MAT425 Lecture Notes

Emily

'26 Winter Semester

Contents

§1 Day 1: Manifolds or Whatever (Jan. 6, 2025)

The textbook is Differential Topology, by Guillemin & Pollack. This course is about smooth manifolds. But this is a very vast subject, so it's more accurate to say that it's all about a specific circle of questions which are related to something called transversality.

What is a manifold? Well a manifold is... a subset of $\mathbb{R}^n$ with some requirements on dimensionality. We don't loose any generality this way (due to Whitney's Embedding Theorem) therefore for our purposes this is sufficient.

Assume that $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ and $f : X \to Y$ is smooth. What does this mean? Well let's recall an easier case. When X and Y are open, then this is just our standard definition of differentiability (locally linear and such). On the other hand, if X and Y aren't open, then it can be much harder—if X and Y are some complicated fractal sets then how do you define it?

We say a map $f : X \to Y$ is smooth at a point $x \in X$ if there exists some open neighborhood U of x and some function $F : U \cup X \to \mathbb{R}^m$ such that the following diagram commutes:

$$\begin{tikzcd} X \cup U \arrow[r, "F"] & \mathbb{R}^m \\ X \arrow[u, hook] \arrow[r, "f"'] & Y \arrow[u, hook] \end{tikzcd}$$

and additionally, F is smooth at x. We say that f is smooth if it is smooth everywhere in its domain.

Definition 1.1. Two sets $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ are diffeomorphic if there is a map $f : X \to Y$ which is bijective, smooth, and f^{-1} is smooth as well.

Definition 1.2. X is a k-dimensional smooth manifold if for all $x \in X$ there exists some open $U \subseteq \mathbb{R}^n$ such that $x \in U$ and $U \cap X$ is diffeomorphic to an open $V \subseteq \mathbb{R}^m$.

Thus, there exists $\varphi : U \cap X \to V$ and $\psi : V \to U \cap X$ with $\varphi^{-1} = \psi$ and φ, ψ are smooth maps.

Definition 1.3. $X \subseteq \mathbb{R}^n$ is a k-dimensional smooth manifold if it is locally diffeomorphic to open sets of $\mathbb{R}^k$.

I have no clue why he wrote this as a separate definition. Assume that $X \subseteq \mathbb{R}^n$ is a k-dimensional manifold. Then there exists a local parameterization $\varphi : U \subseteq \mathbb{R}^k \to V \cap X$. The opposite direction ($V \to U$) is called a coordinate system for our manifold.

Let us now define the tangent spaces. If X is a k-dimensional manifold, the tangent space is given by the local parameterizations. Suppose that $\varphi : U \to V \cap X$ is a parameterization of our manifold. Suppose that $u_0 \in U$ and $f(u_0) = x$. Then we say that the tangent space of X at x is

$$T_x X = d\varphi(u_0)(\mathbb{R}^k) \subseteq \mathbb{R}^n.$$

Notice that the tangent spae is independent of parameterization. He drew a photo of the tangent line to a curve. Also notice that the tangent space is centered at 0, but it's natural to imagine it centered at x (and infinitesimal) instead.

To define the derivative for a smooth manifold naturally, we need two things. First, it should agree with our usual Euclidean definition of derivative. Second, it should satisfy the chain rule. For it to make sense, $df(x)$ should map $T_x X$ to $T_{f(x)} Y$.

Let U and V be open neighborhoods, and ϕ. and ψ. be parameterizations and charts of their manifolds. Define h such that the following diagram commutes.

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, shift left, "\psi_X"] & Y \arrow[d, shift left, "\psi_Y"] \\ U \arrow[r, "h"] \arrow[u, shift left, "\varphi_X"] & V \arrow[u, shift left, "\varphi_Y"] \end{tikzcd}$$

Namely, $h = \psi_Y \circ f \circ \varphi_X$. Suppose (for simplicity) that $0 \in U, V$, $\varphi_X(0) = x$, and $\psi_Y(f(x)) = 0$. Then we have that

$$df(x) = d\varphi_Y(0) \circ dh(0) \circ d\psi_X(x).$$

[This part is quite dense in the notation, so draw a picture to make sure you're clear with the notation—it helped me a ton to work through it].

Because our manifolds are smooth we can discuss general position. What does this mean? Well consider two lines. They might intersect at one point, no points, or infinitely many points. But if we slightly perturb our lines, they will only intersect at one point. Therefore, we say that they generally intersect once.

We can do the exact same thing on a torus. It's possible to find two lines which intersect infinitely many times, but we can find a general number of times they intersect. Using these ideas we can formulate all sorts of interesting topological invariants. This is all about transversality (he didn't define).

We can formulate these ideas in a general setting. Sard's theorem is most at home in Euclidean space (telling you that the critical points of a function form a measure zero set), but we can also do it for manifolds.

Definition 1.4. A critical point $x \in X$ of a smooth map $f : X \to Y$ between manifolds is a point where $df(x)$ does not have full rank. Otherwise, x is regular.

Definition 1.5. A critical value $y \in Y$ of a smooth map $f : X \to Y$ is a point which is the image of some critical point. Otherwise, y is regular.

Sard's theorem tells us that the set of critical values of our manifold has measure 0. This is important because we want to perturb into regular values.

Next class we will cover the implicit function theorem (i might have misheard, he might have said inverse), normal maps, Sard's theorem, degrees of maps, Transversality, topological invariants, then some bonus if there's time (Morse theory).

§2 Day 2: IFT, Immersions, Submersions (Jan. 8, 2025)

Review: $M \subseteq \mathbb{R}^n$ is a k-manifold if it is locally diffeomorphic to $\mathbb{R}^k$. Our manifolds are locally parameterized by k-dimensional space. This means that for every $x \in M$ there exists a φ such that $\varphi(0) = x$, and $\varphi : U \subseteq \mathbb{R}^k \to M$ is a local diffeomorphism.

We can think of φ as a collection of n functions from $\mathbb{R}^k$ to $\mathbb{R}$.

When we think about the differential, sometimes it's nice to imagine a curve through 0 $\alpha : \mathbb{R} \to \mathbb{R}^k$, and then we have

$$d(\varphi \circ \alpha) = d\varphi(0) \cdot \alpha'(0).$$

But otherwise, we can just do what we did last time.

