Differential Topology **Spring 2023**

Foundations of Manifolds

Professor Daniel Allcock Abhishek Shivkumar

Manifolds and Maps

The basic object in differential topology is a smooth manifold, as opposed to C^k, topological, real-analytic, complex analytic (etc.) manifolds. We'll start with a Calc III-esque study of surfaces in space, curves in $\mathbb{R}^{2+}$, before moving onto abstract manifolds without respect to any specific embedding. Viewing manifolds as embedded in $\mathbb{R}^n$ eases some psychological difficulties but many manifolds do not come with natural embeddings into $\mathbb{R}^n$. The natural first example of such a manifold is $\mathbb{RP}^2 = S^2/\{\pm 1\}$.

This one is not a real definition yet. Need to unpack terms; also Hausdorffness, second countability are missing.

Definition 1.1.1: Manifolds

Let M be a topological space with an open cover U_α and homeomorphisms $\varphi_\alpha : U_\alpha \to V_\alpha \subseteq \mathbb{R}^n$. When U_α and U_β overlap, we require that

$$\varphi_\beta(U_\alpha \cap U_\beta) \xrightarrow{\varphi_\alpha \circ \varphi_\beta^{-1}} \varphi_\alpha(U_\alpha \cap U_\beta)$$

is a diffeomorphism, that is, a bijective map between two subsets of $\mathbb{R}^n$ that is C^∞ with C^∞ inverse.

The symbolic complexity of this definition (in contrast to its intuitive simplicity) is essentially tied to the fact that we have to bootstrap the definition of a diffeomorphism from the one place where we know what it means: open subsets of $\mathbb{R}^n$.

Example 1.1.2

The standard example of a map that is bijective and smooth but not a diffeomorphism is $f(x) = x^3$ from $\mathbb{R}$ to itself, since $f^{-1}(y) = \sqrt[3]{y}$ is not even differentiable at 0 (so f is not even a C^1 homeomorphism).

C^∞ functions require some more discussion. For functions from $\mathbb{R}$ to itself, the standard definition suffices, e.g, f is differentiable, f' is differentiable, etc. For $n \geq 2$, C^∞ requires more than just existence of partial derivatives to all orders.

I think all that needs to be done here is add "and are continuous" to every "partial derivatives exist" to make the definition work out. The exposition here gets a little confusing.

Example 1.1.3

Consider

$$f(x,y) = \begin{cases} \frac{xy}{x^2+y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$$

One can check that f has partial derivatives of all orders because it vanishes identically along the x and y axes. However, f is not even continuous. Along the line $y = mx$, $f(x, y) = \frac{mx^2}{(1+m^2)x^2} = \frac{m}{1+m^2}$, so along each line through the origin, f takes a different, constant value (and therefore the limit at 0 is undefined). So the existence of partials of all orders is an artifact of the choice of coordinate system with which we described f.

Definition 1.1.4: Differentiability

Let $f : \mathbb{R}^m \to \mathbb{R}^n$, we say that f is differentiable at $x \in \mathbb{R}^m$ if there exists a linear function $\lambda : \mathbb{R}^m \to \mathbb{R}^n$ such that

$$\frac{f(x + \Delta x) - (f(x) + \lambda(\Delta x))}{\|\Delta x\|} \to 0$$

as $\|\Delta x\| \to 0$. If such a λ exists, then it is unique, and it is called the derivative of f at x.

Definition 1.1.5: C^k

f is differentiable on an open subset U of $\mathbb{R}^n$ if f is differentiable at each $x \in U$. Consider the map $(x \mapsto \lambda_x) : U \to \mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^n)$ from points of U to linear maps (essentially matrices); we say f is C^1 if this assignment is continuous. Since $\mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^n) \cong \mathbb{R}^{mn}$, we can now ask if the assignment $(x \mapsto \lambda_x)$ is differentiable, in which case f is C^2, etc.

All the following definition is doing is using the natural scaling operation on vector spaces to express the local linearity condition somewhat more cleanly, and without as much formal baggage.

Definition 1.1.6

Suppose $f : V \to W$ is a function between real vector spaces, and suppose that (perhaps after translation) $f(0) = 0$. Then f is differentiable at 0 if there exists a linear map $\lambda : V \to W$ such that for any bounded neighborhood A of 0 in V and any neighborhood B of 0 in W, for all $\epsilon > 0$, there exists $t_0 > 0$ such that $f(tv) - \lambda(tv) \in t(\epsilon B)$ for all $0 < t < t_0$, all $v \in A$.

Fixing A, B as above, and $g : V \to W$, we say that g vanishes faster than linearly at 0 if for all $\epsilon > 0$, there exists $t_0 > 0$ such that $g(tA) \subseteq t(\epsilon B)$ for all $0 < t < t_0$. Then f is differentiable with derivative λ when $f - \lambda$ vanishes faster than linearly at 0.

Theorem 1.1.7

If λ as above exists, it is unique, and is called the derivative of f at 0.

Notation for the derivative varies: Df, df, Tf all appear in the literature.

