M382D NOTES: DIFFERENTIAL TOPOLOGY

ARUN DEBRAY
JANUARY 29, 2016

These notes were taken in UT Austin's Math 382D (Differential Topology) class in Spring 2016, taught by Lorenzo Sadun. 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. Thanks to Adrian Clough, Parker Hund, and Thérèse Wu for finding and correcting a few mistakes.

CONTENTS

Lecture 1.

The Inverse and Implicit Function Theorems: 1/20/16

"The most important lesson of the start of this class is the proper pronunciation of my name [Sadun]: it rhymes with 'balloon.' "

We're basically going to march through the textbook (Guillemin and Pollack), with a little more in the beginning and a little more in the end; however, we're going to be a bit more abstract, talking about manifolds more abstractly, rather than just embedding them in $\mathbb{R}^n$, though the theorems are mostly the same. At the beginning, we'll discuss the analytic underpinnings to differential topology in more detail, and at the end, we'll hopefully have time to discuss de Rham cohomology.

Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$. Its derivative is $\mathrm{d}f$; what exactly is this? There are several possible answers.

- It's the best linear approximation to f at a given point.
- It's the matrix of partial derivatives.

What we need to do is make good, rigorous sense of this, moreso than in multivariable calculus, and relate the two notions.

Definition. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is *differentiable* at an $a \in \mathbb{R}^n$ if there exists a linear map $L : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{h\to 0} \frac{|f(a+h)-f(a)-L(h)|}{|h|} = 0. \tag{1.1}$$

In this case, L is called the *differential* of f at a, written $\mathrm{d}f|_a$.

Note that $h \in \mathbb{R}^n$ and the numerator is in $\mathbb{R}^m$, so it's quite important to have the magnitudes there, or else it would make no sense.

Another way to rewrite this is that $f(a+h) = f(a)+L(h)+o(\text{small})$, i.e. along with some small error (whatever that means). This makes sense of the first notion: L is a linear approximation to f near a. Now, let's make sense of the second notion.

Theorem 1.1. *If f is differentiable at a, then $\mathrm{d}f$ is given by the matrix $\left(\frac{\partial f^i}{\partial x^j}\right)$.*

Proof. The idea: if f is differentiable at a, then (1.1) holds for $h \to 0$ along any path!

So let's take $\mathbf{e}_j$ be a unit vector and $h = t\mathbf{e}_j$ as $t \to 0$ in $\mathbb{R}$. Then, (1.1) reduces to

$$L(t\mathbf{e}_j) = \frac{f(a_1, a_2, \dots, a_j + t, a_{j+1}, \dots, a_n) - f(a)}{t},$$

and as $t \to 0$, this shows $L(\mathbf{e}_j)^i = \frac{\partial f^i}{\partial x^j}$. ⊠

In particular, if f is differentiable, then all partial derivatives exist. The converse is *false*: there exist functions whose partial derivatives exist at a point a, but are not differentiable. In fact, one can construct a function whose directional derivatives all exist, but is not differentiable! There will be an example on the first homework. The idea is that directional derivatives record linear paths, but differentiability requires all paths, and so making things fail along, say, a quadratic, will produce these strange counterexamples.

Nonetheless, if all partial derivatives exist, then we're almost there.

Theorem 1.2. *Suppose all partial derivatives of f exist at a and are continuous on a neighborhood of a; then, f is differentiable at a.*

In calculus, one can formulate several "guiding" ideas, e.g. the whole change is the sum of the individual changes, the whole is the (possibly infinite) sum of the parts, and so forth. One particular one is: *one variable at a time*. This principle will guide the proof of this theorem.

Proof. The proof will be given for $m = 2$ and $n = 1$, but you can figure out the small details needed to generalize it; for larger n, just repeat the argument for each component.

We want to compute

$$\begin{aligned} &f(a_1 + h_1, a_2 + h_2) - f(a_1, a_2) \\ &\qquad = f(a_1 + h_1, a_2 + h_2) - f(a_1 + h_1, a_2) + f(a_1 + h_1, a_2) - f(a_1, a_2) \end{aligned}$$

Regrouping, this is two single-variable questions. In particular, we can apply the mean value theorem: there exist $c_1, c_2 \in \mathbb{R}$ such that

$$\begin{aligned} &= \left.\frac{\partial f}{\partial x^2}\right|_{(a_1+h_1, a_2+c_2)} h_2 + \left.\frac{\partial f}{\partial x^1}\right|_{(a_1+c_1, a_2)} h_1 \\ &= \left(\left.\frac{\partial f}{\partial x^1}\right|_{a_1+c_1, a_2} - \left.\frac{\partial f}{\partial x^1}\right|_a \right) h_1 + \left(\left.\frac{\partial f}{\partial x^2}\right|_{a_1+h_1, a_2+c_2} - \left.\frac{\partial f}{\partial x^2}\right|_a \right) h_2 + \left(\left.\frac{\partial f}{\partial x^1}\right|_a, \left.\frac{\partial f}{\partial x^2}\right|_a \right) \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}, \end{aligned}$$

but since the partials are continuous, the left two terms go to 0, and since the last term is linear, it goes to 0 as $h \to 0$. ⊠

We'll often talk about *smooth* functions in this class, which are functions for which all higher-order derivatives exist and are continuous. Thus, they don't have the problems that one counterexample had.

Since we're going to be making linear approximations to maps, then we should discuss what happens when you perturb linear maps a little bit. Recall that if $L : \mathbb{R}^n \to \mathbb{R}^m$ is linear, then its image $\mathrm{Im}(L) \subset \mathbb{R}^m$ and its kernel $\ker(L) \subset \mathbb{R}^n$.

