M382D NOTES: DIFFERENTIAL TOPOLOGY

ARUN DEBRAY
JANUARY 22, 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.

CONTENTS

1. The Inverse and Implicit Function Theorems: 1/20/15 1
2. The Contraction Mapping Theorem: 1/22/15 3

Lecture 1.

The Inverse and Implicit Function Theorems: 1/20/15

"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

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

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 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) *In general, 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 $n < m$, there's a neighborhood U of a such that $f(U)$ is the graph of a smooth $g : \mathbb{R}^n \to \mathbb{R}^{n-m}$.*

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: 1/22/15

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.

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

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

We can bound the integral below:

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

[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* $\|A\| = \sup\limits_{v \in S^{n-1}} |Av|$.

Thus, since $x_2 - x_1 = v$, then their sum 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.

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 $2|y|$. 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 $d(T(x), T(y)) \le c d(x, y)$ 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 $x \neq x'$, then $d(T(x), T(x')) \le c d(x, x') = d(T(x), T(x'))$, 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, $d(x_m, x_n) \le c^{|n-m|} d(x_n, x_{n-1})$, 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,

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

Since we assumed that $\mathrm{d}f|_0 = I$, then in this neighborhood, this integral is small in x, so $y = x + o(|x|)$, and in particular $y = x + o(|y|)$, or rearranging, $x = f^{-1}(y) = y + o(|y|)$. Thus, f^{-1} is differentiable once at 0. Moreover, this is true in a neighborhood of $0 \in Y$, because we can repeat this argument for sufficiently close $y \in Y$.

To see why this is smooth, let $A(t)$ be an invertible function, so that $A(t)A(t)^{-1} = I$. Thus, $A'A^{-1} = A(A^{-1})' = 0$, so rearranging, $(A^{-1})' = A^{-1}A'A^{-1}$. Thus, you can differentiate the inverse as many times as you can differentiate the original function, so f^{-1} is smooth, and (after some playing about with the chain rule) $\mathrm{d}(f^{-1})|_a = (\mathrm{d}f|_a)^{-1}$. ⊠

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

Proof of Theorem 1.3, parts (2) and (3). 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 $U = f^{-1}(f(a)) = U \cap \{g(y), y\}$ 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 (its determinant is $\det A \neq 0$). Thus, we apply the inverse function theorme to F to conclude that a smooth F^{-1} exists, and so if π_1 denotes projection onot the first component, $x = \pi_1 \circ F^{-1}(0, y) = g(y)$. ⊠

[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.