Theorem 1.1.8: Chain Rule

If we have a sequence $V \xrightarrow{f} W \xrightarrow{g} X$ with f and g differentiable at $v_0 \in V$ and $f(v_0) \in W$ resp., then so is their composition at v_0, with derivative $D_{f(v_0)}g \circ D_{v_0}f$.

Corollary 1.1.9

If f, g as above are C^k on some open set in V, and g is C^k on the image of that open set (under f), then $g \circ f$ is C^k on it as well.

One can show this by induction on k and the chain rule.

Theorem 1.1.10: Inverse Function Theorem

If $f : V \to W$ is differentiable at v_0, and $D_{v_0}f : V \to W$ is an isomorphism, then f is bijective from some neighborhood of v_0 in V to some neighborhood of $f(v_0)$ in W, and f^{-1} is differentiable as well. Since $f^{-1} \circ f$ is the identity, the chain rule tells us that $D_{f(v_0)}f^{-1} = (D_{v_0}f)^{-1}$, and thus f is actually a diffeomorphism on some neighborhood of v_0.

It is traditional to skip this proof, although apparently there's a very clean proof (or at least clean exposition) by Terry Tao on MO.

We are now ready to formally define manifolds (again?).

Definition 1.1.11: Smooth Manifolds

A smooth manifold M is a Hausdorff topological space equipped with charts $(U_\alpha, \varphi_\alpha : U_\alpha \xrightarrow{\sim} V_\alpha)$ where V_α is an open set in a real vector space, such that the U_α cover M and the charts agree on overlaps, i.e,

$$\varphi_\beta(U_\alpha \cap U_\beta) \xrightarrow{\varphi_\alpha \circ \varphi_\beta^{-1}} \varphi_\alpha(U_\alpha \cap U_\beta)$$

is C^∞ with everywhere invertible derivative. We also require that each component of M has a countable basis for its topology (e.g, is second countable).

There is now a problem, that our manifolds depend explicitly on given charts, so there are many different (redundant) representatives of what we would think of as the same manifold. We will resolve this by insisting that all charts compatible with given charts are included in our set of charts, e.g, by taking a maximal atlas.

Found it mildly confusing that we didn't say that the φ_α themselves were C^∞, just their compositions on overlaps, but this is because, as above, we have to bootstrap what it means to be C^∞ from $\mathbb{R}^m \to \mathbb{R}^n$ functions, which is what $\varphi_\beta^{-1} \circ \varphi_\alpha$ is. Saying that φ_α itself is C^∞ doesn't actually mean anything at this stage.

Many authors require that M itself is second countable, which only rules out the case where M has uncountably many components. Some natural (for some definition of natural) manifolds do have uncountably many components: consider the 2-torus foliated by lines of irrational slope. Every such line is dense in T^2. One can define a topology on T^2 (the "leaf topology") that makes each leaf into a copy of $\mathbb{R}$ and all leaves separate components.

As an example for why Hausdorffness should be imposed, consider the "line with two origins," the topological space obtained by gluing two copies of $\mathbb{R}$ along $\mathbb{R} \setminus \{0\}$, which is not Hausdorff since open neighborhoods of both origins will always intersect.

The two technical criteria, Hausdorffness and that every component has a countable basis, can be expressed together as paracompactness.

As far as I know, paracompactness does not usually include Hausdorffness.

To actually specify a manifold M, you choose an underlying space and a few charts satisfying the overlap conditions, and then add in all possible charts that are compatible with the given charts. One must verify that any two charts obtained this way are compatible with each other, e.g, that "the" maximal atlas is well-defined, i.e unique.

Here Daniel introduces the following slogan, which we should not regard as actually true: "Any concept you can define for open subsets in real vector spaces that is invariant under diffeomorphisms makes sense for manifolds as well."

However, it is quite rare to actually define a manifold with charts given the huge amount of data to track. When we need to actually use local charts, we will generally just choose them on the fly rather than having them all defined at the outset. Choosing local charts around a point p for a manifold M amounts to a choice of chart around p, and functions $x_1, \cdots, x_n : M \to \mathbb{R}$ which, when taken together as a tuple, give a diffeomorphism from the chart to $\mathbb{R}^n$.

For example, a manifold M with charts $(U_\alpha, \varphi_\alpha)$, $f : M \to \mathbb{R}$ is smooth if for all $x \in M$, there exists a neighborhood U_α of x in the charts such that $f \circ \varphi_\alpha^{-1}$ is smooth as a function from some open neighborhood in $\mathbb{R}^n$ to $\mathbb{R}$.

Skipping some examples here of explicit definitions by charts, open subsets of $\mathbb{R}^n$, spheres via stereographic projection from the poles, etc.

Definition 1.1.12: Smooth Functions

A function $f : M \to \mathbb{R}$ is smooth if for all $p \in M$, there exists a chart $(U_\alpha, \varphi_\alpha)$ containing p s.t for all $U_\beta \ni p$,

$$f \circ \varphi_\beta^{-1} = (f \circ \varphi_\alpha^{-1}) \circ (\varphi_\alpha \circ \varphi_\beta^{-1})$$

where $f \circ \varphi_\alpha^{-1}$ is smooth by assumption, as is $\varphi_\alpha \circ \varphi_\beta^{-1}$ by the compatibility criterion for charts.