We say that if $N^k \subseteq M^n \subseteq \mathbb{R}^N$, and M^n is a manifold in $\mathbb{R}^N$, then N^k is a submanifold of M^n.[1]

§2.1 Inverse and Implicit Function Theorem

Suppose that $f : M \to N$ is a smooth map, and that at $x_0 \in M$, we have that $df(x_0) : TM_{x_0} \to TN_{f(x_0)}$ is an isomorphism. Then we have that f is a local diffeomorphism near x_0. That is, there exists an open subset $U \subseteq M$, and $x_0 \in U$, and an open set $V \subseteq N$, with $f(x_0) \in V$ such that $f|_U$ is a diffeomorphism. Notice that because the differential is an isomorphism, M and N have the same dimension.

This version of the theorem can be derived by the regular inverse function theorem in Euclidean space. This is because given parameterizations $\varphi : U \to M^k$ and $\psi : V \to N^k$,

$$\begin{tikzcd} M^k \arrow[r, "f"] & N^k \\ U \arrow[u, "\varphi"] \arrow[r, "\psi^{-1}\circ f\circ\varphi"'] & V \arrow[u, "\psi"'] \end{tikzcd}$$

we can use the regular inverse function theorem on $\psi^{-1} \circ f \circ \varphi$, to get an inverse function $\varphi^{-1} \circ f^{-1} \circ \psi$ locally, and then just compose with ψ^{-1} and φ to get our local inverse function for f. Nabutovsky drew a picture, but it was not at all helpful so I will not include it.

Nabutovsky then went over the regular inverse function theorem. This is because nobody in the class could recall the proof of the inverse function theorem from memory. Nabutovsky gave an interesting proof. Notice that

$$f(\vec{x}) = f(\vec{x}_0) + Df(\vec{x}_0)(\vec{x} - \vec{x}_0) + O(||\vec{x} - \vec{x}_0||^2),$$

therefore it's possible to define a contraction mapping and then use this to get a unique inverse point. If we can show this point varies continuously on our output point, then chain rule shows it's differentiable and gives the formula. The contraction mapping was blocked by Nabutovsky's body, but if we say $f(x^*) = y$, and assume without loss of generality that $f(\vec{x}_0) = 0$ and $x_0 = 0$, then

$$\Gamma(x^*)(p) = df(0)^{-1}(f(p) - y) + x^*.$$

[1] Nabutovsky seems to be using the notation M^n to refer to an n-dimensional manifold. It's not the product.

Then I believe Banach gives us that x^* is the unique fixed point, and we can use another argument to get continuity. Nabutovsky however said that there is an entirely elementary proof of the inverse function theorem, which we can prove by first proving the implicit function theorem, then showing their equivalence.

Let $\vec{x}_0 \in \Omega \subseteq \mathbb{R}^{n+m}$, and $f : \Omega \to \mathbb{R}^n$. Then $Df(\vec{x}_0)$ is an $n \times (n+m)$ matrix

$$\begin{pmatrix} \partial_1 f_1 & \cdots & \partial_{n+m} f_1 \\ \vdots & \ddots & \vdots \\ \partial_1 f_n & \cdots & \partial_{n+m} f_n \end{pmatrix}.$$

If we assume that $Df|_{(x_1,\ldots,x_n)}(\vec{x}_n)$ is invertible, then there exists some $\Omega_1 \subseteq \mathbb{R}^n$, $\Omega_2 \subseteq \mathbb{R}^m$ (both open), and a map $\varphi = \begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_m \end{pmatrix}$ with $\varphi : \Omega_1 \to \Omega_2$ such that

$$f(x_1, \ldots, x_n, \varphi_1(x_1, \ldots, x_n), \ldots, \varphi_m(x_1, \ldots, x_n)) = f(\vec{x}_0),$$

for $(x_1, \ldots, x_n) \in \Omega_1$.

We can actually prove the inverse function theorem from the implicit function theorem (the other implication is in Spivak, but he did do it explicitly in class), by defining

$$F(\vec{x}, \vec{b}) = f(\vec{x}) - \vec{b},$$

and then if $F(\vec{x}, \vec{b}) = 0$, we can find $\vec{x}$ as a function of $\vec{b}$. Therefore to prove all our theorems, we need only prove the implicit function theorem. But there is an elementary proof.

Proof (Implicit Function Theorem). Suppose that $f : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$. Then, suppose without loss of generality that $f(0,0) = 0$, and $Df_{(x_1,\ldots,x_n)}(0,0)$ is invertible.

For the purposes of illustration, let us first consider the case where $n = 1$. Then in this case we have that $\partial_1 f(0,0) \neq 0$. This implies that on some neighborhood U of 0, $\partial_1 f(0,0)$ all has the same sign. Let us assume without loss of generality that it is positive. This implies that along the x-axis our function is strictly increasing. But since $f(0,0) = 0$, this implies that for $(x,0) \in U$, $f(x,0) < 0$ when $x < 0$ and $f(x,0) > 0$, when $x > 0$.

Fix some $a > 0$ such that $(-a,a) \times \{0\} \subseteq U$. Now, find an $\varepsilon > 0$ such that for $(x,y) \in\in B_\varepsilon(-a,0)$, $f(x,y) < 0$, and for $(x,y) \in B_\varepsilon(a,0)$, $f(x,y) > 0$, and

$$[-a,a] \times (-\varepsilon, \varepsilon) \subseteq U.$$

It follows that for any $y_0 \in (-\varepsilon, \varepsilon)$,

$$f(-a, y_0) < 0, \qquad f(a, y_0) > 0,$$

and so by the intermediate value theorem, there exists $x_0 \in (-a,a)$ such that $f(x_0, y_0) = 0$.

This implies that each y_0 has a unique zero along it. We also know it depends continuously on y_0 by the continuity of f, and therefore we can handwave away any requirements of differentiability or invertibility (Nabutovsky said it's easy—so it probably isn't[2]).

