is a linear map, this means it must have full rank, which is a bit easier to check.

Submanifolds can be thought of as manifolds that exist inside some space $\mathbb{R}^n$ instead of abstractly, which makes them a little easier to visualize: as surfaces that are locally Euclidean or are locally the graph of a smooth function. These are quite important examples of manifolds.

4. There are also manifolds that don't have a natural embedding into an ambient space, such as quotient spaces. For example, the two-dimensional torus $T^2 = \mathbb{R}^2/\mathbb{Z}^2$ is defined as the set of equivalence classes of elements of $\mathbb{R}^2$, where $(x, y) \sim (x + m, y + n)$ for $x, y \in \mathbb{R}$ and $m, n \in \mathbb{Z}$. Each equivalence class can be thought of as a lattice in the plane.

   To obtain charts for this object, pick an $(x_0, y_0) \in \mathbb{R}^2$ and let $U_{(x_0,y_0)} = (x_0 - 1/2, x_0 + 1/2) \times (y_0 - 1/2, y_0 + 1/2)$, which is an open set, so consider the map $\pi : \mathbb{R}^2 \to \mathbb{R}^2/\mathbb{Z}^2$ that sends a point to its equivalence class. Since U has side length 1 and is open, then $\pi|_U$ is one-to-one, so if $\overline{U}_{(x_0,y_0)} = \pi(U_{(x_0,y_0)})$, then $\varphi : \overline{U}_{(x_0,y_0)} \to \mathbb{R}^2 = \mathrm{id} \circ \pi^{-1}$.

   This is a common construction: if a group acts on a manifold without fixed points, then one obtains a quotient manifold.

There are lots of other examples in the textbook.

Definition. If M is a manifold, then $U \subseteq M$ is open if for all charts (φ, U'), the set $\varphi(U \cap U')$ is open in $\mathbb{R}^n$.

One can think of this in the following way: in $\mathbb{R}^n$, a set U is open if for every $\mathbf{x} \in U$, there's an open ball around $\mathbf{x}$ contained in U. The same thing is going on here, but the open ball is given by φ^{-1}. The collection of open sets gives a topology to the manifold, in which all of the chart maps are continuous.[3] This topology determines various properties of the manifold, such as whether it is compact.

In this class, some assumptions will be made on this topology, though only one will be introduced now.

H1. M is Hausdorff; that is, if $p_1, p_2 \in M$ are distinct, then there exist open sets $U_1, U_2 \subset M$ such that $p_1 \in U_1$, $p_2 \in U_2$, and $U_1 \cap U_2 = \emptyset$. That is, every two points have some disjoint open neighborhoods.

While this might seem intuitive or obvious, it doesn't actually follow from the other axioms as given.

Definition. A manifold is compact if every open cover has a finite subcover (i.e. that if one has a collection $\{U_i\}_{i\in I}$ of sets such that $\bigcup_i U_i = M$, then there is some finite set $U_1', \ldots, U_n'$ such that $\bigcup_{j=1}^n U_j' = M$ as well. This can be thought of as akin to finiteness).

Definition. A map $F : M \to N$ of manifolds is continuous if for every open set U of N, its preimage $F^{-1}(U)$ is open in M. A continuous function is smooth if for all $p \in M$, charts (φ, U) such that $p \in U$, and charts (ψ, V) with $F(p) \in V$, it's possible to choose an open neighborhood $U_1 \subseteq U$ of p such that $F(U_1) \subseteq V$ (which is always true because F is required to be continuous), but then that $\psi \circ F \circ \varphi^{-1} : \varphi(U_1) \to \varphi(V)$ is smooth as a map $\mathbb{R}^m \to \mathbb{R}^n$. This is well-defined if one takes different charts, because change-of-chart maps are also smooth.

For some extremal cases, a smooth curve is a map $c : (a, b) \to M$ and a smooth function is a map $f : M \to \mathbb{R}$. For the latter, establish the following conventions: let $\varphi : U \to \mathbb{R}^n$ be a chart and $u^1, \ldots, u^n$ be coordinates in $\mathbb{R}^n$ and $x^1, \ldots, x^n$ be local coordinates on U, i.e. $x^i = u^i \circ \varphi$ (here, coordinates are treated as functions; the coordinate function takes a point and returns the value of that coordinate at that point, e.g. $x(1, 2) = 1$). The notation $\left.\frac{\partial f}{\partial x^i}\right|_p$ means $\left.\frac{\partial (f\circ\varphi^{-1})}{\partial u^i}\right|_{f(p)}$ for $p \in U$. In some sense, the function is transplanted to a function $\mathbb{R}^n \to \mathbb{R}$.

The next subject will be that of tangent vectors and tangent spaces, which will be discussed more than tangentially next lecture. For submanifolds, which are embedded in an ambient space, the notion is reasonably intuitive, but how should one define the tangent space of an abstract manifold? The tangent space of a $p \in M$ is an n-dimensional vector space, the set of tangent vectors at M, so the question reduces to finding tangent vectors. There are three equivalent and useful definitions.

The geometric definition considers the set of curves $c : (-\varepsilon, \varepsilon) \to M$ for some $\varepsilon > 0$ and $c(0) = p$. Then, create an equivalence relation on this set by $c \sim \widetilde{c}$ if $\varphi \circ c$ and $\varphi \circ \widetilde{c}$ agree to first order for all charts φ (i.e. their difference vanishes, as does its first derivative). Then, the tangent space is this set of equivalence classes. This definition is pretty, and it's important and valid, but it's not immediately clear how to add tangent vectors.

A more computational definition will invoke the notion of the directional derivative.

2. Tangent and Cotangent Spaces: 1/9/14

Last time, we mentioned manifolds and defined open sets on a manifold M as those sets U where for all charts (φ, U'), $\varphi(U \cap U')$ is open in $\mathbb{R}^n$. This implies that if (φ, U) is a chart, then U is open in M (though this isn't completely obvious), and that if $V \subseteq \varphi(U)$ is open in $\mathbb{R}^n$, then $\varphi^{-1}(V)$ is open.

[3] If you are unfamiliar with point-set topology, and therefore the general, abstract definitions of topological spaces and continuous maps, check out the first few chapters of Munkres' textbook or a similar source, or come ask me.

Recall that in $\mathbb{R}^n$, a set V is open if for all $\mathbf{p} \in V$, there exists an $\varepsilon > 0$ such that $B_\varepsilon(\mathbf{p}) = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x} - \mathbf{p}\| < \varepsilon\} \subseteq V$. This also works for manifolds: a set $U \subseteq M$ is open if for all $p \in U$ and charts (φ, U') with $p \in U$, then $B_\varepsilon(x_0) \subseteq U$ for some $\varepsilon > 0$, where $x^1, \dots, x^n$ are local coordinates induced by (φ, U') in U' and $x_0 = (x^1(p), \dots, x^n(p))$. In $\mathbb{R}^n$, one can show that this is equivalent to to the other definition. Though these definitions may seem a bit mechanical, all they mean is that if one introduces local coordinates on a manifold, an open set contains some neighborhood of every one of its points.

Definition. If M and N are manifolds (m- and n-dimensional, respectively), then $F : M \to N$ is smooth if whenever (φ, U) is a chart in M and (ψ, V) is a chart in N such that $F(U) \subseteq V$, then $\psi \circ F \circ \varphi^{-1} : \varphi(U) \to \mathbb{R}^n$ is smooth.

A more compact (heh) way to say this is that if one views the local coordinates as living in $\mathbb{R}^n$, then the resulting map is smooth.

In general, manifolds are trickier to deal with, because geometrical notions need to be coordinate-independent, but there is no one canonical set of global coordinates.

Tangent Spaces. If M is an n-dimensional manifold and $p \in M$, then the tangent space T_pM is an n-dimensional real vector space (though this won't be immediately obvious from the definition).

Definition. Place an equivalence relation on the set of smooth (i.e. C^∞), real-valued functions on M in which $f_1 \sim f_2$ if $f_1 = f_2$ within a neighborhood of p. The set of equivalence classes $\mathscr{F}_p(M) = \{[f]; f \in C^\infty(M)\}$ is called the set of germs of smooth functions at p. (Here $[f]$ denotes the equivalence class of f.)

Observe that adding and multiplying functions and multiplying them by real numbers respects the equivalence relation, so $\mathscr{F}_p(M)$ is an $\mathbb{R}$-algebra.

Definition. The tangent vectors of M at p are the derivations $X : \mathscr{F}_p(M) \to \mathbb{R}$; that is, they must satisfy the following two properties:

1. X must be linear: $X(c_1 f_1 + c_2 f_2) = c_1 X(f_1) + c_2 X(f_2)$ for any $c_1, c_2 \in \mathbb{R}$ and $f_1, f_2 \in C^\infty(M)$.[4]
2. X must satisfy the Leibniz property: $X(f_1 f_2) = f_1 X(f_2) + X(f_1) f_2$.

These are basically derivative-like operators on the germs of functions.

Then, the tangent space T_pM is the set of tangent vectors at p.

The tangent space is a real vector space because derivations can be added and multiplied by scalars. However, its dimension is not apparent.

If $p \in M$ and $c : (-\varepsilon, \varepsilon) \to M$ is a smooth curve (one says that it's parameterized, so that it's a smooth map from an interval into M) such that $c(0) = p$, then there is a natural tangent vector $c'(0) \in T_p(M)$ given by

$$c'(0)(f) = \frac{\mathrm{d}}{\mathrm{d}t}(f \circ c(t))\bigg|_{t=0}.$$

Basically, how does the function change along a curve? It's not hard to check that $c'(0)$ is a derivation.

Let $x^1, \dots, x^n$ be local coordinates near p. Then, some additional tangent vectors are notated

$$\frac{\partial}{\partial x^i}\bigg|_p (f) = \frac{\partial f}{\partial x^i}\bigg|_p = \frac{\partial f \circ \varphi^{-1}}{\partial u^i}\bigg|_{\varphi(p)},$$

where $u^1, \dots, u^n$ are coordinates in $\mathbb{R}^n$, so that $x^i = u^i \circ \varphi$. These are also derivations.

Theorem 2.1. *Given $x^1, \dots, x^n$ local coordinates near p,*

$$\left\{ \frac{\partial}{\partial x^1}\bigg|_p, \dots, \frac{\partial}{\partial x^n}\bigg|_p \right\}$$

is a basis for T_pM. In particular, T_pM is an n-dimensional vector space.

Once again, the message is the same: there are local coordinates, but no canonical choice of them, and certainly no global coordinates. Similarly, there is no canonical basis for the tangent space, but there certainly is a basis. Change of basis happens a lot in geometry, because every useful geometric notion must be independent of coordinates.

Proof of Theorem 2.1. This proof will rely on Taylor's theorem.

[4] Technically, X acts on the equivalence classes of f_1 and f_2, but the end result is the same.

Choose a chart (φ, U) such that $p \in U$, and without loss of generality assume $\varphi(p) = 0$ (since if not, then the chart can be translated). Then, take a small ball around the origin and points $\mathbf{u} = (u^1, \dots, u^n)$ in the ball, and let f be a smooth function on a neighborhood of p and $h = f \circ \varphi^{-1}$. Now, define n functions

$$h_i(\mathbf{u}) = \int_0^1 \frac{\partial h}{\partial u^i}(t\mathbf{u})\,\mathrm{d}t,$$

where $h_i(0) = \frac{\partial h}{\partial u^i}(\mathbf{0})$.

Claim. Then, $h(\mathbf{u}) - h(\mathbf{0}) = \displaystyle\sum_{i=1}^n h_i(\mathbf{u})u^i$.

Proof. Well, by the Fundamental Theorem of Calculus and the Chain rule, this can be simplified to

$$h(\mathbf{u}) - h(\mathbf{0}) = \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} h(t\mathbf{u})\,\mathrm{d}t = \sum_i \frac{\partial h}{\partial u^i}(t\mathbf{u})u^i\,\mathrm{d}t. \qquad \boxtimes$$

The professor described this as sort-of an exact first-order approximation.

Now, this can be transplanted to the manifold: write $h_i = f_i \circ \varphi^{-1}$ and $u^i = x^i \circ \varphi^{-1}$. Let $q = \varphi^{-1}(\mathbf{u})$, so that $f(q) - f(p) = \sum_{i=1}^n f_i(q)x^i(q)$. Thus, in a neighborhood of p, $f = f(p) + \sum f_i x^i$, and at p, $x^i(p) = 0$ and $f_i(p) = h_i(\mathbf{0}) = \left.\frac{\partial f}{\partial x^i}\right|_p$.

Now, given an $X \in T_pM$, one can apply the Leibniz rule:

$$X(f) = X(f(p)) + \sum_{i=1}^n (X(f_i)x^i(p) + f_i(p)X(x^i)).$$

However, the derivation of the constant $f(p)$ must be zero (which can be proven by appealing again to the Leibniz rule). Thus,

$$X(f) = \sum_{i=1}^n X(x^i) \cdot \left.\frac{\partial}{\partial x^i}\right|_p (f),$$

and therefore

$$X = \sum_{i=1}^n X(x^i) \left.\frac{\partial}{\partial x^i}\right|_p.$$

These $\left.\frac{\partial}{\partial x^i}\right|_p$ are linearly independent because $\frac{\partial}{\partial x^i}(x^j) = \delta_i^j$, so they do indeed form a basis. $\boxtimes$

This proof also gave the coefficients $X(x^i)$ for the linear combination, which is useful. It also shows that this abstract definition is concretely what we want as the "right thing."

Using this, one can make some more useful constructions, appealing to the general principle that from one vector space one can obtain a lot of other useful vector spaces.

Definition. The cotangent space T_p^*M is the dual space to T_pM, i.e. $L(T_pM, \mathbb{R})$, the space of linear real-valued functions on T_pM (which is a construction that can be made on any vector space).

T_p^*M is also an n-dimensional vector space, and given a basis $v_1, \dots, v_n$ for T_pM, there is an associated dual basis $\omega^1, \dots, \omega^n$ for T_p^*M given by $\omega^i(v_j) = \delta_j^i$ (i.e. the Kronecker delta, equal to 1 if $i = j$ and 0 otherwise). Notice the notation: the indices on basis vectors go down, but those for the dual basis go up.

The dual basis to the basis $\left.\frac{\partial}{\partial x^i}\right|_p$ has the special notation $\mathrm{d}x^1|_p, \dots, \mathrm{d}x^n|_p$, which means that $\mathrm{d}x^i\big(\frac{\partial}{\partial x^j}\big) = \delta_j^i$. Thus, coordinates induce a basis for both the tangent and cotangent spaces.

Vector Fields and One-Forms. A vector field is a map $X : p \mapsto X_p \in T_pM$, assigning each p to a point in its tangent space, and is required to be smooth (i.e. that when it is expressed in coordinates, the coordinates themselves are smooth). More concretely, if (φ, U) is a chart with $p \in U$, then

$$X_p = \sum_{i=1}^n \xi^i(p) \left.\frac{\partial}{\partial x^i}\right|_p$$

for some smooth functions $\xi^1, \dots, \xi^n$ on U. This is akin to the definitions seen in physics.

A vector field is a special case of a more general object called a tensor field. Specifically, it is a contravariant tensor field of degree 1, one of the simpler cases.

A one-form is a map $p \mapsto \omega_p \in T_p^*M$ that is also required to be smooth: $\omega_p = \sum_{i=1}^n a_i(p)\,\mathrm{d}x^i|_p$. For notational consistency, the coordinates have lower indices, and these $a_1, \dots, a_n$ are also required to be smooth. The name is because a one-form induces a line integral on curves. These are also tensor fields, specifically covariant tensor fields of degree 1.

The most important tensor fields in this course are Riemannian metrics. Let $\mathrm{Bil}(T_pM, \mathbb{R})$ denote the space of bilinear forms on M,[5] i.e. α such that $\alpha(v, w) \in \mathbb{R}$ for $v, w \in T_pM$ and $\alpha(c_1v_1 + c_2v_2, w) = c_1\alpha(v_1, w) + c_2\alpha(v_2, w)$ (and similarly for the second argument). $\mathrm{Bil}(T_pM, \mathbb{R})$ is a vector space, and again it's true that a basis for the tangent space $v_1, \dots, v_n$ and the induced dual basis $\omega^1, \dots, \omega^n$ induce a basis $\omega^i \otimes \omega^j \in \mathrm{Bil}(T_pM, \mathbb{R})$, given by $\omega^i \otimes \omega^j(v_1, v_2) = \omega^i(v_1)\omega^i(v_2)$, which is clearly bilinear. These $\{\omega^i \otimes \omega^j \mid 1 \leq i, j \leq n\}$ form a basis for $\mathrm{Bil}(T_pM, \mathbb{R})$, so it has dimension n^2. Inner products in this space lead to Riemannian metrics.

3. Vector Fields, One-Forms, and Riemannian Metrics: 1/14/14

Definition. Let M and N be smooth manifolds of dimensions m and n, respectively, and let $F : M \to N$ be a smooth map. Then, the differential of F at a point $p \in M$ is the linearization of F at p: $DF|_p : T_pM \to T_{F(p)}N$ given by $DF|_p(X)(f) = X(f \circ F)$.

From this definition, intuition isn't incredibly clear, but it's easier to see that the differential is a linear map.

Exercise 3.1. Suppose that $x^1, \dots, x^m$ are coordinates in M near p, and $y^1, \dots, y^n$ are coordinates in N near $F(p)$, where $f : y^i = y^i(x^1, \dots, x^n)$. If $\frac{\partial}{\partial x^i}\big|_p$ and $\frac{\partial}{\partial y^i}\Big|_{F(p)}$ are the induced coordinates on the tangent spaces, then the matrix $DF|_p$ is the Jacobian, i.e. it is an $n \times m$ matrix whose ij^{th} entry is $\frac{\partial y^i}{\partial x^j}$.

This exercise convinces us that the differential is really the right thing, so to speak.

Theorem 3.1 (Chain Rule). *Let $F : M_1 \to M_2$ and $G : M_2 \to M_3$ be smooth maps of smooth manifolds. Then, for $p \in M_1$, $D(G \circ F)|_p = DG|_{F(p)} \circ DF|_p$.*

Proof. Let $X \in T_pM_1$ and f be smooth near p. Then,

$$\begin{aligned} D(G \circ F)|_p(X)(f) &= X(f \circ (G \circ F)) \\ &= X((f \circ G) \circ F) \quad \text{because composition is associative.} \\ &= DF|_p(X)(f \circ G) \\ &= DG|_{F(p)}(DF_p(X))(f) = (DG|_{F(p)}DF|_p)(X)(f), \end{aligned}$$

because matrix composition is just multiplication. ⊠

Basically, the proof boils down to a careful application of the definition.

On the homework, one is provided with the definition of the tangent bundle TM to a manifold M. This allows one to define a vector field somewhat snazzily as a smooth assignment into the tangent field that projects to the identity, but it's generally easier to think of a vector field as $X = \sum_{i=1}^n \xi^i(x) \frac{\partial}{\partial x^i}\big|_p$, where $x^1, \dots, x^n$ are local coordinates and the ξ^i are smooth real-valued functions. Because coordinate changes are smooth, this implies it's smooth in all possible charts.

Similarly, a smooth one-form, or a cotangent vector field, is a smooth (in local coordinates) assignment $p \mapsto \omega_p \in T_p^*M$. Thus, in local coordinates, a one-form looks like $\omega = \sum_{i=1}^n a_i(x)\,\mathrm{d}x^i$, where the a_i are smooth.

Last lecture, it was briefly mentioned that one-forms can be integrated along curves. This means that if $c : [a, b] \to M$ is a smooth curve, then the integral of a one-form ω along the curve c is

$$\int_c \omega = \int_a^b \omega(c'(t))\,\mathrm{d}t.$$

One can integrate higher-degree forms over more general submanifolds, but that's not really the point of this class.

Natural extensions of this lead to the notion of a Riemannian metric, i.e. a 2-covariant tensor field. The set of bilinear functions from $T_pM \to \mathbb{R}$ is a vector space of dimension n^2, where $n = \dim M$. Given a coordinate chart $x^1, \dots, x^n$ near p, one has dual vectors $\mathrm{d}x^1, \dots, \mathrm{d}x^n$. Then, $\mathrm{d}x^i \otimes \mathrm{d}x^j \in \mathrm{Bil}(T_pM, \mathbb{R})$ is given by

$$\mathrm{d}x^i \otimes \mathrm{d}x^j\left(\frac{\partial}{\partial x^k}\bigg|_p, \frac{\partial}{\partial x^\ell}\bigg|_p\right) = \delta^i_k\delta^j_\ell.$$

Then, this extends linearly: if $v = \sum a^i \frac{\partial}{\partial x^i}\big|_p$ and $w = \sum b^j \frac{\partial}{\partial x^j}\big|_p$, then $\mathrm{d}x^i|_p \otimes \mathrm{d}x^j|_p(v, w) = a^ib^j$.

[5] The professor used the notation $L^2(T_pM; \mathbb{R})$, which usually means something else.

Claim. These $\mathrm{d}x^i \otimes \mathrm{d}x^j$ form a basis; if $\alpha \in \mathrm{Bil}(T_pM, \mathbb{R})$, then

$$\alpha = \sum_{i,j=1}^{n} \alpha_{ij}\, \mathrm{d}x^i|_p \otimes \mathrm{d}x^j|_p. \tag{1}$$

Proof. That they're independent is fairly clear, so for any given α, let $\alpha_{ij} = \alpha(\frac{\partial}{\partial x^i}\big|_p, \frac{\partial}{\partial x^j}\big|_p)$, so that the goal is to show (1) holds for this choice of α_{ij}. Thus, just check on all of the basis vectors:

$$\sum_{k,\ell=1}^{n} \alpha_{k\ell}\, \mathrm{d}x^k \otimes \mathrm{d}x^\ell \left(\frac{\partial}{\partial x^i}\bigg|_p, \frac{\partial}{\partial x^j}\bigg|_p \right) = \sum_{k,\ell=1}^{n} \delta_k^i \delta_\ell^j \alpha_{k\ell} = \alpha_{ij}. \qquad \boxtimes$$

Now, with this space of bilinear maps (also called quadratic forms), one can talk about 2-covariant tensor fields: assignments $p \mapsto \alpha_p \in \mathrm{Bil}(T_pM, \mathbb{R})$ that are smooth in local coordinates near p, i.e. if α is given by α_{ij} as in (1), then α_{ij} is smooth in x.