Suppose $n \le m$; then, L is said to have *full rank* if $\operatorname{rank} L = n$. This is an open condition: every full-rank linear function can be perturbed a little bit and stay linear. This will be very useful: if a (possibly nonlinear) function's differential has full rank, then one can say some interesting things about it.

If $n \ge m$, then full rank means rank m. This is once again stable (an open condition): such a linear map can be written $L = (A \mid B)$, where A is an invertible $m \times m$ matrix, and invertibility is an open condition (since it's given by the determinant, which is a continuous function).

To actually figure out whether a linear map has full rank, write down its matrix and row-reduce it, using Gaussian elimination. Then, you can read off a basis for the kernel, determining the free variables and the relations determining the other variables. In general, for a k-dimensional subspace of $\mathbb{R}^n$, you can pick k variables arbitrarily and these force the remaining $n - k$ variables. The point is: *the subspace is the graph of a function*.

Now, we can apply this to more general smooth functions.

Theorem 1.3. *Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is smooth, $a \in \mathbb{R}^n$, and $\mathrm{d}f|_a$ has full rank.*

(1) (Inverse function theorem) *If $n = m$, then there is a neighborhood U of a such that $f|_U$ is invertible, with a smooth inverse.*
(2) (Implicit function theorem) *If $n \geq m$, there is a neighborhood U of a such that $U \cap f^{-1}(f(a))$ is the graph of some smooth function $g : \mathbb{R}^{n-m} \to \mathbb{R}^m$ (up to permutation of indices).*
(3) (Immersion theorem) *If $n \leq m$, there's a neighborhood U of a such that $f(U)$ is the graph of a smooth $g : \mathbb{R}^n \to \mathbb{R}^m$.*

This time, the results are local rather than global, but once again, full rank means (local) invertibility when $m = n$, and more generally means that we can write all the points sent to $f(a)$ (analogous to a kernel) as the graph of a smooth function.

It's possible to sharpen these theorems slightly: instead of maximal rank, you can use that if $\mathrm{d}f|_a$ has block form with the square block invertible, then similar statements hold.

The content of these theorems, the way to think of them, is that in these cases, smooth functions locally behave like linear ones. But this is not too much of a surprise: differentiability means exactly that a function can be locally well approximated by a linear function. The point of the proof is that the higher-order terms also vanish.

For example, if $m = n = 1$, then full rank means the derivative is nonzero at a. In this case, it's increasing or decreasing in a neighborhood of a, and therefore invertible. On the other hand, if the derivative is 0, then bad things happen, because it's controlled by the higher-order derivatives, so one can have a noninvertible function (e.g. a constant) or an invertible function whose inverse isn't smooth (e.g. $y = x^3$ at $x = 0$).

This is not the last time in this class that maximal rank implies nice analytic results.

We're going to prove (2); then, as linear-algebraic corollaries, we'll recover the other two.

Lecture 2.

The Contraction Mapping Theorem: 1/22/16

Today, we're going to prove the generalized inverse function theorem, Theorem 1.3. We'll start with the case where $m = n$, which is also the simplest in the linear case (full rank means invertible, almost tautologically).

Theorem 2.1. *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be smooth. If $\mathrm{d}f|_a$ is invertible, then*

(1) f is invertible on a neighborhood of a,
(2) f^{-1} is smooth on a neighborhood of a, and
(3) $\mathrm{d}(f^{-1})|_{f(a)} = (\mathrm{d}f|_a)^{-1}$.

Proof of part (1). Without loss of generality, we can assume that $a = f(a) = 0$ by translating. We can also assume that $\mathrm{d}f|_a = I$, by precomposing with $\mathrm{d}f|_a^{-1}$:

$$\begin{tikzcd} \mathbb{R}^n \arrow[r, "f"] & \mathbb{R}^n \\ \mathbb{R}^n \arrow[u, "\mathrm{d}f|_a^{-1}"] \arrow[ur] & \end{tikzcd}$$

If we prove the result for the diagonal arrow, then it is also true for f. Since the domain and codomain of f are different in this proof, we're going to call the former X and the latter Y, so $f : X \to Y$.

Now, since f is smooth, its derivative is continuous, so there's a neighborhood of a in X given by the x such that $\|\mathrm{d}f|_x - I\| < 1/2$.[1] And by shrinking this neighborhood, we can assume that it is a closed ball C.

[1]There are many different norms on the space of $n \times n$ matrices, but since this is a finite-dimensional vector space, they are all equivalent. However, for this proof we're going to take the *operator norm* $\|A\| = \sup_{v \in S^{n-1}} |Av|$.

On C, f is injective: if $x_1, x_2 \in C$, then since C is convex, then there's a line $\gamma(t) = x_1 + tv$ (where $v = x_2 - x_1$) joining x_1 to x_2, and $\frac{\mathrm{d}f}{\mathrm{d}t} = (\mathrm{d}f|_{\gamma(t)})v$. Therefore

$$\begin{aligned} f(x_2) - f(x_1) &= \left(\int_0^1 \mathrm{d}f|_{\gamma(t)}\,\mathrm{d}t\right)v \\ &= \int_0^1 \big((\mathrm{d}f|_{\gamma(t)} - I) + I\big)v\,\mathrm{d}t \\ &= x_2 - x_1 + \int_0^1 (\mathrm{d}f|_{\gamma(t)} - I)v\,\mathrm{d}t. \end{aligned}$$

We can bound the integral:

$$\left|\int_0^1 \big(\mathrm{d}f|_{\gamma(t)} - I\big)v\right| \le \int_0^1 \left|(\mathrm{d}f|_{\gamma(t)} - I)v\right|\mathrm{d}t \le \int_0^1 \frac{1}{2}|v|\,\mathrm{d}t = \frac{|v|}{2}.$$

Thus, since $x_2 - x_1 = v$, then $f(x_2) - f(x_1)$ has magnitude at least $v/2$, so in particular it can't be zero. Thus, f is injective on C. The point is, since $\mathrm{d}f$ is close to the identity on C, we get an error term that we can make small.