More examples of explicit charts that I'm not backfilling since I missed the lecture anyway. Summary is that you might need actual charts if you're a general relativist.

All this is to say that $\mathbb{R}$-valued functions on manifolds "make sense," despite the oppressive verbosity that God has required to make formal "makes sense." We play a similar game to define smooth functions $f : M \to N$ between manifolds; in particular, to sketch the main idea, given a chart $(U_\alpha, \varphi_\alpha)$ about $p \in M$, and a chart (V_β, ψ_β) about $f(p) \in N$, the function we want to consider is $\psi_\beta \circ f \circ \varphi_\alpha^{-1} : \mathbb{R}^m \to \mathbb{R}^n$, from which we can bootstrap smoothness as above from the established definition on $\mathbb{R}^n$.

Submersions

Theorem 1.2.1: Implicit Function Theorem

Suppose $f : M \to N$ is differentiable at $x \in M$. Then if the linear map $T_x M \to T_{f(x)} N$ is surjective (i.e if f is a *submersion* at x), then $f^{-1}(f(x))$ is a manifold near x. Moreover, we can choose local coordinates x_i around x and $f(x)$ such that f is the canonical surjection: $f(x_1, \cdots, x_m) = (x_1, \cdots, x_n, 0, \cdots, 0)$ with $m \geq n$.

The phrasing of the implicit function theorem in the lecture is a little confusing and doesn't fully match the statement in Guillemin-Pollack (where it is called the local submersion theorem), but hopefully the general idea is correct.

We can derive this result from the earlier Theorem 1.1.10 (the inverse function theorem) by doing some linear algebra to massage submersions (surjective differentials) and immersions (injective differentials) into local bijections.

First, let us consider the case of submersions, and suppose that $D_u f$ is surjective for $f : U \to V$. Then the complement of $\ker D_u f$ maps isomorphically onto $T_{f(u)} V$. Write $T_u U = \ker D_u f \oplus W$, where by moderate abuse of notation we can write $W = T_{f(u)} V$ and that f is the identity on W.

Is there anything funky going on with choosing a complement? Implicitly we're picking an inner product.

From this map we can construct a new map $c : U \to T_u U$ (perhaps after shrinking U to a coordinate chart where the above splitting of the tangent space can be extended) given by $(k \in K, v \in V) \mapsto k + f(v)$; clearly c is an isomorphism, so by the inverse function theorem, c is a diffeomorphism near u.

This whole bit is reconstructed from other people's notes and is pretty incomprehensible to me; the analogous statement below for immersions is more comprehensible. Also omitted an example showing that $\mathrm{SL}_n \mathbb{R}$ is a smooth manifold using the implicit function theorem applied to det.

We will apply the implicit function theorem to prove $O(n)$ (the group of orthogonal $n \times n$ matrices) is a smooth manifold. Recall that if M is a symmetric $n \times n$ matrix, then it represents a symmetric bilinear form $(x, y) \mapsto x^T M y$. $\mathrm{GL}_n \mathbb{R}$ acts on $\mathbb{R}^n$ by left multiplication, so $\mathrm{GL}_n \mathbb{R}$ acts on the set S of symmetric bilinear forms by $(M \cdot g)(x, y) := M(gx, gy)$ which, in matrices, is the map $(M, g) \mapsto g^T M g$ for g symmetric.

The orthogonal group of M are simply the elements of GL_n preserving the inner product. In matrices, $O(M) = \{g \in \mathrm{GL}_n \mathbb{R} : g^T M G = M\}$. in particular, there exists a function $\mathrm{GL}_n \mathbb{R} \to S$ given by $g \mapsto g^T M g$, and $O(M)$ is the preimage of $\{M\}$ under this map. Since $\mathrm{GL}_n \mathbb{R} \to S$ is smooth, we expect the preimage of M (which is $O(M)$) to be a manifold, which will hold if the derivative of the map is surjective at every point of $O(M)$. Let's compute the derivative at $g = I$. Let ϵ be an $n \times n$ matrix very close (in the standard norm) to zero. Then the image of $(I + \epsilon)$ in S is

$$(I + \epsilon)^T M (I + \epsilon) = M + \epsilon^T M + M\epsilon + \epsilon^T M \epsilon$$

Setting the quadratically vanishing term on the right to zero, so $\epsilon^T M + M\epsilon$ is the derivative at I. Now, taking M to be the identity (so that $O(M) = O(n)$), the derivative is simply $\epsilon \mapsto \epsilon^T + \epsilon$; we want to show that this surjects to S, i.e, if it is true that every symmetric matrix can be written in this form. Let P be a symmetric matrix, and note that $P = \frac{1}{2}P^T + \frac{1}{2}P$, from which surjectivity follows.