The proof for $n \neq 1$ is the exact same except with parallelepipeds instead of rectangles. To prove it for $m \neq 1$, use induction on m to reduce it into showing it for $m-1$ dimensions and 1 dimension independently. This is fairly clear. $\square$

[2] Many of the above details were filled in by me, he mostly just gave intuition and a proof sketch

x) So let's return to the inverse function theorem on manifolds. Given X, Y manifolds of the same dimension k, if $df(x_0) : T_{x_0}X^k \to T_{f(x_0)}Y^k(x_0)$ is an isomorphism, then by the inverse function theorem in $\mathbb{R}^n$, we've proven the inverse function theorem for manifolds.

We say that a function $f : X^n \to Y^m$ is an immersion if for all $x \in X^n$, $df(x)$ is injective. If $n > m$ then it is clear that f will never be an immersion. Thus we have the counterpart of a submersion, which is a map such that $df(x)$ is surjective, for all $x \in X^n$.

What the inverse function theorem for manifolds proves is that if f is an immersion and a submersion, then f is a *local* diffeomorphism. There is a related notion. We say that $x \in X^n$ is a regular point of X if $df(x)(T_xX^n) = T_{f(x)}Y^m$. Therefore, f is a submersion if and only if every point of X^n is regular.

We say that x is critical if it is not regular. If $f : X^n \to Y^m$ and $n < m$, then every point of X^n is a critical point.

If x is a critical point, then we call $f(x)$ a critical value. A regular value $y \in Y^m$ is a value which is not critical—for all critical point $x \in X^n$, $f(x) \neq y$. It follows that if $n < m$, then $f(X^n)$ are the critical values of Y^m. Notice that $Y^m \setminus f(X^n)$ are all regular values.[3]

Now how do we actually apply the implicit and inverse function theorem? Well consider $f : \mathbb{R}^{n+m} \to \mathbb{R}^n$ and f is a submersion, with $f(\vec{x}_0) = \vec{0} \in \mathbb{R}^n$. Suppose that

$$f(x_1, \ldots, x_n, x_{n+1}, \ldots, x_{n+m}) = \vec{a}.$$

Then a priori we know that $\{\vec{a} : f(\vec{x}) = \vec{a}\}$ is nonempty, but not much else. On the other hand, with the implicit function theorem, we know that this set is an m-dimensional manifold.

If $df(\vec{x}) : T_{\vec{x}}\mathbb{R}^{n+m} \to T_{\vec{a}}\mathbb{R}^n$ is a surjection. Then it has full rank when we consider it as a matrix, so we can use the implicit function theorem to get φ which parameterizes the image locally.

Let us consider a generalization: If $f : X^n \to Y^m$ is a submersion (we know therefore that $n \geq m$), then for all $y \in f(X^n) \subseteq Y^m$, then $f^{-1}(y)$ is an $(m-n)$-manifold. Even further, if y is regular, then $f^{-1}(y)$ is an $(m-n)$-manifold. This is the preimage theorem.

Think about our domain as being X^n and our range being $\mathbb{R}^n$, with $f(x) = x$ for simplicity. Then we can just use our local parameterization to get our theorem, because the implicit function theorem is a local statement.

Let's look more at immersions and submersions. If f is an immersion, then the canonical immersion is from $\mathbb{R}^n \to \mathbb{R}^m$ where

$$(x_1, \ldots, x_n) \mapsto (x_1, \ldots, x_n, \overbrace{0, \ldots, 0}^{n-m \text{ 0's}})$$

If $f : X^n \to Y^m$ is an immersion, then for all $x \in X^n$, there exists some open sets $U \subseteq \mathbb{R}^n$, $V \subseteq \mathbb{R}^m$, where $x \in U$, $f(x) \in V$, with paramaterizations $\varphi : U \to X^n$, $\psi : V \to Y^m$, such that

$$\begin{tikzcd} X^n \arrow[r, "f"] & Y^m \\ U \arrow[u, "\varphi"] \arrow[r, "\text{canon}"'] & V \arrow[u, "\psi"'] \end{tikzcd}.$$

Analogously, the canonical submersion maps $(x_1, \ldots, x_m, \ldots, x_n)$ to $(x_1, \ldots, x_m)$. This is just the orthogonal projection. The submersion theorem shows us that the exact same commutative diagram commutes, except it's the canonical surjection at the bottom this time. This implies the preimage theorem when f is a submersion immediately.

[3] cf. Sard's Theorem

§3 Day 3: Submersion, Immersion, Preimage Theorems (Jan. 13, 2025)

As a notation (which Nabutovsky did not use) I will say that π is our canonical submersion and ι is our canonical immersion. This is because Nabutovsky wrote it out in plain text every time he referred to it, which is harder to fit into latex. Use this notation at your own risk.

Let $f : X^k \to Y^\ell$ be a function (where $k \leq \ell$). Suppose that for some $x_0 \in X^k$, $df(x_0) : TX^k \to TY^\ell$ is injective. We want to show that there are open sets $U \subseteq \mathbb{R}^k$, $V \subseteq \mathbb{R}^\ell$, with $x_0 \in U$, and parameterizations $\varphi : U \to X^k$ and $\psi : V \to Y^\ell$ such that the following diagram commutes.

$$\begin{tikzcd} X \arrow[r, "f"] & Y \\ U \arrow[u, "\varphi"] \arrow[r, "\iota"] & V \arrow[u, "\psi"'] \end{tikzcd}$$

Recall that our canonical immersion $\iota : \mathbb{R}^k \to \mathbb{R}^\ell$ is just

$$\iota(x_1, \ldots, x_k) = (x_1, \ldots, x_k, \underbrace{0, \ldots, 0}_{\ell - k \text{ zeroes}}).$$

Despite Nabutovsky's incomprehensible summary, the idea behind this is actually quite simple. Choose arbitrary parameterizations $\varphi : U \to X^k$ and $\psi : U' \to Y^\ell$ such that $x_0 \in \operatorname{im} \varphi$, and the following diagram commutes for some $g : U \to U'$

$$\begin{tikzcd} X \arrow[r, "f"] & Y \\ U \arrow[u, "\varphi"] \arrow[r, "g"] & U' \arrow[u, "\psi"'] \end{tikzcd}$$