To construct an inverse, we need to make it surjective on a neighborhood of $f(a)$ in Y. The way to do this is called the contraction mapping principle, but we'll do it by hand for now and recover the general principle later.

To be precise, we'll iterate with a "poor-man's Newton's method:" if $y \in Y$, then given x_n, let $x_{n+1} = x_0 - (f(x_0) - y) = y + x_0 - f(x_0)$ (since we're using the derivative at the origin instead of at x, and this is just the identity). A fixed point of this iteration is a preimage of y. Specifically, we'll want $x_0 = a$, since we're trying to bound the distance of our fixed point from a.

Since

$$x_{n+1} - x_n = y + x_n - f(x_n) - (y + x_{n-1} - f(x_{n-1})) = (x_n - x_{n-1}) - (f(x_n) - f(x_{n-1})),$$

then $|x_{n+1} - x_n| < (1/2)|x_n - x_{n-1}|$, so in particular, this is a Cauchy sequence! Thus, it must converge, and to a value with magnitude no more than $2|y|$ (since $f(x_0) = f(a) = 0$). Thus, if C has radius R, then for any y in the ball of radius $1/2$ from the origin (in Y), y has a preimage x, so f is surjective on this neighborhood. ⊠

Now, we can discuss the contraction mapping principle more generally.

Definition. Let X be a complete metric space and $T : X \to X$ be a continuous map such that $d(T(x), T(y)) \le cd(x, y)$ for all $x, y \in X$ and some $c \in [0, 1)$. Then, T is called a *contraction mapping*.

Theorem 2.2 (Contraction mapping principle)**.** *If X is a complete metric space and T a contraction mapping on X, then there's a unique fixed point x (i.e. $T(x) = x$).*

Proof. Uniqueness is pretty simple: if T has two fixed points x and x' such that $x \ne x'$, then $d(T(x), T(x')) \le cd(x, x') = d(T(x), T(x'))$, and $c < 1$, so this is a contradiction, so $x = x'$.

Existence is basically the proof we just saw: pick an arbitrary $x_0 \in X$ and let $x_{n+1} = T(x_n)$. Then, $d(x_m, x_n) \le c^{|n-m-1|}d(x_n, x_{n-1})$, so this sequence is Cauchy, and has a limit x. Then, since T is continuous, $T(x) = x$. ⊠

Now, back to the theorem.

Proof of Theorem 2.1, part (2). Once again, we assume $f(0) = 0$. By the fundamental theorem of calculus, on our neighborhood of 0,

$$y = f(x) = \int_0^1 \mathrm{d}f|_{tx}(x)\,\mathrm{d}t.$$

Since we assumed $\mathrm{d}f|_0 = I$, and f is smooth, then $\mathrm{d}f$ is continuous, so for any $\varepsilon > 0$, there's a neighborhood U of 0 such that for all $x \in U$, $\mathrm{d}f|_x = I + A$, where $\|A\| < \varepsilon$. When we integrate this, this means $y = x + o(|x|)$: $\mathrm{d}f$ is "small in x." Hence, $|x| - \varepsilon < |y| < |x| + \varepsilon$, so since U is bounded, this puts a bound on x in terms of y, too; in other words, $x = y + o(|y|)$ (this is little-o, because we can do this for any $\varepsilon > 0$, though the neighborhood may change). This is exactly what it means for f^{-1} to be differentiable at $y = f(0)$, and its derivative is the identity! In general, if $\mathrm{d}f|_0 \ne I$, but is still invertible, then we get that $\mathrm{d}f^{-1}|_{f(0)} = (\mathrm{d}f|_0)^{-1}$.

We'd like this to extend to a neighborhood of the origin. Since $\mathrm{d}f|_0$ is invertible, and $\mathrm{d}f$ is continuous, then locally a neighborhood of 0 corresponds to a neighborhood of $\mathrm{d}f|_0$ in the space of $n \times n$ matrices, and vice versa.

But the set of invertible matrices is open in the space of matrices, so there's a neighborhood V of 0 such that $\mathrm{d}f|_x$ is invertible for all $x \in V$, so for each $x \in V$, $\mathrm{d}f^{-1}|_{f(x)} = (\mathrm{d}f|_x)^{-1}$. Then, matrix inversion is a continuous function on the subspace of invertible matrices, so this means $\mathrm{d}f^{-1}$ is continuous in a neighborhood of $f(0)$.

This gives us one derivative; we wanted infinitely many. Using the chain rule,

$$\frac{\partial(\mathrm{d}f^{-1})}{\partial y} = \frac{\partial(\mathrm{d}f)^{-1}}{\partial x}\frac{\partial x}{\partial y},$$

and $\frac{\partial x}{\partial y} = (\mathrm{d}f)^{-1}$. So we want to understand derivatives of matrices. Let A be some invertible matrix-valued function, so that $AA^{-1} = I$. Thus, using the product rule, $A'A^{-1} = A(A^{-1})' = 0$, so rearranging, $(A^{-1})' = A^{-1}A'A^{-1}$. That is, the derivative inverse can be specified in terms of the inverse and the derivative of A. In particular, this means $\frac{\partial(\mathrm{d}f^{-1})}{\partial y}$ is a product of continuous functions ($\frac{\partial(\mathrm{d}f)}{\partial x}$ and $(\mathrm{d}f)^{-1}$), so it is continuous. By the same argument, so is the partial derivative in the x-direction, so by Theorem 1.2, $\mathrm{d}f^{-1}$ is differentiable. This can be repeated as an inductive argument to show that $\mathrm{d}f^{-1}$ is differentiable as many times as $\mathrm{d}f$ is, and by smoothness, this is infinitely often. ⊠

We can use this to recover the rest of Theorem 1.3 as corollaries.