Thus we can conclude that the orbit map $\mathrm{GL}_n \mathbb{R} \to S$ of the standard inner product $I \in S$ has surjective derivative $M_n\mathbb{R} \to S$ at the identity in $\mathrm{GL}_n \mathbb{R}$, so $O(n)$ is a manifold near $I \in \mathrm{GL}_n \mathbb{R}$. Smoothness everywhere in $O(n)$ follows by the homogeneity trick; if $g \in O(n)$, then multiplication by g^{-1} identifies a neighborhood of $g \in \mathrm{GL}_n \mathbb{R}$ with a neighborhood of I in $\mathrm{GL}_n \mathbb{R}$ and preserves $O(n)$

This is a standard proof pattern for Lie groups; use the fact that the group multiplication operation is smooth to translate open sets around.

Immersions

Recall that $f : M \to N$ has derivative injective at $x \in M$, then f is called an immersion at x. Any embedding (e.g $\mathbb{R}^n \hookrightarrow \mathbb{R}^{m+n}$) is an immersion.

Theorem 1.3.1

Every immersion is locally equivalent to the immersion $\mathbb{R}^m \hookrightarrow \mathbb{R}^{n \geq m}$.

By abuse of notation, we shrink M and N to small coordinate charts without changing their names.

PROOF : Let $f : M \to N$ be an immersion. Choosing charts, may suppose M is a neighborhood of 0 in $\mathbb{R}^m$, $x = 0$, N is a neighborhood of 0 in $\mathbb{R}^n$, $f(0) = 0$. We assume that $D_0 f : \mathbb{R}^m \to \mathbb{R}^n$ is injective; by following f by an element of $\mathrm{GL}_n \mathbb{R}$, may suppose that $D_0 f$ is in fact just the inclusion $\mathbb{R}^m \hookrightarrow \mathbb{R}^n$. Now we use the inverse function theorem: consider $M \times \mathbb{R}^{n-m} \xrightarrow{F} \mathbb{R}^n$ with

$$F(p, x_{m+1}, \cdots, x_n) = f(p) + (0, \cdots, 0, x_{m+1}, \cdots, x_n)$$

Now $D_0 F = D_0 f + \mathrm{id}_{\mathbb{R}^{n-m}}$, which is surjective. Thus F must be a diffeomorphism in a small neighborhood of 0, and identifying a neighborhood of $f(x) \in N$ with $\mathbb{R}^n$ via F, you get that $f : M \to N$ is given in coordinates by

$$f(p) = (p, 0, \cdots, 0) \in \mathbb{R}^m \times \mathbb{R}^{n-m}$$

■

We can say a lot about submersions and immersions, but for $f : M \to N$ neither a submersion nor immersion, the local structure can be quite a bit more complicated.

As an aside, he says that singularity theory is the branch of math devoted to finding "normal forms" of degenerate functions, but that we will usually try to stick with nicer functions. He gave a few examples I couldn't really describe since they were mostly drawings. One of them showed a map from $\mathbb{R}^2$ to a surface in $\mathbb{R}^3$ where there's a fold in the sheet, and he draws a curve along that fold that when projected down to $\mathbb{R}^2$ is cuspidal, e.g, something rhyming with a blowup. This is called the elementary catastrophe, as in catastrophe theory (which, fun fact, Dalí was a fan of).

Example 1.3.2

Let $f(x_1, \cdots, x_n) = \sum_i x_i^2$ which "looks" like a paraboloid. The derivative at 0 is the zero map; however, for $g(x_1, \cdots, x_n) = x_1^2 + \cdots + x_k^2 - x_{k+1}^2 - \cdots - x_n^2$, the derivative at the origin is again the zero map, even though the two functions locally look nothing alike (f is a local minimum at 0, g is some kind of complicated saddle).

Tangent Spaces

Recall that a manifold M is covered by charts $(U_\alpha, \varphi_\alpha)$ with codomain V_α an open set in a vector space, with C^∞ (hence locally diffeomorphic, since $\varphi_\alpha \circ \varphi_\beta^{-1}$ and $\varphi_\beta \circ \varphi_\alpha^{-1}$ are inverse diffeomorphisms) transition maps. So we can transfer any concept invariant under diffeomorphisms from vector spaces to manifolds. As we've already seen, one such concept is the notion of $f : M \to \mathbb{R}$ being smooth, which is bootstrapped from charts, since if f is smooth at a point in one chart, it is smooth in all charts containing that point since the transition maps are diffeomorphisms. The same process for defining smoothness of functions $f : M \to N$ also makes sense, by choosing charts around $m \in M$ and $f(m) \in N$ and inspecting the only composition that gives a function between (an open subset of) $\mathbb{R}^m$ and $\mathbb{R}^n$.

Lots of nice pictures here that I unfortunately cannot draw. They're basically the natural pictures one would draw, with commutative diagrams of disks as open sets.

We will similarly define the tangent space by bootstrapping. The tangent space at $x \in M$ is a vector space T_xM attached to the point x. If $f : M \to N$, then the derivative $df : T_xM \to T_{f(y)}N$ has a natural pushforward action on tangent spaces. We have seen already that if df_x is surjective (i.e, if f is a submersion at x), then there are coordinates for which f (near x) is the "canonical" projection, and similarly when df_x is injective (f immersive at x) there are coordinates where f is the "canonical" inclusion of a linear space.

