M382D NOTES: DIFFERENTIAL TOPOLOGY

ARUN DEBRAY
FEBRUARY 10, 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* $\|A\| = \sup\limits_{v \in S^{n-1}} |Av|$.

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, $\mathrm{d}f|_a=\left(\begin{smallmatrix}A\\B\end{smallmatrix}\right)$, 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 $\operatorname{Im}(\mathrm{d}g_1|_{g_1^{-1}(p)})$; 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 have spaces like this one, so we require manifolds to be Hausdorff.

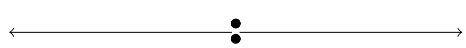

FIGURE 1. Depiction of the line with two origins. Note, however, that the two origins are technically infinitely close together.

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?

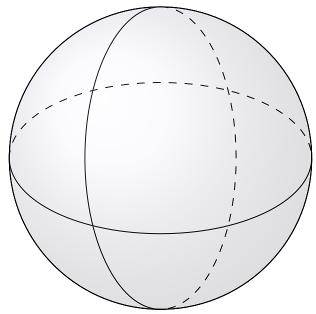

FIGURE 2. The 2-sphere, an example of a manifold.

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.

Lecture 6.

Smooth Maps Between Manifolds: 2/1/16

We're going to talk more about tangent spaces today. We've already talked about what they are in $\mathbb{R}^n$, but in order to talk about them for abstract manifolds, we'll transfer the notion from $\mathbb{R}^n$. This is very general: since manifolds are defined to locally look like Euclidean space, everything we do with manifolds will involve constructing a notion in $\mathbb{R}^n$ and showing that it still works when one passes to manifolds.

At an $x \in \mathbb{R}^n$, the tangent space $T_x\mathbb{R}^n$ can be thought of arrows based at x, or as velocities of smooth paths through x, or as derivations[5] at x (the equivalence of these was a problem on the last homework). Then, the tanget bundle is $T\mathbb{R}^n = \{(x, v) \mid v \in T_x\mathbb{R}^n\}$, which is isomorphic (as vector spaces) to $\mathbb{R}^n \times \mathbb{R}^n$; thus, we can give it the topology of $\mathbb{R}^{2n}$: two vectors are close if either their basepoints or their directions are close.

First, we generalize this slightly to a k-dimensional manifold $X \subset \mathbb{R}^N$. If $x \in X$, then x is in the image of a chart $U \subset \mathbb{R}^k$ under the chart map ϕ. Let a be the preimage of x; then, we defined $T_xX = \mathrm{Im}\,\mathrm{d}\phi|_a \subset T_x\mathbb{R}^N$, and we showed that this was independent of the chart used to construct this, because change-of-charts maps are smooth. This is also the space of velocities of paths through X, or the derivations at x on X (i.e. using $C^\infty(X)$ instead of $C^\infty(\mathbb{R}^N)$; this is the same as $C^\infty(\mathbb{R}^k)$ through ϕ). This is a little more work than we had to do for $\mathbb{R}^n$, but everything is still the same, because everything (derivations, paths) is the same in $\mathbb{R}^k$ and X, at least near x. Then, the tangent bundle is $TX = \{(x, v) : x \in X, v \in T_xX\} \subset T\mathbb{R}^N$, which is a $2k$-dimensional manifold.

[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.

[5]Recall that you can also think of derivations as directional derivatives.

So from this perspective, do we even need $\mathbb{R}^N$? Not really: if you're working in an abstract manifold, pulling derivations back to a chart in $\mathbb{R}^k$ still works, so one can define tangent vectors and tangent bundles on abstract manifolds, which have the same properties (though an abstract tangent manifold doesn't naturally sit inside $T\mathbb{R}^N$).

Now, we want to talk about maps between manifolds, and what derivatives of those maps mean. If we're inside $\mathbb{R}^N$, this is easy: a smooth function on a manifold inside $\mathbb{R}^N$ is the restriction of a smooth function on a neighborhood in $\mathbb{R}^N$; courtesy of the inverse function theorem, you could construct these, but generally don't. Instead, you use charts: a map between manifolds $f : X \to Y$ (where X is k-dimensional and Y is ℓ-dimensional) can be defined in terms of neighborhoods. If $U \subset \mathbb{R}^k$ and $V \subset \mathbb{R}^\ell$ are neighborhoods with charts $\phi : U \to X$ and $\psi : V \to Y$ such that $\phi(a) = p$ and $\psi(b) = f(p)$, then f can be understood on $\mathbb{R}^k$ and $\mathbb{R}^\ell$; let $h = \psi^{-1} \circ f \circ \phi$, which fits into the commutative diagram

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, shift right, "\phi^{-1}"'] & Y \arrow[d, shift right, "\psi^{-1}"'] \\ U \arrow[r, "h"] \arrow[u, shift right, "\phi"'] & V. \arrow[u, shift right, "\psi"'] \end{tikzcd} \tag{6.1}$$

We say that f is *smooth* if h is smooth. One has to show that this is independent of the choice of charts (which it is, for the reason that the change-of-charts map is smooth, and compositions of smooth functions are smooth), and that this agrees with the definition given above (which is a homework exercise).

