Introduction to Differential Topology

Matthew G. Brin

Department of Mathematical Sciences
State University of New York at Binghamton
Binghamton, NY 13902-6000

Spring, 1994

Contents

0. Introduction.

This is a quick set of notes on basic differential topology. It gets sketchier as it goes on. The last few sections are only to introduce the terminology and some of the concepts. These notes were written faster than I can read and may make no sense in spots. Were I to do them again, the first few topics would be rearranged into a different order. I am told that there are many misprints.

The notes were designed to give a quick and dirty, half semester introduction to differential topology to students that had finished going through almost all of *Topology: A first course* by James R. Munkres. There are references to this book as "Munkres" in these notes. The notes were written so that all of the material could be presented by the students in class. This explains various exhortations to "presenters" that occur periodically throughout the notes.

I cribbed from three main sources:

(1) Serge Lang, *Differential manifolds*, Addison Wesley, 1972,
(2) Morris W. Hirsch, *Differential topology*, Springer-Verlag, 1976, and
(3) Michael Spivak, *Calculus on manifolds*, Benjamin, 1965.

The last is a particularly pretty book that unfortunately seems to be out of print. I also stole from a few pages in

(4) James R. Munkres, *Elementary differential topology*, Princeton, 1966

whose title does not mean what it seems to mean. I do not identify the sources for the various pieces that show up in the notes. Other sources that might be interesting are

(5) Th. Bröcker & K. Jänich, *Introduction to differential topology*, Cambridge, 1982,
(6) John W. Milnor, *Topology from the differentiable viewpoint*, Virginia, 1965, and
(7) Andrew Wallace, *Differential topology: first steps*, Benjamin, 1968.

Milnor's book covers an amazing amount of ground in remarkably few pages. Wallace's takes an independent path and sets some of the machinery needed for discussion of surgery on manifolds.

1. Basics.

Let U be an open subset of $\mathbf{R}^m$. Let $f : U \to \mathbf{R}^n$ be a map. Note that for each $x \in U$ we have that $f(x)$ is an element of $\mathbf{R}^n$ so that $f(x)$ is an n-tuple or $f(x) = (f_1(x), \dots, f_n(x))$. The functions $f_i(x)$ are the coordinate functions of f. Note that each $x \in U$ is an m-tuple and can be written $x = (x_1, \dots, x_m)$.

We can now write down the partial derivatives of f if they exist. They are the derivatives

$$\frac{\partial f_i}{\partial x_j}.$$

We say that f is *differentiable* of class C^1 (short for continuous first derivatives) or just that f is C^1 if all of the first partial derivatives exist and are continuous at all points of U. We say that f is *smooth* or differentiable of class C^∞ or just C^∞ if all partial derivatives of all orders exist and are continuous at all points of U. (We define C^r by requiring that partial derivatives up to order r exist and be continuous. We can even define class C^0 by just requiring that the function f be continuous and make no mention of derivatives.) Later, we will replace the definition of C^1 by another one that is not tied to the calculation of partial derivatives.

We can now try to apply these definitions to spaces that are modeled on Euclidean spaces — namely manifolds.

Recall the definition of an n-manifold. We say that M is an n-manifold if M is a separable, metric space so that every point $x \in M$ has a neighborhood U in M with a homeomorphism $\theta_U : U \to \mathbf{R}^n$. Note that the homeomorphism θ_U gives each point $y \in U$ a set of coordinate values (by reading off the coordinates of $\theta_U(y)$ in $\mathbf{R}^n$). Thus the functions θ_U are called *coordinate functions*. The open set U is called a coordinate patch. Note that the coordinate patches form an open cover of M. (We will sometimes refer to the pair (U, θ_U) as a *coordinate chart*.) An alternative wording for the definition of an n-manifold is that it is a separable, metric space with an open cover of sets homeomorphic to $\mathbf{R}^n$. Note that the topology of M is determined by the open cover in that a set $A \subseteq M$ is open in M if and only if $A \cap U$ is open in U (i.e., $\theta_U(A \cap U)$ is open in $\mathbf{R}^n$) for every U in the open cover. We will use this later in a certain situation to determine a topology from a cover of coordinate patches.

Coordinate functions can be used to transfer activities taking place in one or more manifolds to activities taking place in one or more Euclidean spaces. Consider the following.