Without loss of generality, we can assume that the top $\ell \times \ell$ minor of Dg (that is, the first ℓ rows of Dg) is invertible. Define $G : \mathbb{R}^k \times \mathbb{R}^{\ell - k} \to \mathbb{R}^\ell$ by

$$G(x, z) = g(x) + z.$$

Then we have that

$$DG = \left(\begin{array}{c|c} Dg & \begin{array}{c} 0 \\ \hline I_{k-\ell} \end{array} \end{array} \right),$$

so G is a local diffeomorphism. Why do we care? Let $V = \iota(U)$. Then we have that the following diagram commutes:

$$\begin{tikzcd} V \arrow[dr, "G"'] & U \arrow[l, "\iota"'] \arrow[r, "\varphi"] \arrow[d, "g"] & X \arrow[d, "f"] \\ & U' \arrow[r, "\psi"] & Y \end{tikzcd}$$

Therefore removing the extraneous pieces, the following diagram commutes

$$\begin{tikzcd} X \arrow[r, "f"] & Y \\ U \arrow[u, "\varphi"] \arrow[r, "\iota"] & V \arrow[u, "\psi \circ G"'] \end{tikzcd}$$

Since G is a local diffeomorphism, by shrinking our sets U and V sufficiently (but ensuring that that $x_0 \in U$), we can get that $G|_U$ is a global diffeomorphism and therefore $\psi \circ G|_U$ is a parameterization, proving the theorem.

For the submersion theorem, we do almost the exact same thing. We want that if $f : X^k \to Y^\ell$ and $x_0 \in X^k$ is such that $df(x_0)$ is surjective, then there exists parameterizations $\varphi : U \to X^k$, $\psi : V \to Y^\ell$ such that $x_0 \in \operatorname{im}\varphi$, andthe following diagram commutes:

$$\begin{tikzcd} X \arrow[r, "f"] & Y \\ U \arrow[u, "\varphi"] \arrow[r, "\pi"] & V \arrow[u, "\psi"'] \end{tikzcd}.$$

For this purpose, choose parameterizations $\varphi : U \to X^k$ and $\psi : U' \to Y^\ell$ such that the following commutes.

$$\begin{tikzcd} X \arrow[r, "f"] & Y \\ U \arrow[u, "\varphi"] \arrow[r, "g"] & U' \arrow[u, "\psi"'] \end{tikzcd}.$$

If we set $G : U \to U' \times \mathbb{R}^{k-\ell}$ and

$$G(x_1, \ldots, x_k) = (g(x_1, \ldots, x_k), x_{\ell+1}, \ldots, x_k)\,,$$

$$DG = \left(\begin{array}{c|c} \multicolumn{2}{c}{Dg} \\ \hline 0 & I_{k-\ell} \end{array}\right),$$

therefore (once again, up to a change of basis) DG is invertible, and so G is a local diffeomorphism. If we set $V = G(U)$, then the following diagram commutes

$$\begin{tikzcd} V \arrow[dr, "\pi"'] & U \arrow[l, "G"'] \arrow[r, "\varphi"] \arrow[d, "g"] & X \arrow[d, "f"] \\ & U' \arrow[r, "\psi"] & Y \end{tikzcd}$$

which implies that so too does

$$\begin{tikzcd} X \arrow[r, "f"] & Y \\ U \arrow[u, "\varphi\circ G^{-1}"] \arrow[r, "\pi"] & V \arrow[u, "\psi"'] \end{tikzcd}$$

which once more yields our desired result by shrinking U and V.

So we are done the immersion and submersion theorems. Let us look at the preimage theorem. If $y \in f(X)$ is regular for $f : X^k \to Y^\ell$, then $f^{-1}(y)$ is locally diffeomorphic to $\mathbb{R}^{k-\ell}$.

This is fairly intuitive in light of the submersion theorem. This is because a direct application submersion theorem yields our parameterization at any given point of X^k. Given $x_0 \in f^{-1}(y)$ (such that f is a submersion at x_0), we know that there exists φ, ψ such that the following diagram commutes

$$\begin{tikzcd} X \arrow[r, "f"] & Y \\ U \arrow[u, "\varphi"] \arrow[r, "\pi"] & V \arrow[u, "\psi"'] \end{tikzcd}$$

where $\varphi^{-1}(x_0) \in U$ and $\psi^{-1}(y) \in V$. But then $\pi^{-1}(\psi^{-1}(y))$ gives a subspace of U which parameterizes $f^{-1}(y)$ via the map φ.

Why do we care? All of this is kind of nebulous, so it might be nice to check out a concrete example. $M_{n\times n}$, the $n \times n$ matrices, can be thought of as basically $\mathbb{R}^{n^2}$. If we set $\mathrm{Sym}_{n\times n}$ to be the $n \times n$ symmetric matrices, then we have a subspace (it can be identified with $\mathbb{R}^{\frac{n(n+1)}{2}}$ by only controlling the entries of the lower diagonal). Now, if we consider the map $f : M_{n\times n} \to \mathrm{Sym}_{n\times n}$

$$f(A) = A^\top A,$$

then $f^{-1}(I)$ will be all the orthogonal matrices $O_{n\times n}$. To show this is a manifold, we can consider df, and show that $df(A)$ is surjective whenever A is orthogonal.

$$\begin{aligned}
\lim_{h\to 0} \frac{f(A+hB) - f(A)}{h} &= \lim_{h\to 0} \frac{A^\top A + hB^\top A + hA^\top B + h^2 B^\top B - A^\top A}{h} \\
&= \lim_{h\to 0} B^\top A + hA^\top B + hB^\top B \\
&= B^\top A + A^\top B.
\end{aligned}$$

Fix some symmetric matrix C. Then we want to construct B such that $B^\top A + A^\top B = C$. But we have that $C = C^\top$ by its orthogonality, so

$$\begin{aligned}
B^\top A + A^\top B &= C \\
B^\top A + A^\top B &= \frac{1}{2}\left(C^\top + C\right) \\
A^\top B &= \frac{C}{2} \\
B &= \left(A^\top\right)^{-1} \frac{C}{2} \\
B &= \frac{AC}{2}
\end{aligned}$$

therefore for all orthogonal A and symmetric C there exists some matrix B such that $df(A)(B) = C$, which implies that $df(A)$ is a surjection at every orthogonal A, and therefore I is a regular value of f. This implies that $f^{-1}(I) = O_{n\times n}$ is a manifold. Furthermore, its dimension will be $\dim M_{n\times n} - \dim S_{n\times n} = \frac{n(n-1)}{2}$.