Next, derivativs. We can take a derivative $\mathrm{d}h|_a : T_a\mathbb{R}^k \to T_b\mathbb{R}^\ell$, and we want to turn this into a map $\mathrm{d}f|_p : T_pX \to T_{f(p)}Y$, or $\mathrm{d}f : TX \to TY$. What this means depends on your definition of tangent vector, so we'll give a few definitions. It's important to prove that they're equivalent, but this follows from the chain rule.

- First, let's suppose v is a derivation on X at p; we'd like $\mathrm{d}f|_p(v)$ to be a derivation at $f(p)$; hence, if $g \in C^\infty(Y)$, then we can pull it back to X: $g \circ f \in C^\infty(X)$, so we can define $(\mathrm{d}f|_p(v))(g) = v(g \circ f)$.
- Next, suppose v is the velocity vector of a $\gamma : \mathbb{R} \to X$. Then, $f \circ \gamma$ is a path in Y, so we can let $\mathrm{d}f_P(v)$ be the velocity of $f \circ \gamma$. Again, we compose with f, but it's a little strange that in one case, we pull back, and in the other case, we pull back. This is an example of a useful mantra: *vectors push forward; functions pull back*. This will come back when we talk about differential forms later.
- The arrow definition is stranger: suppose $v = \mathrm{d}\phi|_a(w)$ for a $w \in T_a\mathbb{R}^\ell$. We don't know anything about abstract arrows, but we can push it forward with $\mathrm{d}h|_a$: $\mathrm{d}h|_a(w) \in T_b\mathbb{R}^\ell$ corresponds through ψ to a tangent vector at $f(p)$. In other words, $\mathrm{d}f|_p(v) = \mathrm{d}\psi_b \circ \mathrm{d}h|_a \circ \mathrm{d}\phi^{-1}|_p(v)$, and you can check that this is independent of choice of charts. That is: there's a commutative diageam (6.1) of spaces, and the tangent spaces also form a commutative diagram!

Exercise. Prove that these notions of derivative are all the same (using the chain rule).

We're going to move interchangeably between these pictures, so it's important to know how to translate between them.

Now that we've translated the notion of derivative to smooth maps between manifolds, we can translate all the nice theorems about them too.

Theorem 6.1 (Inverse function theorem)**.** *Suppose X and Y are k-dimensional manifolds. If $f : X \to Y$ is smooth and $\mathrm{d}f|_p$ is invertible, then there's a neighborhood $U \subset X$ of p such that $f|_U$ is a diffeomorphism onto its image.*

In other words, f *is locally a diffeomorphism* in a neighborhood of p.

Proof. Recall our commutative diagram (6.1). Since $\mathrm{d}\phi|_a$ and $\mathrm{d}\psi|_b$ are invertible, then $\mathrm{d}f|_p$ is invertible iff $\mathrm{d}h|_a$ is. Hence, h is locally a diffeomorphism $\mathbb{R}^k \to \mathbb{R}^k$, so since ϕ and ψ are, then f is. ⊠

We've already done the unpleasant analysis, so now we can just do definition chasing. Similarly, using this diagram, you can define the inverse of f locally, by chasing it across the commutative diagram (as h^{-1} already exists).

The next question is what happens when X and Y have different dimensions. If Y is ℓ-dimensional, with $k < \ell$, then $\mathrm{d}f|_p$ is a skinny matrix, with block from $\binom{A}{B}$, and A is invertible. Then, there are diffeomorphisms

$\phi : \mathbb{R}^k \to \mathbb{R}^k$ and $\psi : \mathbb{R}^\ell \to \mathbb{R}^\ell$ such that the following diagram commutes.

$$\begin{tikzcd} \mathbb{R}^k \arrow[r, "f"] \arrow[d, leftrightarrow, "\phi"] & \mathbb{R}^\ell \arrow[d, leftrightarrow, "\psi"] \\ \mathbb{R}^k \arrow[r, "x \mapsto (x,0)"] & \mathbb{R}^\ell \end{tikzcd} \tag{6.2}$$

The map $h(x) = (x,0)$ on the bottom is known as the *canonical immersion*, and is the simplest way to put $\mathbb{R}^k$ into $\mathbb{R}^\ell$.

Why is this true? We know the image of f is a graph of points $(x, g(x))$ for some smooth g. Thus, $\psi(x,y) = (x, y - g(x))$, so

$$\mathrm{d}\psi|_{f(p)} = \left(\begin{array}{c|c} I & 0 \\ \hline -\mathrm{d}g^{\mathrm{T}} & I \end{array}\right).$$

Thus, this is invertible, so we can use the inverse function theorem om $\psi \circ f$.

In other words, if π_1 denotes ptojection onto the first coordinate (the $\mathbb{R}^k$ one), then $\mathrm{d}\pi_1 \circ f)|_a = A$. This is invertible, so $\pi_1 \circ f$ is locally a diffeomorphism! Thus, we let it be ϕ in (6.2), and thus, the map along the bottom really is the canonical immersion. In other words, we've sketched the proof of the following theorem.

Theorem 6.2. *If $k < \ell$ and $\mathrm{d}f|_p$ has rank k, then there are coordinates such that $h(x) = (x,0)$.*

And this translates to manifolds in exactly the same way as before. This kind of argument (working in local coordinates and using it to translate things from $\mathbb{R}^k$ to manifolds) is very common in this subject, and can be summarized as "think locally, act globally."

Lecture 7.

Immersions and Submersions: 2/3/16

Immersions.