Let M be an m-manifold, let $x \in M$ and let N be an n-manifold. Let $f : M \to N$ be a map taking x to $y \in N$. Let U be a coordinate patch about x and V be a coordinate patch about y. Then $f^{-1}(V)$ is open in M and intersects U in an open set. Thus there are open sets $W \subseteq \mathbf{R}^m$ and $W' \subseteq \mathbf{R}^n$ so that $\theta_V \circ f \circ \theta_U^{-1}$ is defined from W to W' after making suitable restrictions. Thus the function f between M and N has been turned into a function between open subsets of Euclidean spaces. Various phrases are attached to this process. The function $\theta_V \circ f \circ \theta_U^{-1}$ is said to be an *expression of f in local coordinates* or *f expressed in local coordinates*.

It is tempting to say that f is C^1 (or smooth or C^r) at x if $\theta_V \circ f \circ \theta_U^{-1}$ is C^1 (or smooth or C^r) and that the partial derivatives of f are just the partial derivatives of $\theta_V \circ f \circ \theta_U^{-1}$. However there are problems with this that we will go into. The problem of consistently determining when a function f is differentiable requires a certain amount of work. The problem of determining exactly what the derivative of f should be turns out to need even more work.

What are the problems? Consider the following homeomorphisms from $\mathbf{R}$ to itself. Let

$$\alpha(x) = x, \quad \text{and}$$
$$\beta(x) = \begin{cases} x, & x \leq 0 \\ 2x, & x \geq 0. \end{cases}$$

The space $\mathbf{R}$ is a 1-manifold because each $x \in \mathbf{R}$ has a neighborhood (namely $\mathbf{R}$ itself) that is homeomorphic to $\mathbf{R}$. The functions α and β are possible choices for such a homeomorphism. Now let M and N be the 1-manifolds whose underlying space is $\mathbf{R}$, where $\mathbf{R}$ is the only coordinate patch for each of M and N, and where M uses α as its coordinate function and N uses β for its coordinate function. Consider the identity map f from $\mathbf{R}$ to itself. This can be viewed as a map from M to M, from M to N, from N to M and from N to N. Now we note that the maps $\alpha \circ f \circ \alpha^{-1}$ and $\beta \circ f \circ \beta^{-1}$ are differentiable but $\alpha \circ f \circ \beta^{-1}$ and $\beta \circ f \circ \alpha^{-1}$ are not. Thus f is differentiable as a map from M to M and from N to N, but not from M to N and not from N to M.

The problem arises now if we use both α and β as choices for coordinate functions for a single 1-manifold. (Such choices are almost never avoidable since an n-manifold will usually have to be covered by overlapping open sets with homeomorphisms to $\mathbf{R}^n$. Consider a collection of open sets that demonstrates that the circle is a 1-manifold.) Multiple choices of coordinate functions mean that there are multiple ways to express a function in local coordinates. For example, if both α and β are available as coordinate functions, then the answer to the question as to whether the identity from $\mathbf{R}$ to itself is differentiable will depend on the coordinate functions used. We need a way to insure that a choice of coordinate functions does not make the question of differentiability ambiguous.

We can now give a definition of a differentiable n-manifold. The definition of an n-manifold is imitated but with a couple of changes. One is for convenience, and the other is to make the notion of differentiability unambiguous. A separable, metric space M is a differentiable n-manifold of class C^r (or just a C^r n-manifold), $0 \leq r \leq \infty$, if there is an open cover O of M so that each $U \in O$ has a homeomorphism $\theta_U : U \to U'$ where U' is an open subset of $\mathbf{R}^n$ and so that for each U and V in O with $U \cap V \neq \emptyset$,

$$\left(\theta_V|_{(U\cap V)}\right) \circ \left(\theta_U|_{(U\cap V)}\right)^{-1} : \theta_U(U \cap V) \to \theta_V(U \cap V)$$

is C^r. The function $\left(\theta_V|_{(U\cap V)}\right) \circ \left(\theta_U|_{(U\cap V)}\right)^{-1}$ is known as an overlap map. The definition requires that all overlap maps be C^r. We will add one more condition later when it becomes convenient to have it and when the reasons for it become more apparent. The new condition will not change the definition and what we have so far will do.

If we regard $\mathbf{R}$ as a 1-manifold and use α above as its only coordinate map, then $\mathbf{R}$ is a C^∞ manifold. It is also a C^∞ manifold if we use β as its only coordinate function. However, if we use both α and β as coordinate functions, then we only get a C^0 manifold.