Proof of Theorem 1.3, part (2). First, for the implicit function theorem, let $n > m$ and $f : \mathbb{R}^n \to \mathbb{R}^m$ be smooth with full rank, and choose a basis in which $\mathrm{d}f|_a = (A \mid B)$ in block form, where A is an invertible $m \times m$ matrix. The theorem statement is that we can write the first m coordinates as a function of the last $n-m$ coordinates: specifically, that there exists a neighborhood U of a such that $U \cap f^{-1}(f(a)) = U \cap \{g(y), y\}$ for some smooth $g : \mathbb{R}^{n-m} \to \mathbb{R}^m$.[2]

Now, the proof. Let $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^{n-m}$, and let

$$F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} f(x,y) \\ y \end{pmatrix}.$$

Hence,

$$\mathrm{d}F|_a = \left(\begin{array}{c|c} A & B \\ \hline 0 & I \end{array}\right).$$

This is invertible, since A is: $\det(\mathrm{d}F|_a) = \det(A) \neq 0$. Thus, we apply the inverse function theorem to F to conclude that a smooth F^{-1} exists, and so if π_1 denotes projection onto the first component, $x = \pi_1 \circ F^{-1}(0, y) = g(y)$. ⊠

Lecture 3.

Manifolds: 1/25/16

"Erase any notes you have of the last eight minutes! But the first 40 minutes were okay."

Recall that we've been discussion Theorem 1.3, a collection of results called the inverse function theorem, the implicit function theorem, and the immersion theorem. These are local (not global) results, and generalize similar results for linear maps: not all matrices are square, but if a matrix has full rank, it can be written in two blocks, one of which is invertible. Using this with $\mathrm{d}f|_a$ as our matrix is the idea behind proving Theorem 1.3: the first several variables determine the remaining variables.

However, we don't know which variables they are: you may have to permute $x_1, \dots, x_n$ to get the last variables as smooth functions of the first ones. For example, for a circle, the tangent line is horizontal sometimes (so we can't always parameterize in terms of x_2) and vertical at other times (so we can't only use x_1).

Before we prove the immersion theorem (part (3) of Theorem 1.3), let's recall what tools we use to talk about curves in the plane.

(1) A common technique is using a *parameterized curve*, the image of a smooth $\gamma(t) : \mathbb{R} \to \mathbb{R}^2$ whose derivative is never zero (to avoid singularities). For example, $f(t) = (t^2, t^3)$ has a zero at the origin, but the curve one obtains is $y = \pm x^{3/2}$, which has a cusp at $(0,0)$. This is the content of the immersion theorem.

(2) Another way to describe curves is as level sets: $f(x,y) = c$, most famously the circle. This is the content of the implicit function theorem: this looks like a graph-like curve locally.

(3) This brings us to the most simple method: graphs of functions, just like in calculus.

[2]For example, if $n = 2$ and $m = 1$, consider $f(x) = |x|^2 - 1$, and $a = (\cos\theta, \sin\theta)$. Then, $f^{-1}(f(a))$ is the unit circle, so the implicit function is telling us that locally, the circle is a function of x_1 in terms of x_2, or vice versa.

And the point of Theorem 1.3 is that these three approaches give you the same sets, *up to permutation of variables* (and that a curve is the graph of a function only locally). We have these three pictures of what higher-dimensional surfaces look like.

And that means that when we talk about manifolds, which are the analogue of higher-dimensional surfaces, we should keep these things in mind: a manifold may be defined abstractly, but we understand manifolds through these three visualizations.

Proof of Theorem 1.3, part (3). We're going to prove the equivalent statement that if the first n rows of $\mathrm{d}f|_a$ are linearly independent, then the remaining $m-n$ variables are smooth functions in the first n.

Since $f : \mathbb{R}^n \to \mathbb{R}^m$, then let π_1 denote projection onto the first n coordinates, so we have a commutative diagram

$$\begin{tikzcd} \mathbb{R}^n \arrow[r, "f"] \arrow[dr, "\pi_1 \circ f"'] & \mathbb{R}^m \arrow[d, "\pi_1"] \\ & \mathbb{R}^n. \end{tikzcd}$$

In block form, $\mathrm{d}f|_a = \binom{A}{B}$, where A is invertible, and therefore $\mathrm{d}(\pi_1 \circ f)|_a = A$. This is invertible, so $(\pi_1 \circ f)^{-1}$ has an inverse in a neighborhood of a, by the inverse function theorem. Thus, if π_2 denotes projection onto the last $m-n$ coordinates, then $g = \pi_2 \circ f \circ (\pi_1 \circ f)^{-1}$ writes the last $m-n$ coordinates in terms of the first n, as desired. ⊠

Now, we're ready to talk about smooth manifolds.

Definition. A *k-manifold* X in $\mathbb{R}^n$ is a set that locally looks like one of the descriptions (1), (2), or (3) for a smooth surface. That is, it satisfies one of the following descriptions.

(1) For every $p \in X$, there's a neighborhood U of p where one can write $N-k$ variables in smooth functions of the remaining k variables, i.e. there is a neighborhood $V \subset \mathbb{R}^k$ and a smooth $g : V \to \mathbb{R}^{N-k}$ such that $X \cap U = \{(x, g(x)) : x \in V\}$ (up to permutation).