Next up we ask when immersions yield manifolds. If $f : X^k \to Y^\ell$ is an immersion, is $f(X^k)$ a manifold? Well not necessarily, consider something like $X = \mathbb{R}$ and $f(x) = (\sin t, \sin t \cos t)$. This will be a lemniscate (looks a bit like ∞) which is not smooth at 0. What if we require f is injective? Then is it a manifold? Still not necessarily. Consider something like $f(x) = (x, \alpha x)$ for some irrational α, then consider the map from $\mathbb{R}^2 \to S^1 \times S^1$ given by

$$f(s,t) = (\cos s, \sin s, \cos t, \sin t)\,.$$

The intuition is that this mapping will be dense in $S^1 \times S^1$ (by a Diophantine approximation argument), but its image should only be one dimensional because f is smooth, which is a contradiction.

Next time we will cover one way to ensure our image is a manifold.

§4 Day 4: Embeddings and Measure (Jan. 22, 2025)

We recapped why our previous construction of the immersion of the line into the torus was not a submanifold. I'm sure you can deduce the argument—no neighborhood locally looks like an interval because they are dense in the torus, but they don't cover any open subset of the torus so it doesn't yield a higher dimensional manifold (removing a single point disconnects our immersion of the line).

Therefore, we need a better notion of a map to make it so that when $f\colon X \to Y$ is *this kind* of map, its image is a submanifold of Y. To this end, we define a preliminary (purely topological) property that maps can have/

Definition 4.1. A map $\varphi : X \to Y$ is proper if for all compact subsets $C \subseteq Y$, $\varphi^{-1}(C)$ is compact as well.

The intuition is that we want the preimage of bounded sets to be bounded.[4]

Definition 4.2. $f : X \to Y$ is an embedding if f is an injective proper immersion.

The obvious question raised is whether this is sufficient to accomplish our goal. Unsurprisingly,[5] it is.

Theorem 4.3. If X, Y are manifolds and $f : X \to Y$ is an embedding, then $f(X)$ is a submanifold of Y.

Proof. If $x \in X$ and W is a sufficiently small neighborhood of x. We need to show first that $f(W)$ is open in $f(X)$. So suppose by way of contradiction that it is not. Then there exists a sequence $\{y_n\}_{n\in\mathbb{N}} \subseteq f(X)$ such that $y_n \in f(X) \setminus f(W)$, but $\lim_{n\to\infty} y_n = y$ has $y \in f(W)$.

We know that $\{y_1, y_2, \dots\} \cup \{y\}$ is compact, so $f^{-1}(\{y_1, y_2, \dots\} \cup \{y\})$ is compact. By shrinking W to be sufficiently small such that the local immersion theorem holds, we have that if we set $x_n = f^{-1}(y_n)$, then we can assume (wlog by taking a subsequence—since $\{x_1, x_2, \dots\} \cup \{f^{-1}(y)\}$ is compact there exists a suitable convergent subsequence) that $\lim_{n\to\infty} x_n = z$ for some $z \in X$.

From here, since $\{x_1, x_2, \dots\} \subseteq X \setminus W$, $z \in X \setminus W$, as $X \setminus W$ is closed. On the other hand, we have that since f is continuous,

$$y = \lim_{n\to\infty} y_n = \lim_{n\to\infty} f(x_n) = f\left(\lim_{n\to\infty} x_n\right) = f(z),$$

so $z = f^{-1}(y) \in W$, which is a contradiction. Therefore, $f(W)$ must be open.

Since it is open, we can shrink W to be sufficiently small for the local immersion theorem to hold, and therefore we have parameterized Y locally. □

§4.1 Sard's Theorem

The most complicated parts of the course have already happened by now. The book would typically go to transversality and then to homotopy and stability, but we will skip to Sard's theorem. Many of the things from differential topology require algebraic topology to prove. Since we don't know algebraic topology, we have to take a more roundabout route to prove many things. I don't remember what the point of this discussion by Nabutovsky was.

[4] In a metric space with the Heine-Borel property, this is indeed equivalent. Since manifolds by our definition are metrizable—they all live in $\mathbb{R}^n$ so use the restriction of the metric—this is really what we care about.

[5] It is unsurprising from a pedagogical point of view; Nabutovsky would not have defined it this way otherwise.

Definition 4.4. A subset $A \subseteq \mathbb{R}^n$ is measure zero if for all $\varepsilon > 0$ there exists a cover $\{U_k\}_{k\in\mathbb{N}}$ of A by parallelepipeds such that

$$\sum_{k\in\mathbb{N}} |U_k| < \varepsilon.$$

Here, we say that if $U_k = (a_1, b_1) \times \cdots \times (a_n, b_n)$, then

$$|U_k| = \prod_{i=1}^{n} b_i - a_i.$$

Notice that given a countable collection of measure zero sets, their union has measure zero. As such, we have a natural characterization of measure zero sets of manifolds which says that $A \subseteq X$ is measure zero if for every parameterization φ of X, $\varphi^{-1}(A)$ has measure zero (notice this follows by manifolds being second-countable).

Theorem 4.5. $[0,1]^n$ has measure 1. In other words, that any covering of $[0,1]^n$ by parallelepipeds has volume of at least 1.

Proof. While one could use the typical proof we learn, there's a much more clever argument that Nabutovsky presented. Suppose that $\{U_n\}$ is a cover of $[0,1]^n$ by parallelepipeds. Since $[0,1]^n$ is compact, without loss of generality we can assume that this set is $\{U_1, \ldots, U_N\}$. Let $\lambda \in \mathbb{R}_{>0}$ and define $T_\lambda x = \lambda x$ for all $x \in \mathbb{R}^n$. Then we have that $T_\lambda([0,1]^n) = [0,\lambda]^n$, and $|T_\lambda(U_n)| = \lambda^n |U_n|$.