We can now attack the idea of differentiable function between C^r manifolds. Almost as before, let M be a C^r m-manifold, let $x \in M$, let N be a C^r n-manifold, let $f : M \to N$ be a map taking x to $y \in N$, let U be a coordinate patch about x, and let V be a coordinate patch about y. We say that f is differentiable of class C^s, $s \leq r$, at x if $\theta_V \circ f \circ \theta_U^{-1}$ (with suitable restrictions) is a C^s map from an open set in $\mathbf{R}^m$ containing $\theta_U(x)$ to an open set in $\mathbf{R}^n$. We say that f is differentiable of class C^s if f is differentiable of class C^s at every $x \in M$.

We accept as a temporary black box: A composition of C^r maps between open sets in Euclidean spaces is C^r. We use this to verify: Whether the function f of the previous paragraph is discovered to be C^s at x is independent of the coordinate patches and functions used. [Presenters: Check it out.] Thus a function is C^s if every expression of f in local coordinates is C^s.

The actual derivative of a differentiable function is another matter. Consider $\mathbf{R}$ as a 1-manifold with $\theta_1(x) = x$ and $\theta_2(x) = 2x$ as the available coordinate functions. It is easily checked that the (only two) overlap maps are C^∞. Thus $\mathbf{R}$ with these coordinate functions is a C^∞ 1-manifold. Now consider the identity function f from $\mathbf{R}$ to itself. We might consider $\theta_1 \circ f \circ \theta_1^{-1}$, or $\theta_1 \circ f \circ \theta_2^{-1}$, or $\theta_2 \circ f \circ \theta_1^{-1}$, or $\theta_2 \circ f \circ \theta_2^{-1}$ to try to discuss the derivative of f at a given point. However, the four expressions above give three possbible candidates for the value of f' at any given point.

An attempt can be made to get around this in the same way that we got around ambiguities in the notion of differentiability. We could try to restrict the overlap maps even further. The requirement could be that the overlap maps introduce no stretching. This can be done but it turns out to be incredibly restrictive. Some manifolds, such as S^1 and products of S^1 with itself, can be given such structures, but infinitely many others can not. Another approach is used.

The calculation of derivative for functions from $\mathbf{R}^m$ to $\mathbf{R}^n$ make use of the fact that Euclidean spaces are vector spaces and that a "calculus of displacement" is available. Displacement is done with vectors. Vectors have the properties of length and direction which can be exploited. In a manifold, the notions of length and direction are handled by tools that can be adapted to the manifold and that don't depend on a notion of straightness. Specifically, we will use curves — differentiable functions from $\mathbf{R}$ to the manifold. If we knew what the derivative of a curve was, then we would say that the derivative at a point was giving us a direction and speed (the norm of the derivative) was giving a length. It turns out that a workable system can be invented even if the derivative of a curve is not known. All you need to know is when two curves "deserve the same derivative" and how to form equivalence classes.

As preparation, we review derivatives of curves into $\mathbf{R}^n$. Let $f : \mathbf{R} \to \mathbf{R}^n$

have coordinate functions $(f_1, \ldots, f_n)$. Then $f' = (f_1', \ldots, f_n')$ and, for a given x, $f'(x) = (f_1'(x), \ldots, f_n'(x))$ which is regarded as a vector that is tangent to the curve f at $f(x)$. For example, the straight line tangent to f at $f(x)$ can be formed as $T(t) = f(x) + t(f'(x))$. The point of tangency is at $T(0) = f(x)$.

We are now ready for some definitions. Let M be a C^r n-manifold, $r \geq 1$, let $x \in M$ and let U be a coordinate patch containing x. Let $C(x)$ be the set of all $f : V \to U$ so that $V \subseteq \mathbf{R}$ is open, $0 \in V$, f is C^1 and $f(0) = x$. (Why is $C(x)$ not empty?) We define a relation on $C(x)$ by saying that $f \sim g$ if $(\theta_U \circ f)'(0) = (\theta_U \circ g)'(0)$. [Presenters: show that this does not depend on the coordinate patch U, and show that this is an equivalence relation. This assumes a chain rule for maps between open subsets of Euclidean space. Such a chain rule is written out in the next section.]

We define T_x to be the set of equivalence classes and call it the the *tangent space* to M at x. Elements of T_x are called *tangent vectors* at x. Of course, the word "vector" is not yet justified.

We note that $\hat{\theta}_U : T_x \to \mathbf{R}^n$ defined by $[f] \mapsto (\theta_U \circ f)'(0)$ is well defined and one to one because of the way the classes of T_x are defined. We claim that it is also a surjection. Let d be a vector in $\mathbf{R}^n$. We can form the straight line $l : \mathbf{R} \to \mathbf{R}^n$ by $l(t) = \theta_U(x) + td$. There is an open set V in $\mathbf{R}$ containing 0 so that $f = \theta_U^{-1} \circ l$ is defined on V. Also, $f(0) = x$ and f is C^1 since $\theta_U \circ f = l$ is C^1. (In the last claim, we used the identity coordinate function from $\mathbf{R}$ to itself in regarding $\mathbf{R}$ as a 1-manifold.) Now $\hat{\theta}_U[f] = l'(0) = d$, so $\hat{\theta}_U$ is onto.