Definition. Let X be a k-dimensional manifold and Y be an ℓ-dimensional manifold.

- A smooth map $f : X \to Y$ is an *immersion* if $\mathrm{d}f$ has full rank k everywhere.
- f is a *local immersion* at an $a \in X$ if $\mathrm{d}f|_a$ has rank k (which means it has full rank on a neighborhood of a).

Notice that either of these forces $k \le \ell$.

If $f : X \to Y$ has rank k at a, then there are coordinate charts $\phi : U \to X$ and $\psi : V \to Y$ (with $U \subset \mathbb{R}^k$ and $V \subset \mathbb{R}^\ell$) such that $h = \psi^{-1} \circ f \circ \phi$ looks like the canonical immersion $x \mapsto (x,0)$. Thus, locally, an immersion has a pretty nice image. Moreover, since $\mathrm{d}(f \circ \phi) = \mathrm{d}f \circ \mathrm{d}\phi$, and both $\mathrm{d}f$ and $\mathrm{d}\phi$ are injective, then $\mathrm{d}(f \circ \phi)$ is also injective. So $f \circ \phi$ looks suspiciously like a coordinate chart.

The big question is, *if f is an immersion, is its image a manifold?*

Just because $\mathrm{d}f$ is injective everywhere does not imply f is. For example, you could map $S^1 \to \mathbb{R}^2$ as a figure-8; then, at the intersection point, the manifold locally looks like a pair of coordinate axes, which is not a manifold (it doesn't look like $\mathbb{R}^n$ locally). Okay, great, so if f is an injective immersion, is its image a manifold?

What we'd like to say is that a neighborhood of $f(a)$ comes from a neighborhood of a. However, we'll still need another condition.

Example 7.1. The torus T^2 can be realized as a rectangle with opposite edges identified, as in Figure 3. Thus, we can smoothly map $\mathbb{R} \hookrightarrow T^2$ as a line in this rectangle (wrapping around the identifications), but if the slope of this line is irrational, then there will be countably many disjoint intervals in each neighborhood of any point, and this means that the image isn't a manifold.

One way to work around this is to restrict to immersions that are homeomorphisms onto their image. But another way to think of this: the issue with $\mathbb{R} \hookrightarrow T^2$ was that very distant points ended up nearby. There's a nice way to formalize this.

Definition. A map $f : X \to Y$ of topological spaces is *proper* if for every compact $K \subset Y$, $f^{-1}(K)$ is compact in X.

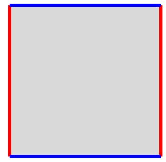

FIGURE 3. The torus can be realized as a rectangle with opposite sides identified, so glue the red sides togather and the blue sides together.

Note that proper maps need not be immersions: the double cover map $\theta \mapsto 2\theta : S^1 \to S^1$ is smooth and proper, but every point has two images.

But a proper injective immersion is sufficient.

Definition. A smooth map $f : X \to Y$ of manifolds is an *embedding* if it is a proper injective immersion.

Remark. A proper injective map is sometimes called a *topological embedding*. This might be enough to imply that it's an immersion (though the textbook sticks with requiring that f is an immersion).

The quality of being proper is sometimes called *properness*, but *propriety* sounds better.

Theorem 7.2. *Let $f : X \to Y$ be an embedding. Then,* $\operatorname{Im}(f)$ *is a submanifold of Y.*

Proof sketch. For any $a \in X$, consider neighborhoods of $f(a)$. Since f is proper, there's a neighborhood of $f(a)$ that is the image of only finitely many neighborhoods in X, and since f is injective, then they all must be positive distances from each other. Thus, we can shrink our neighborhood to one that only contains the neighborhood for $f(a)$, and then since f is an embedding, a chart for a makes a chart for $f(a)$, so we win. ⊠

Much of the time, we're going to be looking at compact manifolds, for which propriety is redundant: if X is compact, then any continuous map $X \to Y$ (where Y is Hausdorff) is proper (since the preimage of a closed set under a continuous map is closed, and a closed subset of a compact space is compact).

Submersions. Immersions aren't the only way full rank can happen; since full rank is such a nice condition, let's look at another case of it.

Definition. Let X be a k-dimensional manifold, Y be an ℓ-dimensional manifold, and $f : X \to Y$ be smooth.

- f is a *submersion* if $\mathrm{d}f$ has full rank ℓ everywhere.
- f is a *local submersion* near an $a \in X$ if $\mathrm{d}f$ has full rank ℓ on a neighborhood of a (equivalently, at a).

This time, these imply that $k \geq \ell$.

Just as immersions locally look like the canonical immersion, submersions locally look like the *canonical submersion* $\pi : \mathbb{R}^k \to \mathbb{R}^\ell$ sending $(x_1, \dots, x_k) \mapsto (x_1, \dots, x_\ell)$.

Theorem 7.3 (Local submersion theorem)**.** *Let $f : X \to Y$ be a local submersion near a. Then, there are coordinate charts $\phi : U \to X$ and $\psi : V \to Y$ such that in these coordinates, f looks like the canonical submersion, i.e. $h = \psi^{-1} \circ f \circ \phi$ sends $(x_1, \dots, x_k) \mapsto (x_1, \dots, x_\ell)$.*

Proof. Start with any coordinate charts $\phi : U \to X$ and $\psi : V \to Y$.