Define now $N(A) = |A \cap \mathbb{Z}^n|$ for all $A \subseteq \mathbb{R}^n$. Then we know that for any parallelepiped U,

$$\lim_{\lambda\to\infty} \frac{N(T_\lambda(U))}{\lambda^n} = |U|.$$

But now notice that

$$\begin{aligned}
\sum_{n=1}^{N} |U_n| &= \sum_{n=1}^{N} \lim_{\lambda\to\infty} \frac{N(T_\lambda(U_n))}{\lambda^n} \\
&= \lim_{\lambda\to\infty} \sum_{n=1}^{N} \frac{N(T_\lambda(U_n))}{\lambda^n} \\
&\geq \lim_{\lambda\to\infty} N\left(\bigcup_{n=1}^{N} T_\lambda(U_n)\right) \frac{1}{\lambda^n} \\
&\geq \lim_{\lambda\to\infty} N\left([0,\lambda]^n\right) \frac{1}{\lambda^n} \\
&= 1
\end{aligned}$$

and therefore we are done. □

§5 Day 5: Sard's Theorem (Jan 22. 2025)

In Euclidean space, we say that A is of measure 0 if for all $\varepsilon > 0$ there exists a cover $\{U_i\}$ of A such that $\sum |U_i| < \varepsilon$, where U_i are parallelepipeds.[6] Notice that $\mathbb{R}^k \subseteq \mathbb{R}^n$ is of measure zero. To do this, cover $\mathbb{R}^k$ by

$$[-N, N]^k \times \left[\frac{-\varepsilon}{N^k 2^N}, \frac{\varepsilon}{N^k 2^N}\right]$$

You can modify this formula to make it sum to exactly ε but that is too much work for me.

Theorem 5.1. Let $U \subseteq \mathbb{R}^n$ and $f\colon U \to \mathbb{R}^n$ be smooth. If $A \subseteq U \subseteq \mathbb{R}^n$ is closed and A has measure zero, then $f(A)$ has measure zero.

Proof. Nabutovsky's proof was quite overcomplicated, so after class I asked him about a different argument and he said that my argument was better and during the Sard's theorem proof he thought of it independently too. I will put this argument rather than Nabutovsky's original one.

First of all, there is a problem which may not be immediately obvious when you start trying to prove this. Namely, just because A has arbitrarily small covers in $\mathbb{R}^n$, it doesn't mean that A necessarily has arbitrarily small covers in U. To resolve this, we can use an idea reminiscent of MAT335:

Define a grid of hyper-cubes with sidelengths δ of the form

$$\mathcal{B}_\delta = \left\{ \left[\frac{k_1}{\delta}, \frac{k_1+1}{\delta}\right] \times \cdots \times \left[\frac{k_n}{\delta}, \frac{k_n+1}{\delta}\right] : k_1, \dots, k_n \in \mathbb{Z} \right\}.$$

Then these will clearly cover $\mathbb{R}^n$. Now, notice that if we take all of the cubes of this form containing a (rectilinear) parallelepiped P, they will be contained in the $\delta\sqrt{n}$-enlargement of P—namely,

$$E_{\delta\sqrt{n}}(P) = \left\{ x \in \mathbb{R}^n : d(x, P) \leq \delta\sqrt{n} \right\},$$

where $E_\bullet$ provides our enlargement function. From here, $|E_{\delta\sqrt{n}}(P)| \leq (1 + 2\delta\sqrt{n})^n |P|$.

Now, notice that since A is closed, there exists some $r > 0$ such that $r < d(A, U^c)$. This is because U^c is closed, so the distance is well-defined. Fix $\varepsilon > 0$. Choose a cover $\{U_i : U_i \in \mathbb{N}\}$ of A such that

$$\sum_{i \in \mathbb{N}} |U_i| < \varepsilon.$$

Choose $\delta > 0$ such that $\delta\sqrt{n} < r$ and

$$\left(1 + 2\delta\sqrt{n}\right)^n \cdot \sum_{i \in \mathbb{N}} |U_i| < \varepsilon.$$

It follows that the cover

$$\mathcal{U} = \{B \in \mathcal{B}_\delta : B \cap A \neq \varnothing\}$$

satisfies two important properties:

1. Since $\operatorname{diam}(B) = \delta\sqrt{n}$ for all $B \in \mathcal{B}_\delta$, we have that our cover $\mathcal{U}$ consists of only rectangles contained in U.

[6] I *think* Nabutovsky said they must be rectilinear, but I don't know. Probably stick to that on midterms.

2. For any $B_1, B_2 \in \mathcal{U}$, $B_1 \cap B_2$ is of measure zero (it's contained in some lower dimensional hyperplane), and for all $B \in \mathcal{U}$,
$$B \subseteq \bigcup_{i \in \mathbb{N}} E_{\delta\sqrt{n}}(U_i).$$
This implies that
$$\sum_{B \in \mathcal{U}} |B| \leq \sum_{i \in \mathbb{N}} (1 + \delta\sqrt{n}) \cdot |U_i|,$$
and so
$$\sum_{B \in \mathcal{U}} |B| \leq (1 + \delta\sqrt{n}) \cdot \sum_{i \in \mathbb{N}} |U_i| < \varepsilon.$$

Thus, $\mathcal{U}$ provides of cover of A of cubes contained within U with volume less than ε.

Now that we have this cover, we need to show that the volume is preserved under our function. To do this we can suppose without loss of generality that A is bounded, and by shrinking our U a little bit we can suppose that f is lipschitz with lipschitz constant K. From here, $\operatorname{diam}(f(B)) \leq K \cdot \operatorname{diam}(B)$ for any set $B \subseteq U$. Since K is fixed, we are done. □

Definition 5.2. A set $A \subseteq M$ has measure 0 if for every parameterization $\varphi\colon U \to M$, $\varphi^{-1}(A)$ has measure 0. This is sensible by the above theorem.

Theorem 5.3 (Mini-Sard's)**.** *If $f\colon M^m \to N^n$, and $m < n$, then the critical values $C(f)$ is a set of measure zero.*

Proof. Consider the following diagram

$$\begin{tikzcd}
 & M \arrow[r, "f"] & N \\
U \arrow[ur, "\varphi"] \arrow[r, "\iota"] & U \times \{0\}^{n-m} \arrow[u, "\varphi\times 0"'] & V \arrow[u, "\psi"]
\end{tikzcd}$$

where ι is the canonical immersion. By definition we have that we can extend $\varphi \times 0$ to some neighborhood of $U \times \{0\}^{n-m}$ smoothly (this is our definition of $\varphi \times 0$ being smooth). It follows that the map $\psi^{-1} \circ f \circ \varphi \times 0 : U \times \{0\}^{n-m} \to V$ is a smooth map, and therefore we can extend it to some open set $\tilde{U}$ containing $U \times \{0\}^{n-m}$. I will call this extended map $\tilde{f}$. Then $U \times \{0\}^{n-m}$ has measure zero in $\tilde{U}$, so by our above theorem, $\tilde{f}(U \times \{0\}^{n-m}) = f(M \cap \varphi(U))$ will have measure zero.