We already know what the tangent space T_xM is for M an open subset of a vector space, namely V itself, so, as above, given a chart $(U_\alpha, \varphi_\alpha)$ on M with $x \in U_\alpha$, we can just define $T_xM := T_{\varphi_\alpha(x)}V_\alpha$. Now we just need to check that this definition doesn't depend on the choice of chart.

Here Daniel defines a tangent vector as the collection of all possible representatives in chart tangent spaces, and runs the same argument to show they are consistent. This seems like a largely philosophical difference to me, and I think I prefer to think in terms of "Pick a chart; your choice didn't matter." Maybe I'm wrong and there's really some essential difference I'm missing here.

Note that for $x \in U_\alpha \cap U_\beta$, the diffeomorphism $\varphi_\beta \circ \varphi_\alpha^{-1} : V_\alpha \to V_\beta$ gives us a canonical isomorphism $d(\varphi_\beta \circ \varphi_\alpha^{-1})_{\varphi(x)}$ of tangent spaces, whereby we may use the compatibility conditions on transition maps to pass between different choices of representatives for a given tangent vector in different chart tangent spaces.

Now suppose $f : M \to M'$. We want $df_x : T_xM \to T_{f(x)}M'$ to be defined similarly, so letting $x \in U_\alpha \cap U_\beta \subseteq M$, $f(x) \in U'_\gamma \cap U'_\delta \subseteq M'$, we define df_x as the *family* of maps $d(\varphi'_\gamma \circ f \circ \varphi_\alpha^{-1})_{\varphi_\alpha(x)} : V_\alpha \to V'_\gamma$ for all charts U_α around x, all charts U'_γ around $f(x)$. We already know that this map is defined, so we just have to check compatibility:

$$\varphi'_\delta \circ f \circ \varphi_\beta^{-1} = (\varphi'_\delta \circ \varphi_\gamma'^{-1}) \circ \varphi'_\gamma \circ f \circ \varphi_\alpha^{-1} \circ (\varphi_\alpha \circ \varphi_\beta^{-1})$$

So, this construction identifies a tuple $(v_\alpha \in V_\alpha) \in T_xM$ satisfying compatibility to a tuple

$$d(\varphi_\beta \circ f \circ \varphi_\alpha^{-1})_{\varphi_\alpha(x)}(v_\alpha) \in T_{f(x)}M'$$

Then the submersion and immersion theorems follow formally from the vector space versions, which are much easier.

Missed this lecture because I like sleep more than manifolds apparently. Notes taken from Isaac.

Aside on Global Embeddings

Theorem 1.4.1

Any compact manifold M embeds in some $\mathbb{R}^n$, i.e, there exists an injective immersion $M \to \mathbb{R}^n$ that is homeomorphic onto its image.

PROOF : The proof will use bump functions. For all $x \in M$, open neighborhoods $K \subseteq \overline{K} \subseteq U$ of x, there exists a C^∞ function $f : M \to \mathbb{R}$ that is identically 1 on $\overline{K}$ and 0 outside U.

We didn't really prove the fact about bump functions, but we did a homework problem about a function related to $e^{-\frac{1}{x^2}}$ that is a bump function on concentric open balls. There's probably some argument to organize (and renormalize) these on the arbitrary union of open balls. Maybe compactness comes into play.

For all $x \in M$, there exists a chart $(U_\alpha, \varphi_\alpha)$ around x. Choose a smaller closed ball (by the definition of a neighborhood) $\overline{K}_x$ within U_α and let $f_x : M \to \mathbb{R}$ be a bump function that is 1 on $\overline{K}_x$ and 0 outside U_α.

Let $\varphi_\alpha : U_\alpha \to \mathbb{R}^m$, and take the associated map $U_\alpha \to \mathbb{R}^{m+1}$ given by $y \mapsto (f_x(y) \cdot \varphi_\alpha(y), f_x(y))$ which agrees (after restricting to the first m coordinates) with φ_α on $\overline{K}_x$. Extending this construction to all of M by making the function vanish on the complement of U_α clearly gives a C^∞ function F_x.

Urysohn's lemma gives an at least continuous bump function from the fact that manifolds are normal (in fact, stronger adjectives hold). Wonder if it's possible to smooth these.

Having done this around every point of M, pass to a finite subcover (via compactness) centered around $x_1, \cdots, x_k$ and the corresponding maps can be concatenated together to a C^∞ map $F : M \to \mathbb{R}^{k(m+1)}$ which is an embedding.