Since f has full rank at a, then $\mathrm{d}f|_a = \binom{A}{L}$, where L is a fat matrix with full rank. Since it's linear, there's a smooth map $H : U \to V \times \mathbb{R}^{\ell - k}$ sending $x \mapsto (h(x), L(x))$. Thus, $\mathrm{d}H|_a = \binom{\mathrm{d}h|_a}{L}$, and each of these blocks has full rank, so $\mathrm{d}H|_a$ does too. Thus, since H is square, it's locally invertible, and $\psi^{-1} \circ f \circ \phi' = h \circ H^{-1}$, so using a new coordinate chart ϕ', $h \circ H^{-1}$ is our change-of-charts map, and it's the canonical submersion. ⊠

Theorem 7.4. *Let $f : X \to Y$ be a submersion and $y \in Y$. Then, $f^{-1}(y)$ is a submanifold of X of codimension equal to* $\dim Y$.

Proof sketch. Again, we can check in neighborhoods: let $a \in f^{-1}(y)$; thus, in a neighborhood U of a in X, f looks like the canonical submersion, by Theorem 7.3. In particular, composing with the canonical submersion in a chart for a gives a chart for $U \cap f^{-1}(a)$. ⊠

Because Theorem 7.3 provides a neighborhood in X, rather than in Y, the nuance between embeddings and immersions doesn't come up for submersions.

We can get a stronger result: consider $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x_1, x_2) = x_1^2 + x_2^2$. Yes, something bad happens at 0, but for the preimage of 1, we don't really care. We can formalize this.

Definition. Let $f : X \to Y$ be a smooth map of manifolds and $y \in Y$.

- y is a *regular value* of f if $\mathrm{d}f|_a$ has full rank for every $a \in f^{-1}(y)$.
- Otherwise, y is called a *critical value*.

Regular values are extremely important.

Theorem 7.5. *Let y be a regular value for $f : X \to Y$; then, $f^{-1}(y)$ is a submanifold of X with codimension equal to* $\dim Y$.

The proof is exactly the same as for Theorem 7.4, since that proof only required local data.

Example 7.6. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x_1, x_2) = x_1^2 - x_2^2$, so that $\mathrm{d}f = \left(\begin{smallmatrix} 2x_1 \\ -2x_2 \end{smallmatrix}\right)$. Thus, $\mathrm{d}f|_{(x_1,x_2)}$ is surjective whenever $(x_1, x_2) \neq (0,0)$, so these are regular values, but at the origin, $\mathrm{d}f|_{(0,0)}$ isn't full rank (it's the zero matrix). Hence, 0 is the only critical value. And lo, the preimage of 0 isn't a manifold, though the preimage everywhere else is.

One interesting nuance is that there are many points in $f^{-1}(0)$ where f is locally a submersion (in fact, all but the origin); but it only takes one bad point to make a set not a manifold.

One important thing to keep in mind is that critical values live in Y, the codomain. We'll hear "points" for things in X and "values" for things in Y, as in the following definition. Be careful to keep them separate!

Definition. Let $f : X \to Y$ be a smooth map of manifolds and $x \in X$.

- If $\mathrm{d}f|_x$ doesn't have full rank, then x is a *critical point*.
- Otherwise, x is a *regular point*.

In Example 7.6, the origin is the only critical point.

There's a nice theorem from real analysis about this, which we will not prove.

Theorem 7.7 (Sard)**.** *If $f : X \to Y$ is smooth, then the set of critical* values *of f has measure zero.*

You might wonder: what measure are we using? Well, that's a tricky question: the standard measure on $\mathbb{R}^n$ isn't preserved by change-of-charts maps. However, the condition of having measure zero is preserved, so a set having measure zero in a manifold is well-defined.

Also, another caveat: the critical *points* in X may not have measure zero (e.g. the zero map $\mathbb{R}^m \to \mathbb{R}^n$ — points not in the image of f are regular, since the condition is vacuously satisfied). The point is: there are lots of regular values, which is the aspect of Sard's theorem that we'll use.

Lecture 8.

Transversality: 2/5/16

Note: I missed the first eight minutes of lecture today; I'll fill in any missing details later.

Recall that if $f : X \to Y$ is smooth and y is a regular value for f, then $f^{-1}(y)$ is a submanifold of X. We want to understand a generalization: if $Z \subset Y$ is a submanifold, when is $f^{-1}(Z)$ a submanifold of X? Locally, we know Z is the zero set of a smooth function $g : Y \to \mathbb{R}^{\ell-k}$ (where X is k-dimensional and Y is ℓ-dimensional). In particular, $f^{-1}(Z) = f^{-1}(g^{-1}(0)) = (g \circ f)^{-1}(0)$. Thus, $f^{-1}(Z)$ is a submanifold when 0 is a regular value of $g \circ f$. In particular, this forces $\mathrm{d}(g \circ f)|_x$ to be surjective.

This motivates an extremely important definition.

Definition. Let $f : X \to Y$ be smooth and $Z \subset Y$ be a submanifold. Then, f is *transverse* to Z, written $f \pitchfork G$, if for all $x \in f^{-1}(Z)$, $\mathrm{Im}(\mathrm{d}f|_x) + T_{f(z)}Y = T_{f(y)}Y$.