Definition. A Riemannian metric g is a symmetric, positive definite 2-covariant tensor field. That is,

(1) $g_p(X, Y) = g_p(Y, X)$, and
(2) $g_p(X, X) \geq 0$ for all p and X, and this is 0 iff $X = 0$.

This is the same thing as requiring g_p to be an inner product for all p, or that the matrix g_{ij} at every p is symmetric and positive definite.

Riemannian metrics are, like functions in general, a dime a dozen.

Example 3.1. Suppose M is a manifold covered by a single chart. Then, any matrix-valued function that is symmetric and positive definite is a Riemannian metric. Some of these are interesting, others aren't.

More interestingly, consider hyperbolic space H^n, with $x^1, \dots, x^n \in \mathbb{R}^n$, but $x^1 > 0$. The hyperbolic metric is $g_{ij}(x) = (1/x_n^2)\delta_{ij}$. This is a conformal metric, because it's a multiple by a function of the standard Euclidean metric δ_{ij}. It also has constant curvature, which will be particularly important later in the class.

Another example can be given on an n-dimensional submanifold M inside $\mathbb{R}^k$. Then, the metric on $\mathbb{R}^k$ has an inner product $\langle , \rangle$, so a Riemannian metric on M is just $\langle , \rangle|_{T_pM}$. It has to be written down why this is smooth, but it is. In Math 143, this was called the first fundamental form.

To do computations with this metric, one can introduce local coordinates $u^1, \dots, u^n$ given by a chart map φ, with $\varphi^{-1} = F$; then, $g_{ij}(x) = \left\langle \frac{\partial F}{\partial u^i}, \frac{\partial F}{\partial u^j} \right\rangle \circ \varphi$. This can be thought of as having the u^i in $\mathbb{R}^n$, so that $u^i \circ \varphi$ is the coordinate on M. These induced metrics are important examples, and often very complicated.

Of course, this is all fine if M is covered by a single chart, but does every smooth manifold in general admit a Riemannian metric? Right now, no; we will need to make a stronger assumption on these manifolds, leading to questions in point-set topology. The assumption is:

3. Assume there exists a countable atlas for M.

This doesn't follow from the other two axioms, and is related to questions of logic and higher cardinals. Every manifold one might encounter in practice satisfies this condition, though. One counterexample is known as the "long line." Some equivalent assumptions in topology are second-countability (i.e. there exists a countable basis) or metrizability (there exists a metric which realizes the topology).

Claim. If M is compact and satisfies the given assumptions, then M has a Riemannian metric.

Second-countability is a generalization of compactness, but restricting the proof to this case gives the general idea and makes the details easier.

Proof. There can be lots of local metrics, and the trick is patching them together using partitions of unity.

For any $p \in M$, pick a chart (φ, U) such that $p \in U$, and pick an open neighborhood V of p such that $V \subset U$, and its closure $\overline{V} \subset U$. Then, there exists a smooth function[6] h such that $h = 1$ on V and $h = 0$ outside of U.

Then, take any metric in the chart, and let $x^1, \dots, x^n$ be local coordinates for this chart. For simplicity, in fact, it suffices to take the Euclidean metric $g_{ij}(x) = \delta_{ij}$ in U. Now, patch them together: these V give an open covering of M, so, because M is compact, then there's a finite subcovering $V_1, \dots, V_N$. Then, there are the associated $U_1, \dots, U_N$ and $h_1, \dots, h_N$ as part of the construction above, so that $V_i \subset U_i$ and so on, and metrics $g_{ij}^{(1)}, \dots, g_{ij}^{(n)}$ in these U_i.

Let $f_i = h_i / \sum_{j=1}^n h_j$: since every point within p lies within some V_i, then $\sum h_j \geq 1$, and at every point, $\sum f_i = 1$ (which is why it's called a partition of unity). Then, let $g|_p = \sum_{i=1}^n f_i g^{(i)}|_p$, where $f_i g^{(i)}$ is understood to be 0 outside U_i. This will be shown to be a Riemannian metric.

[6] This is decidedly *un*true for analytic functions, but is true for C^∞ functions.

If V is a vector space, then the collection of inner products on V is a convex cone within the space of bilinear functions, so taking a convex combination of them still results in an inner product: if $a_1, \dots, a_k \in [0,1]$, such that $\sum a_i = 1$, then $\sum a_i g^{(i)}$ is still an inner product. Thus, this is also true of Riemannian metrics.[7] $\boxtimes$

There are lots of metrics, because the construction in the proof offers lots of freedom. Partitions of unity are in general good ways to patch stuff together, but they don't work for Lorentz metrics, which don't have the same convexity, and in fact stronger topological assumptions are necessary for a manifold to admit a Lorentz metric.

4. The Lie Bracket and Riemannian Connections: 1/16/14

If g is a Riemannian metric, then the assignment $p \mapsto g_p$ produces an inner product on T_pM. For example, if $U \subseteq \mathbb{R}^n$ is open, then U is a manifold, and one can choose global coordinates $x^1, \dots, x^n$. Then, $x \mapsto g_x = \sum g_{ij}(x)\,\mathrm{d}x^i \otimes \mathrm{d}x^j$, where $(g_{ij}(x))$ is symmetric, positive definite, and smooth. Then, if $v = \sum v^i \frac{\partial}{\partial x^i}$ and $w = \sum w^i \frac{\partial}{\partial x^i}$, then $g_x(v,w) = \sum_{i,j=1}^n g_{ij}(x) v^i w^j$. A special case of this is the Euclidean metric $g_{ij}(x) = \delta_{ij}$.

The point of the above example is that it's not hard to construct Riemannian metrics locally.

For another example, consider an n-dimensional submanifold in $\mathbb{R}^{n+1}$: locally, $x^{n+1} = u(x^1, \dots, x^n)$ (i.e., it's the graph of a function). This is one of the (not too common) cases where there are natural coordinates on the manifold, $x^1, \dots, x^n$. The metric is induced from the Euclidean metric, and is unsurprisingly called a Euclidean metric: $p \mapsto g_p$ is the restriction of the dot product to T_pM. One can find a formula for this: take the coordinate curve $c_i(t) = (x^1, \dots, x^{i-1}, t, x^{i+1}, \dots, x^n)$, i.e. only varying the i^{th} term. Then, $\frac{\partial}{\partial x^i}$ is tangent to c_i at that point, but since $x^{n+1} = u(x^1, \dots, x^n)$, then $\frac{\partial}{\partial x^i} = (1, 0, 0, \dots, 0, \frac{\partial u}{\partial x^i})$, and so on. From these vectors, one can write down

$$g_{ij} = g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = \frac{\partial}{\partial x^i} \cdot \frac{\partial}{\partial x^j} = \delta_{ij} + \frac{\partial u}{\partial x^i}\frac{\partial u}{\partial x^j}.$$

One can see that this is a smooth, positive definite, symmetric matrix.

Definition. Given a smooth map $M \xrightarrow{F} \tilde{M}$ and a Riemannian metric $\tilde{g}$ on $\tilde{M}$, the pullback of $\tilde{g}$ by F is $F^*\tilde{g}$, a 2-tensor field given by $(F^*\tilde{g})|_p(v,w) = \tilde{g}_{F(p)}(F(v), F(w))$ for any $v, w \in T_pM$.

Notice that, though this is always a 2-tensor field, it is *not* generally the case that $F^*\tilde{g}$ is a Riemannian metric.

Definition. If $\dim M = \dim \tilde{M}$ and $F^*\tilde{g} = g$ for some Riemannian metric g of M, then F is called a local isometry. If in addition F is a diffeomorphism, then F is called a (global) isometry.

Isometries are the equivalence relation on Riemannian manifolds; if two manifolds are isometric, one can think of them as equivalent.

Example 4.1. Consider the torus $T^2 = \mathbb{R}^2/\mathbb{Z}^2$, i.e. $(x,y) \sim (x+m, y+n)$ for any $x, y \in \mathbb{R}$ and $m, n \in \mathbb{Z}$. The map $(x,y) \mapsto (x+m, y+n)$ is an isometry, so one is identifying points given by an isometry. This means that T^2 inherits a metric, called the flat metric (or the flat torus), because it's locally isometric to Euclidean space. There's more generally a family of geometrically distinct flat tori.

In summary, one can take a smooth manifold with symmetry and quotient by such a symmetry, yielding a quotient manifold with an inherited metric.

Example 4.2. Another example is the real projective plane $\mathbb{R}P^2$, which can be thought of in several ways, e.g. lines through the origin in $\mathbb{R}^3$. it might be easier to see it as $\mathbb{R}P^2 = S^2/\sim$, where $x \sim -x$ for $x \in S^2$. Since this is an isometry, then the real projective plane inherits a metric.

The geometry of a Riemannian manifold is coordinate-independent, so it's unaffected by a diffeomorphism. Thus, the basic problem of Riemannian geometry, which Riemann actually partially solved, is: how can one tell if a Riemannian metric g is locally isomorphic to the Euclidean metric? That is, do there exist coordinates $\tilde{x}^1, \dots, \tilde{x}^n$ such that $\tilde{g}_{ij} = \delta_{ij}$? Coordinates can be very complicated, so this isn't easy. Alternatively, one has some coordinate system $x^1, \dots, x^n$ and wants to end up with the $\tilde{x}^i$ above. Let $x^i = x^i(\tilde{x}^1, \dots, x^n)$; then,

$$\mathrm{d}x^p = \sum_{i=1}^n \frac{\partial x^p}{\partial \tilde{x}^i}\,\mathrm{d}\tilde{x}^i,$$

i.e. the coordinate basis for the tangent space changes by the Chain Rule, which is left as an exercise. Then, the metric itself transforms as

$$g = \sum_{i,j=1}^n \tilde{g}_{ij}(x)\,\mathrm{d}\tilde{x}^i \otimes \mathrm{d}\tilde{x}^j = \sum_{p,q,i,j=1}^n g_{pq}\frac{\partial x^p}{\partial \tilde{x}^i}\frac{\partial x^q}{\partial \tilde{x}^j}\,\mathrm{d}\tilde{x}^i \otimes \mathrm{d}\tilde{x}^j,$$

[7] The actual statement is ever so slightly different, since some of the values could be zero on some of the sets. But it still works.

so finding the coordinates is equivalent to solving the PDE

$$\tilde{g}_{ij} = \sum_{p,q} g_{pq}(x) \frac{\partial x^p}{\partial \tilde{x}^i} \frac{\partial x^q}{\partial \tilde{x}^j} = \delta_{ij}.$$

There are n unknowns and $n(n+1)/2$ equations, so this is overdetermined. Intuitively, this works only if g satisfies $n(n+1)/2 = \binom{n}{2}$ compatibility conditions, and in fact these compatibility conditions are exactly that the Riemann curvature tensor $R = 0$, but we'll get to that. The thing that makes this difficult is the diffeomorphism-invariance. Once again, it'll be necessary to express these compatibility conditions in a coordinate-free way. We can in fact solve this problem, thanks to lots of hindsight.

Riemann did this directly, by brute force. But between the metric and the curvature is an important object called the Riemannian connection. But first, vector fields.

Lie Bracket. Consider arbitrary vector fields X and Y. Everything will eventually be done coordinate-independently, but to motivate it, coordinates will be introduced, so suppose $X = \sum \xi^i \frac{\partial}{\partial x^i}$ and $Y = \sum \eta^j \frac{\partial}{\partial x^j}$. Then, these are just operators, so they can be composed:

$$XY(f) = X\left(\sum_{j=1}^n \eta^j \frac{\partial f}{\partial x^j}\right) = \sum_{i,j} \xi^i \frac{\partial \eta^j}{\partial x^i} \frac{\partial f}{\partial x^j} + \sum_{i,j} \xi^i \eta^j \frac{\partial^2 f}{\partial x^i \partial x^j}.$$

But $YX(f)$ has the same second-order term:

$$YX(f) = \sum \eta^j \frac{\partial \xi^i}{\partial x^j} \frac{\partial f}{\partial x^i} + \sum \xi^i \eta^j \frac{\partial^2 f}{\partial x^i \partial x^j},$$

so $XY - YX$, an *a priori* 2^{nd}-order operator, is actually first-order, and therefore a vector field! In coordinates,

$$(XY - YX)(f) = \sum_{i,j} \left(\xi^i \frac{\partial \eta^i}{\partial x^j} - \eta^j \frac{\partial \xi^i}{\partial x^i}\right) \frac{\partial f}{\partial x^i}.$$

This vector field is called the Lie bracket of X and Y, denoted $[X, Y]$.[8] It has several important properties, which are listed in the book and left as exercises.

1. It's anti-symmetric, so $[X, Y] = -[Y, X]$. This is probably the most obvious property.
2. It's $\mathbb{R}$-linear in each slot: $[a_1 X_1 + a_2 X_2, Y] = a_1[X_1, Y] + a_2[X_2, Y]$, and similarly for the Y slot, where $a_1, a_2 \in \mathbb{R}$. This is not true when the scalars are replaced with more general functions.
3. $[fX, hY] = fh[X, Y] + f(Xh)Y - h(Yf)X$.
4. Most importantly, the Jacobi identity $[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0$.
5. In some sense, the Lie bracket distinguishes general vector fields from coordinate vector fields: $\left[\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right] = 0$, because the second-order terms vanish; thus, these operators commute. This is not true for vector fields in general.

This is related to the idea that vector fields generate diffeomorphisms (e.g. as the velocity vector field of some fluid flow), and that a similar structure exists on Lie algebras and Lie groups.

The Riemannian Connection. If $p, q \in M$, then T_pM and T_qM are generally in no way related, no matter how close p and q are to each other. In Euclidean geometry, one isn't used to vectors having a fixed location like this, as there is a global parallelism that allows one to move them back and forth between tangent spaces.

In general, T_pM and T_qM are both n-dimensional real vector spaces, which means there are lots of isomorphisms between them, but none are canonical. The Riemannian connection will address this, and be very useful; it's what allows one to define the directional derivative of vectors in Euclidean space in terms of each other:

$$\nabla_X Y = \lim_{h \to 0} \frac{Y(p + hX(p)) - Y(p)}{h}.$$

Writing this in coordinates, one obtains the directional derivatives as the component functions. But this requires vectors to have no fixed location, which in general requires a connection.

Definition. An (affine) connection ∇ is a structure imposed on a manifold: given two vector fields X and Y on M, $\nabla_X Y$ is a vector field such that

(1) ∇ is $\mathbb{R}$-linear in X and Y: $\nabla_{a_1 X_1 + a_2 Y_2} Y = a_1 \nabla_{X_1} Y + a_2 \nabla_{X_2} Y$, and similarly $\nabla_X(b_1 Y_1 + b_2 Y_2) = b_1 \nabla_X Y_1 + b_2 \nabla_X Y_2$ for $a_1, a_2, b_1, b_2 \in \mathbb{R}$.
(2) ∇ is C^∞-linear in X; that is, $\nabla_{f_1 X_1 + f_2 X_2} Y = f_1 \nabla_{X_1} Y + f_2 \nabla_{X_2} Y$ for any smooth functions f_1 and f_2.

[8] It's also possible to define $[X, Y] = XY - YX$ in a coordinate-free way, though this makes showing it's a vector field a little more involved. In particular, one wants to pay attention to the Leibniz rule $[X, Y](fh) = h[X, Y](f) + f[X, Y](h)$, and linearity isn't that hard. In some sense, all that needs to be proven is that it's a derivation of a function at a point.

(3) Multiplying Y by a function is a little more complicated (since it intuitively gets differentiated): if f is a smooth function, then $\nabla_X(fY) = (Xf)Y + f\nabla_X Y$. This is another form of the Leibniz rule.

Notice that all of these properties hold for the standard directional derivative in $\mathbb{R}^n$. In general, there is always a connection on a manifold, and there are often very many. However, in the context of a Riemannian metric, there is a distinguished connection.

Theorem 4.1 (Fundamental Theorem of Riemannian Geometry)**.** *On a Riemannian (or Lorentz) manifold (M, g), there exists a unique affine connection ∇ that satisfies the following two properties:*

(1) ∇ is metric-compatible, i.e. if X, Y, and Z are vector fields, so that $\langle X, Y\rangle$ (i.e. $g(X, Y)$) is a smooth function, then $X(\langle X, Z\rangle) = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle$.
(2) ∇ is torsion-free (or symmetric), i.e. $\nabla_X Y - \nabla_Y X = [X, Y]$.

This connection is called the Riemannian connection.

The first condition should be thought of as yet another Leibniz rule, this time in terms of the metric. In Euclidean space, thinking of these as derivatives makes them more reasonable, though in this case $\nabla_X Y \sim XY$.

The proof of Theorem 4.1 will be given next lecture, albeit through a calculation rather than abstractly. One can choose local coordinates $x^1, \dots, x^n$, and therefore induce the basis $\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^n}$ of the tangent space. If ∇ is some affine connection, then it can be determined from families of functions Γ_{ij}^k, called Christoffel symbols or connection coefficients, given by

$$\nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} = \sum_k \Gamma_{ij}^k(x) \frac{\partial}{\partial x^k}. \tag{2}$$

If we know the Γ_{ij}^k, then we also know the connection by doing a calculation: if $X = \sum \xi^i \frac{\partial}{\partial x^i}$ and $Y = \sum \eta^j \frac{\partial}{\partial x^j}$, then

$$\nabla_X Y = \sum_{i,j} \xi^i \nabla_{\frac{\partial}{\partial x^j}} \left(\eta^j \frac{\partial}{\partial x^j} \right),$$

so knowing the Γ_{ij}^k shows how to calculate the connection in coordinates. These coefficients illustrate what the connection looks like locally. In the special case where the connection is given by the metric, there will exist an explicit formula in terms of the metric.

5. Existence and Uniqueness of the Riemannian Connection: 1/21/14

"In fact, I can use my magical eraser!"

The point of having an affine connection is to, given vector fields X and Y, understand $\nabla_X Y$, the directional derivative of Y (in the X-direction). However, just given the smooth structure, there isn't really enough information, because the tangent spaces aren't related. Connections aren't special, much like Riemannian metrics, and for a similar reason involving partitions of unity.

In $\mathbb{R}^n$ there is the much nicer case of

$$\nabla_X Y(p) = \lim_{h \to 0} \frac{Y(p + hX(p)) - Y(p)}{h},$$

but this cannot be done on arbitrary manifolds; specifically the minus sign.

Remark. Turning back to Theorem 4.1 and writing in local coordindates as discussed at the end of the last lecture, the torsion-free property of a Riemannian connection can be restated in terms of the Christoffel symbols as $\Gamma_{ij}^k = \Gamma_{ji}^k$ for all i, j, k, because the coordinate vector fields commute:

$$\nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} - \nabla_{\frac{\partial}{\partial x^j}} \frac{\partial}{\partial x^i} = \left[\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \right] = 0.$$

This is only true in coordinate bases, not all bases.

Intuition for metric compatibility, the other condition on a Riemannian connection, is that it's akin to the Leibniz rule. In $\mathbb{R}^n$, X and ∇_X are the same operator, but this is distinctly untrue elsewhere.

The proof given below will work for any scalar product, i.e. an assignment $p \mapsto \mathrm{Bil}(T_pM, \mathbb{R})$ whose component matrix is everywhere invertible. The positive-definiteness of a Riemannian metric, or the indefiniteness of the Lorentz metric, do not come into play.

Proof of Theorem 4.1. This proof will be constructive. First, as for uniqueness: choose 3 vector fields X, Y, and Z. The metric compatibility implies

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

Add the first two and subtract the third, but by the second property, $\nabla_Y X = \nabla_X Y + [X, Y]$, and so on, so that

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

Be careful with signs in this equation; the Lie bracket is anti-symmetric!

But now, it is possible to determine $\langle \nabla_X Y, Z\rangle$ as a formula in terms of things that are already known, i.e. the metric and the Lie bracket. That is, one has the following, known as the *Kozul formula*:

$$2\langle \nabla_X Y, Z\rangle = X\langle Y, Z\rangle + Y\langle Z, X\rangle - Z\langle X, Y\rangle - \langle X, [Y, Z]\rangle - \langle Y, [X, Z]\rangle - \langle Z, [Y, X]\rangle. \tag{3}$$

Now we have this formula, which implies uniqueness. Specifically, if we know the inner product of a vector field with every other vector field, then we can recover the original vector field.

In some special cases, some of the terms in (3) go away. For example, in local coordinates $x^1, \dots, x^n$, one also has induced coordinates $\frac{\partial}{\partial x^i}$. Then,

$$2\left\langle \nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j}, \frac{\partial}{\partial x^k}\right\rangle = g_{jk,i} + g_{ki,j} + g_{ij,k},$$

where the above uses the classical notation $g_{ij,k} = \frac{\partial g_{ij}}{\partial x^k}$. This is because $g_{ij} = \left\langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right\rangle$, and then one applies $Z = \frac{\partial}{\partial x^k}$. On the left-hand side, one can use the formula (2) for the Christoffel symbols to obtain

$$2\sum_p \Gamma_{ij}^p g_{pk} = g_{jk,i} + g_{ki,j} + g_{ij,k}.$$

Another notational thing: if (g_{ij}) is a matrix, then $(g_{ij})^{-1} = g^{ij}$ is understood to be the matrix inverse. Then,

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

This uses *another* notation called the summation convention, i.e. if the same index i appears as an upper index and a lower index in an expression, then the expression is treated as a sum, even if it isn't explicitly written. Thus, the above formula actually means

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

Sometimes it's easier to instead work in an orthonormal basis, so that the inner products drop out and it just deals with the brackets. This is more common in the study of Lie groups.

Now, with uniqueness settled, it still remains to show that (3) actually gives a Riemannian connection. First, why is it a connection?