We now have a bijection $\hat{\theta}_U$ between T_x and the vector space $\mathbf{R}^n$. We can use this to define a vector space structure on T_x by saying that $[f] + [g] = \hat{\theta}_U^{-1}(\hat{\theta}_U[f] + \hat{\theta}_U[g])$ and $r[f] = \hat{\theta}_U^{-1}(r\hat{\theta}_U[f])$. Not only does this give us a vector space structure on T_x but it makes $\hat{\theta}_U$ an isomorphism. We will make use of this isomorphism later, so it is worth summarizing in a lemma.

Lemma 1.1. *Let $\theta_U : U \to \mathbf{R}^n$ be a coordinate function and $x \in U$. Then $\hat{\theta}_U : T_x \to \mathbf{R}^n$ defined by $[f] \mapsto (\theta_U \circ f)'(0)$ is an isomorphism.*

Let M be a C^r m-manifold and let N be a C^s n-manifold, r and s at least 1. We are now ready to talk derivatives. Let $f : M \to N$ be a C^1 map. Let x be in M with $y = f(x)$. We will define a function from T_x to T_y. Let g be a curve representing a tangent vector at x. Then we define $Df_x([g]) = [f \circ g]$. [Presenters: this is well defined and is a linear function from the vector space T_x to the vector space T_y.]

Proposition 1.2 (The chain rule). *Let M, N and P be differentiable manifolds of class at least C^1. Let $f : M \to N$ and $h : N \to P$ be differentiable of class at least C^1. Let $x \in M$ and let $y = f(x)$. Then $D(h \circ f)_x = (Dh_y) \circ (Df_x)$.*

Proof: [Presenters: ....]

The chain rule is actually one step in a construction designed to make the derivative a functor. It is not very interesting when applied only to the tangent space at one point, but it is a start. The other half of this start is the following trivial lemma.

LEMMA 1.3. *Let M be a C^r m-manifold, $r \geq 1$, and let $i : M \to M$ be the identity map. Then for any $x \in M$, $Di_x : T_x \to T_x$ is the identity.*

COROLLARY 1.3.1. *Let M and N be C^r m-manifolds, $r \geq 1$, and let h be a C^1 homeomorphism between them whose inverse is C^1. Then for any $x \in M$, $Dh_x : T_x \to T_{h(x)}$ is an isomorphism.*

The approach taken here is not the only approach to tangent vectors and tangent spaces. There are at least three approaches (and possibly more) that appear quite different, but which give structures with identical behavior.

The next topic will fill in the black box mentioned above: compositions of C^r maps between open sets in Euclidean spaces are C^r maps. Even further, we will derive a chain rule for maps between Euclidean spaces. This will then be used to put a structure on the collection of all T_x, $x \in M$.

2. Derivative and Chain rule in Euclidean spaces.

If $f : \mathbf{R} \to \mathbf{R}$ is a function, then its derivative at x is defined by

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}.$$

If we try to generalize to functions $f : \mathbf{R}^m \to \mathbf{R}^n$, then we run into the problem of dividing by a vector.

If we return to the case of $f : \mathbf{R} \to \mathbf{R}$, then the definition of derivative can be reinterpreted to say that f is differentiable at x and that its derivative at x has the value $f'(x)$ if

$$\lim_{h\to 0} \frac{f(x+h) - f(x) - f'(x)h}{h} = 0.$$

The function $h \mapsto f'(x)h$ is a linear function from $\mathbf{R}$ to $\mathbf{R}$. If we call this linear function λ, then we have that f is differentiable at x if there is a linear function $\lambda : \mathbf{R} \to \mathbf{R}$ so that

$$\lim_{h\to 0} \frac{f(x+h) - f(x) - \lambda(h)}{h} = 0.$$

The number $f'(x)$ is just the slope of the linear function λ. Instead of defining the derivative of f at x to be the slope of the linear function λ we can define the derivative of f at x to be the linear function λ itself. This gives a setting that can be imitated in higher dimensions. Note that since the definition involves a limit at a specific point, we only need to have f defined on an open set containing the point. This will be reflected in the setting of the defintion.