An important special case is when both X and Z are submanifolds of Y and $f : X \to Y$ is inclusion. Then, $f^{-1}(Z) = Z \cap X$, and this is a submanifold if $f \pitchfork Z$. The derivative of inclusion is also inclusion on tangent spaces, so this condition means that $T_xX + T_xZ = T_xY$. In this case, one simply says X is transverse to Z, written $X \pitchfork Z$.

Intuitively, transversality means that the infinitesimal angle of intersection is not parallel: if it is, then they share tangent vectors, and so we don't get the entire tangent space.

Suppose $p(x)$ is a 17^{th}-order polynomial. Then, we know some conditions on how it intersects the x-axis: it must intersect at least once, and in fact an odd number of times, if the intersection is transverse (no multiple roots). However, if it's not transverse, we have a multiple real root, and it can intersect an even number of times. Strange things happen when you perturb a double root slightly: it can become two real roots, or two complex roots. However, we're going to prove that if you start with a transverse intersection of submanifolds, it's stable under slight perturbations (the number of intersections is the same).

Generally, two curves in $\mathbb{R}^3$ cannot intersect transversely... unless they never intersect at all, in which case they vacuously satisfy the definition. But this set of 0 intersection points is stable, after all. The way to gain intuition about transversality is to think of it in terms of this stability of intersections.

To summarize, the following are equivalent for a smooth map $f : X \to Y$ and a submanifold $Z \subset X$.

- f is transverse to Z.
- $\mathrm{Im}(\mathrm{d}f|_x) + T_{f(x)}Z = T_{f(x)}Y$ for all $x \in f^{-1}(Z)$.
- Locally, 0 is a regular value of $g \circ f$, where g is a local submersion $Y \to \mathbb{R}^{\ell-k}$ defined on a neighborhood, and on this neighborhood $Z = g^{-1}(0)$.

Each of these implies that $f^{-1}(Z)$ is a submanifold of X, but the converse is not true: the submanifolds $y = x^2$ and $y = 0$ intersect non-transversely at 0, but a point is a zero-dimensional manifold. However, there do exist non-transverse intersections where the intersection is not a manifold.

Homotopy. We want to make precise this fuzzy notion that if you mess with an intersection a little bit, transversality guarantees its stability. The way to slightly change a submanifold is a homotopy.

Definition. Let X and Y be topological spaces and $f_0, f_1 : X \rightrightarrows Y$ be two continuous functions. Then, a *homotopy* from f_0 to f_1 is a continuous map $F : [0,1] \times X \to Y$ such that $F(0,x) = f_0(x)$ and $F(1,x) = f_1(x)$. If there exists a homotopy between f_0 and f_1, one says that they're *homotopic*, and writes $f_0 \sim f_1$.

This is a topological notion: starting with two functions, we generate a whole family of them interpolating between F_0 and f_1: for every $t \in [0,1]$, we have the interpolator $f_t(x) = F(t,x)$.

For example, if $f_0, f_1 : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$ are given by $f_0(x) = 0$ and $f_1(x) = x$, then $F(t,x) = tx$ is a homotopy between them.

We would like to introduce smoothness to this definition, but $[0,1] \times X$ is not a manifold: for any $x \in X$, $(0,x)$ doesn't have a neighborhood diffeomorphic to a Euclidean space. So we don't know what it means to be smooth on the boundary.

There are two inequivalent ways to make this precise if f_0 and f_1 are smooth.

- We could require that F is smooth on the manifold $(0,1) \times X$ and continuous on $[0,1] \times X$. Since we knew f_0 and f_1 are smooth, this seems reasonable.
- A stronger notion of smooth homotopy is that F can be extended $(-\varepsilon, 1+\varepsilon) \times X \to Y$.

For the most part, we'll only need the weaker definition of smooth homotopy. The homotopy $F(t,x) = \sqrt{t}x$ between $f_0(x) = 0$ and $f_1(x) = x$ satisfies the weaker definition, but not the stronger one.

For various properties of maps, we want to know whether they're preserved under this notion. Specifically, let X and Y be smooth manifolds, and P be a property of maps $X \to Y$ (e.g. immersion, submersion, proper, embedding, injective, rank is at most 3, it's smooth, it's analytic, ...). If $f_0 : X \to Y$ and F is a homotopy, does f_t have the property P for all sufficiently small t? This is what we mean by stability; if this is the case for all homotopies, P is said to be *stable*.

The first thing we'll see is that it's very hard to preserve *any* properties if X isn't compact; for example, one could define a homotopy that changes things more and more as one goes out to infinity. So this is generally studied when X is compact, and indeed, under this assumption, a whole bunch of properties are stable, including transversality.

One example is that when X and Y are vector spaces, a *linear homotopy* (a homotopy of linear maps for which all the intermediate maps are linear) locally preserves full rank: this is a stable property. Not having full rank is not stable, however.

Now, the flipside is that certain properties are generic, i.e. if a map doesn't have the property, you can bump it a little bit and make it have that property.

Definition. A property P is *generic* if for any f_0, there's a homotopy F for f_0 and an $\varepsilon > 0$ such that f_t has P for all $t \in (0, \varepsilon)$.

This is existence: the constant homotopy might not work if f_0 doesn't have property P.

The best properties are both generic and stable: you can change a map a little bit and it has the property. And the big punchline is: *transversality is both generic and stable*. We cannot prove this yet, but it's a major stop on this highway. Next time, we'll be able to prove that a lot of properties are stable, and talk about genericity.