(1) It's pretty obviously $\mathbb{R}$-linear in X and Y, because the right-hand side of (3) is.
(2) To check that $\nabla_{fX} Y = f\nabla_X Y$, take the difference of these terms. Then, $X\langle Y, Z\rangle$ cancels out, because it isn't changed since X isn't differentiated, and a couple other terms that don't differentiate X go away. The terms that do change come out to

$$2\langle \nabla_{fX} Y, Z\rangle - 2f\langle \nabla_X Y, Z\rangle = (Yf)\langle Z, X\rangle - (Zf)\langle X, Y\rangle + (Zf)\langle X, Y\rangle - (Yf)\langle X, Z\rangle,$$

which clearly goes to zero, because the inner product is symmetric.
(3) $\nabla_X(fY) = X(f)Y + f\nabla_X Y$. Essentially the same idea: after ignoring the terms where Y isn't differentiated in (3),

$$2\langle \nabla_X fY, Z\rangle - 2X(f)\langle Y, Z\rangle - 2f\langle \nabla_X Y, Z\rangle = (Zf)\langle X, Y\rangle - (Zf)\langle X, Y\rangle,$$

albeit after a little bit of thinking. It's somewhat hard to follow on the board, but you can plug and chug.

Then, the Riemannian properties follow from the symmetries of (3):

(4) For metric compatibility, we want that $X\langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle$. But when subtracting these two, this equation is symmetric in terms of Y and Z, so anything antisymmetric in Y and Z cancels out, such as $\langle X, [Y, Z]\rangle$, $\langle Y, [X, Z]\rangle - \langle Z, [X, Y]\rangle$, and $Y\langle Z, X\rangle - Z\langle X, Y\rangle$. Then, everything ends up working, which is no surprise since metric compatibility was used to generate (3).

(5) For the torsion-free property, all of the terms symmetric in X and Y drop out, including $X\langle Y,Z\rangle + Y\langle Z,X\rangle$, $Z\langle X,Y\rangle$, and $\langle X,[Y,Z]\rangle + \langle Y,[X,Z]\rangle$. Then,

$$2\langle \nabla_X Y, Z\rangle - 2\langle \nabla_Y X, Z\rangle = -2\langle Z,[Y,X]\rangle = 2\langle [X,Y],Z\rangle. \qquad \boxtimes$$

Again, the proof of existence shouldn't be too surprising, given that the formula was essentially derived from the properties that one wants to prove.

The connection allows for various geometrical notions. For example, given a Riemannian manifold (M,g) and a connection ∇, we have parallel transport, or parallel displacement. This is in general not as absolute as $\mathbb{R}^n$.

Given $p,q \in M$ and a curve $c(t) : [0,1] \to M$ with $c(0) = p$ and $c(1) = q$, a vector field along c is a smooth assignment $Y : (a,b) \to TM$ that projects back down to the curve, i.e. if $\pi : TM \to M$ is the natural projection $\pi(p,V) = p$, then $\pi \circ Y = c$. The smoothness condition means that in local coordinates, there exist smooth functions η^i such that

$$Y(t) = \sum_{i=1}^n \eta^i(t) \left.\frac{\partial}{\partial x^i}\right|_{c(t)}.$$

This is different from the notion of a vector field in general! If c is self-intersecting, so that $c(t_1) = c(t_2)$ for some distinct $t_1 \neq t_2$, then it's perfectly possible to have $Y(t_1) \neq Y(t_2)$, while this is not possible for vector fields in general. Thus, this is a broader notion than the restriction of a vector field to a curve.

Then, using the connection, we can differentiate Y: let $\dot{c} = \sum \dot{x}^j \frac{\partial}{\partial x^j}$. Then,

$$\nabla_{\dot{c}(t)} Y = \sum_{i=1}^n \left(\frac{\mathrm{d}\eta^i}{\mathrm{d}t} \frac{\partial}{\partial x^i} + \eta^i \nabla_{\dot{c}} \frac{\partial}{\partial x^i} \right).$$

The connection can be used to define the covariant derivative along a curve. Then,

$$= \sum_i \frac{\mathrm{d}\eta^i}{\mathrm{d}t} \frac{\partial}{\partial x^i} + \sum_{i,j,k} \Gamma_{ij}^k(c(t)) \dot{x}^j \eta^i \frac{\partial}{\partial x^k}.$$

Rewriting this, it becomes

$$\nabla_{\dot{c}(t)} Y = \left(\frac{\mathrm{d}\eta^i}{\mathrm{d}t} + \Gamma_{kj}^i(c(t)) \dot{x}^j \eta^k \right) \frac{\partial}{\partial x^i}(c(t)).$$

From this calculation, we see that given an initial vector $Y_0 \in T_pM$ this is just an ODE, so there is a unique vector field $Y(t)$ such that $\nabla_{\dot{c}} Y(t) = 0$ for all t and $Y(0) = Y_0$. In fact, since it's a linear, first-order system with smooth coefficients, the solution is global; it exists for all time. The Γ_{ij}^k illustrate that it depends on the connection, and the $\dot{x}^j$ that it depends on the curve.

In essence, one can start with a vector Y_0 at p and uniquely drag it along the curve to $Y(1) \in T_qM$. Thus, one can work around the notion that tangent vectors are unrelated in different tangent spaces.

The specific ODE that one must solve is

$$\frac{\mathrm{d}\eta^i}{\mathrm{d}t} + \Gamma_{jk}^i(c(t)) \dot{x}^j \eta^k(t) = 0 \tag{4}$$

for $i = 1, \ldots, n$, and subject to the initial conditions that $\eta^i(0) = \eta_0^i$ and $Y_0 = \sum \eta_0^i \left.\frac{\partial}{\partial x^i}\right|_p$.[9]

Now, there is a parallel transport map $P_c : T_pM \to T_qM$. It ends up being a linear map, and since we have the inverse map given by reversing the orientation of the curve, it's an isomorphism. Also, because ∇ has metric compatibility, then P_c also preservers the metric: if Y_1 and Y_2 are both parallel, then

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle Y_1(t), Y_2(t)\rangle = \langle \nabla_{\dot{c}} Y_1, Y_2\rangle + \langle Y_1, \nabla_{\dot{c}} Y_2\rangle = 0.$$

Thus, P_c is a linear isometry between tangent spaces!

In $\mathbb{R}^n$, parallel transport is path-independent, and in particular, if c is a loop, then $P_c = \mathrm{id}$. This is distinctly not true on general manifolds, but more on that next lecture.

Connections can also be used to define geodesics.

Definition. A (parameterized) geodesic is a curve c for which $\dot{c}$ is parallel along c, i.e. $\nabla_{\dot{c}} \dot{c} = 0$.

[9] Theoretically, one would have to deal with coordinate changes, but it happens to be true that every curve can be covered with a single chart.

One can think of this as having a curve with zero acceleration. It makes the ODE (4) into a second-order equation called the geodesic equation:

$$\frac{\mathrm{d}^2x^i}{\mathrm{d}t^2} + \Gamma_{ij}^k(c(t))\frac{\mathrm{d}x^i}{\mathrm{d}t}\frac{\mathrm{d}x^j}{\mathrm{d}t} = 0.$$

With two initial conditions (since it's second-order), one can solve this. The values of $c(0)$ and $\dot{c}(0)$ will do the trick. However, since Γ_{ij}^k is nonlinear, then the geodesic equation isn't necessarily solvable for all time — solutions can blow up.

6. Tensor Fields, Parallel Transport, and Holonomy: 1/23/14

Though we already saw the directional derivative as an example of a Riemannian connection in $\mathbb{R}^n$, this can be generalized somewhat to provide another example. Let M be an n-dimensional submanifold of $\mathbb{R}^k$, so that $n < k$. Then, D_XY still makes sense, since one only needs to know Y along a curve tangent to X, so take the tangential component. That is, $D_XY = (D_XY)^\top + (D_XY)^\perp$, where the first term is the tangential component and the second the normal part. This is possible because the tangent space is a vector space.

Then, the Riemannian connection is $\nabla_XY = (D_XY)^\top$. That ∇ is in fact a Riemannian connection is easy to check, though it helps to know that the Lie bracket of two tangent vector fields is also a tangent vector field. This is another nice, concrete example of the Riemannian connection; one can easily calculate its Christoffel symbols.

Recall the notion of parallel transport from last lecture; though there it was called P_c, today it is P^c. Since this isn't an index in the sense of the summation convention, it makes no difference. The map itself was given by a solution to an ODE; though the theory guarantees that such a solution exists, it's in practice rather hard to explicitly solve and write down.

P^c has some nice properties; to wit, it's a linear isometry $T_pM \to T_qM$. Linearity follows from the fact that the equations to solve are linear, so plugging in the solution shows that it must respect linearity.[10] It's an isometry because if $Y(t)$ and $Z(t)$ are parallel along c, then

$$c\langle Y, Z\rangle = \frac{\mathrm{d}}{\mathrm{d}t}\langle Y(t), Z(t)\rangle = \langle \nabla_{\dot{c}}Y, Z\rangle + \langle Y, \nabla_{\dot{c}}Z\rangle = 0,$$

so $\langle P^c(Y(q)), P^c(Z(q))\rangle = \langle Y_0, Z_0\rangle$.

Holonomy. This is particularly nice if c is a smooth closed curve, since then it's an automorphism of a vector space (i.e. $c(0) = c(1) = p$), and is also an isometry. This curve has an orientation given by increasing values of t, and it's possible to define a curve $-c$ with the opposite orientation (i.e. $(-c)(t) = c(1-t)$). Then, $P^{-c} = (P^c)^{-1}$. This is because when one looks at the ODE, reversing the sign reverses the solution, and thus $P^{-c}(P^c(Y_0)) = Y_0$.

Furthermore, if c_1 and c_2 are closed curves, let c_2c_1 be the curve given by tracing c_1 and then c_2.[11] Then, $P^{c_2c_1} = P^{c_2} \circ P^{c_1}$. c_2c_1 isn't necessarily smooth at p, but it's smooth everywhere else, which is OK, because it's possible to define parallel transport on piecewise smooth curves and so this works out.

What this means is that the set of parallel-transport isometries under closed curves is a group, called the holonomy group. This is a subgroup (often a strict subgroup) of the orthogonal group of all linear isometries of T_pM.

Example 6.1. The easiest example is $\mathbb{R}^n$ with the standard connection D. Then, for any $p \in \mathbb{R}^n$ and closed curve c through P, $P^c = \mathrm{id}$, because $D_{\dot{c}}Y = 0$ iff $\frac{\mathrm{d}}{\mathrm{d}t}Y^i = 0$, and therefore $Y^i(0) = Y^i(1)$. In this case, there's this notion of global parallelism.

Remark. If one has a metric (U, g) that's isometric to an open set of $\mathbb{R}^n$ with the Euclidean metric (i.e. pulled back with a diffeomorphism, which could look very complicated), then $P^c = \mathrm{id}$ for all loops $c \subset U$ at p. This is an interesting way of showing that a given metric is *not* isometric to the Euclidean one.[12]

Example 6.2. Consider the sphere S^2 and let p be the north pole. Then, let c be a curve that goes down from p to the equation along a great circle, along the equator for some angle θ, and then back up along a great-circle path to p. This ends up being a rotation: let X_0 be the unit tangent vector of c. Then, since the great circles are geodesics, parallel transport of X_0 along the first sector of c makes it still the unit tangent vector, so at the equator, it still looks like $(0, 0, -1)$. Along the equator, it doesn't change at all, so that when it begins the third sector it's rotated from the unit tangent vector and ends up back at p rotated by θ. In particular, the holonomy group is nontrivial, so the sphere isn't isometric to $\mathbb{R}^2$. This is why there are no perfect maps of the sphere on paper: there must always be some distortion.[13]

---

[10] This also depends on the fact that the solution to the initial value problem is unique.

[11] This is technically parameterized on $[0, 2]$ rather than $[0, 1]$, but this is easy to fix.

[12] It turns out this implication goes both ways, but that's a story for later.

[13] The fact that this is piecewise smooth might not appeal to everyone, but this can be made into a smooth curve by slowing down and speeding up the acceleration (or suitably scaling the parameter) in order to smooth it out.

Tensor Fields. The next topic will be the curvature tensor, but first one needs to discuss tensor analysis more formally.

Definition. Let M be an n-dimensional manifold and $r, s \in \mathbb{N}$.

- A $(0,s)$-tensor field (an s-covariant tensor field) A is a smooth assignment
$$p \in M \longmapsto A_p : \underbrace{T_pM \times \cdots \times T_pM}_{s \text{ times}} \to \mathbb{R},$$
where A_p is required to be s-linear; that is, linear in each slot if the others are fixed.
If one has coordinates $x^1, \dots, x^n$ near p, this is written
$$A = \sum_{i_1,\dots,i_s=1}^{n} A_{i_1 i_2 \cdots i_s}(x)\,\mathrm{d}x^{i_1} \otimes \cdots \otimes \mathrm{d}x^{i_s},$$
where the $A_{i_1 \cdots i_s}$ are required to be smooth and the basis elements are the functions
$$\mathrm{d}x^{i_1} \otimes \cdots \mathrm{d}x^{i_s} : (v_1, \dots, v_s) \longmapsto \mathrm{d}x^{i_1}(v_1)\,\mathrm{d}x^{i_2}(v_2) \cdots \mathrm{d}x^{i_s}(v_s).$$
Each of these is clearly s-linear, and they form a basis on the n^s-dimensional subspace of s-linear functions (adding symmetries gives interesting, lower-dimensional subspaces).
For a few special cases, if $s = 1$, one has a vector field, and it's convenient to think of a smooth function as a $(0,0)$-tensor, in that it associates each $p \in M$ with a real number.
- An $(r,0)$-tensor field (r-contravariant) can be quickly defined because a finite-dimensional real vector space is self-dual, so $T_pM = (T_p^*M)^*$. This sounds silly, but it turns out to work really well. Let $v : T_p^*M \to \mathbb{R}$ be defined as $v(\omega) = \omega(v)$ (so that now, we vary ω instead of v), so $v \in (T_p^*M)^*$, allowing the canonical identification with the double-dual.
Then, an $(r,0)$-tensor field is a smooth assignment
$$p \in M \longmapsto B_p : \underbrace{T_p^*M \times \cdots \times T_p^*M}_{r \text{ times}} \to \mathbb{R}$$
that is required to be r-linear. The smoothness condition once again means that in local coordinates, it's given by smooth functions
$$B = \sum_{j_1,\dots,j_r=1}^{n} B^{j_1 j_2 \cdots j_r} \frac{\partial}{\partial x^{j_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{j_r}}.$$
Here, $\frac{\partial}{\partial x^i}$ is the i^{th} component of the standard basis for T_pM.
- In general, a tensor field of type (r,s), also known as an (r,s)-tensor field, is a smooth assignment
$$p \longmapsto A_p : \underbrace{T_p^*M \times \cdots \times T_p^*M}_{r \text{ times}} \times \underbrace{T_pM \times \cdots \times T_pM}_{s \text{ times}} \to \mathbb{R},$$
where the smoothness conditions are the same as before.

In geometry, one often makes big tensor fields, so having some properties and examples will be useful. We have already been introduced to $(0,0)$-tensors (smooth functions), $(1,0)$-tensors (vector fields), and $(0,1)$-tensors (one-forms). Riemannian metrics are examples of $(0,2)$-tensors (though there are the additional conditions of symmetry and positive definiteness).

For another example, consider an assignment $p \mapsto A_p : T_pM \to T_pM$ that is smooth in the above sense. This is a $(1,1)$-tensor, because if $v \in T_pM$, then one can define $A_p(v, \omega) = \omega(A_p(v))$. The coefficients in terms of a local chart are
$$A = \sum_{i,j=1}^{n} a^i_j \frac{\partial}{\partial x^i} \otimes \mathrm{d}x^j,$$
where a^i_j is the matrix of A with respect to the coordinate basis. This is a fairly common construction in geometry. In this sense, tensor fields generalize matrix theory: higher-order tensors look like matrices in more than two dimensions.

As an example of this, if ∇ is a connection and Y is a vector field, then $A_p(X) = \nabla_X Y(p)$ is a $(1,1)$-tensor.

There are various algebraic operations one can use to send tensors of one type to tensors of another type, such as the trace or the tensor product.

"When I was an undergraduate, I thought all functions had converging power series."

On this week's homework, it will be useful to know that if (M_1, g_1) and (M_2, g_2) are manifolds and one has an isometry $F^*(f_2) = g_1$, then if $c(T)$ is a geodesic in M_1, then $F \circ c(t)$ is a geodesic in M_2, because ∇ and everything else necessary is preserved.

Recall that we had an assignment $p \mapsto L_p : T_pM \to T_pM$ linear and smooth in coordinates. This is a $(1,1)$-tensor field $A \in \mathcal{T}_{1,1}$: if $v \in T_pM$ and $\omega \in T_p^*M$, then $A_p(v, \omega) = \omega(L_p(v)) \in \mathbb{R}$. It is also possible to describe A in coordinates, as was done last lecture.

Soon we will construct the most important tensor in this course, but first we will see how to recognize tensors. Recall more generally that $A \in \mathcal{T}_{r,s}$ is an assignment

$$p \mapsto A_p : \underbrace{T_p^*M \times \cdots \times T_p^*M}_{r \text{ times}} \times \underbrace{T_pM \times \cdots \times T_pM}_{s \text{ times}} \to \mathbb{R}.$$

This can also be thought of as a multilinear operation on vector fields and one-forms: if $\mathcal{X}$ is the space of vector fields o M and $\mathcal{E}$ the space of (smooth) one-forms, then one has

$$A : \underbrace{\mathcal{E} \times \cdots \times \mathcal{E}}_{r \text{ times}} \times \underbrace{\mathcal{X} \times \cdots \times \mathcal{X}}_{s \text{ times}} \to C^\infty(M),$$

i.e. to the space of C^∞ functions $M \to \mathbb{R}$. This is given by

$$A(\omega_1, \ldots, \omega_r, X_1, \ldots, X_s)(p) = A_p(\omega_1(p), \cdots, \omega_r(p), X_1(p), \ldots, X_s(p)).$$

Moreover, this operation is C^∞-linear in each slot: it's obviously $\mathbb{R}$-linear, but if one takes, for example,

$$A(\omega_1, \ldots, \omega_{i-1}, f \cdot \omega_i, \omega_{i+1}, \ldots, \omega_r, X_1, \ldots, X_s) = f \cdot A(\omega_1, \ldots, \omega_r, X_1, \ldots, X_s).$$

This is because at a point p, $(f \cdot \omega_i)(p) = f(p) \cdot \omega_i(p)$, but then all that one has to pull out is the real number $f(p)$. Recall that $\nabla_X Y$ is also C^∞-linear in the X slot; this linearity is sometimes called *tensorial.* This property characterizes tensor fields, and in fact one often constructs tensor fields by cooking up multilinear functions on $\mathcal{E}$ and $\mathcal{X}$ and showing that they're C^∞-linear.

Lemma 7.1. *If $A : \mathcal{E} \times \cdots \times \mathcal{E} \times \mathcal{X} \times \cdots \times \mathcal{X} \to C^\infty(M)$ is $\mathbb{R}$-multilinear, then A defines a tensor field iff A is C^∞-linear in each slot.*

Example 7.1. Recall that the Christoffel symbols looked like $(1,2)$-tensors, so if ∇ is the Riemannian connection, is $A(\omega, X, Y) = \omega \nabla_X Y$ a $(1,2)$-tensor? It's certainly $\mathbb{R}$-linear, and it's C^∞-linear in ω and X, but not in Y: $A(\omega, X, fY) = \omega(\nabla_X fY) \neq f\omega(\nabla_X Y)$ in general.

The tensor is an algebraic operator: its value only depends on the values of other things at that point. But the connection is a differential operator: it also depends on the local behavior of Y. In operator terminology, the tensor is a 0^{th}-order operator.

Proof of Lemma 7.1. The forward direction was just shown above, so the content is in the converse. Suppose A is C^∞-linear in each slot. Then, we need to define A_p pointwise, so pick $\omega_1, \ldots, \omega_r \in T_p^*M$ and $X_1, \ldots, X_s \in T_pM$. Then, define

$$A_p(\omega_1, \ldots, \omega_r, X_1, \ldots, X_s) = A(\widehat{\omega}_1, \ldots, \widehat{\omega}_r, \widehat{X}_1, \ldots, \widehat{X}_s),$$

where $\widehat{\omega}_i \in \mathcal{E}$ and $\widehat{X}_i \in \mathcal{X}$ are smooth extensions, i.e. $\widehat{\omega}_i(p) = \omega_i$ and $\widehat{X}_i(p) = X_i$. Now, all that remains is to show this is well-defined independently of the choice of the extension. Let $\widetilde{\omega}_i$ and $\widetilde{X}_i$ be some other extensions that agree at p. Thus, since A is $\mathbb{R}$-multilinear, it suffices to show that $A(\omega_1, \ldots, \omega_r, X_1, \ldots, X_s)(p) = 0$ if any ω_i or $X_j = 0$ at p. But then, assuming that (which will be proven later), there is always a way to break up the difference of any two extensions into terms which vanish at p:

$$\begin{aligned} A(\widehat{\omega}_1, \ldots, \widehat{\omega}_r - \widehat{X}_1, \ldots, \widehat{X}_s) - A(\widetilde{\omega}_1, \ldots, \widetilde{\omega}_r, \widetilde{X}_1, \ldots, \widetilde{X}_s) &= A(\widehat{\omega}_1 - \widetilde{\omega}_1, \widehat{\omega}_2, \ldots, \widehat{\omega}_r, \widehat{X}_1, \ldots, \widehat{X}_s) \\ &\quad + A(\widehat{\omega}_1, \widetilde{\omega}_2, \ldots, \widetilde{\omega}_r, \widetilde{X}_1, \ldots, \widetilde{X}_s) \\ &\quad - A(\widetilde{\omega}_1, \widetilde{\omega}_2, \ldots, \widetilde{\omega}_r, \widetilde{X}_1, \ldots, \widetilde{X}_s). \end{aligned}$$

Then, this can be repeated in the second slot, to carry the difference further down, and thus eventually terminate it. It makes much more sense in small tensors as an example:

$$\begin{aligned} A(\widehat{X}_1, \widehat{X}_2) - A(\widetilde{X}_1, , \widetilde{X}_2) &= A(\widehat{X}_1 - \widetilde{X}_1, \widehat{X}_2) - A(\widetilde{X}_1, \widehat{X}_2) - A(\widetilde{X}_1, \widetilde{X}_2) \\ &= A(\widehat{X}_1 - \widetilde{X}_1, \widehat{X}_2) - A(\widetilde{X}_1, \widehat{X}_2 - \widetilde{X}_2). \end{aligned}$$

In short, there's an inductive way to break it up so that each term goes to zero at p. This only uses the $\mathbb{R}$-linearity that was already known of A.