(2) X is locally the image of a smooth map, i.e. for every $p \in X$, there's a neighborhood U of p and a smooth $f : \mathbb{R}^k \to \mathbb{R}^N$ with full rank such that the image of f in U is $X \cap U$. This is the "parameterized curve" analogue.

(3) Locally, X is the level set of a smooth map $f : \mathbb{R}^N \to \mathbb{R}^{N-k}$ with full rank.

If k is understood from context (or not important), X will also be called a *manifold*.

The big theorem is that these three conditions are equivalent, and this follows directly from Theorem 1.3.

For example, suppose we have the graph of a smooth function $y = x^2$. How can we write this as the image of a smooth map? Well, $(x, y) = (t, t^2)$ has nonzero derivative, and we can do exactly the same thing (locally) for a manifold in general. And it's the level set $f(x, y) = 0$, where $f(x, y) = y - x^2$, and the same thing works (locally) for manifolds: for a general graph $\mathbf{y} = g(\mathbf{x})$, this is the level set of $f(\mathbf{y}, \mathbf{x}) = \mathbf{y} - g(\mathbf{x})$, whose derivative $\mathrm{d}f$ has block matrix form $(I \mid -\mathrm{d}g)$, which has full rank. Neat.

And perhaps most useful for now, something that's locally a graph is really easy to visualize: it's the bedrock on which one first defined curves and surfaces.

Now, that's a manifold in $\mathbb{R}^n$. As far as Guillemin and Pollack are concerned, that's the only kind of manifold there is, but we want to talk about abstract manifolds, but that means we'll need one more important property.

Suppose $X \subset \mathbb{R}^N$ is a manifold, and $p \in X$. We're going to look at a neighborhood of p as the image of a smooth $g_1 : \mathbb{R}^k \to \mathbb{R}^N$; this is the most common and most fundamental description of a manifold. However, this is not in general unique; suppose $g_2 : \mathbb{R}^k \to \mathbb{R}^N$ lands in a different neighborhood of p — though, by restricting to their intersection, we can assume we have two smooth maps (sometimes called *charts*) into the same neighborhood, and they both have inverses, so we have a well-defined function $g_2^{-1} \circ g_1 : \mathbb{R}^k \to \mathbb{R}^k$. Is it smooth?

Theorem 3.1. *$g_2^{-1} \circ g_1$ is smooth.*

The key assumption here is that $\mathrm{d}g_1$ and $\mathrm{d}g_2$ both have maximal rank.

Definition. The *tangent space* to X at p, denoted T_pX, is $\mathrm{Im}(\mathrm{d}g_1|_{g_1^{-1}(p)})$; it is a k-dimensional subspace of $\mathbb{R}^N$.

This is the set of velocity vectors of paths through p, which makes sense, because such a path must come from a path downstairs in $\mathbb{R}^k$, since g_1 is locally invertible.

Lemma 3.2. *The tangent space is independent of choice of g_1.*

The idea is that any velocity vector must come from a path in both $\mathrm{Im}(\mathrm{d}g_1|_{g_1^{-1}(p)})$ and $\mathrm{Im}(\mathrm{d}g_2|_{g_2^{-1}(p)})$, so these two images are the same.

Then, we'll punt the proof of Theorem 3.1 to next lecture.

Lecture 4.

Abstract Manifolds: 1/27/16

Last time, we were talking about change of variables, but we were missing a lemma that's important for the proof, but not really the right way to view manifolds.

Let X be a k-dimensional manifold in $\mathbb{R}^n$, so for any $p \in X$, there's a map ϕ from the neighborhood of the origin in $\mathbb{R}^k$ to a neighborhood of p in X, where $\phi(0) = p$ and $\mathrm{d}\phi|_0$ has rank k. We'd like a local inverse to ϕ, which we'll call F; it's a map from a neighborhood of $\mathbb{R}^n$ to a neighborhood of $\mathbb{R}^k$. We'd like F to be smooth, and we want $F \circ \phi = \mathrm{id}|_{\mathbb{R}^k}$.

By permuting coordinates, we can assume that the first k rows of $\mathrm{d}\phi$ are linearly independent. That is, $\mathrm{d}\phi|_0$ has block form $\left(\begin{smallmatrix} A \\ B \end{smallmatrix}\right)$, where A is invertible. Then, define $\widetilde{\phi} : \mathbb{R}^k \times \mathbb{R}^{n-k} \to \mathbb{R}^n$ sending $(x, y)^\mathrm{T} \to \phi(x) + (0, y)^\mathrm{T}$,[3] so that $\widetilde{\phi}(x, 0)) = \phi(x)$. ϕ and $\widetilde{\phi}$ fit into the following diagram.

$$\begin{tikzcd} \mathbb{R}^k \arrow[r, hook, "x \mapsto (x,0)"] \arrow[rr, bend right, "\phi"'] & \mathbb{R}^n \arrow[r, "\widetilde{\phi}"] & \mathbb{R}^n \end{tikzcd}$$

Thus, by the chain rule,

$$\mathrm{d}\widetilde{\phi}|_0 = \left(\begin{array}{c|c} A & 0 \\ \hline B & I \end{array}\right),$$

so $\mathrm{d}\widetilde{\phi}|_0$ has full rank! Thus, in a neighborhood of p, it has an inverse, and certainly the inclusion $\mathbb{R}^k \hookrightarrow \mathbb{R}^n$ has a left inverse π (projection onto the first k coordinates), so we can let $F = \pi \circ \widetilde{\phi}^{-1}$, because

$$F \circ \phi(x) = F \circ \widetilde{\phi}(x, 0) = \pi \circ \widetilde{\phi}^{-1} \circ \widetilde{\phi}((x, 0)) = \pi(x, 0) = x.$$

Likewise, $\phi \circ F = \mathrm{id}|_X$, since every point in our neighborhood is in the image of ϕ.