Let $f: U \to \mathbf{R}^n$ be a function where U is an open subset of $\mathbf{R}^m$. We say that f is differentiable at $x \in U$ if there is a linear function $\lambda: \mathbf{R}^m \to \mathbf{R}^n$ so that

$$\lim_{h \to 0} \frac{\|f(x+h) - f(x) - \lambda(h)\|}{\|h\|} = 0.$$

We could also say

$$\lim_{h \to 0} \frac{f(x+h) - f(x) - \lambda(h)}{\|h\|} = 0$$

since a vector goes to zero if and only if its length goes to zero. We say that the derivative of f at x is λ and denote it Df_x. The quotients make sense since the denominators are real numbers. Note that the "domain" of the limit is $U - x = \{u - x | u \in U\}$ which is the translation of the open set U that carries x to 0 and is thus an open set in $\mathbf{R}^m$ containing 0. In (ϵ, δ) form, the limit statement reads: for any $\epsilon > 0$, there is a $\delta > 0$ so that for any $h \neq 0$ in the δ-ball about 0 in $\mathbf{R}^m$, we have that

$$\frac{\|f(x+h) - f(x) - \lambda(h)\|}{\|h\|} < \epsilon.$$

Or, in other words,

$$\|f(x+h) - f(x) - \lambda(h)\| < \epsilon \|h\|.$$

Proposition 2.1. *Let $f: U \to \mathbf{R}^n$ be differentiable at x where U is an open set in $\mathbf{R}^m$. Then Df_x is unique.*

Proof: Suppose that linear $\lambda_i: \mathbf{R}^m \to \mathbf{R}^n$, $i = 1, 2$ both satisfy

$$\lim_{h \to 0} \frac{\|f(x+h) - f(x) - \lambda_i(h)\|}{\|h\|} = 0.$$

Thus for $\epsilon > 0$ and restriction of h to a suitable δ-ball we can make

$$\|f(x+h) - f(x) - \lambda_i(h)\| < \frac{\epsilon}{2} \|h\|.$$

Now,

$$\begin{aligned}
\|\lambda_1(h) - \lambda_2(h)\| &= \|\lambda_1(h) - f(x+h) + f(x) + f(x+h) - f(x) - \lambda_2(h)\| \\
&\leq \|\lambda_1(h) - f(x+h) + f(x)\| + \|f(x+h) - f(x) - \lambda_2(h)\| \\
&< \epsilon \|h\|.
\end{aligned}$$

This gives the not surprising statement that the λ_i do not differ by much on small vectors. But the λ_i are linear and we can use this and the inequality above to show that they do not differ by much on any vector. Let $v \in \mathbf{R}^m$ be arbitrary and let $t > 0$ be small enough so that tv is in the δ-ball. Then

$$\begin{aligned} t\epsilon\|v\| &= \epsilon\|tv\| \\ &> \|\lambda_1(tv) - \lambda_2(tv)\| \\ &= \|t\lambda_1(v) - t\lambda_2(v)\| \\ &= t\|\lambda_1(v) - \lambda_2(v)\|. \end{aligned}$$

So

$$\|\lambda_1(v) - \lambda_2(v)\| < \epsilon\|v\|.$$

But this can be done for this v and any $\epsilon > 0$. So $\|\lambda_1(v) - \lambda_2(v)\| = 0$ and $\lambda_1 = \lambda_2$.

The next result, the chain rule, fills in the "black box" from the previous section. In its proof, we will need the continuity of certain linear functions. This is straightforward but not trivial in the finite dimensional setting that we are in if we use the usual topology on the Euclidean spaces. It is false in infinite dimensions for most topologies that are put on the vector spaces.

We will need the notion of the norm of a linear map. Let $\lambda : \mathbf{R}^m \to \mathbf{R}^n$ be a linear map. Let B be the closed unit ball in $\mathbf{R}^m$ and let $\|\lambda\|$ be the maximum distance from 0 to a point in $f(B)$. This exists and is finite since B is compact. It may be zero if f is the zero linear map. Let $v \in \mathbf{R}^m$. We have the following inequality:

$$\|\lambda(v)\| \;=\; \|v\| \cdot \|\lambda\left(\frac{v}{\|v\|}\right)\| \;\leq\; \|v\| \cdot \|\lambda\|.$$

The finiteness of $\|\lambda\|$ depends on the continuity of λ. As mentioned above, linear maps with finite dimensional domains are continuous. In an infinite dimensional setting, the finiteness of $\|\lambda\|$ is equivalent to the continuity of λ.