The derivative is injective everywhere since F_{x_i} is just a coordinate chart around x_i shifted to a hyperplane in $\mathbb{R}^{m+1}$. F is also injective since if $F(y) = F(z)$, then, specifically, $F_{x_i}(y) = F_{x_i}(z)$ for some i, so x and y are in the same U_α for one of the x_i, and $f_{x_i}(y) = f_{x_i}(z) = 1$, so

$$\varphi_\alpha(y) f_{x_i}(y) = \varphi_\alpha(z) f_{x_i}(z) \implies \varphi_\alpha(y) = \varphi_\alpha(z) \implies y = z$$

where the last implication is from the fact that φ_α is a chart. F is homeomorphic onto its image since a continuous bijection from a compact space to a Hausdorff space is a homeomorphism. ■

In the post proof remarks, it's not clear to me whether these claims refer to strengthenings of this proof or just general statements of this form. The former seems almost certainly false since it's impossible to get bounds on the size of a finite subcover in general.

Note that the embedding obtained by running this proof is almost certainly not optimal, dimensionally. M being compact was not necessary, second countability would have sufficed, and one can show that $n = 2m + 1$ is possible. Whitney showed that you can take $n = 2m$, and that this is optimal since $\mathbb{RP}^2$ doesn't embed in $\mathbb{R}^3$.

Tangent Vectors, Redux

If a particle is moving on a manifold, the tangent vectors at a point are the possible directions for the particle to move. The actual trajectory of the particle is a smooth curve $\gamma : R \to M$, which should always have well-defined tangent vectors. Based on this idea, one can (roughly) define tangent vectors and tangent spaces by looking at the tangent vectors of all possible curves on the manifold in question.

Definition 1.4.2

Let $\gamma, \delta : (-\epsilon, \epsilon) \to M$ satisfy $\gamma(0) = \delta(0) = x$. If

$$\frac{d}{dt}\bigg|_{t=0} (f \circ \gamma(t)) = \frac{d}{dt}\bigg|_{t=0} (f \circ \delta(t))$$

for all smooth f from a neighborhood of x to $\mathbb{R}$, we say that γ and δ are tangent vectors to each other.

We need to check that these vectors form a vector space. Given $v \in T_x M$, represented by $\gamma : (-\epsilon, \epsilon) \to M$, λv is represented by $\gamma(\lambda \cdot -)$, e.g, the same image curve, but going λ times as "fast" via the parameterization.

If we have γ, δ, two curves at x with tangent classes γ', δ', then $\gamma' + \delta'$ is defined as follows: pick a chart (U, φ) around x and (by abuse of notation) regard γ and δ as functions from $\mathbb{R}$ to $\mathbb{R}^n$, and define $\gamma' + \delta'$ to be the tangency class of the curve $t \mapsto \gamma(t) + \delta(t)$. There is some chart independence of this definition to check, which is left as an exercise.

Linear Approximations as Covectors

Definition 1.5.1: Cotangent Spaces

The dual of the tangent space $T_x M$ is called the *cotangent space* $T_x^* M$, and its elements are called *covectors*

It is dangerous and incorrect to think of covectors as essentially the same as vectors. For example, the gradient from multivariable calculus is not a vector, it is a covector. In particular, we are allowed to think of gradients as vectors due to the metric (Riemannian) structure on $\mathbb{R}^n$ that allows us to convert vectors to covectors and vice versa. The intrinsic definition of a gradient, in contrast, makes sense for any manifold M (as we will see), not necessarily possessing a natural metric or Riemannian structure.

I omit some discussion here of visualizing covectors on vector spaces via level sets since it was largely a series of pictures I can't convey well here.

The *differential* of a smooth function on (say) $\mathbb{R}^n$ is a linear function $df_x : T_x\mathbb{R}^n \to \mathbb{R}$ (i.e a covector) to be defined, and the gradient vector in $\mathbb{R}^n$

is given by the dot product isomorphism $T_x\mathbb{R}^n \cong T_x^*\mathbb{R}^n$. If we choose different charts this identification will still exist, but the transition maps will typically not be a Euclidean isometry, so ∇f will not be preserved, e.g, the two versions of ∇f are not equal.

Definition 1.5.2

Suppose f is a C^∞ $\mathbb{R}$-valued function defined on a neighborhood of $x \in M$. Then df_x is the linear function $T_xM \to \mathbb{R}$ defined as follows: for all curves γ through x, the derivative of $f \circ \gamma$ is the usual derivative of the function $(-\epsilon, \epsilon) \to \mathbb{R}$, so

$$df_x : \{\text{curves } \gamma \text{ through } x\} \to \mathbb{R}$$

given by $\gamma \mapsto \frac{d}{dt}\Big|_{t=0} f \circ \gamma$. One then checks that curves in the same tangency class get the same number, so that df_x is well-defined as a map from T_xM to $\mathbb{R}$. One also checks that this is a linear function and that $df_x = D_xf : T_xM \to T_{f(x)}\mathbb{R} = \mathbb{R}$ in the notation we sometimes used above.

An even more intrinsic way to get T_xM and T_x^*M is as follows: for $x \in M$, consider the ring of all $\mathbb{R}$-valued C^∞ functions defined on some neighborhood of x. This has an ideal $\mathfrak{m}_x$ of functions vanishing at x.

Lemma 1.5.3: (Hadamard)

$\mathfrak{m}_x$ is generated as an ideal by the coordinate functions $x_1, \cdots, x_n$ in some chart.