Now, it's reduced to the case where $\omega_i(p) = 0$, and the goal is to show $A(\omega_1, \dots, \omega_r, X_1, \dots, X_s) = 0$. (The same idea applies for any other slot, even the contravariant ones). Work in local coordinates $x^1, \dots, x^n$ near p given by a chart U. Then, choose some cutoff function $\zeta \in C^\infty(M)$ and a V open in U such that $\overline{V} \subset U$, and $\zeta = 1$ on V but $\zeta = 0$ near ∂U. Then, by C^∞-linearity,

$$\begin{aligned}A(\omega_1, \dots, \omega_r, X_1, \dots, X_s)(p) - A(\zeta\omega_1, \omega_2, \dots, \omega_r, X_1, \dots, X_s)(p) &= A((1-\zeta)\omega_1, \omega_2, \dots, \omega_r, X_1, \dots, X_s)(p) \\ &= (1-\zeta(p))A(\omega_1, \dots, \omega_r, X_1, \dots, X_s) = 0.\end{aligned}$$

Great, so now we've reduced to showing that $A(\zeta\omega_1, \omega_2, \dots, \omega_r, X_1, \dots, X_s) = 0$. Well,

$$\zeta\omega_1 = \sum_{i=1}^n \zeta a_i \, \mathrm{d}x^i$$

in U, where the a_i are globally defined functions that are 0 outside of U. Assume for simplicity that $x(p)$ is the original; then, this forces $a_i(0) = 0$. Therefore, again by C^∞-linearity,

$$\begin{aligned}A(\zeta\omega_1, \omega_2, \dots, \omega_r, X_1, \dots, X_s) &= A\left(\sum_{i=1}^m a_i\zeta \, \mathrm{d}x^i, \omega_2, \dots, \omega_r, X_1, \dots, X_s\right) \\ &= \sum_{i=1}^n a_i(0) A(\zeta \, \mathrm{d}x^i, \omega_2, \dots, \omega_r, X_1, \dots, X_s) \\ &= 0.\end{aligned}$$

⊠

Cutoff functions were originally counterexamples to the notion that C^∞ functions must have compact support. For example, on $\mathbb{R}$, one could take

$$f(t) = \begin{cases} 0, & t \le 0 \\ e^{-1/t^2}, & t > 0. \end{cases}$$

There are lots of such examples, but they tend to be complicated to write down. The trick is necessary so that one can do this everywhere (i.e. make $\omega_1 = 0$ outside of U).

The term C^∞-linear in this discussion appears because one can perceive M as a $C^\infty(M)$-module.

The Riemann Curvature Tensor. Now, for the most interesting tensor in this course. Let M be a manifold and ∇ be an affine (so, not necessarily Riemannian) connection. Then, the curvature tensor of ∇ is the $(1,3)$-tensor field

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

Behind the scenes we have $A(\omega, X, Y, Z) = \omega(R(X,Y)Z) \in C^\infty(M)$, but usually we'll use the first formula.

The curious notation is to reflect that $R(X,Y)-$ is a linear transformation of Z.

Proposition 7.2. *R is in fact a $(1,3)$-tensor field.*

Proof. The proof will show that R is C^∞-multilinear, and then use the previous lemma. R is pretty clearly $\mathbb{R}$-multilinear because ∇ is. As $A(\omega, \dots)$, it's clearly C^∞-linear, because the function is already on the outside. Thus, it's necessary to check this in terms of X, Y, and Z. But since R is anti-symmetric in X and Y, C^∞-linearity in one implies it in the other, so there are in fact only two things to check.

On to the calculation. The only interesting terms in the first calculation will be those where fX is differentiated, i.e. the Lie bracket.

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

Thus, this is tensorial in the X (and therefore Y) slots. Z is a little harder, since it involves two derivatives of f.

$$\begin{aligned}R(X,Y)(FZ) &= fR(X,Y)Z + (Yf)\nabla_X Z + (XYf)Z + (Xf)\nabla_Y Z \\ &\quad - (Xf)\nabla_Y Z - (YXf)Z - (Yf)\nabla_X Z - ([X,Y]f)Z \\ &= (XYf)Z - (YXf)Z - ([X,Y]f)Z = 0,\end{aligned}$$

so this is also tensorial in Z. ⊠

R is really a $(1,3)$-tensor, an algebraic object even as it was defined as a differential-like object. This is kind of magical. Its meaning will be mentioned today, proven in the next lecture.

Definition. let M be a manifold and ∇ a connection.

- ∇ is locally trivial if for all $p \in M$ there exists a local basis $E_1, \dots, E_n$ of vector fields in a neighborhood of p that are parallel, i.e. $\nabla_X E_i = 0$ for all i.
- The induced curvature tensor R is called flat if $R = 0$ everywhere.

Then, we will show the following theorem. It is a local theorem, as it speaks about things happening in neighborhoods of a point.

Theorem 7.3. *A connection is flat iff it is locally trivial.*

This tells us what the curvature actually is, i.e. an obstruction to parallelism. The Riemannian connection has special properties (i.e. locally trivial implies locally Euclidean).

Theorem 7.4. *If g is a Riemannian (or Lorentz) metric and ∇ is metric-compatible, then if $R = 0$, then one can take $E_1, \dots, E_n$ to be an orthonormal basis.*

In particular, if g is a Riemannian metric and ∇ its connection, then $g(E_i, E_j) = \delta_{ij}$ if $R = 0$. The final step will be making these locally Euclidean — which is precisely the torsion-free condition.

8. Flatness: 1/30/14

Recall that the Riemannian curvature tensor is a $(1,3)$-tensor field defined as $R(X,Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z$. In local coordinates, one writes

$$R\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right)\frac{\partial}{\partial x^k} = R^\ell_{kij}\frac{\partial}{\partial x^\ell}.$$

(Remember the summation convention!)

Proposition 8.1. *∇ is locally trivial iff $R = 0$.*

Definition. A connection is trivial if there exists a trivialization, i.e. a basis of vector fields $E_1, \dots, E_n$ such that $\nabla_{E_i} E_j = 0$ for all i and j.

Proof of Proposition 8.1. The forward direction is easy, because in that basis one can just calculate R and see that it is zero.

In the reverse direction, suppose $R = 0$; then, choose coordinates $x^1, \dots, x^n$ centered at p (i.e. $x^i(p) = 0$). Work in the cube $C = \{x : -a < x^i < a, i = 1, \dots, n\}$, so that we have the x^1-axis $\{x \in C : x = (x^1, 0, \dots, 0)\}$. Let $e_1, \dots, e_n$ be a basis for T_pM, and let $E_1, \dots, E_n$ be the vector fields obtained from $e_1, \dots, e_n$, respectively, by parallel displacement along the x^1-axis; thus, on the x^1-axis, $\nabla_{\frac{\partial}{\partial x^1}} E_i = 0$.

Now, extend one dimension higher, to the x^1x^2-plane. Extend E_i to $(x^1, x^2, 0 \dots, 0)$ by parallel transport along the curve $c(t) = (x^1, t, 0, \dots, 0)$. Then, the E_i are still a basis, because parallel transport is invertible, and now $\nabla_{\frac{\partial}{\partial x^2}} E_i = 0$ and $\nabla_{\frac{\partial}{\partial x^1}} = 0$ when $x^2 = 0$. But using the flatness of R, this can be extended to the entire plane: $\nabla_{\frac{\partial}{\partial x^1}} E_i = 0$ on the entire x^1x^2-plane. Since $R = 0$ and the Lie bracket of two coordinate vector fields is 0, then

$$\nabla_{\frac{\partial}{\partial x^2}} \nabla_{\frac{\partial}{\partial x^1}} E_i = 0,$$

and therefore $\nabla_{\frac{\partial}{\partial x^1}} E_i$ is parallel in the x^2-direction. But since it was zero on the x^1-axis, then it must stay zero on the entire plane, because the parallel transport of the zero vector is always the zero vector.

This gives an inductive process for finding a parallel basis on the whole cube. Assume $E_1, \dots, E_n$ are parallel in the $x^1 \dots x^p$-plane for some $p < n$. Then, one can extend $E_1, \dots, E_n$ to a parallel basis in the $x^1 \cdots x^{p+1}$-plane: extend the E_i by parallel transport along $c(t) = (x^1, \dots, x^p, 0, \dots, 0)$, so that $\nabla_{\frac{\partial}{\partial x^{p+1}}} E_i = 0$ on the whole $x^1 \cdots x^{p+1}$-plane. Then, using flatness, for any $j = 1, \dots, p+1$,

$$\nabla_{\frac{\partial}{\partial x^j}} \nabla_{\frac{\partial}{\partial x^{p+1}}} E_i = 0, \text{ so } \nabla_{\frac{\partial}{\partial x^{p+1}}}\left(\nabla_{\frac{\partial}{\partial x^j}} E_i\right) = 0, \tag{5}$$

and thus it's parallel in the x^{p+1}-direction, so the parallel transport preserves the zero vector. ⊠

The key to the proof is that in (5), you can switch the order of the derivatives because the space is flat. In some sense, the curvature measures the failure of the covariant derivatives to commute. Notice also that this proof works for any scalar product and induced connection, e.g. Lorentz metrics.

Theorem 8.2. *A Riemannian manifold (M, g) is locally isometric to $\mathbb{R}^n$ iff $R = 0$, where R is the curvature tensor given by the Riemannian connection ∇.*

(M, g) is locally isometric to $\mathbb{R}^n$ means that for all $p \in M$, there is a chart (φ, U) and induced local coordinates $x^1, \dots, x^n$ such that $\varphi : (U, g) \to (\varphi(U), \delta_{ij})$ is an isometry, or $g_{ij} = \delta_{ij}$ in these local coordinates.

Proof of Theorem 8.2. The proof will almost entirely follow from Proposition 8.1. Again, it's clear that if (M, g) is locally isometric to $\mathbb{R}^n$, then the cuvature tensor is flat.

In the other direction, suppose $R = 0$. Then, for any $p \in M$, choose an orthonormal basis $e_1, \dots, e_n$ for T_pM; thus, there exists a locally parallel frame (i.e. a trivialization of the connection) $E_1, \dots, E_n$ in U such that $\nabla_{E_i} E_j = 0$ and $E_i(p) = p_i$. Then, these E_i are still orthonormal: for any vector field X, by metric compatibility,

$$X\langle E_i, E_j\rangle = \langle \nabla_X E_i, E_j\rangle + \langle E_i, \nabla_X E_j\rangle = 0.$$

We haven't used the torsion-free condition, and in fact the theorem is false without it. This tells us that the vector fields commute:

$$[E_i, E_j] = \nabla_{E_i} E_j - \nabla_{E_j} E_i = 0.$$

As a consequence of the rectification theorem from the homework, if one takes k commuting vector fields which are independent at all points, then there exists a set of local coordinates $x^1, \dots, x^n$ such that $E_i = \frac{\partial}{\partial x^i}$ for $i = 1, \dots, k$. But since the E_i produced above satisfy these conditions (where $k = n$), then they become coordinate vector fields. But they're still orthonormal, so in U,

$$g_{ij} = \left\langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \right\rangle = \delta_{ij}.$$

The fact above (the generalized rectification theorem) deserves at least a proof sketch. The basic lemma is that if one takes two smooth vector fields X and Y, then $[X, Y] = 0$ iff their local flows commute: if X gives local flow ψ_t and Y gives the local flow φ_s, then $\varphi_s \circ \psi_t = \psi_t \circ \varphi_s$. Then, one can define a system of coordinates $\tilde{x}^1, \dots, \tilde{x}^n$ such that $\tilde{x}^i$ is in the direction of the flow of E_i, and since they commute, then one can do this everywhere.

Now, why is the lemma true? The flow and vector field uniquely determine each other, so given some flow φ_s, one can conjugate it, and get a new flow $\psi_{-t} \circ \varphi_s \circ \psi_t$ given by s, so that $s_1 + s_2 \mapsto \psi_{-t} \circ \varphi_{s_1} \circ \varphi_{s_2} \circ \psi_t = (\psi_{-t} \circ \varphi_{s_1} \circ \psi_t) \circ (\psi_{-t} \circ \varphi_{s_2} \circ \psi_t)$. Let Z be the vector field with this flow; then, $Z(p) = D\psi_{-t}(Y_{\psi_t(p)})$. Then, the goal is to show that $Z = Y$; assume X is nonzero (so that this problem is interesting), so that one can rectify and choose coordinates in which $X = \frac{\partial}{\partial x^1}$, and $\psi_t(x^1, \dots, x^n) = (x^1 + t, x^2, \dots, x^n)$. Then, one can explicitly calculate Z in these coordinates, and get Y. ⊠

Structure of the Riemann Curvature Tensor. The way the curvature tensor is set up gives it an interesting algebraic structure. Though as defined it is a $(1, 3)$-tensor, the metric it induces can make it into a $(0, 4)$-tensor, defined as follows for four vector fields X, Y, Z, and V:

$$(X, Y, Z, V) \mapsto \langle R(X, Y)Z, V\rangle.$$

In some sense, one uses the metric to lower one of the indices. This can be nicer, because each index is on equal footing, so to speak.

Here are some zero-order symmetries, i.e. properties of the tensor that are algebraic in nature.

(1) Like any curvature tensor, R is anti-symmetric in X and Y: $R(X, Y)Z = -R(Y, X)Z$.
(2) The first Bianchi identity: $R(X, Y)Z + R(Y, Z)X + R(Z, X)Y = 0$. Not all curvature tensors of all connections satisfy this, but the Riemann curvature tensor does.
(3) The $(0, 4)$-tensor is anti-symmetric in Z and V: $\langle R(X, Y)Z, V\rangle = -\langle R(X, Y)V, Z\rangle$. This is not obvious, and follows from metric compatibility.
(4) $\langle R(X, Y)Z, V\rangle = \langle R(Z, V)X, Y\rangle$.

These symmetries make it easier to see the geometric structure of the tensor: it's a huge object, with four indices, so these make life easier to understand. For example, in general relativity, Einstein's equations use the Ricci curvature, which is derived from these symmetries. Similarly, one might use these in differential geometry to understand sectional curvature on two-dimensional slices of a manifold.

These properties aren't too hard to check: (1) is trivial. For (2), add the following together:

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

Since ∇ is torsion-free, $\nabla_Y Z - \nabla_Z Y = [Y, Z]$, so the sum becomes

$$\nabla_X [Y, Z] + \nabla_Y [Z, X] + \nabla_Z [X, Y] - \nabla_{[X,Y]} Z - \nabla_{[Y,Z]} X - \nabla_{[Z,X]} Y,$$

which using metric compatibility again, becomes $[X,[Y,Z]]+[Y,[Z,X]]+[Z,[X,Y]]$, which is zero by the Jacobi identity.

For part (3), one can use the polarization trick, that showing a bilinear form is antisymmetric is equivalent to showing it's 0 on the diagonal, because then one can use linearity to get the full anti-symmetry. Thus, the goal is to show that for any X, Y, and Z, $\langle R(X,Y)Z,Z\rangle = 0$. Then,

$$\begin{aligned} XY\langle Z,Z\rangle &= X(2\langle\nabla_Y Z,Z\rangle) = 2\langle\nabla_X\nabla_Y Z,Z\rangle + 2\langle\nabla_Y Z,\nabla_X Z\rangle. \\ \implies [X,Y]\langle Z,Z\rangle &= 2\langle\nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z, Z\rangle, \end{aligned}$$

where the second follows from the symmetries in the first. Then, since this is also equal to $2\big\langle\nabla_{[X,Y]}Z,Z\big\rangle$, then one can collect the terms and discover $\langle R(X,Z)Z,Z\rangle = 0$.

9. Symmetries of the Curvature Tensor: 2/4/14

Recall that given a Riemannian manifold (M,g), one obtains a Riemannian connection ∇, and from that the Riemann curvature tensor, either the $(1,3)$-tensor $R(X,Y)Z = \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z$ or the $(0,4)$-tensor $\hat{R}(X,Y,Z,V) = \langle R(X,Y)Z,V\rangle$. Then, we showed the following identities:

- $R(X,Y)Z = -R(Y,X)Z$.
- The first Bianchi identity: $R(X,Y)Z + R(Y,Z)X + R(Z,X)Y = 0$.
- $\langle R(X,Y)Z,V\rangle = -\langle R(X,Y)V,Z\rangle$.
- $\langle R(X,Y)Z,V\rangle = \langle R(Z,V)X,Y\rangle$.

These are the 0^{th}-order identities of the curvature tensor; there are other symmetries, but they involve derivatives.

We haven't yet proved the last one, so let's do that: using the third identity,

$$\begin{aligned} \langle R(X,Y)Z,V\rangle &= -\langle R(Y,X)Z,V\rangle \\ &= \langle R(X,Z)Y,V\rangle + \langle R(Z,Y)X,V\rangle \end{aligned}$$

by the 1^{st} Bianchi identity.

$$\begin{aligned} \langle R(X,Y)Z,V\rangle &= -\langle R(X,Y)V,Z\rangle \\ &= \langle R(Y,V)X,Z\rangle + \langle R(V,X)Y,Z\rangle. \end{aligned}$$

Taking the sum,

$$2\langle R(X,Y)Z,V\rangle = \langle R(X,Z)Y,V\rangle + \langle R(Y,V)X,Z\rangle + \langle R(Z,Y)X,V\rangle + \langle R(V,X)Z,Y\rangle.$$

Sectional Curvature. Let $\Pi \subseteq T_pM$ be a two-dimensional subspace, and $\{v_1,v_2\}$ be a basis for Π.

Definition. The sectional curvature of Π is

$$K(\Pi)_p = \frac{\langle R(v_1,v_2)v_2,v_1\rangle}{\det(\langle v_i,v_j\rangle)}.$$

In two dimensions (i.e. $n=2$, so that $\Pi = T_pM$), this is called the Gauss curvature.

This quantity is a smooth function of the point p. It's also independent of the choice of basis, which follows from the symmetries discussed above. Let $\{\tilde{v}_1,\tilde{v}_2\}$ be another basis for Π, so that

$$v_i = \sum_{j=1}^{2} a_{ij}\tilde{v}_j.$$

Let $A = (a_{ij})$. Then,

$$\begin{aligned} \langle R(v_1,v_2)v_2,v_1\rangle &= \langle R(a_{11}\tilde{v}_1 + a_{12}\tilde{v}_2, a_{21}\tilde{v}_1 + a_{22}\tilde{v}_2)v_2,v_1\rangle \\ &= \underbrace{(a_{11}a_{22} - a_{12}a_{21})}_{\det A}\langle R(\tilde{v}_1,\tilde{v}_2)v_2,v_1\rangle \\ &= (\det A)^2\langle R(\tilde{v}_1,\tilde{v}_2)\tilde{v}_2,\tilde{v}_1\rangle. \end{aligned}$$

The change is by the determinant squared, but fortunately, this goes for the denominator, too: $(\langle v_i,v_j\rangle) = A(\langle\tilde{v}_i,\tilde{v}_j\rangle)A^{\mathrm{T}}$, so when one takes the determinant, there's an extra factor of $\det(A)\det(A^{\mathrm{T}}) = (\det A)^2$. Thus, the changes cancel out, and the sectional curvature is independent of basis.

The idea of sectional curvature is that a two-dimensional section of T_pM traces out a two-dimensional submanifold of M, and the sectional curvature is just the Gauss curvature of that submanifold.

Curvature of Hypersurfaces. Let M be an n-dimensional submanifold of an $(n+1)$-dimensional manifold N. Let $\overline{g}$ be a Riemannian metric on N and g be the induced metric on M, i.e. $g_p = \overline{g}|_{T_pM}$. If ∇ is the induced Riemannian connection on M and $\overline{\nabla}$ is that on N, then $\nabla_X Y = (\overline{\nabla}_X Y)^\top = \overline{\nabla}_X Y - \langle \nabla_X Y, \nu\rangle\nu$, where ν is a unit normal. (Note that this does require a choice of a unit normal at every point.)

Definition. The second fundamental form on a Riemannian manifold $(M, \langle\cdot,\cdot\rangle)$ with Riemannian connection ∇ is $h(X,Y) = \langle \nabla_X Y, \nu\rangle$, where ν is normal to X and Y. Sometimes, in physics, this is called the extrinsic curvature.

This is a $(0,2)$-tensor, and is symmetric: $h(X,Y) - h(Y,X) = \langle [X,Y], \nu\rangle = 0$. Moreover,

$$h(X, fY) = \langle f\nabla_X Y + (Xf)Y, \nu\rangle = f\langle \nabla_X Y, \nu\rangle = fh(X,Y).$$

The second fundamental form isn't really an intrinsic quantity, and thus doesn't come up much in Riemannian geometry. But it's useful in some cases, e.g. deriving the Gauss equation.