Theorem 2.2 (Chain Rule on Euclidean spaces). *If $U \subseteq \mathbf{R}^m$ and $V \subseteq \mathbf{R}^n$ are open sets and $f : U \to \mathbf{R}^n$ and $g : V \to \mathbf{R}^p$ are differentiable at $a \in U$ and $b = f(a) \in V$ respectively, then $g \circ f : U \to \mathbf{R}^p$ is differentiable at a and*

$$D(g \circ f)_a = Dg_b \circ Df_a.$$

Proof: Another way to interpret the definition of the derivative of f at x is to say that if we define

$$E(h) = f(x + h) - f(x) - Df_x(h),$$

then for any $\epsilon > 0$, there is a $\delta > 0$ so that $\|h\| < \delta$ implies $\|E(h)| < \epsilon\|h\|$. Note that $E(0) = 0$ so that we do not have to say $0 < \|h\| < \delta$.

Let $\lambda = Df_a$ and $\mu = Dg_b$. We have

$$\begin{aligned}
&\|g(f(x+h)) - g(f(x)) - \mu(\lambda(h))\| \\
&\qquad \le \|g\big(f(x) + \lambda(h) + E(h)\big) - g(f(x)) - \mu(\lambda(h) + E(h))\| \\
&\qquad\qquad + \|\mu(\lambda(h) + E(h)) - \mu(\lambda(h))\| \\
&\qquad = \|g\big(f(x) + \lambda(h) + E(h)\big) - g(f(x)) - \mu(\lambda(h) + E(h))\| \\
&\qquad\qquad + \|\mu(E(h))\|
\end{aligned}$$

where the equality follows from the linearity of μ. We will be done if for a given $\epsilon > 0$ we can find a $\delta > 0$ so that $\|h\| < \delta$ makes

$$\|g\big(f(x) + \lambda(h) + E(h)\big) - g(f(x)) - \mu(\lambda(h) + E(h))\| < \frac{\epsilon}{2}\|h\| \tag{1}$$

and

$$\|\mu(E(h))\| < \frac{\epsilon}{2}\|h\|. \tag{2}$$

We have

$$\|g\big(f(x) + \lambda(h) + E(h)\big) - g(f(x)) - \mu(\lambda(h) + E(h))\| < \epsilon_1\|\lambda(h) + E(h)\|$$

if

$$\|\lambda(h) + E(h)\| < \delta_1. \tag{3}$$

Now

$$\begin{aligned}
\|\lambda(h) + E(h)\| &\le \|\lambda(h)\| + \|E(h)\| \\
&< \|\lambda\| \cdot \|h\| + \|\epsilon_2\|h\| \\
&= \big(\|\lambda\| + \epsilon_2\big)\|h\|
\end{aligned} \tag{4}$$

for

$$\|h\| < \delta_2 \tag{5}$$

so

$$\begin{aligned}
\epsilon_1\|\lambda(h) + E(h)\| &< \big(\epsilon_1\|\lambda\| + \epsilon_1\epsilon_2\big)\|h\| \\
&< \frac{\epsilon}{2}\|h\|
\end{aligned}$$

if all of

$$\epsilon_1 < \frac{\epsilon}{4}, \qquad \epsilon_1 < \frac{\epsilon}{4\|\lambda\|}, \qquad \epsilon_2 < 1 \tag{6}$$

hold. Thus we get (1) if we can satisfy all of (6). Now

$$\begin{aligned}\|\mu(E(h))\| &\le \|\mu\| \cdot \|E(h)\| \\ &< \epsilon_2\|\mu\| \cdot \|h\| \\ &< \frac{\epsilon}{2}\|h\|\end{aligned}$$

if

$$\epsilon_2 < \frac{\epsilon}{2\|\mu\|}. \tag{7}$$

Thus we get (2) if we can satisfy (7).

So given ϵ, we determine ϵ_1 and ϵ_2 from (6) and (7). This determines δ_1 and δ_2 which puts our first restriction $\delta \le \delta_2$ on δ because of (5). We must deal with (3). But we can get this from (4) by putting the resriction

$$\delta < \frac{\delta_1}{\|\lambda\| + \epsilon_2}$$

on δ. This finishes the proof.

We give two easily computed derivatives.

LEMMA 2.3. *Let $f : \mathbf{R}^m \to \mathbf{R}^n$ be a linear mapping. Then for all $x \in \mathbf{R}^m$, $Df_x = f$.*

PROOF: With f linear, $f(x+h) = f(x) + f(h)$ so

$$\lim_{h\to 0} \frac{f(x+h) - f(x) - f(h)}{\|h\|} = 0.$$

Since we need a linear function of h that gives the above limit and the linear f does the trick, f must be the derivative.