Since we let our parameterizations be arbitrary, this proves the theorem.[7] □

Lemma 5.4 (Fubini's theorem)**.** *Let $A \subseteq \mathbb{R}^{n+1}$ be closed.*[8] *Define*
$$A_c = A \cap \{(x_1, \ldots, x_{n+1}) \in \mathbb{R}^n : x_{n+1} = c\}.$$
Then we have that A has measure zero if every $\pi(A_c)$ has measure zero, where $\pi\colon \mathbb{R}^{n+1} \to \mathbb{R}^n$ is the orthogonal projection along the $n+1$th coordinate.

[7] I might have fucked up some details, but I'm sure you see the idea.

[8] This theorem is more generally true. If we know that A is Lebesgue-measurable, then it holds. Closed sets are contained in our measurable sets of $\mathbb{R}^{n+1}$ (if we can establish that all the sets in the Borel σ-algebra are measurable then this is immediate). But be careful not to apply it accidentally without the closedness condition.

Proof. We can reduce to the case where A is bounded. Since we can cover $\mathbb{R}^n$ by countably many parallelepipeds, if we know that $A \cap P$ has measure zero for every parallelepiped, then let $\{P_1, \ldots\}$ be our cover of $\mathbb{R}^n$ by (countable) parallelepipeds, and cover $A \cap P_k$ by a cover with volume $\varepsilon \cdot 2^{-k}$. It follows that this will be a cover of A with volume ε.

As such, consider a bounded closed set satisfying the above property. Then A is compact. Suppose that $A \subseteq \mathbb{R}^n \times [a, b]$ for some $a, b \in \mathbb{R}$. Now, let V be an open neighborhood of A_c for some $c \in [a, b]$. I claim that there exists some $I \subseteq [a, b]$ such that

$$A_{c_0} \subseteq V \times I, \quad c_0 \in I.$$

Otherwise would yield a sequence of points $\{x_n\}$ with $x_n \in A_{c_n}$ where $c_n \to c$, but $x_n \notin V$ for all $n \in \mathbb{N}$, This would imply that $\lim x_n \notin A_c$, which would imply that A is not closed.

From here, it's just a matter of choosing open neighborhoods V_c of each A_c and open intervals I_c such that $c \in I_c$, $A \cap U \times I \subseteq V_c \times I_c$, and ensuring that V_c has measure less than ε.[9] Then it's a matter of taking a finite subcover of A by $V_c \times I_c$, and potentially modifying these finitely many I_c such that their lengths sum to less than $2(b - a)$. This is all quite simple so I will leave the details to you rather than writing out the endless details necessary. $\square$

Remark 5.5. This is not an if and only if. Consider $A_c = \mathbb{R}^n \times \{0\}$. I almost fell for this.

Theorem 5.6 (Sard's Theorem)**.** Given a map $f : M \to N$, the critical values C of f have measure zero.

Remark 5.7. What's the intuition here? Well we'll use an identity to yield the most important case: If μ denotes the Lebesgue measure of a set, then we have that for all $A \subseteq \mathbb{R}^n$,

$$\mu(A) \leq C_n \cdot \operatorname{diam}(A)^n,$$

where C_n is the volume of a unit diameter ball in $\mathbb{R}^n$, with equality holding of A is a ball in $\mathbb{R}^n$. As such, if $f : \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz with constant K,

$$\mu(f(A)) \leq C_m \cdot \operatorname{diam}(f(A))^m \leq C_m \cdot K^m \cdot \operatorname{diam}(A)^m \leq \frac{C_m K^m}{C_n} \mu(B)^{\frac{m}{n}},$$

where B is a ball containing A. As such, when we are mapping from low dimensions to higher dimensions, we have that $\frac{m}{n} > 1$, so as we cover A with smaller and smaller balls, this bounds the volume of $f(A)$ by smaller and smaller quantities. But when $m \leq n$, we need a better remainder.

Bu imagine for a moment that rather than just Lipschitz, we had the condition that for some $\alpha > \frac{n}{m}$

$$\operatorname{diam}(f(A)) \leq C \cdot \operatorname{diam}(A)^\alpha.$$

Then our inequalities would look like

$$\mu(f(A)) \leq C_m \cdot \operatorname{diam}(f(A))^{\alpha \cdot m} \leq C_m \cdot K^m \cdot \operatorname{diam}(A)^{\alpha \cdot m} \leq \frac{C_m K^m}{C_n} \mu(B)^{\alpha \cdot \frac{m}{n}},$$

which would yield itself to the same proof as above. Unfortunately requiring such a condition globally is equivalent to requiring that our function is constant (it's equivalent

[9] It's not necessary to cite the full strength of Lebesgue measure here—for open sets the outer measure equals the Lebesgue measure.

to α-Hölder, and we're considering the case of $\alpha > 1$). But if we only want it at a single point, then we can acheive this by assuming that all of the kth derivatives, where $k \leq \frac{n}{m}$ vanish. Then our function would locally look like some homogeneous function. Specifically, given $v \in \mathbb{R}^n$ with and $h \in \mathbb{R}$, we would have

$$\lim_{h\to 0} \frac{f(x+hv) - f(x)}{\sqrt[m]{h^n}} = 0,$$

which is precisely what we want. Then we could essentially just repeat the proof of mini-Sard verbatim and conclude. But why are we even considering such a dream scenario? After all, we don't even know that the first derivatives vanish. Well in Nabutovsky's words, "This is precisely the drama of the proof." Let me move on to proving it.

Proof. Notice that a zero dimensional manifold is just a discrete collection of points, so its image will always have measure zero, as its image will be countable. So we will take this as a base case and use strong induction to prove the theorem.