This is how we talk about smoothness on manifolds: we don't know what smoothness means on some arbitrary submanifold, so we'll use the fact that we can locally pretend we're in $\mathbb{R}^n$ to talk about smoothness.

Suppose $\phi, \psi : \mathbb{R}^k \rightrightarrows X$ are two such smooth coordinate maps; we'd like to find a smooth function g from a neighborhood in $\mathbb{R}^k$ to a neighborhood in $\mathbb{R}^k$ relating them (again, locally). But we have a local inverse to ψ called F, so since we want $\psi = \phi \circ g$, then define $g = F \circ \psi$, because $\phi \circ g = \phi \circ F \circ \psi = \psi$. And g is the composition of two smooth functions, so it's smooth (this is Theorem 3.1). This is our change-of-coordinates operation.

Theorem 4.1. *A function $g : X \to \mathbb{R}^m$ can be extended to a smooth map G on a neighborhood of p in $\mathbb{R}^n$ iff $g \circ \phi$ is smooth.*

This is another notion of smooth: the first one determines smoothness by coordinates, and the second says that smooth functions on a submanifold are restrictions of smooth functions $\mathbb{R}^n \to \mathbb{R}^m$. But the theorem says that they're totally equivalent.

Proof. Suppose such a smooth extension G exists; since $G|_X = g$ and $\mathrm{Im}(\phi) \subset X$, then $G \circ \phi = g \circ \phi$. G and ϕ are smooth, so $G \circ \phi = g \circ \phi$ is smooth.

Conversely, if $g \circ \phi$ is smooth, then let $G = g \circ \phi \circ F$, which is a smooth map (since it's a composition of two smooth functions) out of a neighborhood of p in $\mathbb{R}^n$. ⊠

Note that this extrinsic definition is the one Guillemin and Pollack use throughout their book; the other notion doesn't depend on an embedding into $\mathbb{R}^n$, but we had to check that it was independent of change of coordinates (which by Theorem 3.1 is smooth, so we're OK). This means we can make the following definition.

Definition.

- A *chart* $\mathbb{R}^k \to X$ for a topological space X is a continuous map that's a homeomorphism onto its image.

[3] $\widetilde{\phi}$ is pronounced "phi-twiddle."

- An *(abstract) smooth k-manifold* is a Hausdorff space X equipped with charts $\varphi_\alpha : \mathbb{R}^k \to X$ such that
  (1) every point in X is in the image of some chart, and
  (2) for every pair of overlapping charts φ_α and φ_β, the change-of-coordinates map $\varphi_\beta^{-1} \circ \varphi_\alpha : \mathbb{R}^k \to \mathbb{R}^k$ is smooth.

The definition is sometimes written in terms of neighborhoods in $\mathbb{R}^k$, so each chart is a map $U \to X$, where $U \subset \mathbb{R}^k$, but this is completely equivalent to the given definition, since $\tan : (-\pi/2, \pi/2) \to \mathbb{R}$ is a diffeomorphism (and there are many others, e.g. $e^x/(1+e^x)$). The point is that every point has a neighborhood homeomorphic to $\mathbb{R}^k$, even if we think of neighborhoods as little balls much of the time.

There are lots of different categories of manifolds: a C^n *manifold* has the same definition, but we require the change-of-coordinates maps to merely be C^n (n times continuously differentiable); an *analytic manifold* requires the change-of-coordinates maps to be analytic; and in the same way one can define *complex-analytic manifolds* (holomorphic change-of-coordinates maps) and *algebraic manifolds*. For a *topological manifold* we just require the change-of-coordinates maps to be continuous, which is always true for a covering of charts. But in this class, the degree of regularity we care about is smoothness.

Definition. Let X be a manifold and $f : X \to \mathbb{R}^n$ be continuous. Then, f is *smooth* if for every chart $\varphi_\alpha : \mathbb{R}^k \to X$, the composition $f \circ \varphi_\alpha$ is smooth.

This is just like the definition of smoothness for manifolds living in $\mathbb{R}^n$.

Example 4.2. Let X be the set of lines in $\mathbb{R}^2$ (*not* just the set of lines through the origin). This is a manifold, but we want to show this. Using point-slope form, we can define a map $\phi_1 : \mathbb{R}^2 \to X$ sending $(a,b) \mapsto \{(x,y) : y = ax+b\}$, which covers all lines that aren't vertical. We need to handle the vertical lines with another chart, $\phi_2 : \mathbb{R}^2 \to X$ sending $(c,d) \mapsto x = cy + d$.

These charts intersect for all lines that are neither vertical nor horizontal, so the change-of-coordinates map describes $c = 1/a$ and $d = -b/a$, i.e. $g(a) = (1/a, -b/a)$. And since we're restricted to non-vertical lines, $a \neq 0$, so this is smooth, and $g^{-1}(c,d) = (1/c, -d/c)$, which is also smooth (since we're not looking at horizontal lines). Thus, we're described X as a manifold.

It turns out that X is a Möbius band. A line may be described by a direction (an angle coordinate) and an offset (intersection with the x-axis, heading in the specified direction). However, there are two descriptions, given by flipping the direction: $(\theta, D) \sim (\theta + \pi, -D)$. Thus, this is the quotient of an infinitely long cylinder by half a rotation and a twist, giving us a Möbius band.