Proof : We can show the more general statement that for any function f defined on a neighborhood of $x \in \mathbb{R}^n$, there exist C^∞ functions $g_1, \cdots, g_n$ nonvanishing at x such that $f(y) = f(x) + \sum_{i=1}^n y_i g_i(y)$ for all y close enough to x.

If this holds, then the functions vanishing at x locally have the form $\sum_{i=1}^n y_i g_i(y)$ which lies in the ideal generated by the y_i.

Let's take $x = 0$, $f(x) = 0$, then for all y close to 0,

$$f(y) = \int_0^1 \frac{d(t \mapsto f(ty))}{dt} dt = \int_0^1 \sum_{i=1}^n \frac{\partial f}{\partial y_i}\Big|_{ty} \frac{dy_i t}{dt}\Big|_t dt$$

where the first equality is just the fundamental theorem of calculus and the second is the chain rule. $\frac{dy_i t}{dt}\Big|_t = y_i$, so the above becomes

$$\sum_{i=1}^n y_i \int_0^1 \frac{\partial f}{\partial y_i}(ty_1, \cdots, ty_n) dt$$

and these are the C^∞ functions we want. ■

We can then define $T_x^* M$ as $\mathfrak{m}_x/\mathfrak{m}_x^2$. Intuitively, whatever the definition of a derivative is, it should roughly obey some relation like

$$f(x + \Delta x) = f(x) + df_x(\Delta x) + O((\Delta x)^2)$$

so df_x is a linear function that vanishes at x, and we only care about the linear degree of vanishing. This says essentially that df_x is an element of $\mathfrak{m}_x/\mathfrak{m}_x^2$ represented by $f - f(x)$. $\mathfrak{m}_x/\mathfrak{m}_x^2$ is automatically a vector space, consisting of linear approximations to smooth functions vanishing at x.

Defining $T_x^* M$ this way, we can also recover $T_x M$ by dualizing. The advantage of this definition is the total avoidance of charts and coordinates, along with the fact that $\mathfrak{m}_x/\mathfrak{m}_x^2$ generalizes more nicely to (for example) schemes.

Note that the integrals in the above proof can in fact be ill-defined if (say) the chosen open neighborhood is non-convex (so the path of the integral leaves the neighborhood), so what we wrote last time really only shows that there exists a smaller neighborhood on which the formulas hold.

Why is shrinking the neighborhood a problem? Is it just an aesthetic problem?

The notion that will help us clarify this proof is the notion of germs:

Definition 1.5.4: Germs

If M is a manifold, $x \in M$, a *germ* of the C^∞ functions at x is an equivalence class of functions f defined on neighborhoods of x, where 2 such functions are equivalent if they agree identically on some neighborhood where they are defined.

It is clear that this is an equivalence relation, where transitivity follows by shrinking neighborhoods.

With this definition, we can define $\mathcal{O}_{M,x} = \varinjlim_{U \ni x} C^\infty(U)$ (the direct limit of C^∞ functions on neighborhoods of x) as the set of germs of smooth functions at x, which has a natural ring structure (as one can check), and $\mathfrak{m}_x$ the maximal ideal of $\mathcal{O}_{M,x}$ of germs vanishing at x.

Thus, we can restart Hadamard's lemma as follows: if $\in \mathcal{O}_{M,x}$ then there exist $g_i \in \mathcal{O}_{M,x}$ such that

$$f = f(x) + \sum_{i=1}^{n} x_i g_i$$

and the proof works as follows: suppose $[f] \in \mathcal{O}_{M,x}$, so there exists a neighborhood of U of x and some function f on U representing $[f]$. By shrinking U, we can suppose that it is a ball, without changing the germ $[f]$, and then we integrate along radial segments as before.

We in fact did not even avoid shrinking the neighborhood by using germs, but Daniel says that the notion of germs makes the "shrinking neighborhoods" bit more natural/baked in. This seems like a bit of the whole "making choices" vs "working with the moduli space of all possible choices" philosophy that keeps cropping up.

We now have three definitions of $T_x M$:

1. One vector space for each chart $(U_\alpha, \varphi_\alpha : U_\alpha \to V_\alpha$ containing x, all identified with each other via derivatives of transition maps.
2. Tangency classes of curves through x.
3. The dual vector space to $\mathfrak{m}_x/\mathfrak{m}_x^2$.

One would hope that these are all the same: that the first and second definitions are equivalent amounts to the statement that two curves through the same point in $\mathbb{R}^n$ are in the same tangency class iff their derivatives agree at 0. If some derivative $\dot{\gamma}_i \neq \dot{\delta}_i$, then the coordinate function x_i gives a different rate of change:

$$\frac{d(x_i \circ \gamma)}{dt} \neq \frac{d(x_i \circ \delta)}{dt}$$

so these will not be in the same tangency class. For the other direction, after subtracting off a linear function with row vector matrix $(\dot{\gamma}_1, \cdots, \dot{\gamma}_n) = (\dot{\delta}_1, \cdots, \dot{\delta}_n)$, it is enough to prove that $f \circ \gamma$ has 0 derivative for all f iff $x_i \circ \gamma$ has 0 derivative for all i.