Lecture 9.

Properties Stable Under Homotopy: 2/8/16

"Welcome to UT! I hope I won't do anything to scare you away."

We're in the middle of talking about smooth homotopies $f_0 \sim f_1$ of manifolds, which are smooth maps $F : I \times X \to Y$ such that $F(0, x) = f_0(x)$ and $F(1, x) = f_1(X)$. Then, we defined $f_t(x) = F(t, x)$. There are two nuances to this.

- Guillemin and Pollack define this as a map $X \times I \to Y$. *A priori*, this makes no difference whatsoever, but when we begin to talk about oriented manifolds, it will be easier to orient this if we use the convention $I \times X$.
- What does "smooth" mean on the boundary? To Guillemin and Pollack, all manifolds live in some ambient space, so this really means it can be extended to an open neighborhood of the boundary. But we find it more useful to require the partial derivative in x to not vanish.

  As an example, let $X = Y = \mathbb{R}$, $f_0(x) = x$, and $f_1(x) = x + \sin x$. As continuous maps, these are clearly homotopic, and one example of the homotopy is

$$F(t, x) = \begin{cases} x + t \sin\left(\dfrac{x}{t^2}\right), & t \neq 0 \\ x, & ts = -0. \end{cases}$$

  Is this smooth? Well, what do you want smoothness to be? We're looking for a stability condition on transversality, but this homotopy sends something transverse to the real line to something not transverse to it, no matter how short you travel along it. And indeed, $\frac{\partial F}{\partial x}$ isn't continuous in t. Hence, for the purposes of stability, we'll require that a smooth homotopy have all partial derivatives of x continuous in t.

Under this definition, we do have some nice stability (i.e. if f_0 has a property, then so does f_t for $t > 0$ sufficiently small).

Theorem 9.1. *Let X be a compact smooth manifold, Y be a smooth manifold. Then, the following properties are stable under smooth homotopies $F : I \times X \to Y$:*

(1) local diffeomorphisms,
(2) immersions,
(3) submersions,
(4) embeddings,
(5) transversality with respect to a fixed closed submanifold $Z \subset Y$, and
(6) diffeomorphisms.

Partial proof. Suppose f_0 is a local diffeomorphism, so for any $a \in X$, $\mathrm{d}f_0|_a$ is invertible. This is true in a neighborhood of a, because the derivative having full rank is an open condition. Thus, for each $a \in X$, there's a neighborhood $U_a \subset X$ of a and a $\varepsilon_a > 0$ such that on $U_a \times (0, \varepsilon)$, $\mathrm{d}f$ has full rank. However, since X is compact, we can cover it by only finitely many of these U_a, and then take ε to be the minimum of those finitely many ε_a; thus, on the interval $(0, \varepsilon)$ and every $x \in X$, $\mathrm{d}f_t|_x$ has full rank; this proves (1).

Since the conditions on immersions and submersions are that the derivative has full rank, the same proof applies, *mutatis mutandis*, to prove (2) and (3).

Now, let's look at (5). We defined transversality to mean that for all $x \in f^{-1}(Z)$, $\mathrm{Im}(\mathrm{d}f|_x) + T_{f(x)}Z = T_{f(x)}Y$. We proved there's a map $g : Y \to \mathbb{R}^{\dim Y - \dim Z}$ that sends a neighborhood of $f(x)$ in Z to 0, and such that $f \circ g$ is a submersion. Thus $f_t \circ g$ is a submersion for sufficiently small t, and so $f_t \pitchfork Z$.

The final two, (4) and (6), depend on global topological behavior, and so we'll leave them to be exercises, but the proofs are not dissimilar. $\boxtimes$

For part (5), the stipulation that Z is closed is important: an open submanifold can be infinitesimally close to another submanifold without intersecting it (e.g. the distance between $(0,1)$ and $[1,2]$ is 0). Another important thing we depend on is that the derivatives with respect to x are continuous in t, because that allowed us to prove the first three parts. We had an explicit counterexample for (6),[6] but there are also counterexamples for the other five parts if you don't have the right notion of smoothness.

Next, let's talk about Sard's theorem, Theorem 7.7, which states that if $f : X \to Y$ is smooth, then its set of critical values has measure zero.

Partial proof of Theorem 7.7. First, we can reduce this to a statement about neighborhoods in X and Y: if we know it in charts, then we can take a countable union of charts in X (which exists because X is second countable), and a countable union of sets with measure zero still has measure zero.

Hence, we may assume without loss of generality that $X = (0,1)^k$ and $Y = (0,1)^\ell$. If $k = \ell$, let C be the set of critical points, so $f(C)$ is the set of critical values. Since C is the points where $\det \mathrm{d}f|_x = 0$, let $C_\varepsilon = \{x \in X : \det \mathrm{d}f_x < \varepsilon\}$. Then, $|f(C)| \le |f(C_\varepsilon)| < \varepsilon$ for each $\varepsilon > 0$, so $|f(C)| = 0$. This estimate comes from the fact that the determinant of the derivative measures how much f changes volume locally, so small determinants in the unit cube squish their image into a small space. The idea here is that there may be a lot of critical points, but they're squashed together.