LEMMA 2.4. *If f is a constant, then all Df_x are the zero tranformation.*

PROOF: The linear map 0 works in

$$\lim_{h\to 0} \frac{f(x+h) - f(x) - 0(h)}{\|h\|} = 0.$$

We end with a lemma that we will use to relate two of the notions of derivative that we have used so far. We assume the usual notation that if $\alpha : A \to C$ and $\beta : B \to D$ are functions, then the notation $\alpha \times \beta$ refers to the function from $A \times B$ to $C \times D$ defined by $(\alpha \times \beta)(a,b) = (\alpha(a), \beta(b))$. We also invent a notation that if $\gamma : A \to B$ and $\delta : A \to C$ are given, then (γ, δ) refers to the function from A to $B \times C$ defined by $(\gamma,\delta)(a) = (\gamma(a), \delta(a))$.

Lemma 2.5. *If $U \in \mathbf{R}^m$ and $V \in \mathbf{R}^s$ are open sets and $f : U \to \mathbf{R}^n$ and $g : V \to \mathbf{R}^t$ are differentiable at $a \in U$ and $b \in V$ respectively, then $f \times g : U \times V \to \mathbf{R}^n \times \mathbf{R}^t$ is differentiable at (a,b) and the derivative there is $Df_a \times Dg_b$. If, in addition, $h : U \to \mathbf{R}^q$ is differentiable at a, then (f,h) is differentiable at a and the derivative there is (Df_a, Dh_a).*

Proof: Consider

$$\begin{aligned} &\|(f \times g)(a+h_1, b+h_2) - (f \times g)(a,b) - (Df_a \times Dg_b)(h_1,h_2)\| \\ =&\|\big(f(a+h_1), g(b+h_2)\big) - \big(f(a), g(b)\big) - \big(Df_a(h_1), Dg_b(h_2)\big)\| \\ =&\|\big(f(a+h_1) - f(a) - Df_a(h_1)\ ,\ g(b+h_2) - g(b) - Dg_b(h_2)\big)\|. \end{aligned} \tag{8}$$

The i-th coordinate, $i = 1,2$, in (8) can be kept less than $\epsilon\|h_i\|$ by confining h_i to some δ_i-ball. So if

$$\|(h_1,h_2)\| = \max\{\|h_1\|, \|h_2\|\} < \min\{\delta_1, \delta_2\},$$

then both coordinates in (8) are less than

$$\epsilon \max\{\|h_1\|, \|h_2\|\} = \epsilon\|(h_1,h_2)\|.$$

This proves the first part.

Now consider the diagonal map $d : U \to \mathbf{R}^m \times \mathbf{R}^m$ defined by $d(u) = (u,u)$. This is linear so $Dd = d$. Note that $(f,h) = (f \times h) \circ d$. Now $D(f,h) = D(f \times h) \circ Dd = (Df \times Dh) \circ d = (Df, Dh)$.

We can use this to relate the standard notion of the derivative of a curve, to the notion of a derivative as developed in this section. Recall that if f is a function from $\mathbf{R}$ to $\mathbf{R}$, then $f'(x)$ gives the slope of Df_x. Thus for f and g from $\mathbf{R}$ to $\mathbf{R}$, we have $f'(x) = g'(x)$ if and only if $Df_x = Dg_x$. Even more, we can recover $f'(x)$ from Df_x. Since $f'(x)$ is the slope of the linear map $Df_x : \mathbf{R} \to \mathbf{R}$, we must have $f'(x) = Df_x(1)$.

Now if we have $f : \mathbf{R} \to \mathbf{R}^n$, we have $f = (f_1, \ldots, f_n)$. By Lemma 2.5, we have $Df = (Df_1, \ldots, Df_n)$. If $g : \mathbf{R} \to \mathbf{R}^n$ is given, then we also have $f'(x) = g'(x)$ if and only if $Df_x = Dg_x$. And further,

$$\begin{aligned} f'(x) &= \big(f_1'(x), \ldots, f_n'(x)\big) \\ &= \big(D(f_1)_x(1), \ldots, D(f_n)_x(1)\big) \\ &= Df_x(1). \end{aligned}$$

Going back to the setting of Section 1, we can now say that two curves f and g represent the same tangent vector if $D(\theta_U \circ f)_0 = D(\theta_U \circ g)_0$.

We leave as easy exercises the fact that the derivative is a linear operator on functions. Specifically, $D(f+g)_x = Df_x + Dg_x$ and $D(rf)_x = rDf_x$.