Example 9.1. A nice example of this is a graph, which has natural global coordinates: suppose $M = \{(x, u(x)) : x \in \Omega\}$, where $\Omega \subseteq \mathbb{R}^n$ is open. Then, $x^{n+1} = u(x^1, \dots, x^n)$, so one can take the coordinates $x^1, \dots, x^n$ for M. Then,

$$\frac{\partial}{\partial x^i} = \left(0, \dots, 0, 1, 0, \dots, \frac{\partial u}{\partial x^i}\right),$$

with the 1 in the i^{th} position, so

$$g_{ij} = \frac{\partial}{\partial x^i} \cdot \frac{\partial}{\partial x^j} = \delta_{ij} + u_{x^i}u_{x^j}.$$

Now, it's necessary to pick a unit normal, but in this case it's fairly explicit: $\nu = (-\nabla u, 1)/\sqrt{1 + |\nabla u|^2}$ (since $x^{n+1} - u(x^1, \dots, x^n) = 0$ if $(x^1, \dots, x^{n+1}) \in M$, so $\nabla F = (-\nabla u, 1)$ is normal to the level sets of $F(x^1, \dots, x^{n+1}) = x^{n+1} - u(x^1, \dots, x^n)$). Then,

$$\overline{\nabla}_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} \cdot \nu = \left(0, \dots, 0, \frac{\partial^2 u}{\partial x^i \partial x^j}\right) \cdot \nu,$$

so in coordinates, the second fundamental form is given by the matrix

$$h_{ij} = \frac{1}{\sqrt{1 + |\nabla u|^2}} \frac{\partial^2 u}{\partial x^i \partial x^j}.$$

The idea behind the Gauss equation is that the curvature tensor can be written in terms of this second fundamental form, which makes it easier to compute the curvature of a given manifold. Let X, Y, and Z be tangent vector fields to M (which sits inside N, an $(n+1)$-dimensional manifold, as before). Let $\overline{R}$ be the curvature tensor for N and R be that for M; then, the Gauss equation says that

$$(\overline{R}(X,Y)Z)^\top = R(X,Y)Z + h(Y,Z)LX - h(X,Z)LY, \tag{6}$$

where L is the *shape operator* (or *Weingarten map*), which is just the $(1,1)$-tensor version of the second fundamental form, i.e. $\langle LX, Y\rangle = h(X,Y) = \overline{g}(\overline{\nabla}_X Y, \nu) = X\langle Y, \nu\rangle - \langle Y, \overline{\nabla}_X \nu\rangle$. Thus, once things cancel out, one has the formula $LX = -\overline{\nabla}_X \nu$. In other words, at a point $p \in M$, one obtains a linear map $L_p : T_pM \to T_pM$. In some sense, this makes the Gauss equation quadratic in terms of the second fundamental form, and when the $(0,4)$-version of the equation is used, the shape operator disappears entirely.

Here's why the Gauss equation works:

$$(\overline{R}(X,Y)Z)^\top = (\overline{\nabla}_X \overline{\nabla}_Y Z)^\top - (\overline{\nabla}_Y \overline{\nabla}_X Z)^\top - (\overline{\nabla}_{[X,Y]} Z)^\top.$$

Then, the first term simplifies to $(\overline{\nabla}_X (\nabla Y Z + h(Y,Z)\nu))^\top$, which is just $\nabla_X \nabla_Y Z - LX$, and the other two terms follow similarly.

The $(0,4)$-version of the Gauss equation might look a little more familiar: if X, Y, Z, and V are all tangent vector fields to M, then

$$\langle R(X,Y)Z, V\rangle = \left\langle \overline{R}(X,Y)Z, V\right\rangle + h(Y,Z)h(X,V) - h(X,Z)h(Y,V). \tag{7}$$

Now, the fact that it's quadratic in the second fundamental form is more obvious. If e_1, e_2 is a basis for T_pM and M is a two-dimensional submanifold of $\mathbb{R}^3$, then $\langle R(e_1, e_2)e_2, e_1\rangle = h_{22}h_{11} - h_{12}^2 = \det(h_{ij})$. Thus, one has the following consequence, also due to Gauss:

Theorem 9.1 (Theorema Egregium)**.**

$$K = \frac{\langle R(e_1, e_2)e_2, e_1\rangle}{\det(g_{ij})} = \frac{\det(h_{ij})}{\det(g_{ij})}.$$

This indicates some interesting relations between the extrinsic curvature, given by the second fundamental form, and the intrinsic curvature, given by the metric.

Example 9.2. The above results make calculating the curvature a lot less painful; for example, consider $S^n \subseteq \mathbb{R}^{n+1}$ given by $\{x \mid \|x\| = 1\}$. Then, the unit normal is just the position vector, which is a really nice property of the sphere: $\nu = x$, so $LX = -\overline{\nabla}_X x = -X$, and thus $R(X,Y)Z = h(Y,Z)(LX) - h(X,Z)(LY) = \langle Y, Z\rangle X - \langle X, Z\rangle Y$, because $h(X,Y) = \langle X, Y\rangle$. Thus the curvature is

$$K = \frac{\langle R(X,Y)Y, X\rangle}{\|X\|^2\|Y\|^2 - \langle X, Y\rangle^2} = \frac{\|X\|^2\|Y\|^2 - \langle X, Y\rangle^2}{\|X\|^2\|Y\|^2 - \langle X, Y\rangle^2} = 1.$$

A useful reference tensor on any Riemannian manifold (M, g) is $R_1(X,Y)Z = \langle Y, Z\rangle X - \langle X, Z\rangle Y$. This is akin to the curvature tensor, but with a constant curvature of 1.

Another interesting fact is that the sectional curvatures algebraically completely determine the curvature tensor. The book explains this in gory detail (even though most don't). It's an interesting formula, if not particularly useful.

Theorem 9.2. *All entries in $\langle R(X,Y)Z, V\rangle$ can be expressed as linear combinations of entries of the form $\langle R(A,B)B, A\rangle$.*

Proof. Let R and $\tilde{R}$ be curvature tensors (i.e. they satisfy the four algebraic symmetries outlined above) and $\langle R(A,B)B, A\rangle$ and $\langle \tilde{R}(A,B)B, A\rangle$ for all $A, B \in \mathcal{X}(M)$, then one can show that $R = \tilde{R}$ by subtracting $R - \tilde{R}$ and doing some algebraic manipulation. ⊠

10. Traces and Sectional Curvature: 2/6/14

"So in particular, if I'm Einstein. . . "

Theorem 10.1. *The sectional curvature determines the full curvature tensor.*

Proof. Let R and $\tilde{R}$ be curvature-type tensors, i.e. the four algebraic identities outlined in the previous lecture all hold. Then, if for all X, Y, $\langle R(X,Y)Y, X\rangle = \left\langle \tilde{R}(,Y)Y, X\right\rangle$, then $R = \tilde{R}$, by the Gauss equation. Each tensor can be written as linear combinations of such elements, albeit in a complicated way.

Let $Q(X,Y,Z,V) = \langle R(X,Y)Z, V\rangle - \left\langle \tilde{R}(X,Y)Z, V\right\rangle$; then, the goal is to show that $Q(X,Y,X,Y) = 0$. This can be shown by expanding out Q and using the identities for R and $\tilde{R}$ (in particular, the Bianchi identity for cyclic permutation) to show that it goes to zero. ⊠

The sectional curvatures are everything, and are useful for understanding the total curvature. This is more of a philosophical point, but still helpful in practice.

The $(0,4)$-curvature tensor has traces defined for all pairs of slots; for example, the $(1,2)$-trace of Z and V is $\sum_i \langle R(e_i, e_i)Z, V\rangle$, where $e_1, \dots, e_n$ is a basis of T_pM. However, this is anti-symmetric, so it goes to zero. Similarly, the $(3,4)$-trace is anti-symmetric in e_i and e_i, and goes to zero as well.

Then, the remaining traces are equal up to sign: The $(2,3)$-trace is

$$\sum_{i=1}^n \langle R(X, e_i)e_i, Y\rangle = \sum_{i=1}^n \langle R(e_i, X)Y, e_i\rangle,$$

using the curvature identities, but this is the $(1,4)$-trace. And the $(2,4)$-trace is $\sum_{i=1}^n \langle R(e_i, X)e_i, Y\rangle$, which is the negative of the $(1,4)$-trace.

The Ricci tensor is just a choice of one of these, e.g. the $(2,3)$-trace:

$$\operatorname{Ric}(X,Y) = \sum_{i=1}^n \langle R(X, e_i)e_i, Y\rangle.$$

This is a symmetric $(0,2)$-tensor field, somewhat like the Riemannian metric.

The Ricci curvature contains less information than the Riemann curvature tensor, because some entries are summed; thus, the Ricci curvature says useful things about the sectional curvature, but doesn't always completely determine R. In lower dimensions ($n = 2, 3$), though, they do contain the same amount of information.

Definition. The scalar curvature is a real-valued function that is given by the full trace of the Riemann curvature tensor:

$$R = \sum_{i=1}^n \operatorname{Ric}(e_i, e_j).$$

These are all independent of basis, which follows from more general facts about tensor fields, generalization of the notion of a trace of a matrix. For example, consider a $(1,1)$-tensor

$$A = \sum_{i,j=1}^{n} a_j^i \frac{\partial}{\partial x^i} \otimes \mathrm{d}x^j.$$

But one can always contract a covariant entry with a contravariant entry, to get a basis-independent quantity:

$$c(A) = \sum_{i=1}^{n} a_i^i \in C^\infty(M).$$

On $(0,2)$-tensor fields (using the summation convention) $\alpha = \alpha_{ij}\,\mathrm{d}x^i \otimes \mathrm{d}x^i$, one can use the metric to convert a covariant index to a contravariant index, obtaining a $(1,1)$-tensor

$$A = g^i \alpha_{pj} \frac{\partial}{\partial x^i} \otimes \mathrm{d}x^j.$$

Now, it's possible to contract (though in general this requires the metric, so that there are both covariant and contravariant indices).

In general, if A is an (r,s)-tensor field, then one can contract and obtain $c(A)$, an $(r-1, s-1)$-tensor field. This can be done in several ways: if in coordinates

$$A = \sum a_{j_1\dots j_s}^{i_1\dots i_r} \frac{\partial}{\partial x^{i_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{i_r}} \otimes \mathrm{d}x^{j_1} \otimes \cdots \otimes \mathrm{d}x^{j_s}.$$

However, it happens that all of the choices for this are the same. This is the principle that underlies index notation.

The scalar curvature admits a geometric interpretation: suppose $e_1, \dots, e_n$ is an orthonormal basis for T_pM.

$$R = \sum_{i,j=1}^{n} \underbrace{\langle R(e_i, e_j)e_j, e_i\rangle}_{K(\Pi_{ij})},$$

where Π_{ij} is the space spanned by e_i and e_j. This means that R is twice the sum of the sectional curvatures of the coordinate 2-planes. In the case $n = 2$, $R = 2K$, where K is the Gauss curvature again. In some sense, this is the weakest version of the curvature tensor (weakest invariant), except in dimension 2, where they're equally powerful.

Definition. If (M, g) is a Riemannian manifold, then g is Einstein if the Ricci tensor is proportional to the metric: $\mathrm{Ric}(g) = \lambda g$ for some λ.

One can also define an Einsteinian metric for Lorentz metrics, and this is what is used in physics, but the Riemannian case is also interesting in geometry.

Example 10.1.

(1) Any space of constant curvature clearly has an Einstein metric.
(2) Suppose $R = cR_1$, for some constant c. Then, one can calculate that $\mathrm{Ric}(X, Y) = c(n-1)\langle X, Y\rangle$, and $c(n-1)$ is constant. This can be shown by plugging into the definition of the Ricci curvature.

Proposition 10.2. *If $n = 2$ or $n = 3$ and if g is an Einstein metric on M, then (M, g) has constant curvature.*

Proof. In the case $n = 2$, the Ricci tensor is just the Gauss tensor, so if g is Einstein, then K must be constant.

For $n = 3$, the idea is that the sectional curvatures can be recovered from the Ricci curvature: if Π is a 2-plane and e_1 and e_2 are orthonormal on Π, then one can choose an orthonormal e_3 to e_1 and e_2. Then, one can recover the sectional curvature as follows: if K_{ij} is the sectional curvature of the e_ie_j-plane, then $\mathrm{Ric}(e_1, e_1) = K_{12} + K_{13}$, $\mathrm{Ric}(e_2, e_2) = K_{12} + K_{23}$, and $\mathrm{Ric}(e_3, e_3) = K_{13} + K_{23}$.

Thus, one solves these to obtain, e.g. $K_{12} = (R_{11} + R_{22} - R_{33})/2$, so $K(\Pi) = \lambda/2$ if g is Einstein with constant λ. ⊠

This also works for Lorentz metrics: in three dimensions, Einstein metrics have constant curvature. But in neither case (Riemannian nor Lorentz) does this work in $n = 4$; instead, there are plenty of counterexamples. Perhaps the simplest is $S^2 \times S^2$, but that's a story for the problem set.

These are all of the 0^{th}-order properties we need, so let's see how to extend ∇ to accept more general tensor fields. As a starting point, $\nabla_X f = Xf$, and $\nabla_X Y$ is defined for vector fields X and Y. We also have the Leibniz rule, which will really be useful.

Look at a simple $(2,0)$-tensor $Y \otimes Z$, where $Y \otimes Z(w_1, w_2) = w_1(Y)w_2(Z)$. Then, $\nabla_X(Y \otimes Z) = (\nabla_X Y) \otimes Z + Y \otimes \nabla_X Z$ is a natural way of understanding the Leibniz rule. Then, this can be extended, since what has been already defined works for the basis. Thus, if $A = a^{ij}\frac{\partial}{\partial x^i}\frac{\partial}{\partial x^j}$, then

$$\nabla_X A = \sum_{i,j=1}^{n} \nabla_X\left(a^{ij}\frac{\partial}{\partial x^i} \otimes \frac{\partial}{\partial x^j}\right),$$

using the rules defined above.

Then, this can be extended to $(r,0)$-tensors in the same way. But what about covariant tensors? These can be accounted for by declaring that ∇ commutes with contractions. Thus, if ω is a $(1,1)$-tensor, then $c(\nabla_X(Y \otimes \omega)) = \nabla_X Y \otimes \omega + Y \otimes \nabla_X \omega$. But then, since $c(Z \otimes \omega) = w(Z)$, then this simplifies to the following rule:

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

This just follows because we want the Leibniz rule to work, and for the operation to commute with contraction. And now, these formulas allow one to apply ∇ on general (r,s)-tensors.

If A is an (r,s)-tensor field, then $\nabla_X A$ is also an (r,s)-tensor field, but since it's C^∞-linear, and therefore tensorial, in X, then ∇A is an $(r,s+1)$-tensor field, given by

$$\nabla A(\omega^1, \dots, \omega^r, Y_1, \dots, Y_s, X) = (\nabla_X A)(\omega^1, \dots, \omega^r, Y_1, \dots, Y_s).$$

The indices might look like $\nabla_i a^{i_1 \dots i_s}_{j_1 \dots j_s}$, so there's an extra lower index, making it an $(r,s+1)$-tensor.

Note that the direction one differentiates in should always be a vector field; more general tensor fields don't seem to have sense of direction, which is what the covariant derivative actually means.

With this extension of ∇, one can state the second Bianci identity for the $(1,3)$ curvature tensor R,

$$(\nabla_X R)(Y,Z)V + (\nabla_Y R)(Z,X)V + (\nabla_Z R)(X,Y)V = 0.$$

Here. $\nabla_X R$ is also a $(1,3)$-tensor field, so the same notation is used $(\nabla_X R)(Y,Z,V) = (\nabla_X R)(Y,Z)V$.

Then, in coordinates with the Christoffel symbols, one has

$$\nabla_i \, \mathrm{d}x^j = -\sum \Gamma^j_{ik} \, \mathrm{d}x^k.$$

11. The Second Bianchi Identity: 2/11/14

> *"Everyone believes that's a completely rigorous proof, right? You won't go complain to the Dean or anything?"*

Recall that along the way to proving the second Bianchi identity, it was necessary to extend the connection ∇ to general tensor fields. For example, if $x^1, \dots, x^n$ are local coordinates and A is a $(1,1)$-tensor in coordinates

$$A = \sum a^i_j \frac{\partial}{\partial x^i} \otimes \mathrm{d}x^j,$$

then $\nabla_k a^i_j$ is the coefficient of the $(1,2)$-tensor

$$\nabla A = \sum \nabla_k a^i_j \frac{\partial}{\partial x^i} \otimes \mathrm{d}x^k \otimes \mathrm{d}x^j.$$

Expanding out and implicitly summing ℓ, this is

$$\nabla_k a^i_j = \partial_k a^i_j + a^\ell_j \Gamma^i_{\ell k} - a^i_\ell \Gamma^\ell_{jk}.$$

(Here, $\partial_k = \frac{\partial}{\partial x^k}$.) Moreover, using the torsion-free property, if $\nabla \frac{\partial}{\partial x^i} = 0$ at p, then $\nabla_k a^i_j = \partial_k a^i_j$ (which follows from a homework problem).

All of this can also be said about more general tensor fields, but there are just more terms on the right.

Recall that we were proving the second Bianchi identity:

$$(\nabla_X R)(Y,X)V + (\nabla_Y R)(Z,X)V + (\nabla_Z R)(X,Y)V = 0.$$

It's enough to check this in a basis, but in local coordinates we have

$$R\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right)\frac{\partial}{\partial x^k} = \sum_\ell R^\ell_{kij} \frac{\partial}{\partial x^\ell}.$$

So the goal is to show that

$$\nabla_k R^p_{qij} + \nabla_i R^p_{qjk} + \nabla_j R^p_{qki} = 0.$$

It's possible to choose suitable coordinates such that $\nabla_{ij}^k = 0$ at a point P, so ∇ becomes ∂:

$$0 = \partial_k R_{qij}^p + \partial_i R_{qjk}^p + \partial_j R_{qki}^p$$
$$R_{qij}^p = \partial_i \Gamma_{qj}^p - \partial_j \Gamma_{qi}^p + \Gamma^2.$$

But the Γ^2 term vanishes quadratically once the derivative is taken, and $\partial_i \Gamma_{qj}^p = a_i$, so the goal is to calculate $\partial_k\partial_i a_j - \partial_k\partial_j a_i + \partial_i\partial_j a_k - \partial_i\partial_k a_j + \partial_j\partial_k a_i - \partial_j\partial_i a_k$, but partials commute, so this is zero.

This is different than in the book, but is a completely honest proof: choosing the right basis can make a lot of terms go away.

Definition. The Einstein tensor is $G = \operatorname{Ric} - (1/2)Rg$. This is a symmetric, $(0,2)$-tensor.

In this notation, G stands for "gravity."

Definition. The divergence of a $(0,2)$-tensor field G is the $(0,1)$-tensor field

$$\operatorname{div} G(X) = \sum_{i=1}^n (\nabla_{e_i} G)(e_i, X),$$

where $e_1, \dots, e_n$ is an orthonormal basis.

Proposition 11.1 (Twice-contracted 2nd Bianchi identity)**.** *G is divergence-free.*

Proof. Write $Q(X,Y,Z,V) = \langle R(X,Y)Z, V\rangle$; then, by the second Bianchi identity,

$$(\nabla_X Q)(Y,Z,V,W) + (\nabla_Y Q)(Z,X,V,W) + (\nabla_Z Q)(X,Y,V,W) = 0.$$

Then, take the trace in the Z and V slots. This commutes with the covariant derivative because of metric compatibility of ∇, which can be explicitly calculated. Then, after the traces are taken, one obtains

$$(\nabla_X \operatorname{Ric})(Y,W) - (\nabla_Y \operatorname{Ric})(X,W) + (\nabla_{e_i} Q)(X,Y,e_i,W) = 0.$$

Then, take the trace with respect to X and W:

$$\begin{aligned} 0 &= 2(\operatorname{div}\operatorname{Ric})(Y) - Y(R) \\ &= 2(\operatorname{div}\operatorname{Ric})(Y) - \operatorname{div}(Rg)(Y), \end{aligned}$$

so $\operatorname{div}(\operatorname{Ric} - (1/2)Rg) = 0$. ⊠

This version is arguably more useful than the conventional 2nd Bianchi identity.

There's two nice corollaries called the Schur theorems. Oddly enough, the easier one is the stronger one.

Theorem 11.2 (First Schur Theorem)**.** *If $n \geq 3$ and there exists an open connected $U \subseteq M$ such that $\operatorname{Ric}(g)_p = \lambda(p) g_p$ for all $p \in U$, then λ is constant (i.e. g is Einstein).*

This means that if the Ricci curvature is constant o T_pM, then it's constant everywhere. But in dimension 2, the fact that the Ricci curvature is identical to the Gauss curvature creates plenty of counterexamples.

Theorem 11.3 (Second Schur Theorem)**.** *If $n \geq 3$ and there exists a open connected $U \subseteq M$ such that, for all $p \in U$, there exists a $\kappa(p)$ such that $K_p(\Pi_p) = \kappa(p)$ for all $\Pi_p \subseteq T_pM$, then κ is constant and g is a constant curvature metric.*

Proof. This proof will assume the first Schur theorem, which will be proven next. Furthermore, we have

$$\langle R(X,Y)Z, V\rangle_p = \kappa(p)(\langle X,V\rangle\langle Y,Z\rangle - \langle Y,V\rangle\langle X,Z\rangle).$$

Thus, $\operatorname{Ric}(X,V) = (n-1)\kappa(p)g(X,V)$, so let $\lambda(p) = (n-1)\kappa(p)$, and then all that needs to be done is invoking the first Schur theorem. ⊠

The direct proof of this theorem is a little more messy algebra.

Proof of Theorem 11.2. Write

$$G = \operatorname{Ric} - \frac{1}{2}Rg = \left(\frac{1}{n} - \frac{1}{2}\right)Rg,$$

and since $n > 2$, then this constant, $(2-n)/2n$, is nonzero. But I know that $\operatorname{div}(G) = 0$, so $\operatorname{div}(Rg) = 0$ as well, so

$$\begin{aligned} 0 = (\operatorname{div}(Rg))(X) &= \nabla_{e_i}(Rg)(e_i, X) \\ &= (e_i R) g(e_i, X) \\ &= XR, \end{aligned}$$

so R must be constant on U. ⊠

Theorem 11.4. *Suppose $n \geq 3$ and M is a connected n-dimensional submanifold embedded in $\mathbb{R}^{n+1}$. If M is Einstein and $R \neq 0$, then M is a portion of the sphere.*

That the sphere is a hypersurface of constant curvature is no surprise, but it's more interesting that it's the only one.