To apply this when $k \ne \ell$, you have to do some extra linear algebra: if you have a fat matrix without full rank, what does it do to volume, and what does a small perturbation do to volume? The takeaway will be that the image will have proper codimension, and therefore automatically is measure zero. But this isn't topology, so we're not going to dwell on it. ⊠

A Five-Minute Crash Course in Morse Theory. One cool use of Sard's theorem is Morse theory. This will be a short digression.

Let X be a compact manifold (the canonical example is a torus) and $f : X \to \mathbb{R}$ be a smooth function (in the example, a height function). Consider the sets $f^{-1}((-\infty, a))$ for $a \in \mathbb{R}$. Since X is compact, there's a minimum a_0, and for values of a just a little bit greater than a_0, you get the behavior of X in a neighborhood of that minimum, but they're all the same until you get to the donut hole.

That is, at a critical value of f, there's something interesting topologically going on, and nothing topologically happens at the regular values. You need f to have a condition that makes its behavior particularly clean around critical values, but such f exists, but the result is a decomposition of X into pieces associated with its critical values.

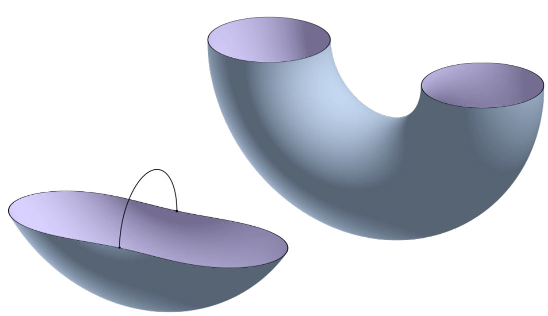

FIGURE 4. Adding a bridge at a critical point of f.

So we need to understand how f behaves around critical values, meaning a power series expansion

$$f(x) = f(a) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\bigg|_a (x_i - a_i) + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}\bigg|_a (x_i - a_i)(x_j - a_j) + o(x^3).$$

[6]Technically, we didn't start with a compact X, but the noncompactness of $\mathbb{R}$ was never needed, and we could replace it with its one-point compactification S^1 without changing the essence of the argument.

If x is a critical point, then the first derivatives vanish, so to make this nondegenerate, we just need that the second derivatives don't vanish at each critical point. Such a function is called a *Morse function*, and a critical point satisfying this is called *nondegenerate*.

The fact that Morse functions exist, and in fact can be made from a perturbation of any function, is a consequence of Sard's theorem.

Lecture 10.

May the Morse Be With You: 2/10/16

Last time, we briefly started talking about Morse theory. Today, we'll slow down and go in more detail.

Definition. A smooth function $f:\mathbb{R}^n\to\mathbb{R}$ is *Morse* if, whenever $\mathrm{d}f|_x$ is smooth, the Hessian at x is invertible.

The awesome fact is that garden-varienty functions are Morse, or, in different words, Morse functions are generic.

Theorem 10.1. *Let $f:\mathbb{R}^n\to\mathbb{R}$ be a smooth function and let $f_a(x)=f(x)+a\cdot x$. Then, for almost all $a\in\mathbb{R}^n$, f_a is Morse.*

"Almost all" means that the statement is true except on a set of measure zero.

Proof. Define $g_a(x)=\nabla f_a=(\partial_1 f_a,\partial_2 f_a,\dots,\partial_n f_a)=a+\nabla f$, and $\mathrm{d}g_a$ is the Hessian of f_a. Hence, f_a is Morse iff 0 is a regular value of g_a iff $-a$ is a regular value of g. By Sard's theorem (Theorem 7.7), regular values have full measure, so almost every $-a$ is a regular value, and therefore almost every f_a is Morse. ⊠

Those were Morse functions on Euclidean space. What about on manifolds?

Definition. If X is a smooth manifold, a smooth $f:X\to\mathbb{R}$ is Morse if whenever $\mathrm{d}f|_x=0$, the Hessian at x in local coordinates is invertible.

At every critical point x, there's a chart $\psi:\mathbb{R}^k\to X$, and this statement is equivalent to $f\circ\psi$ being Morse as a function $\mathbb{R}^k\to\mathbb{R}$.

Remark. Since we chose a chart to make this definition, we need to know that it's independent of choice of charts, so suppose $\phi:\mathbb{R}^k\to X$ is another chart for a neighborhood of x, and let g be the change-of-charts map for ϕ and ψ. The fact that it's a diffeomorphism means that the critical points of $f\circ\phi$ and $f\circ\psi\circ g$ are the same, and using the Chain rule, $H(f\circ\phi)=(\mathrm{d}g)H(f\circ\psi)\mathrm{d}g^{\mathrm{T}}$, and since $\mathrm{d}g$ is invertible, this is linear in $\mathrm{d}(f\circ\psi)$, and therefore one is invertible when the other does. This argument should be fleshed out a bit, but the point is that Morseness doesn't depend on which local coordinates you use.

Now, we can prove an analogue of Theorem 10.1 for submanifolds of $\mathbb{R}^n$. There's an analogue for abstract manifolds, but it's a little harder to state, since we can't take the dot product abstractly.