Suppose $x_i \circ \gamma$ has 0 derivative for all i, and let f be given,

$$f(x) = f(0) + \sum_i x_i g_i(x)$$

for x near 0. Then

$$(f \circ \gamma)(t) = f(0) + \sum_i x_i(\gamma(t)) g_i(\gamma_1(t), \cdots, \gamma_n(t))$$

and the product rule gives the derivative of this with respect to t is equal to the sum of terms, all of which contain x_i or $\frac{dx_i}{dt}$ which are then equal to 0.

To see that the second and third formulations are equivalent, note that in $\mathbb{R}^n$, we have a coordinate system $(x_1, \cdots, x_n)$, and their images in $\mathfrak{m}_x/\mathfrak{m}_x^2$ are a basis for $T_x^*\mathbb{R}^n$ for $x = 0$, the origin. They span $\mathfrak{m}_x/\mathfrak{m}_x^2$ since any $f \in \mathfrak{m}_x$ can be written as

$$f = f(0) + \sum_i x_i g_i(x) = \sum_i x_i g_i(0) + \sum_i x_i(\text{functions vanishing at } 0)$$

by Hadamard's lemma, and the second sum on the right lies in $\mathfrak{m}_x^2$ and therefore vanishes in the quotient. The matrix of partial derivatives of the coordinates x_i is just the identity matrix, so the map $\mathfrak{m}_x/\mathfrak{m}_x^2 \to \mathbb{R}^n$ is onto and therefore the x_i form a basis as claimed.

I was confused here why we need to split up the $g_i(x)$ into constant terms and higher order terms since $\mathfrak{m}_x$ is an ideal over the stalk $\mathcal{O}_{\mathbb{R}^n,x}$ so the coefficients of the x_i are allowed to be arbitrary germs at x (not just constants), which to me seemed to imply that the x_i generate $\mathfrak{m}_x$ itself as an ideal (rather than its quotient by the square of itself), in concert with Hadamard's lemma. I guess we expand things out this way to get the second order terms explicitly and kill them off.

Note that we often refer to an element in some ring and its image in some quotient of that ring by the same symbol f, but here, the notation is df, which the element of $\mathfrak{m}_x/\mathfrak{m}_x^2$ represented by $f - f(x)$.

Was spiritually unwilling to attend a Zoom class, and the recording hasn't been posted yet, so this section of notes is adapted from Vincent Hoffmann's notes.

Thus, by the above argument, we have that, if $x_1, \cdots, x_n$ are coordinates around a point p of a manifold M, then $dx_1, \cdots, dx_n$ form a basis for

$\mathfrak{m}_p/\mathfrak{m}_p^2$. So for any smooth function f defined near p, df_p can be written as a linear combination of the dx_i,

$$df_p = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i$$

Note the subtlety that df_p and the dx_i all have independent meaning, but the $\frac{\partial f}{\partial x_i}$ do not, in the sense that you can only calculate the partials with respect to a full coordinate system, whereas picking a single coordinate function x_i and looking at the corresponding form dx_i is reasonable and allowed. The partials only have meaning when they are taken together.

The above relation holds in any coordinate system. In particular, if φ_α is the chart corresponding to the coordinates x_i, and φ_β is some other chart around p corresponding to coordinates y_i, then $\varphi_\beta \circ \varphi_\alpha^{-1}$ expresses the y_i in terms of the x_i (NB: not necessarily a linear combination), which gives rise to

$$dy_i = \sum_{j=1}^{n} \frac{\partial y_i}{\partial x_j} dx_j$$

Example 1.5.5

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given in polar coordinates, $f(r, \theta) = r^2$. Then

$$df = \frac{\partial f}{\partial r} dr + \frac{\partial f}{\partial \theta} d\theta = 2r dr$$

In rectangular coordinates, $f(x, y) = x^2 + y^2$, so we also have that

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy = 2x dx + 2y dy$$

One can show that $r dr = x dx + y dy$ using the chain rule as above and the equality $r = \sqrt{x^2 + y^2}$, so these two expressions coincide as one would expect.

Example 1.5.6

If $\gamma(t) = (5, t)$ in rectangular coordinates, we normally write $\dot{\gamma}(t) = (0, 1)$. In fact, it is more natural to write $\dot{\gamma} = 0\frac{\partial}{\partial x} + 1\frac{\partial}{\partial y}$, e.g, as a partial differential operator, so that evaluating a form (say, df) on it is more straightforward in coordinates. For example, $\frac{d}{dt} f \circ \gamma$ is equal to the pairing of $df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy$ with $\dot{\gamma} = \frac{\partial}{\partial y}$, which evaluates to $\frac{\partial f}{\partial y}$ since dx and $\frac{\partial}{\partial x}$ are dual by construction, and similarly for y.

Some discussion right at the end about what a vector field is, I assume it'll be covered in detail next time, so I've omitted it.