Corollary 11.5. *No piece of hyperbolic space H^n for $n \geq 3$ can be embedded as a hypersurface.*

Once again, this is false for $n = 2$: there are local embeddings (though by a theorem of Hilbert, no global embeddings), including a surface of revolution called the pseudosphere.

The consequence of this theorem is that Riemannian geometry is much more general than that of hypersurfaces. There is a (hard) theorem of John Nash that every manifold can be globally embedded in sufficiently high-dimensional Euclidean space, but embeddings as hypersurfaces are special. The signature of the metric is important: H^n has a natural embedding in $(n+1)$-dimensional Minkowski space, for example.

The question of finding an isometric embedding $F : M \to \mathbb{R}^n$ is equivalent to solving the system of PDEs $\tilde{g}_{ij} = \frac{\partial F}{\partial x^i} \cdot \frac{\partial F}{\partial x^j} = g_{ij}$ (for given g_{ij}, solving for F, which leads to $\tilde{g}_{ij}$). This is an overdetermined system, so generically there isn't a solution (unless the codimension is high). It's more reasonable for $n = 3$, because then there are fewer constraints, and under a lot of conditions surfaces can be embedded (sometimes only locally) into $\mathbb{R}^3$; the general conjecture is open.

12. Models of Hyperbolic Space: 2/18/14

There are standard examples of simply connected spaces of constant curvature: for positive curvature, the sphere S^n, and for zero curvature, Euclidean space $\mathbb{R}^n$. There are several standard models for hyperbolic space H^n, and they are abstractly isometric. In some cases, one can give explicit isometries between these formulations.

Consider Minkowski space $\mathbb{R}^{n+1}_1$, i.e. $(n+1)$-dimensional space with the Lorentz pairing $\langle v, w\rangle = -v^0 w^0 + \sum_{i=1}^n v^i w^i$. Another way to write this is

$$g = -(\,\mathrm{d}x^0)^2 + \sum_{i=1}^n (\,\mathrm{d}x^i)^2 = \sum_{a,b=0}^n \eta_{ab}\,\mathrm{d}x^a\,\mathrm{d}x^b,$$

where η is the matrix equal to I_n except with a -1 in position $(1,1)$. (It's common to leave the $\otimes$ sign out of metrics in this context.) Minkowski space has some interesting geometry. For a start, vectors fall into three types: spacelike, i.e. $\langle v, v\rangle > 0$, timelike, i.e. $\langle v, v\rangle < 0$, and null, for which $\langle v, v\rangle = 0$.

A more general linear-algebraic fact is that if $\langle\cdot,\cdot\rangle$ is any nondegenerate scalar product and $V \subseteq \mathbb{R}^{n+1}_1$ is a linear subspace of codimension 1, then there exists a normal vector $\nu \neq 0$ such that $V = \{\mathbf{x} : \langle \mathbf{x}, \nu\rangle = 0\}$. In other words, V is the zero set of a linear function. This means one can classify subspaces of codimension 1:

- V is spacelike if its normal vector ν is timelike. In this case, one can always scale ν such that $\langle \nu, \nu\rangle = -1$.
- V is timelike if its normal vector ν is space-like, in which case it can be normalized such that $\langle \nu, \nu\rangle = 1$.
- V is a null plane if ν is a null vector. In this case, $\nu \in V$, which is slightly weird.

One can think of a spacelike plane as a plane with slope less than 1 (relative to everything vs. x^0), roughly, so that its normal vector is inside the null cone (and therefore timelike). A timelike plane has slope greater than 1 (or is vertical), so that its normal vector is spacelike. Then, a null plane must have slope exactly 1, and is contained within the null cone (lightcone).

It is also possible to classify curves: a curve is spacelike if its tangent vector is always spacelike, and so on. We will ignore the cases of curves (and hypersurfaces) where the curve doesn't change.

One can also classify hypersurfaces: if M is an n-dimensional hypersurface in $\mathbb{R}^{n+1}_1$, then:

- M is spacelike if T_pM is always spacelike for all p.
- M is timelike if T_pM is timelike for all p.
- M is null if T_pM is a null plane for all p.

The null cone is singular at the origin, but when that singularity is removed, it becomes a null hypersurface. If $M = \{x^0 = u(x^1, \dots, x^n)\}$, then it is spacelike if $\|\nabla u\| < 1$, timelike if $\|\nabla u\| > 1$, and is timelike if $\|\nabla u\| = 1$ (the eikonal equation), since this determines the slopes of the tangent planes. Null hypersurfaces have more of a structure because this is an equality, rather than an inequality.

Within this Minkowski space, consider the unit sphere $S^n_{-1} = \{\mathbf{x} \in \mathbb{R}^{n+1}_1 : \langle \mathbf{x}, \mathbf{x}\rangle = -1, x^0 > 0\}$. The first equation requires that $(x^0)^2 = 1 + \sum_{i=1}^n (x^i)^2$, which is a hyperboloid of two sheets, and thus has two connected components. S^n_{-1} is one of them. This is isometric to H^n, as it will be shown to have constant curvature of -1. Interestingly, the restriction of the standard Lorentz metric to S^n_{-1} gives a Riemannian metric, which is what allows this to work.

One also has $S^n_{+1} = \{\mathbf{x} \in \mathbb{R}^{n+1}_1 : \langle \mathbf{x}, \mathbf{x}\rangle = 1\}$. This intersects all planes where x^0 is constant, and is a hyperboloid of one sheet. This is a surface of revolution, asymptotic to the lightcone, and simply connected. This is a timelike hypersurface, unlike S^n_{-1}, which is spacelike. Thus, the induced metric onto S^n_{+1} is a Lorentz metric. This induced spacetime is called de Sitter spacetime, and is a well-studied object. This Lorentz manifold has lots of symmetries; it's akin to the sphere, in terms of constant positive curvature, but the fact that it's not compact makes life more interesting.

Since S^n_{-1} has constant negative curvature and is Riemannian, one can compute that the sectional curvature is just -1. But for de Sitter spacetime, it's harder, and one can't even compute cuvature for null hypersurfaces. This can be derived by looking at the Gauss equation: let $M \subseteq \mathbb{R}^{n+1}_1$ be either spacelike or timelike, so that it has a unit normal ν. Let $\varepsilon = \langle \nu, \nu \rangle$, so that it's 1 is M is timelike and -1 if M is spacelike. The induced metric is nondegenerate (Riemannian in the spacelike case, Lorentz in the timelike case; it's degenerate in the null case), so one can compute its connection. This ends up being $\nabla_X Y = D_X Y - \varepsilon \langle D_X Y, \nu \rangle \nu$.

This works, because we want $\langle \nabla_X Y, \nu \rangle = 0$, but this is $\langle D_X Y, \nu \rangle - \varepsilon \langle D_X Y, \nu \rangle \langle \nu, \nu \rangle$, which works because $\varepsilon \langle \nu, \nu \rangle = 1$. The Gauss equation has a similar form, with an ε also appearing:

$$R(X, Y)Z = \varepsilon(h(Y, Z)LX - h(X, Y)LY),$$

where $h(X, Y) = \langle D_X Y, \nu \rangle$. Since $D_X \nu$ is tangential and $\langle Y, \nu \rangle = 0$, then $LX = -D_X \nu$, with the shape operator again a $(1, 1)$-tensor $L_p : T_p M \to T_p M$.

First, the curvature in the ambient space is set to zero:

$$\begin{aligned} 0 &= D_X D_Y Z - D_Y D_X Z - D_{[X,Y]} Z \\ &= (D_X D_Y Z)^\top - (D_Y D_X Z)^\top - (D_{[X,Y]} Z)^\top. \end{aligned}$$

But $D_Y Z = \nabla_Y Z + \varepsilon h(Y, Z)\nu$, so $D_X D_Y Z = \nabla_X \nabla_Y Z - \varepsilon h(Y, Z)LX$. Then,

$$= R(X, Y)Z - \varepsilon h(Y, Z)LX + \varepsilon h(X, Y)LY. \tag{8}$$

The last equation, (8), is the specific form of the Gauss equation in Minkowski space. It's the same idea as before, but with the sign associated with the tangential projection.

Now, one can check the curvature of a hypersurface, such as S^n_{-1}. Just as with the sphere, one can take the normal vector to $\mathbf{x}$ at a point to be $\nu = \mathbf{x}$: if Y is tangent to S^n_{-1}, then $\langle \mathbf{x}, \mathbf{x} \rangle = -1$, which is constant, so $Y(\langle \mathbf{x}, \mathbf{x} \rangle) = 0$, so $2\langle D_Y \mathbf{x}, \mathbf{x} \rangle = 0$. But $D_Y \mathbf{x} = Y$, so $\langle Y, \mathbf{x} \rangle = 0$ for all $Y \in T_p S{-1}^n$.

Then, $LX = -D_X \nu = -X$, i.e. $L = -\operatorname{id}$, and

$$R(X, Y)Z = -(-\langle Y, Z \rangle X + \langle X, Z \rangle(-Y)) = -(\langle Y, Z \rangle X - \langle X, Z \rangle Y) = -R_1(X, Y)Z.$$

Thus, R is a constant multiple of R_1, so it has constant sectional curvature equal to -1. Thus, this is the curvature for the timelike unit sphere S^n_{-1}.

For de Sitter spacetime, the ε changes sign, so one instead obtains $R(X, Y)Z = R_1(X, Y)Z$. This has constant positive curvature, but since it's Lorentz and not Riemannian (unlike the timelike unit sphere), then it's not isometric to S^n. Then, for any nondegenerate two-dimensional $\Pi \subseteq T_p S^n_{+1}$, $K(\pi) = 1$.

The Lorentz and Poincaré Groups. In standard hyperbolic space, the isometries are kind of unpleasant to write down. However, the advantage of the presentation via Minkowski space is that the isometries are all linear operators.

Definition. An $(n+1) \times (n+1)$ matrix A is Lorentz if $\langle Av, Aw \rangle = \langle v, w \rangle$.

Using the matrix

$$\eta = \begin{pmatrix} -1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix}$$

described above, A is Lorentz if $\langle Av, Aw \rangle = (Av)^{\mathrm{T}} \eta (Aw) = v^{\mathrm{T}} \eta w$, so we need $v^{\mathrm{T}}(A^{\mathrm{T}} \eta A)w = v^{\mathrm{T}} \eta w$, so $A^{-1} = (\eta A \eta)^{\mathrm{T}}$. Another characterization is that A is Lorentz iff its columns form a Lorentz basis, i.e. if $A = (\mathbf{e}_0, \dots, \mathbf{e}_n)$, then $\langle \mathbf{e}_a, \mathbf{e}_b \rangle = \eta_{ab}$. This is because the columns of A are the image of the standard basis under some isometry, so this equation should still follow.

The group of Lorentz matrices is denoted $\mathrm{O}(n, 1)$, and there is a subgroup $\mathrm{O}_+(n, 1) = \{A \in \mathrm{O}(n, 1) : a_{00} > 0\}$.

Theorem 12.1. $\mathrm{O}_+(n, 1)$ *is the isometry group of* S^n_{-1}.

Proof. First off, all of these act by isometries: if $A \in \mathrm{O}_+(n, 1)$, then $A(S^n_{-1}) \subseteq S^n_{-1}$, and since $a_{00} > 0$, then it sends $e_0 = (1, 0, \dots, 0)$ to somewhere in the positive component, and thus sends the whole positive component to itself. Then, the fat that it is Lorentz implies that it preserves the metric.

These are all of the isometries, because there are enough of them that they account for all possible symmetries one could have. Suppose $p, q \in S^n_{-1}$ and $e_1, \dots, e_n$ is an orthonormal basis at q and $v_1, \dots, v_n$ is an orthonormal basis at p; then, if one can show that there is an isometry $A \in \mathrm{O}_+(n,1)$ such that $Ap = q$ and $A(e_i) = v_i$ for $i = 1, \dots, n$, then these are all possible transformations. The standard way to piece this apart is to construct an A that sends any p to any q; then, the problem restricts to sending bases to other bases at a single point. First consider $p = e_0 = (1, 0, \dots, 0)$, and try to send $q \to p$. ⊠

13. The Variational Theory of Geodesics: 2/20/14

"I think the ideal course is the one where the students do everything through the homework and the professor goes up and talks about philosophy."

Recall that we defined $S^n_{-1} = \{x \in \mathbb{R}^{n+1}_1 : \langle x, x\rangle = -1 \text{ and } x_0 > 0\}$. This is the upper component of the hyperboloid that contains $e_0 = (1, 0, \dots, 0)$ and is asymptotic to the lightcone. We also defined the Lorentz group $\mathrm{O}(n, 1)$ and $\mathrm{O}_+(n + 1)$; the latter acts transitively on S^n_{-1}.

Claim. If $x, y \in S^n_{-1}$, $e_1, \dots, e_n$ is an orthonormal basis at x, and $v_1, \dots, v_n$ is an orthonormal at y, then there exists an $A \in \mathrm{O}_+(n, 1)$ such that $Ax = y$ and $Ae_i = v_i$ for each i.

This will in effect show that every Lorentz basis (i.e. a basis $w_1, \dots, w_n$ such that $\langle w_1, w_1\rangle < 0$ and the rest are space-like) is given by a matrix in $\mathrm{O}_+(n, 1)$.

The Lorentz group is much more interesting than similar groups, such as the orthogonal group. For example, it's noncompact (in the metric induced from $\mathbb{R}^{n^2}$).

Turning to the variational theory of geodesics, let (M, g) be a Riemannian manifold. The metric allows one to measure lots of things, in particular a curve: if $c : [a, b] \to M$, then its length is

$$L(c) = \int_a^b \|c'(t)\| \,\mathrm{d}t.$$

This is independent of parametermization. Once can look at the space of curves, and try to find critical points (especially minima) of this curve, akin to studying critical points of a function in calculus. The tricky thing is that a curve can be varied in an infinite number of dimensions (i.e. the space of curves is infinite-dimensional), unlike the finite-dimensional questions asked in calculus.

But then, one can restrict the dimension. For example, one can take a path of curves, a one-parameter family $c_s(t)$ for different s. Thus, one has a map $C : [a, b] \times (-\varepsilon, \varepsilon) \to M$ sending $(t, s) \mapsto c_s(t)$, such that $c(t, 0) = c_0(t) = c(t)$. But now, the lengths of these curves is a function of one variable: $s \mapsto L(c_s)$ is a smooth function of s. In fact, since the length is independent of parameterization, then one can assume that $c(t)$ is parameterized by arc length (i.e. $\|c'(t)\| = 1$ for all $t \in [a, b)$).

It's also possible to create curves $C(t_0, s)$ for a fixed t_0 and varying s. Then, there are two kinds of tangent vectors: let $T = \frac{\partial C}{\partial t} = c_s'(t)$, i.e. (in a slight abuse of notation) the tangent vector to C as t varies, and $X = \frac{\partial C}{\partial s}$, i.e. as s varies. Then, one has the following, called the First Variational Formula:

$$\left.\frac{\mathrm{d}}{\mathrm{d}s} L(c_s)\right|_{s=0} = \langle X, T\rangle\big|_{t=a}^{t=b} - \int_a^b \langle X, \nabla_T T\rangle \,\mathrm{d}t. \tag{9}$$

Proof. When differentiating $L(c_s)$ with respect to s, one applies the Leibniz rule: $\|T\| = \sqrt{\langle T, T\rangle}$, so

$$\frac{\partial}{\partial s}\|T\| = \frac{\frac{\partial}{\partial s}\langle T, T\rangle}{2\|T\|} = \frac{2\nabla_X T, T}{2\|T\|}.$$

Since $\|T\| = 1$ can be chosen (by parameterizing c by arc length[14]). Since T and X are coordinate vector fields, then $[T, X] = 0$, so $\nabla_X T = \nabla_T X$. Thus,

$$\begin{aligned}\left.\frac{\mathrm{d}L}{\mathrm{d}s}\right|_{s=0} &= \int_a^b \langle \nabla_X T, T\rangle \,\mathrm{d}t = \int_a^b \langle \nabla_T X, T\rangle \,\mathrm{d}T \\ &= \int_a^b (T\langle X, T\rangle - \langle X, \nabla_T T\rangle) \,\mathrm{d}t \\ &= \langle X, T\rangle\big|_{t=a}^{t=b} - \int_a^b \langle X, \nabla_T T\rangle \,\mathrm{d}t\end{aligned}$$

by the Fundamental Theorem of Calculus. ⊠

[14] Though $c = c_0$ is parameterized by arc length, nearby c_ε might not be.

As a consequence, suppose c is a geodesic and $X = 0$ at $t = a$ and $t = b$ (i.e. the endpoints are fixed). Then, the variational length (i.e. $\frac{\partial L}{\partial s}$ at $s = 0$) is zero, which encodes the notion that geodesics are the straightest curves. Conversely, if $X(a) = 0$ and $X(b) = 0$ and the variational length is 0, then c parameterized by arc length is a geodesic, though the proof of this doesn't follow from (9) and will be deferred.

In particular, geodesics are critical points of the arc length functional. However, they aren't necessarily minima: for example, consider $S^2 \subseteq \mathbb{R}^3$. The geodesics on S^2 are the fixed point sets of reflections, i.e. circular arcs of planes through the origin. A short enough geodesic segment minimizes length (which is true of any Riemannian manifold, locally), but a great circle from the north pole to the south pole doesn't uniquely minimize length. Past the south pole, too-long geodesics don't even minimize length!

In order to understand how the function changes at a point (as opposed to that it's just an extremum), one uses the second derivative. This leads to the Second Variation Formula: suppose c is parameterized by arc length. Then,

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2} L(c_s)\bigg|_{s=0} = \langle \nabla_X X\rangle\big|_{t=a}^{t=b} + \int_a^b \Big(\|\nabla_T \widetilde{X}\|^2 - \langle R(T,X)X,T\rangle\Big)\,\mathrm{d}t, \tag{10}$$

where $\widetilde{X} = X - \langle X, T\rangle T$ is the normal component of X to the curve. Notice that $\langle R(T,X)X,T\rangle = \langle R(T,\widetilde{X})\widetilde{X},T\rangle$, so the entire integral can be expressed in termss of $\widetilde{X}$. Here, it's called the index form, $I(\widetilde{X},\widetilde{X})$.

Bring this to H^n, and look at an X where $X(a) = 0$ and $X(b) = 0$. Then,

$$\frac{\mathrm{d}^2 L}{\mathrm{d}s}\bigg|_{s=0} = \int_a^b (\|\nabla_T \widetilde{X}\|^2 - \langle R(T,\widetilde{X})\widetilde{X},T\rangle)\,\mathrm{d}T,$$

so $-1 = \langle R(T,\widetilde{X})\widetilde{X},T\rangle/\|\tilde{X}\|^2$. Thus, c is a local minimum of length (and more sophisticated arguments show it's a global minimum). This is also called stable (i.e. stable geodesics are local minima of the arc length functional).

On surfaces, the normal component is just a function, so the formula simplifies considerably; in particular, $X = \varphi(t)\nu(t)$ for a unit normal $\nu(t)$. Then,

$$\nabla_T \nu = \langle \nabla_T \nu, T\rangle T + \langle \nabla_T \nu, \nu\rangle = 0T\langle \nu, T\rangle - \langle \nu, \nabla_T T\rangle.$$

Thus, on S^2,

$$\frac{\partial^2 L}{\partial s^2}\bigg|_{s=0} = \int_a^b (\|\nabla_T X\|^2 - \langle R(T,X)X,T\rangle)\,\mathrm{d}t = \int_a^b ((\varphi')^2 - \varphi^2)\,\mathrm{d}t = I(\varphi,\varphi),$$

with the boundary condition $\varphi(0) = \varphi(\ell) = 0$. This is hauntingly familiar to the physics majors in the audience. By Fourier series, or Physics 131, or both, one can show that $I(\varphi,\varphi) \geq 0$ for all φ with $\varphi(0) = 0 = \varphi(\ell)$, then this works if $\ell \leq \pi$; after this, there is something in the null space. These variations correspond to taking the geodesic and rotating it. But then, after this, the geodesics are strictly unstable. This equation plays an important role in both Rienannian geometry and relativity: one can ask what it means for a timelike geodesic to maximize length and how this works.

One interesting characterization is how geodesics spread, related to the Jacobi equation in relativity. For example, in spaces of positive curvature, it says that global geodesics initially spread and then eventually converge.

Proof of (10). Using (9), one can compute the second derivative. Since T is a unit vector, then

$$\begin{aligned}
\frac{\partial^2}{\partial s^2}\|T\|\bigg|_{s=0} &= \frac{\partial}{\partial s}\langle \nabla_X T, T\rangle - \frac{\langle \nabla_X T, T\rangle^2}{\|T\|^{3/2}} \\
&= \frac{\partial}{\partial s}\langle \nabla_X T, T\rangle - \langle \nabla_X T, T\rangle^2 \\
&= \langle \nabla_X \nabla_T X, T\rangle + \langle \nabla_T X, \nabla_X T\rangle - \langle \nabla_T X, T\rangle^2.
\end{aligned}$$

Then, introduce a curvature term:

$$\begin{aligned}
&= \langle R(X,T)X,T\rangle + \langle \nabla_T \nabla_X X, T\rangle + \|\nabla_T X\|^2 - \langle \nabla_T X, T\rangle^2 \\
&= \langle R(X,T)X,T\rangle + \langle \nabla_T \nabla_X X, T\rangle + \|\nabla_T \widetilde{X}\|^2 \\
&= -\langle R(X,T)X,T\rangle + \frac{\partial}{\partial t}\langle \nabla_X X, T\rangle.
\end{aligned}$$

⊠

Today's class was taught by the TA, Khoa Nguyen.

Last time, we saw that if (M, g) is a Riemannian manifold and $c[a, b] \to M$ is a curve, then one can take a C^∞ variation of curves c_s with fixed endpoints and such that $c_0 = c$. This is equivalent to a C^∞ vector field $\frac{\mathrm{d}}{\mathrm{d}s}\big|_{s=0} c_s$ along the curve c with value 0 at the endpoints p and q, and in particular any such vector field gives a variation. Then, we defined the arc length functional $L(c) = \int_a^b \|\dot{c}\|^2 \,\mathrm{d}t$, and derived (9). In particular, this implied that if $\frac{\mathrm{d}L}{\mathrm{d}s} = 0$ for all variations c_s, then c is a geodesic, and then took the second derivative, which is (10).