Theorem 10.2. *Let X be a k-dimensional submanifold of $\mathbb{R}^n$ and $f:X\to\mathbb{R}$ be smooth. Then, if $f_a(x)$ is defined as in Theorem 10.1, then for almost every $a\in\mathbb{R}^n$, f_a is Morse.*

Proof. We're going to work in charts: because Euclidean space is separable, X can be covered by countably many charts, or more precisely, every cover of X by charts has a countable subcover. Thus, if we prove that on each chart, the set of a which fail has measure zero, then the total set of such a is a countable union of sets of measure zero, and thus has measure zero. And every point in X has a neighborhood to which the immersion theorem applies, so we can cover X by countably many neighborhoods in which it applies.

We can write $a=(b,c)$, where b denotes the first k coordinates and c denotes the last $n-k$ coordinates. Around a given point p, we can thus write $f_a(x)=f(x)+c\cdot(x_{k+1},\dots,x_n)+b\cdot(x_1,\dots,x_k)$. And because $f(x)+c\cdot(x_{k+1},\dots,x_n)$ is smooth, then f_a is Morse for almost every b. By Fubini's theorem, the set of (b,c) where b doesn't work also has measure zero, so f_a is Morse for almost every a. ⊠

Morse functions are a dime a dozen, if not a dime a countably many! And there are lots of useful things you can do with a Morse function, e.g. looking at the topology of a manifold using preimages of intervals under Morse functions.

Embeddings of Manifolds. We're going to make a series of increasingly strong statements about how to embed abstract manifolds into $\mathbb{R}^N$ for sufficiently large N.

Theorem 10.3 (Whitney embedding theorem)**.** *Let X be an abstract k-dimensional manifold.*

(1) There's an embedding $X \hookrightarrow \mathbb{R}^N$ for some N.
(2) There's an injective immersion $X \hookrightarrow \mathbb{R}^{2k+1}$.
(3) There's an embedding $X \hookrightarrow \mathbb{R}^{2k+1}$.
(4) There's an immersion $X \hookrightarrow \mathbb{R}^{2k}$.
(5) There's an embedding $X \hookrightarrow \mathbb{R}^{2k}$.

One consequence is that the Guillemin-and-Pollack approach to manifolds captures all diffeomorphism classes of manifolds. Of course, this theorem is not in the textbook. Parts (2), (3), (4), and (5) are all due to Whitney.

We'll attack this as follows. First, we'll prove (1) for X compact, and then prove (2), (3), and (4) assuming (1) in generality (the details aren't that different). (5) is extremely difficult to prove.

To prove these statements, we'll rely heavily on the concept of a partition of unity. We'll discuss these more on Friday (and provide a proof of existence).

Definition. Let X be a smooth manifold.

- Let $\rho : X \to \mathbb{R}$ be a smooth function. Then, its *support* is the closed set $\operatorname{supp}\rho = \overline{\{x : \rho(x) \neq 0\}}$ (the closure of where it's nonzero).
- If $U \subset X$ is open and $K \supset U$ is compact, a *bump function* $\rho : X \to \mathbb{R}$ is a smooth function such that $\rho|_U = 1$ and $\operatorname{supp}\rho \subseteq K$.

That is, a bump function is smooth, but if K isn't much bigger than U, it has to change from 1 to 0 smoothly and quickly.

Definition. Let $X \subset \mathbb{R}^n$ be a manifold and $\mathfrak{U} = \{U_i\}_{i\in I}$ be an open cover of X. Then, a collection of smooth functions $\rho_i : X \to \mathbb{R}$ (also indexed by I) is a *partition of unity* if it satisfies the following axioms.

- $\operatorname{supp}(\rho_i) \subset U_i$ for each i.
- For every $x \in X$, there's a neighborhood of x on which only finitely many ρ_i are nonzero.[7]
- $\sum_{i\in I} \rho_i(x) = 1$ for all $x \in X$. (This makes sense, because at each point, it's a finite sum.)

Bump functions can be used to construct these, as we will show at some point.

Proof of Theorem 10.3, part 1. Since X is compact, there's an $s \in \mathbb{N}$ and an open cover $\mathfrak{U}$ of X by s coordinate charts. Let $\{\rho_i\}$ be a partition of unity indexed by $\mathfrak{U}$.

On the chart U_1, we have coordinates $(x_1, \dots, x_k)$, and the function $\widetilde{g}_1 : U_1 \to \mathbb{R}^{k+1}$ sending $x \mapsto (\rho_1, \rho_1 x_1, \dots, \rho_1 x_k)$ is smooth and supported in U_1, so we can extend it to all of X by defining it to be 0 outside of U_1. Thus, this formula defines a smooth $g_1 : X \to \mathbb{R}^{k+1}$. The same construction defines functions $g_2, \dots, g_s : X \to \mathbb{R}^{k+1}$.

Now, our embedding will be $j : X \to \mathbb{R}^{s(k+1)}$, defined by $j(x) = (g_1(x), \dots, g_s(x))$, which is smooth and has full rank (since each point is in a chart U_i, where g_i has full rank k, so j has to have rank k as well). Then, it's injective, because if $j(x_1) = j(x_2)$, then $\rho_i(x_1) = \rho_i(x_2)$ for some i where this quantity is nonzero. Thus, they lie in the same chart, so their coordinates in that chart agree (since $g_i(x_1) = g_i(x_2)$), and therefore j is injective. And since X is compact, it's proper, so j is an embedding. ⊠

[7] This can be thought of as a form of "ρ reduction."