Let $f\colon M^n \to N^m$ be smooth, where $n \geq 0$ and $m \geq 1$.[10] Now, let C be the set of critical values of f, and define C_i to be the critical values of f where all of the first i derivatives vanish.

Let us first reduce the problem to something slightly simpler. Given charts (U, φ) of M and (V, ψ) of N, if we can show that $\psi(C)$ has measure zero, then we are done. Therefore, since U has a countable compact exhaustion, this reduces to showing that $\psi(C \cap f \circ \varphi^{-1}(X))$ has measure zero for every compact $X \subseteq U$. But then this is the same as showing that $f : A \subseteq \mathbb{R}^n \to B \subseteq \mathbb{R}^m$ satisfies Sard's theorem, where A and B are both bounded sets. This makes life easier.

Now, there are three cases we need to consider:

1. First, let $\mathcal{B}_\delta$ be the collection of rectangles of sidelength δ, as in the proof of Theorem 5.1. Now, I claim that for every $\alpha > 0$ there exists sufficiently large k such that for every $x \in C_k$, there exists some sufficiently large δ and some constant c such that for all $B \in \mathcal{B}_\delta$ with $x \in B$,

$$\mu(f(B)) \geq c \cdot \mu(B)^\alpha.$$

To prove this, notice that nearby x, f satisfies

$$\lim_{h\to 0} \frac{f(x+hv) - f(x)}{h^{k-1}} = 0,$$

which implies that on for any small enough neighborhood U of x,

$$\operatorname{diam}(f(U)) < c \cdot \operatorname{diam}(U)^{k-1},$$

which (by the discussion in the remark) yields the claim. Now, since we have that (for small enough δ)

$$\sum_{B \in \mathcal{B}_\delta, B\cap C_k \neq \varnothing} \mu(B) \leq \mu(A),$$

we have that

$$\mu(B) \leq \frac{\mu(A)}{|\mathcal{B}_\delta|}$$

so therefore as $\delta \to 0$, if $\alpha > 1$, then

$$\sum_{B \in \mathcal{B}_\delta, B\cap C_k \neq \varnothing} \mu(f(B)) \leq \sum_{B \in \mathcal{B}_\delta, B\cap C_k \neq \varnothing} c \cdot \mu(B)^\alpha \to 0.$$

[10] I hate that n is the dimension for M and m is the dimension for N.

2. We will show that $C \setminus C_1$ has measure zero. If $y = f(x_0)$, then x_0 is critical, but there exists i, j such that $D_i f^j(x_0) \neq 0$. Suppose that $i = j = 1$ for simplicity and consider the following change of coordinates:

   Let $h\colon B_\delta(x_0) \to \mathbb{R}^n$ where
$$h(x_1, \dots, x_n) = (f(x_1, \dots, x_n), x_2, \dots, x_n)\,.$$
   Then for some nonzero m_l
$$Dh(x_0) = \begin{bmatrix} \lambda & 0 & \cdots & 0 \\ * & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ * & 0 & \cdots & 1 \end{bmatrix},$$
   so Dh is invertible, and thus we can assume (potentially shrinking δ) that h is a diffeomorphism. Define h^{-1} and use it to change coordinates. Then
$$d(f \circ h^{-1}) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ * & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & * \end{bmatrix}$$
   so considering the minor of this matrix attained by removing the top row and the leftmost column, we get a matrix which is singular (otherwise $d(f \circ h^{-1})$ would be invertible and thus we wouldn't have that x_0 is a critical point).

   Suppose that x_0 has first coordinate a and y has first coordinate b. Then applying our induction hypothesis we have that the map $g_a\colon \mathbb{R}^{n-1} \to \{b\} \times \mathbb{R}^{m-1}$ given by
$$g_a(x) = f(a, x)$$
   has a set of critical points which are measure zero. It follows that $C \setminus C_1 \cap \{b\} \times \mathbb{R}^{m-1}$ is measure zero for all $b \in \mathbb{R}$, so by Fubini's theorem, $C \setminus C_1$ has measure zero as well.

3. We claim that $C_i \setminus C_{i+1}$ has measure zero. To show this, consider the coordinate maps $f_1, \dots, f_m$ of f, and notice that one of these will have a nonzero $i + 1$th derivative. As such, consider
$$a = \frac{\partial f_\ell}{\partial x_{j_1} \partial x_{j_2} \cdots \partial x_{j_i}}$$
   Then define $h(x) = (a(x), x_2, \dots, x_n)$. Then a has a nonzero derivative, so we have that dh is invertible, and thus (potentially shrinking the neighborhood once more) h is a diffeomorphism. But then all the critical values of $f \circ h^{-1}$ lie in the plane $\{0\} \times \mathbb{R}^{n-1}$, and therefore by induction this set has measure zero as well.

The three of these together yield that
$$C = C \setminus C_1 \cup C_1 \setminus C_2 \cup \cdots \cup C_{n-1} \setminus C_n \cup C_n$$
is the union of finitely many measure zero sets, so it has measure zero as well. $\square$

Corollary 5.8. Given $\{f_k\}_{k \in \mathbb{N}}$, we have that the union of their critical values has measure zero.

Corollary 5.9. The critical values of a function are meagre topologically—their complement is a dense open set.

Proof. Notice that the determinant is a continuous map and $Df(x)$ is a continuous map. As such $\det Df^{-1}(\{0\})$ is an open subset of $\mathbb{R}^n$. Now, notice that since the critical values have measure zero, they cannot contain any open set, therefore its complement intersects every open set, which is the definition of being dense. □

The important thing is that the critical values can be approximated by regular values always.

Remark 5.10 (Nabutangent)**.** Now really having the critical values meagre isn't good enough in many scenarios. Thinking about invertibility, having a critical value is the same as having an eigenvalue of zero of the Jacobian somewhere. For some things we want to ensure that not only is our matrix *invertible*, but furthermore, we want a bound on the size of the smallest eigenvalue. This leads us to something called the quantitative Sard's theorem. This shows that the values with eigenvalues smaller than some ε has small measure (not zero in general). If we have this then we can generally move to values that are not only not critical, but not *close* to being critical.[11]

[11] I don't know the details here so don't take any of this as set in stone. He didn't prove the quantitative Sard's theorem.