One thing we haven't talked much about is: why do manifolds need to be Hausdorff? This makes our example much less terrible: here's just one creature we avoid with this condition.

Example 4.3 (Line with two origins)**.** Take two copies of $\mathbb{R}^2$, and identify $(x,1) \sim (x,2)$ for all $x \neq 0$. Thus, we seem to have one copy of $\mathbb{R}$, but two different copies of the origin. The charts are perfectly nice: any interval on either copy of $\mathbb{R}$ is a chart for this space, but every neighborhood of one of the origins contains the other, so it isn't Hausdorff (it is T_1, though). See Figure 1 for a (not perfectly accurate) depiction of this space. We don't want to

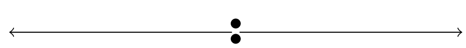

FIGURE 1. Depiction of the line with two origins. Note, however, that the two origins are technically infinitely close together.

have spaces like this one, so we require manifolds to be Hausdorff.

Tune in Friday to learn how to determine when two manifolds are equivalent. Is the same space with different charts a different manifold?

Lecture 5.

Examples of Manifolds and Tangent Vectors: 1/29/16

"How do you make the unit disc into a manifold? With pie charts."

Today, we're going to make the notion of a manifold more familiar by giving some more examples of what structures can arise: specifically, the 2-sphere S^2 and the projective spaces $\mathbb{RP}^n$ and $\mathbb{CP}^n$. Then, we'll move to discussing tangent vectors and how to define smooth maps between manifolds.

Example 5.1 (2-sphere)**.** The concrete 2-*sphere* is $S^2 = \{\mathbf{x} \in \mathbb{R}^3 : |\mathbf{x}|^2 = 1\}$. Why is this a manifold?

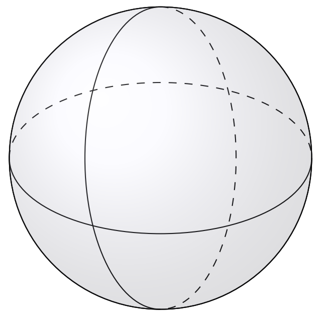

FIGURE 2. The 2-sphere, an example of a manifold.

We can put charts on this surface as follows: if $z > 0$, then we have a chart $(u, v, \sqrt{1-u^2-v^2})$, and if $z < 0$, then the chart is $(u, v, -\sqrt{1-u^2-v^2})$. Similarly, if $y > 0$, we have $(u, \sqrt{1-u^2-v^2}, v)$, and similarly for $y < 0$ and for x. However, since $\mathbf{0} \notin S^2$, then this covers all of S^3, and one can check that the transition maps are smooth and the chart maps have full rank.

Another way to realize this is that if $f : \mathbb{R}^3 \to \mathbb{R}$ is defined by $f(x, y, z) = x^2 + y^2 + z^2$, then f is smooth and $S^2 = f^{-1}(1)$. Thus, S^2 is the level set of a smooth function whose derivative $\mathrm{d}f = (2x, 2y, 2z)$ has full rank, so by the implicit function theorem, it must be a manifold.

That is, you can see S^2 is a manifold using maps into it, or maps out of it.

Example 5.2 (Real projective space)**.** $\mathbb{RP}^n$, *real projective space*, is defined to be the set of lines through the origin in $\mathbb{R}^{n+1}$. Any nonzero point in $\mathbb{R}^{n+1}$ defines a line through the origin, and scaling a point doesn't change this line. Thus, $\mathbb{RP}^n = \{\mathbf{r} \in \mathbb{R}^{n+1} \setminus 0\}/(\mathbf{r} \sim \lambda\mathbf{r}$ for $\lambda \in \mathbb{R} \setminus 0)$. We have coordinates $(x_0, \dots, x_n)$ for $\mathbb{R}^{n+1}$, and want to make coordinates on $\mathbb{RP}^n$.

The set $U_0 = \{\mathbf{x} : x_0 \neq 0\}$ is open, and $(x_0, x_1, \dots, x_n) \sim (1, x_1/x_0, \dots, x_n/x_0)$ in $\mathbb{RP}^n$, so we get a chart on U_0. We're parameterizing non-horizontal lines by their slope (or, well, the reciprocal of it). Thus, we have a map $\psi_0 : \mathbb{R}^n \to \mathbb{RP}^n$ sending $(x_1, \dots, x_n) \mapsto [(1, x_1, \dots, x_n)]$ (where brackets denote the equivalence class in $\mathbb{RP}^n$).

We can do this with every coordinate: let $\psi_1 : \mathbb{R}^n \to \mathbb{RP}^n$ send $(x_1, \dots, x_n) \mapsto [(x_1, 1, x_2, \dots, x_n)]$, and so forth. Then, since every point in $\mathbb{RP}^n$ has a nonzero coordinate, then this covers $\mathbb{RP}^n$. Are the transition maps smooth? $\mathbb{RP}^2$ will illustrate how it works: if $[1, a, b] = [c, 1, d]$, then $c = 1/a$ and $d = b/a$, which is smooth (because in these charts, a and c are nonzero).

By the way, $\mathbb{RP}^1$ is just a circle. More generally, one can also realize $\mathbb{RP}^n$ as the unit sphere with opposite points identified (every vector can be scaled to a unit vector, but then $\mathbf{x} \sim -\mathbf{x}$). However, $\mathbb{RP}^2$, etc., are more interesting spaces.