Definition. Let $\mathcal{X}$ denote the set of vector fields on c that are perpendicular to its tangent $T = \dot{c}$. Then, the index form $\mathrm{Ind} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is given by

$$\mathrm{Ind}(X, Y) = \int_a^b \langle \nabla_T X, \nabla_T Y\rangle - \langle R(X, T)T, Y\rangle \,\mathrm{d}t.$$

Thus, $\frac{\mathrm{d}^2 L}{\mathrm{d}s^2} = \mathrm{Ind}(X, X)$. This is a bilinear form, so one can talk about when it's degenerate. The kernel of this form is defined to be those $X \in \mathcal{X}$ such that $\mathrm{Ind}(X, Y) = 0$ for all $Y \in \mathcal{X}$. Using metric compatibility, this implies that

$$\begin{aligned} 0 = \mathrm{Ind}(X, Y) &= \int_a^b \nabla_T\langle \nabla_T X, Y\rangle - \left\langle \nabla_T^2 X, Y\right\rangle - \langle R(X, T)T, Y\rangle \,\mathrm{d}t \\ &= \int_a^b \left\langle -\nabla_T^2 X - R(X, T)T, Y\right\rangle \mathrm{d}t \end{aligned}$$

for every Y. Thus, it's necessary that

$$\nabla_T^2 X + R(X, T)T = 0. \tag{11}$$

This is known as the Jacobi field equation.

Intuitively, if one has a function $f : W \to \mathbb{R}$ for a manifold W, then the critical points form a submanifold W_{crit}. Then, the Hessian provides information about these points: its kernel $\ker \mathrm{Hess}_p(f)$ contains $T_p W_{\mathrm{crit}}$. In this specific situation, $W = \Omega_{pq}$, the space of paths between p and q, which is slightly different (it's infinite-dimensional, called a Banach manifold; techniques from functional analysis are needed to understand it); the analogy still works, as we're trying to find critical points of the arc length L.

Remark. The idea that's just come up is to look at a manifold by analyzing its critical points under some real-valued function. This leads the way to a subject called Morse theory, which is extremely important in differential and geometric topology. The standard reference is a fantastic book called *Morse Theory*, by Milnor.

The Exponential Mapping. Given a $p \in M$ and $v \in T_pM$, there is a unique geodesic γ_v such that $\gamma_v(0) = p$ and $\gamma_v'(0) = v$, as we have seen before, and if v is sufficiently small, then $\gamma_v(1)$ is well-defined. The notation $\mathcal{N}_M(p)$ denotes a neighborhood of $p \in M$; then, the exponential map $\exp_p : \mathcal{N}_{T_pM}(0) \to \mathcal{N}_M(p)$ sending $v \mapsto \gamma_v(1)$ is well-defined.

Proposition 14.1. *$\exp_p$ is differentiable and $(D\exp_p)_0 = \mathrm{Id} : T_pM \to T_pM$.*

Moreover, by the Inverse Function Theorem, $\exp_p$ is locally a diffeomorphism near 0.

Remark. A line through 0 in T_pM is sent under the exponential map to a geodesic of M. This isn't necessarily true of lines that don't intersect the origin!

Proof of Theorem 14.1. That $\exp_p$ is differentiable follows from the theory of ODEs, but calculating its derivative shows that

$$(D\exp_p)_0(v) = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \exp_p(tv) = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \gamma_{tv}(1) = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \gamma_v(t) = v,$$

so $(D\exp_p)_0 = \mathrm{Id}$. ⊠

Remark. The exponential mapping defines a local chart at p in which the Christoffel symbols vanish at p. These coordinates are called normal coordinates. This is particularly helpful when computing intrinsic quantities; any chart will do, so this one makes some calculations easier.

With the exponential map, one can locally construct geodesic variations: if T is tangent to c and W is normal to it, then one can take combinations of them (e.g. $T + W$) to get a family of straight lines through the origin in T_pM and therefore a family of geodesics through c in M. Specifically, for each s, $c_s(t) = \exp_p(t(T + sW))$ is a geodesic variation of $c = c_0$; then, $Y(t) = \frac{\mathrm{d}}{\mathrm{d}s}\big|_{s=0} c_s(t)$ is a Jacobi field. Technically, their endpoints aren't the same, but it satisfies (11) regardless.

The name "exponential" seems kind of weird, but it comes from Lie theory, where it makes a little more sense.

If c is a geodesic, let J_c denote the space of Jacobi fields along c. This is a vector space, and $T \in J_c$ (i.e. the velocity vector field), because plugging in T for X satisfies (11). Similarly, $tT \in J_c$ for all t. A particularly important property is that if $X, Y \in J_c$, then $\langle \nabla_T X, Y\rangle - \langle \nabla_T Y, X\rangle$ is constant along c. This looks complicated, but in the case $Y = T$, then $\langle \nabla_T X, T\rangle$ is constant, so $\nabla_T \langle X, T\rangle$ is constant. Thus, $\langle X, T\rangle$ is linear in t. This ends up meaning that the exponential map sends circles around the origin to curves perpendicular to all of the geodesics, and for the $Y(t)$ constructed above, $\langle Y(t), T\rangle = 0$ for all t.

Theorem 14.2 (Length distortion)**.** *Assume $W \perp T$ and $\|W\| = \|T\| = 1$. Then, near $t = 0$,*

$$\|Y(t)\|^2 = t^2 - \frac{1}{3}Kt^4 + O(t^6),$$

where K is the sectional curvature of the plane spanned by T and W.

This is in some sense a Taylor expansion (so the proof is just four differentiations followed by the Jacobi equation), placing a bound on how much circles centered at the origin can change under the exponential map. There's another, related theorem that uses the sectional curvature to measure the area of the circle, rather than its length.

15. Uniqueness of Constant-Curvature Spaces: 2/27/14

There's an alternate notation for the exponential map, different from the one used in the book: if $p \in M$, then $\exp_p : T_pM \to M$ is given by $\exp_p(v) = c_v(1)$, where c_v is the geodesic such that $c_v(0) = p$ and $c_v'(0) = p$. If one defines $v = tV$ where $\|V\| = 1$, then it is also true that $\exp_p(v) = c_v(t)$. This is because of the scaling property of geodesics: for any $\lambda > 0$, $c_{\lambda v}(T) = c_v(\lambda t)$, which is true because these are both solutions to the same ODE with the same initial conditions, and thus they must be the same, i.e. if $\gamma(t) = c_v(\lambda t)$, then $\gamma(0) = p$ and $\gamma'(0) = \lambda c_v'(\lambda t)|_{t=0} = \lambda v$.

The terminology comes from Lie groups: if the manifold is also a group, e.g. the orthogonal group, then the exponential map is also the exponential of a matrix. Another simple case is S^1, the unit circle. The identity element's tangent space is a copy of the imaginary line, and the exponential map sends $it \mapsto e^{it}$. We also saw last time that $D\exp_p\big|_0 = \mathrm{id} : T_pM \to T_pM$.

There is also a notion of normal coordinates, given by $\varphi = \exp_p^{-1} : U \to T_pM \cong \mathbb{R}^n$ given by an orthonormal basis $e_1, \dots, e_n$. In these coordinates, $g_{ij} = \delta_{ij} + O(|x|^2)$, and $\Gamma^i_{jk}(0) = 0$. In some sense, this is as close to the Euclidean metric as possible. Furthermore, this map preserves radial length:

$$\sum_{j=1}^n g_{ij}(x)x^j = x^i.$$

This also encodes the Gauss lemma, which says that spherical and radial directions are orthogonal: if v is spherical and x is radial, then

$$\sum_{i,j} g_{ij}x^jv^i = \sum x^iv^i = 0.$$

Additionally, the rays are geodesics, which is more general but not necessarily as important and the Gauss lemma. Another way of stating the lemma is that if $v, w \in T_pM$ are orthogonal, with $\langle v, w\rangle_p = 0$, then

$$\left\langle D\exp_p\big|_v(w), D\exp_p\big|_v(v)\right\rangle_{\exp_p(v)} = 0.$$

Now, what happens to $D\exp|_v : T_pM \to T_{\exp_p(v)}M$ when $v \neq 0$? Calculating this at some w is akin to understanding an arrow $v \to w$, which suggests the natural curve $v + sw$. Thus,

$$D\exp_p\big|_v(w) = \frac{\mathrm{d}}{\mathrm{d}s}\exp_p(t + sw)\bigg|_{s=0}.$$

But now, to actually calculate the exponential map, take rays through the origin, so for $0 \le t \le 1$, we're looking at $\exp_p(t(v + sw)) = C(t, s)$. Now, we can apply variational principles; let $Y = \frac{\partial C}{\partial s}$ and $T = \frac{\partial C}{\partial t}$. Then, the derivative we want to calculate is $Y(1, 0)$, and the goal is to solve the Jacobi equation, the linearization of the geodesic equation:

$$Y''(t) + R(Y, T)T = 0, \tag{12}$$

for $0 \le t \le 1$. Here, the notation is $Y''(T) = \nabla_T\nabla_T Y$. The initial conditions for this are $Y(0) = 0$, encoding that we're starting at v, and $Y'(0) = w$.

The motivating goal here is to use the exponential map to relate Euclidean geometry to that of a more general metric. To compute where a vector goes, one solves (12) to obtain a solution $Y(t) = Y(t, 0)$ (i.e. $s = 0$ is suppressed from the notation) and evaluate it at $t = 1$. Intuitively, the curvature is used to obtain the metric by integrating it twice, even though it's not really possible to do this explicitly in practice. Nonetheless, it is an extremely important tool in Riemannian and Lorentz geometry.

We also had Theorem 14.2. The essence of this theorem is that a family of geodesics with the same initial derivative spread linearly, so the theorem compares the Euclidean spread and that under the metric. In Euclidean space, $\|Y(t)\|^2_{\exp_p(tv)} = t^2$, but more generally a fourth-order correction is given by the sectional curvature: $-(1/3)K_{\{v,w\}}t^4$. Thus, the sectional curvature determines how quickly the geodesics separate: positive curvature gives a focusing effect on geodesics, and negative ones spread them.

The proof is Theorem 7.16 in the book, and is a straight calculation, taking four derivatives.

Let $c(t)$ be some geodesic parameterized by arc length, so that $\|c'(t)\|^2 = 1$, and let Y be a Jacobi field along c. Since Y is given by a solution to a second-order system with n coefficients, the Jacobi fields form a space of dimension $2n$ (parameterized by the initial conditions). Notice that $T = c'(t)$ is a Jacobi field, because $T''(t) = 0$ and $R(T,T) = 0$, so (12) holds, and the same holds true for $tT(t)$; in fact, every Jacobi field Y can be written as $Y = aT + btT + Y^\perp$, where $Y^\perp$ is orthogonal to T. This can be done by choosing a parallel, orthonormal basis of vector fields $E_0(t) = T(t), E_1(t), \dots, E_{n-1}(t)$. Then, denote $Y(t) = \sum_{i=0}^{n-1} a_i(t)E_i(t)$, so

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\begin{pmatrix} a^0 \\ \vdots \\ a^{n-1}\end{pmatrix} + A(t)\begin{pmatrix} a^0 \\ \vdots \\ a^{n-1}\end{pmatrix} = 0$$

for some $n \times n$ matrix A. To understand this better, see that

$$\begin{aligned} 0 &= \langle Y''(t) + R(Y,T)T, E_i\rangle \\ &= \frac{\mathrm{d}^2}{\mathrm{d}t^2}a^i + \sum_j a^j \langle R(E_j,T)T, E_i\rangle. \end{aligned}$$

Thus, $A = (a_{ij})$, where $a_{ij} = \langle R(E_i,T)T, E_i\rangle$. On a general manifold, the curvature tensor is difficult to evaluate along a curve, this is tricky to calculate, but in special cases, this is quite nice.

For example, suppose g is a metric of constant curvature κ, so that $R = \kappa R_1$. Then, a_{ij} can be simplified to

$$a_{ij} = \kappa(\underbrace{\langle T,T\rangle\langle E_i,E_j\rangle - \langle E_j,T\rangle\langle E_i,T\rangle}_{\langle R_1(E_j,T)T,E_i\rangle}).$$

But since $\langle T,T\rangle = 1$, then $a_{ij} = \kappa\delta_{ij}$, and the overall Jacobi system is

$$\frac{\mathrm{d}^2a^i}{\mathrm{d}t^2} + \kappa a^i = 0, \qquad i = 1, \dots, n-1.$$

(When $i = 0$, the curvature tensor vanishes.) This is a simple harmonic oscillator, giving sines and cosines, or hyperbolic sines and cosines if the curvature is negative. Then, $a^i(0) = a_0$ and denote $\frac{\mathrm{d}a^i}{\mathrm{d}t}(0) = \alpha^i$, and let $w = (\alpha^1, \dots, \alpha^n)$. Thus, the overall solution is

$$f_\kappa(r) = \begin{cases} \sin(\sqrt{\kappa}r)/\sqrt{\kappa}, & \text{if } \kappa > 0 \\ r, & \text{if } \kappa = 0 \\ \sinh(\sqrt{-\kappa}r)/\sqrt{-\kappa}, & \text{if } \kappa < 0. \end{cases}$$

The conclusion is that if (M,g) is a manifold with constant curvature κ, then for any $p \in M$ and the normal coordinates $x^1, \dots, x^n$, let

$$r = \sqrt{\sum (x^i)^2};$$

then, the metric can be explicitly written down as

$$g = \mathrm{d}r^2 + f_\kappa(r)^2\,\mathrm{d}s_1^2,$$

where $\mathrm{d}s_1^2$ is the metric on S^{n-1}.

For example, if $n = 2$, then in normal coordinates, (r,θ) looks like $g = \mathrm{d}r^2 + f^2(r,\theta)\,\mathrm{d}\theta^2$, where $f(r,\theta) \to 1$ as $r \to 0$. Then, f satisfies the Gauss equation with respect to r, for each θ, so in two dimensions it's particularly easy to describe. However, the curvature isn't constant over all θ. In some sense, the distortion is given by the second term.

In the more general case, for $n \geq 3$, in normal coordinates, $g = \mathrm{d}r^2 + \tilde{g}(r)$, where $\tilde{g}$ is a family of metric of metrics on S^{n-1} parameterized by r, and $\tilde{g}(r)/r^2$ approaches the standard metric $\mathrm{d}s_1^2$.

Corollary 15.1. *Let (M_1, g_1) and (M_2, g_2) be manifolds of the same constant curvature K. Then, for any $p_1 \in M_1$ and $p_2 \in M_2$ with neighborhoods U_1 and U_2 respectively, there is a local isometry between them, i.e. an isometry $F : U_1 \to U_2$.*

This map F is just given by $\exp_{p_2} \circ I \circ \exp_{p_1}^{-1}$, where I is any linear isometry, so that $DF|_{p_1} = I$, so that we can choose I to fit some constraint. This is the same thing as saying that the metric has the same form when written in normal coordinates, which we saw above.

The global version isn't true in general: $\mathbb{R}^2$ and the torus are locally, but globally, isometric. The following theorem provides some more insight, though its proof requires a little bit of covering spaces.

Theorem 15.2. *Let M_1 and M_2 be complete and simply connected manifolds of constant curvature K. Then, M_1 and M_2 are globally isometric.*

Note that the stipulation of simple-connectedness (i.e. every curve can be contracted to a point) eliminates the torus.

Definition. Here, completeness means that geodesics extend forever (which is a nontrivial condition, because the geodesic equation is nonlinear). Formally, a Riemannian manifold is said to be geodesically complete if all geodesics extend for infinite arc length.

There are other, equivalent notions of completeness defined in different ways. This rules out the possibility of boundaries, but is satisfied by any compact manifold. An equivalent criterion to geodesic completeness is that $\exp_p : T_pM \to M$ is globally defined (though even in these cases, it's not necessarily a global diffeomorphism, e.g. on S^n).

16. The Hilbert-Einstein Action: 3/4/14

"I know a lot of you are interested in physics, and the rest of you are interested in science fiction..."

If M is a smooth manifold, one can assign the *Hilbert-Einstein action* to (Riemannian or Lorentz) metrics of M. This is a functional given by

$$\mathcal{R}(g) = \int_M R(g)\,\mathrm{d}V_g.$$

Then, the goal is to pick g which extremize $\mathcal{R}(\cdot)$. But how does one vary metrics? Let h be a symmetric $(0,2)$-tensor with compact support. Then, for $-\varepsilon < s < \varepsilon$, $g_s = g + sh$ is still a metric.

Theorem 16.1.

$$\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} \mathcal{R}(g_s) = -\int_M \langle G, h\rangle\,\mathrm{d}V_g,$$

where $G = \operatorname{Ric} - (1/2)R \cdot g$ is the Einstein tensor.

This inner product comes from a natural extension of the inner product to tensor spaces: if h and k are $(0,2)$-tensors, then

$$\langle h, k\rangle = \sum_{i,j,k,\ell} g^{ij} g^{k\ell} h_{ik} h_{j\ell}.$$

Corollary 16.2. *For $n = 2$, $G = 0$, so $\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} \mathcal{R}(g_s) = 0$.*

In this case, every metric is a stationary point for $\mathcal{R}$, which we can use in the following.

Theorem 16.3 (Gauss-Bonnet)**.** *Let M be a sphere with γ handles. Then*

$$\int_M K\,\mathrm{d}A = 4\pi(1-\gamma).$$

Proof. This theorem is more typically a Math 143 result (and also can be stated in a bit more generality), and the proof would involve triangulating the surface and using the Euler characteristic. But a more interesting proof can be given here.

Claim. If g is a Riemannian metric on a 2-dimensional manifold M, then $\int_M K\,\mathrm{d}A$ is independent of g.

Proof. Given two metrics g and $\widehat{g}$, let $g_S = (1-s)g + s\widehat{g}$ for $0 \le s \le 1$. Then, by Theorem 16.1, $\frac{\mathrm{d}}{\mathrm{d}s}\mathcal{R}(g_s) = 0$, so $\mathcal{R}(g) = \mathcal{R}(\widehat{g})$. ⊠

Thus, in this case $\mathcal{R}$ is a topological invariant: it doesn't depend on the metric. In particular, to calculate it, one can pick any metric that makes computation easier. For example, in the normal metric,

$$\int_{S^2} K\,\mathrm{d}A = 4\pi,$$

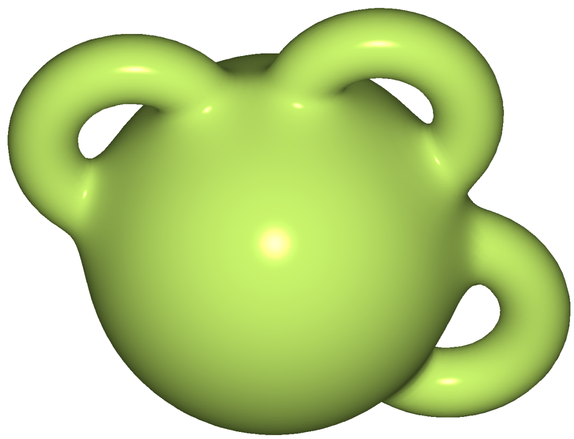

FIGURE 1. A sphere with three handles: all closed, connected, orientable 2-manifolds are homeomorphic to a sphere with γ handles for some $\gamma \geq 0$. Source: `http://en.wikipedia.org/wiki/Handle_decomposition`.

which has $\gamma = 0$, and for the flat torus T^2, for which $\gamma = 1$,

$$\int_{T^2} K \, \mathrm{d}A = 0 = 4\pi(1-\gamma).$$

Now, we want to show that when one adds a handle, the integral decreases by 4π, i.e. if M_γ is a sphere with n handles and $M_{\gamma+1}$ is formed by attaching a handle to M_γ, then

$$\int_{M_{\gamma+1}} K \, \mathrm{d}A = \int_{M_\gamma} K \, \mathrm{d}A - 4\pi = 4\pi(1-(\gamma+1)). \tag{13}$$

Given such an M_γ, one can construct an $M_{\gamma+1}$ in an interesting topological process. One can create a "neck" N by revolving a curve such as $y = 1 - \cos x$ on $[0, \pi]$ around the x-axis, and then use it to join M_γ with T^2, which has one handle.

Then, N has total curvature -4π: suppose it's used to connect two spheres S_1 and S_2. This is topologically a sphere again, so

$$4\pi = \int_{S_1 \cup N \cup S_2} K \, \mathrm{d}A = \int_{S_1} K \, \mathrm{d}A + \int_{S_2} K \, \mathrm{d}A + \int_N K \, \mathrm{d}A = 8\pi + \int_N K \, \mathrm{d}A.$$

Then, inductively create $M_{\gamma+1}$ from M_γ and T^2 by joining a neck. Then,

$$\int_{M_{\gamma+1}} K \, \mathrm{d}A = \int_{M_\gamma} K \, \mathrm{d}A - 4\pi + 0. \qquad \boxtimes$$

As we're used to proofs by metric calculations in this class, this more topological proof seems like it ought to have holes somewhere, but it's perfectly sound, and everything can be explicitly written out if you insist on it.

Now, we can return to the variation of $\mathcal{R}$. The textbook gives a hopelessly complicated proof of Theorem 16.1, but here's a simpler one.