Example 5.3 (Complex projective space)**.** We can also refer to *complex projective space*, $\mathbb{CP}^n$. The idea of "lines through the origin" is the same, but, despite what algebraic geometers call it, a one-dimensional complex subspace looks a lot more like a (real) plane than a real line. In any case, one-dimensional complex subspaces of $\mathbb{C}^{n+1}$ are given by nonzero vectors, so we define $\mathbb{CP}^n = \{\mathbf{r} \in \mathbb{C}^{n+1} \setminus 0\}/(\mathbf{r} \sim \lambda\mathbf{r}, \lambda \in \mathbb{C} \setminus 0)$. Now, the same definitions of charts give us $\psi_k : \mathbb{C}^n \to \mathbb{CP}^n$, but since we know how to map $\mathbb{R}^{2n} \to \mathbb{C}^n$, this works just fine.

In this case, the first interesting complex projective space is $\mathbb{CP}^1$. Our two charts are $[1, a]$ and $[b, 1]$, and their overlap is everything but the two points $[1, 0]$ and $[0, 1]$. In other words, every point is of the form $[z, 1]$ for some $z \in \mathbb{C}$ or $[1, 0]$: that is $[1, 0]$ is a "point at infinity" ∞, whose reciprocal is 0! So $\mathbb{CP}^1$ is the complex numbers plus one extra point. We can actually realize this as S^2 using a map called *stereographic projection*: the sphere sits inside $\mathbb{R}^3$, and the xy-plane can be identified with $\mathbb{C}$. Then, the line between the north pole $(0, 0, 1)$ and a given $(u, v, 0)$ (corresponding to $[u + vi, 0]$) intersects the sphere at a single point, which is defined to be the image of the projection $\mathbb{CP}^1 \to S^2$. However, the point at infinity isn't identified in this way, and neither is the north pole;

thus, the north pole can be made the point at infinity. This is a great exercise to work out yourself, e.g. how it relates to the change of charts if you use the south pole instead. In fact, it will be on the homework![4]

Tangent vectors. In order to discuss tangent vectors concretely, we'll work in $\mathbb{R}^n$ for now. At every point $p \in \mathbb{R}^n$, there's a tangent space $T_p\mathbb{R}^n$ of vectors based at p, which is an n-dimensional vector space. And you can take the union of all of the tangent vectors and call it the *tangent bundle*: these are pairs (p, v), where $p \in \mathbb{R}^n$ and v is a vector originating at p. This is a $2n$-dimensional vector space, and this is cool and all, but it doesn't really tell us anything. We'd like a better way to characterize tangent vectors.

One way to define a tangent vector is the velocity vector of a smooth curve through p, and another way is as a derivation (or, as we saw on the homework, the directional derivatives $v = \sum v^i \partial_i$). These are related in a natural way: if $\gamma : \mathbb{R} \to \mathbb{R}^n$ is smooth and has $\gamma(0) = p$, and $f : \mathbb{R}^n \to \mathbb{R}$ is smooth, then one could ask how fast f changes along the path γ. This is

$$\frac{\mathrm{d}}{\mathrm{d}t}(f \circ \gamma)\bigg|_{t=0} = \sum_{i=1}^n \frac{\mathrm{d}\gamma^i}{\mathrm{d}t}\bigg|_{t=0} \frac{\partial f}{\partial x^i} = v \cdot \nabla f.$$

That is, the space of possible velocities *is* the space of directional derivatives: in the way we just described, curves do act as first-order differential operators. And in coordinates, the tangent vectors are just n-tuples of numbers (like with any basis). You'll need to be used to working with all of these perspectives and switching between them.

Now, let's generalize to an n-dimensional submanifold X of $\mathbb{R}^N$. For any $p \in X$, let $\phi : \mathbb{R}^n \to X$ send $a \mapsto p$; then, we can define the *tangent space* of X at p to be $T_pX = \mathrm{Im}(\mathrm{d}\phi|_a)$, which is necessarily an n-dimensional subspace of $\mathbb{R}^N$, as $\mathrm{d}\phi|_a$ has full rank. These are "vectors living at p," and we'll be able to relate these to velocities and directional derivatives, too.

However, we need to show that this is independent of chart: if $\psi : b \mapsto p$ is another chart for X, we know that in neighborhoods of a, b, and p, the change-of-coordinates is a diffeomorphism $g : b \mapsto a$. Then, $\psi = \phi \circ g$, and these are smooth, so the chain rule says $\mathrm{d}\psi|_b = \mathrm{d}\phi_| a \circ \mathrm{d}g|_b$. But since g is a diffeomorphism, $\mathrm{d}g_b$ is invertible, so its image is all of $\mathbb{R}^n$; thus, $\mathrm{Im}\,\mathrm{d}\psi|_b = \mathrm{d}\phi|_a(\mathbb{R}^n) = \mathrm{Im}(\mathrm{d}\phi|_a)$, and this is indeed independent of coordinates.

Thus, since $T_pX \subset \mathbb{R}^N$, then we can realize the tangent bundle as $TX \subset T\mathbb{R}^N$: $TX = \{(p, v) \mid p \in X \text{ and } v \in T_pX\}$. This tangent bundle sits inside $T\mathbb{R}^N = \mathbb{R}^{2N}$, so we know what it means for it to be a manifold, and can write down charts, and so forth.

Another interesting insight is that smooth curves through p correspond to smooth curves through $a \in \mathbb{R}^n$ through ϕ, and so we can relate the other definitions of tangent vectors to this definition of T_pX. The point is: local coordinates allow us to translate the notions of tangent vectors to submanifolds of $\mathbb{R}^N$; we'll be able to turn this into talking about abstract manifolds and derivatives of maps between manifolds.

[4] Stereographic projection works for the n-sphere and $\mathbb{R}^n$ for all n, so $S^n = \mathbb{R}^n \cup \{\infty\}$, in a sense; however, it won't correspond to projective space in higher dimensions.