Proof of Theorem 16.1. Choose some coordinates $x^1, \dots, x^n$, so that

$$\mathcal{R}(g_s) \, \mathrm{d}V_{g_s} = \sum_{i,j} g_s^{ij} \mathcal{R}_{ij}(g_s) \sqrt{|g_s|} \, \mathrm{d}x,$$

so using $^\bullet$ to denote derivatives with respect to s,

$$\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} \mathcal{R}(g_s) \, \mathrm{d}V_{g_s} = \underbrace{\sum_{i,j} (g^{ij})^\bullet \mathcal{R}_{ij} \sqrt{|g|}}_{I} + \underbrace{\sum_{i,j} g^{ij} \mathcal{R}_{ij}^\bullet \sqrt{|g|}}_{II} + \underbrace{\sum_{i,j} g^{ij} \mathcal{R}_{ij} (\sqrt{|g|})^\bullet}_{III}.$$

But we saw on the homework that $(g^{ij})^\bullet = -h^{ij}$ and that

$$(\log \sqrt{|g|})^\bullet = \frac{(\sqrt{|g|})^\bullet}{\sqrt{|g|}} = \frac{1}{2} \sum_{k,\ell} g^{k\ell} h_{k\ell},$$

and therefore

$$(\sqrt{|g|})^\bullet = \frac{1}{2}\operatorname{Tr}_g(h)\sqrt{|g|}.$$

The key is that *III* is the divergence of a vector field, so the answer is given by

$$\begin{aligned}\int_M (I + II)\,\mathrm{d}x &= -\int_M \left(\langle \mathrm{Ric}, h\rangle - \frac{1}{2}R\langle g, h\rangle\right)\sqrt{|g|}\,\mathrm{d}x \\ &= -\int_M \langle G, h\rangle\,\mathrm{d}V.\end{aligned}$$

Then, we still have to deal with the third component.

Lemma. *There exists a vector field X with compact support such that $\operatorname{div}(X) = \sum_{i,j=1}^n g^{ij}R^\bullet_{ij}$, and therefore $\int_M III = 0$.*

Proof. First, calculate R_{ij}:

$$R_{ij} = \partial_k \dot\Gamma^k_{ij} - \partial_j \dot\Gamma^k_{ik} + \Gamma\Gamma - \Gamma\Gamma \tag{14}$$

(where the last four Christoffel symbols will go away soon enough, so their indices don't matter). This is a $(1,2)$-tensor.

For any $p \in M$, one can choose normal coordinates $x^1, \dots, x^n$, so that $\Gamma^k_{ij} = 0$ (i.e. $\partial_k g_{ij} = 0$) because

$$g_{ij}(0) = \begin{cases} \delta_{ij}, & g \text{ is Riemannian;} \\ \eta_{ij}, & g \text{ is Lorentz.}\end{cases}$$

Thus, the terms quadratic in Γ in (14) do indeed go away, and so when $x = 0$,

$$g^{ij}\dot R_{ij} = \partial_j(g^{ik}\dot\Gamma^j_{ik}) - \partial_j(g^{ij}\dot\Gamma^k_{ik}) = \partial_j(\xi^j),$$

where (using the summation convention) $\xi^j = g^{ik}\dot\Gamma^j_{ik} - g^{ij}\dot\Gamma^k_{ik}$. Thus, if one lets $X = \xi^j\partial_j$, then X has the same support as h, so it's a compactly supported vector field such that $\operatorname{div}(X) = III$. ⊠

This means that the third term of the integral vanishes, finishing the proof. ⊠

Corollary 16.4. *If $n \geq 3$, then g is a critical point of $\mathcal{R}(\cdot)$ iff g satisfies the* vacuum Einstein equation $G(g) = 0$, *or equivalently* $\mathrm{Ric}(g) = 0$ *(which can be shown by expanding out the definition of G and taking the trace).*

Corollary 16.5. *If $n \geq 3$ and g is Riemannian, then g is critical for metrics of fixed volume iff $\mathrm{Ric}(g) = cg$ for some constant c.*

This is associated with the Euler-Lagrange functional.

Proof. Supposing that $\mathrm{Vol}(g_s) = \mathrm{Vol}(g_0)$, then

$$0 = \frac{\mathrm{d}}{\mathrm{d}s}\int_M \sqrt{g}\,\mathrm{d}x = \frac{1}{2}\int_M \operatorname{Tr}_g(\dot g)\,\mathrm{d}V.$$

Thus, when $h = \dot g$ (i.e. $\frac{\mathrm{d}}{\mathrm{d}s}\big|_{s=0} g_s$), then $\int_M \operatorname{Tr}(g_h)\,\mathrm{d}V = 0$, so $\int_M \langle G, h\rangle\,\mathrm{d}V = 0$ for all h.

Thus, in general over the variation,

$$0 = \int_M \operatorname{Tr}_g(h - cg) = \int_M \operatorname{Tr}_g(h) - n\cdot c\,\mathrm{Vol}(M),$$

but in the first integral, $\operatorname{Tr}_g(h - cg) = \langle c, h - cg\rangle$, so

$$\int_M \langle G - \alpha g, h - cg\rangle\,\mathrm{d}V = 0$$

for all α, and thus it holds true for the unique α for which $\int_M \langle g, G - \alpha g\rangle = 0$. Then, $cg = 0$, so $\int_M \langle G - \alpha g, h\rangle\,\mathrm{d}V = c$ for a constant c. Thus, $\mathrm{Ric} - (1/2)Rg = \alpha g$, so $R - (n/2)R = \alpha$, and thus R is constant and $\mathrm{Ric} = cg$, so g is Einstein. ⊠

In relativity, this constant c is called the cosmological constant.

17. The Schwarzschild Solution: 3/6/14

18. The Maximally Extended Schwarzschild Metric: 3/11/14

"Es ist immer angenehm, über strenge Lösungen einfacher Form zu verfügen." [It is always pleasant to have exact solutions in simple form at your disposal.] - Karl Schwarzschild

The Schwarzschild metric and other similar ones (e.g. the Kerr metric) are local solutions, asymptotically flat. Next lecture, we will deal with metrics and ideas that are supposed to model the entire universe.

The idea of null coordinates is to consider the Minkowski metric $\mathrm{d}s^2 = -\mathrm{d}t^2 + \mathrm{d}x^2$ on $\mathbb{R}^2_1$. Suppose $(t(s), x(s))$ is null, so that $-(\dot{t})^2 + (\dot{x})^2 = 0$. Thus, $\dot{t} = \pm\dot{x}$, or $\mathrm{d}t = \pm\,\mathrm{d}x$. This can be solved by integration: $t \pm x = k$ for a constant k. Thus, we introduce *null coordinates* $u = t - x$ and $v = t + x$. In these coordinates, the diagonal terms of the metric are zero, because the curve is null, so $\mathrm{d}s^2 = -\,\mathrm{d}u\,\mathrm{d}v$.

This notation might be a little unintuitive: usually, one writes $g_{ij}\,\mathrm{d}x^i \otimes \mathrm{d}x^j$, and specifies that $g_{ij} = g_{ji}$ (where $g_{ij} = \langle\partial_i, \partial_j\rangle$). But then, one could write this as $g_{ij} = \mathrm{d}x^i\,\mathrm{d}x^j$, which doesn't have to be symmetric. Thus, it's necessary to symmetrize, so

$$\mathrm{d}u\,\mathrm{d}v = \frac{1}{2}(\,\mathrm{d}u \otimes \mathrm{d}v + \mathrm{d}v \otimes \mathrm{d}u).$$

Thus, $\langle\partial_u, \partial_u\rangle = \langle\partial_v, \partial_v\rangle = 0$ and $\langle\partial_u, \partial_v\rangle = 0$.

This choice of null coordinates is much more general than the Minkowski metric. In particular, we can adapt them to the Schwarzschild metric

$$\mathrm{d}s^2 = -\left(1 - \frac{2m}{\rho}\right)\mathrm{d}t^2 + \left(1 - \frac{2m}{\rho}\right)^{-1}\mathrm{d}\rho^2,$$

defined on the surface of a sphere for $-\infty < t < \infty$ and $\rho > 2m$ (so that it's nonsingular). Setting this to zero, one needs to integrate

$$\mathrm{d}t = \pm\frac{\mathrm{d}\rho}{1 - 2m/\rho}.$$

Thus,

$$\begin{aligned}
\rho_* &= \int \frac{1}{1 - 2m/\rho}\,\mathrm{d}\rho = \int \frac{\rho}{\rho - 2m}\,\mathrm{d}\rho \\
&= \int \left(\frac{\rho - 2m}{\rho - 2m} + \frac{2m}{\rho - 2m}\right)\mathrm{d}\rho \\
&= \rho + 2m\log\left(\frac{\rho}{2m} - 1\right).
\end{aligned}$$

Thus, assign $u = t - \rho_*$ and $v = t + \rho_*$. In these coordinates, the metric looks like

$$\mathrm{d}s^2 = -\left(1 - \frac{2m}{\rho}\right)\mathrm{d}u\,\mathrm{d}v.$$

But this still has a ρ in it, so substitute in

$$\begin{aligned}
\frac{v - u}{4m} &= \frac{\rho}{2m} + \log\left(\frac{\rho}{2m} - 1\right). \\
e^{(v-u)/4m} &= e^{\rho/2m}\left(\frac{\rho}{2m} - 1\right) = \frac{\rho}{2m}e^{\rho/2m}\left(1 - \frac{2m}{\rho}\right),
\end{aligned}$$

so the metric becomes

$$\begin{aligned}
\mathrm{d}s^2 &= -\frac{2m}{\rho}e^{-\rho/2m}\left(e^{(v-u)/4m}\,\mathrm{d}u\,\mathrm{d}v\right) \\
&= -\frac{2m}{\rho}e^{-\rho/2m}\,\mathrm{d}U\,\mathrm{d}V,
\end{aligned}$$

where $U = -4me^{-u/4m}$ and $V = 4me^{-v/4m}$, so that

$$\mathrm{d}U\,\mathrm{d}V = \frac{1}{16m^2}e^{(v-u)/4m}\,\mathrm{d}u\,\mathrm{d}v.$$

Now, the metric, which was previously only defined for $\rho > 2m$, makes sense for any $\rho > 0$. Plugging this back into the metric, we realize that

$$\mathrm{d}s^2 = -32m^3\frac{e^{-\rho/2m}}{\rho}\,\mathrm{d}U\,\mathrm{d}V$$

for $U, V > 0$. This is defined for all U and V as long as $\rho > 0$. This looks somewhat like the Minkowski metric, so let $T = (U+V)/2$ and $X = (V-U)/2$, so that $U = T - X$ and $V = T + X$. Thus,

$$\left(\frac{\rho}{2m} - 1\right)e^{\rho/2m} = X^2 - T^2$$

and

$$\frac{t}{2m} = \log\left(\frac{X+T}{X-T}\right),$$

where in the second equation, $X + T > 0$ and $X - T > 0$. Now, we have $-UV = X^2 - T^2 > -1$, so the maximally extended Schwarzschild metric is

$$g = \frac{32m^3 e^{-\rho/2m}}{\rho}\left(-\,\mathrm{d}T^2 + \mathrm{d}X^2\right) + \rho^2 g_{S^2},$$

where g_{S^2} is the spherical metric and $X^2 - T^2 > -1$. It's not immediately clear why it's maximal, but this is it.

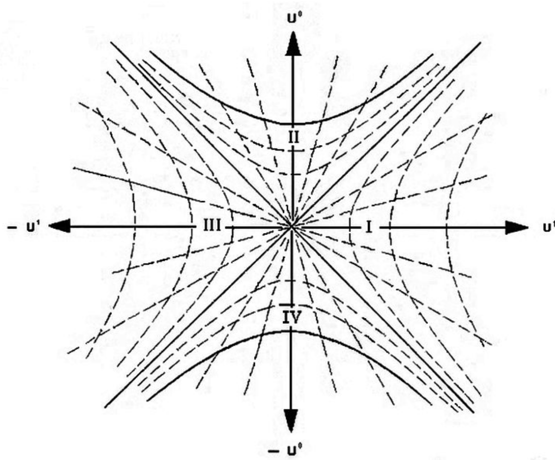

FIGURE 2. Regions of the extended Schwarzschild metric (suppressing the ϕ and θ directions). The horizontal axis is the X-axis, and the vertical axis is the T-axis.

Looking at Figure 2, region I is the original metric, for which $\rho > 2m$, and all four regions correspond to the new metric (the hyperbola is $\rho = 0$, i.e. $X^2 - T^2 = -1$. Furthermore, regions I and III are isometric, as are II and IV. Region II is called the black hole region, and $\rho = 2m$ is called the event horizon.

Where this gets interesting is that in the Lorentz metric, there's causality and futures and such. Two points can communicate if their future lightcones intersect (so that in the Minkowski metric, any two points can communicate). But here, any point within region II cannot communicate with the outside world, and will probably end up at the singularity! However, it is possible for a point outside of this to end up inside of it (i.e. there's a timelike curve that starts outside of region II, crosses the event horizon, and then is in region II). Basically, it's a one-way gate. But if someone traveling into the black hole emits signals to an observer outside of it at regular intervals, then to the observer the signals seem to get farther and farther apart in time, and after a finite number they stop. But it would probably be unpleasant to fall into the singularity — the curvature goes to infinity.

Region IV is called the "white hole," and is the reflection of Region II. Regions III and IV are more science-fictiony; the former is in some sense an alternate universe, and there wouldn't be any communication between two points in these regions unless they elected to both go into the black hole. The line $t = 0$ is asymptotic to Minkowski space, becoming very flat asymptotically, and moving into region III near $T = 0$, it forms a wormhole or Einstein-Rosen bridge. Again, how much of this is science fiction? Of course, though, since this is a spacelike hypersurface, you can't just move along it. But regions I and II should be taken seriously: if one takes some matter field and contracts it, with everything else in a vacuum, then one ends up with the Schwarzschild metric. Interestingly, this dynamical collapse (e.g. of a star) leads to the black hole which doesn't change. This can be done mathematically rigorously.[15]

There's a more general set of solutions corresponding to rotating systems, e.g. the Kerr metric of a stationary, rotating black hole. If one takes a general star which might not be rotationally symmetric, but axially symmetric, it

[15] A lot of this depends on the assumption that the universe follows the Einstein equations and Lorentz geometry, but that's one of the core assumptions of relativity, and region I at least has been well tested, though the dynamics haven't been that well tested. Inside the black hole, there should be quantum effects too. . .

will collapse into a rotating black hole. But it's a general mathematical conjecture that arbitrary collapsing matter collapses into one of these states, and it's going to be open for some time.

19. Friedman-Robertson-Walker Spaces: 3/13/14

"Any other questions I can't answer?"

Friedman-Robertson-Walker spaces are actually a relatively simple class of manifolds, but are very important in cosmology and general relativity. The idea is that the metric is dictated by symmetries, and is homogeneous (i.e. there is no special location in spacetime, or more formally there's an isometry from any point to any other point) and isotropic (i.e. an observer sees the same thing in all directions). Friedman-Robinson-Walker spaces are homogeneous and isotropic spatially, i.e. $\mathcal{S} = I \times M$ for Riemannian manifolds M (spacelike hypersurfaces) that are homogeneous and isotropic.

Definition. Let (M, g) be a Riemannian manifold.

- M is homogeneous if for any $p, q \in M$, there is an isometry F of M such that $F(p) = q$.
- M is isotropic at a $p \in M$ if for all $v, w \in T_pM$ with $\|v\| = \|w\| = 1$, there exists an isometry F such that $F(p) = p$ and $DF_p(v) = w$.

Theorem 19.1. *For $n = 2$ and $n = 3$, a homogeneous isotropic Riemannian manifold has constant curvature.*

Proof. For $n = 2$, it's already true that homogeneous implies constant curvature, because if F is an isometry, then the Gauss curvature satisfies $K(F(p)) = K(p)$, so it's constant.

For $n = 3$, this isn't true, e.g. $S^1 \times S^2$ in the product metric, which is homogeneous but doesn't have constant curvature (it's homogeneous because S^1 and S^2 are, so isometries between pairs of points are given by pairs of isometries). And when $n \geq 4$, there are homogeneous, isotropic Riemannian manifolds that aren't of constant curvature, which follows because there are subgroups of the orthogonal group that act isotropically (i.e. transitively), but aren't the whole group. For example, looking at $\mathrm{U}(2) \subsetneq \mathrm{SO}(4)$, the former is the set of complex linear transformations of $\mathbb{C}^2$, and the latter is the set of (generally) linear transformations that preserve the dot product. Then, one could choose some space (e.g. $\mathbb{C}P^2$) on which this leads to a counterexample.

Notice that for any n, homogeneous and isotropic implies (Riemannian) Einstein, i.e. $\mathrm{Ric}(g) = cg$, because by the Spectral Theorem, there's an orthonormal basis $e_1, \dots, e_n$ in which $R_{ij} = \lambda_i \delta_{ij}$, but since the space is isotropic, then $\lambda_i = \lambda_j$ (there's a transformation preserving the Ricci tensor that sends any coordinate to any other coordinate).

But for $n = 3$, Einsteinian metrics have constant curvature, as we saw. ⊠

Conveniently enough, since cosmologists are interested in four-dimensional spacetime, then $\mathcal{S} = I \times M$ means M is three-dimensional. Thus, if it's homogeneous and isotropic, then it must have constant curvature κ (which can be scaled to $0, 1$, or -1). Thus, in a Friedman-Robinson-Walker space, the metric has the frm $g = -\,\mathrm{d}\tau^2 + a^2(\tau) g_\kappa$, where g_κ has constant curvature $\kappa \in \{-1, 0, 1\}$.

There's still some ambiguity as to what M looks like, but within cosmology it is often assumed that M has no fundamental group, i.e. is simply connected. Thus, $M = M_\kappa$ is one of $\mathbb{R}^n$, S^n, or H^n (for $\kappa = 1$, 0, and -1, respectively). (More generally, you might have the quotient by something.)

These aren't vacuum solutions, so there's a matter field, which is also induced from the symmetries of the metric. Thus, in this case the Einstein equation is $G = 8\pi T$ for the stress-energy tensor T (and where $G = \mathrm{Ric}(g) - (1/2)Rg$ as usual). T encodes the density of the matter and energy fields that are present. Since energy and momentum go together into one vector (and which is which is observer-dependent), one can use this stress-energy tensor. If an observer is at vector $v = \sum_{b=0}^{3} v^b e_b$, and the mass density is $\sum_b T_{ab} v^b v^a$, which represents the mass density as seen by the observer at v, and if one took the perpendicular components, one gets the momentum density.

Some matter fields require adding more data (e.g. if one uses electromagnetism, the Maxwell equations have to be factored in somehow). But if this isn't the case, this is called a perfect fluid, and is akin to modeling the universe as a fluid. In general, a fluid is determined by a velocity vector field, so let u be a unit timelike vector field, ρ be the density, and P be pressure. Thus, in these coordinates, the stress-energy tensor has the form

$$T_{ab} = \rho u_a u_b + P(g_{ab} u_a u_b).$$

This is equivalent to saying that if an observer moves along a fluid path, then in the timelike direction the component is ρ, and in the spacelike directions is P (so the identity matrix times ρ in the first entry, and P in all of the others).

Returning to the formula for the FRW metric, there are two components to the Einstein tensor, in the timelike and spatial directions; the latter will be by symmetry a multiple of the metric. The remaining terms vanish (which can be a homework exercise, though wasn't this time around). Thus,

$$G_{\tau\tau} = R_{\tau\tau} + \frac{1}{2}R = 3\frac{(\dot{a})^2}{a^2} + \frac{3K}{a^2} = 8\pi\rho$$
$$G_{\text{sp}} = R_{\text{sp}} - \frac{1}{2}R = -2\frac{\ddot{a}}{a} - \frac{(\dot{a})^2}{a^2} - \frac{\kappa}{a^2} = 8\pi P.$$

But this looks just like the equation for a perfect fluid, so it seems like we'd need to automatically satisfy the Euler equations, but these follow from the Einstein ones, so it's OK. In particular, for a perfect fluid, $\operatorname{div} G = 0$, so $\operatorname{div} T = 0$ as well. This latter equality is the content of the relativistic Euler equations. We want to add some physically plausible conditions, i.e. $P, \rho \geq 0$. But there are still three unknowns, a, P, and ρ, and not enough equations. So one imposes something called the equation of state, a way of relating the pressure and the density, $P = P(\rho)$, which takes a specific value depending on the type of fluid. For example, a fluid with $P = 0$ is called dust, and $P = \rho/3$, which held early on in the universe, is called radiation.

Now, we can solve the Euler equations, using this equation of state, so the system becomes

$$3\frac{(\dot{a})^2}{a^2} + \frac{3\kappa}{a^2} = 8\pi\rho$$
$$3\frac{\ddot{a}}{a} = -4\pi(\rho + 3P).$$

The behavior of these models depends heavily on the sign of κ; for example, if $\kappa = 1$, so that the universe is closed, then the universe expands and then contracts in a Big Crunch; in the flat and negatively curved cases, it always lies above, and grows ($O(\tau^{2/3})$ for the flat case, and faster for the hyperbolic case). Explicitly solving all of the ODEs is not entirely pleasant; there's a conserved quantity for the system which is not quite obvious.

When $P = 0$, the equation simplifies to

$$3\dot{a}^2 + 3\kappa = 8\pi\rho a^2,$$

so after differentiating,

$$6\dot{a}\ddot{a} = 16\pi\rho a\dot{a} + 8\pi\dot{\rho}a^2.$$

This further simplifies to $-\dot{a}\rho = 2\rho\dot{a} + \dot{\rho}a$, so in particular $\dot{\rho}a + 3\rho\dot{a} = 0$. Since a is a positive function, then this implies that $\frac{\mathrm{d}}{\mathrm{d}t}(a^3\rho) = 0$, and thus that $a^3\rho$ is constant! This can be used to turn one of the above equations into a first-order equation, which can then be integrated. (In the radiation case, it's the same, except for $a^4\rho$ being constant instead.) One can also argue that these are constant from physical principles; these are in some sense rest masses of the systems.

These Friedman-Robinson-Walker models are the currently accepted models of the universe, though sometimes there's a cosmological constant added. They are used more for explaining data rather than predicting it, but the data is generally consistent with this model. This is pretty impressive, given how simple this is, and how it adds matter (unlike the Schwarzschild metric). But recently detected accelerating expansion, which isn't completely mathematically happy, requires adding a cosmological constant to factor in dark energy: $G = 8\pi T + \Lambda g$, where Λ is called the cosmological constant. There's a lot of open research and interesting, unanswered questions here.