3. Three derivatives.

We have been exposed to three kinds of derivatives. One is the usual Calculus I–III derivative and has shown up in

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}$$

for a function from $\mathbf{R}$ to $\mathbf{R}$, and in

$$(f_1, \ldots, f_n)' = (f_1', \ldots, f_n')$$

for a function from $\mathbf{R}$ to $\mathbf{R}^n$. The second kind is the "advanced calculus" derivative defined in the previous section as the best linear approximation to a function from $\mathbf{R}^m$ to $\mathbf{R}^n$. The third kind was defined in the first section as a linear function on a tangent space. We would like to combine these three notions as much as possible, expecially as we have used the same notation Df_x for the last two of them. Because of this, we will agree *for this section only* to use $\overline{D}$ for the "advanced calculus" derivative (best linear approximation).

The use of f' has only been used in these notes to define classes of curves to build tangent spaces and for the isomorphism of Lemma 1.1. In the previous section, we showed that the use of f' can be eliminated from definition of classes in tangent spaces. That still leaves the use of f' in the isomorphism of Lemma 1.1. We will try to eliminate as many references to f' as possible by filtering all such references through an application of Lemma 1.1.

We now concentrate on D and $\overline{D}$. We cannot eliminate $\overline{D}$ since it is essential in defining the notion of differentiable for functions between Euclidean spaces. However, what we can aim for is to show such a strong equivalence between D and $\overline{D}$ that distinctions between them become unimportant.

Here is the first lemma to try to blur some distinctions.

LEMMA 3.1. *Let $U \subseteq \mathbf{R}^m$ be an open set with $u \in U$. Let $f : U \to \mathbf{R}^n$ be C^1 and let $v = f(u)$. Let $i : U \to \mathbf{R}^m$ be inclusion and let $j : \mathbf{R}^n \to \mathbf{R}^n$ be the identity. In the following diagram, $\hat{\imath}$ and $\hat{\jmath}$ are the isomorphisms of Lemma 1.1.*

$$\begin{tikzcd} T_u \arrow[r, "\hat{\imath}"] \arrow[d, "Df_u"'] & \mathbf{R}^m \arrow[d, "h = \hat{\jmath} \circ Df_u \circ \hat{\imath}^{-1}"] \\ T_v \arrow[r, "\hat{\jmath}"] & \mathbf{R}^n \end{tikzcd}$$

If h is defined as shown in the diagram, then $h = \overline{D}f_u$.

Proof: We consider $(\hat\jmath \circ Df_u \circ \hat\imath^{-1})(d)$ for some d in $\mathbf{R}^m$. We start with $\hat\imath^{-1}(d)$. For $l : \mathbf{R} \to \mathbf{R}^n$ defined by $l(t) = i(u) + td = u + td$, we have $\hat\imath^{-1}(d) = [i^{-1} \circ l] = [l]$. So

$$\begin{aligned}
(\hat\jmath \circ Df_u \circ \hat\imath^{-1})(d) &= (\hat\jmath \circ Df_u)[l] \\
&= \hat\jmath([f \circ l]) \\
&= (f \circ l)'(0) \\
&= \overline{D}(f \circ l)_0(1) \\
&= \left(\overline{D}f_{l(0)} \circ \overline{D}l_0\right)(1) \\
&= \overline{D}f_u(\overline{D}l_0(1)) \\
&= \overline{D}f_u(l'(0)) \\
&= \overline{D}f_u(d).
\end{aligned}$$

This says that the two notions of derivative behave the same for functions between Euclidean spaces. Now we bring in manifolds. In the statement we simplify the notation for the coordinate function on a patch U by dropping the subscript U and write θ instead of θ_U. This is to keep the notation from exploding.

Lemma 3.2. *Let U be a coordinate patch in a C^r m-manifold M with coordinate function θ and let $u \in U$. Let $V = \theta(U)$ regarded as an m-manifold with one coordinate patch V whose coordinate function is the inclusion map $i : V \to \mathbf{R}^m$. Then the following is a commutative diagram of isomorphisms.*

$$\begin{tikzcd}
T_u \arrow[rr, "\hat\theta"] \arrow[dr, "D\theta_u"'] & & \mathbf{R}^m \\
& T_{\theta(u)} \arrow[ur, "\hat\imath"'] &
\end{tikzcd}$$

Proof: We know from Lemma 1.1 that $\hat\theta$ and $\hat\imath$ are isomorphisms. If the diagram commutes, then $D\theta_u$ will be an isomorphism. To see that the diagram commutes, let $[f]$ be in T_u. We have $\hat\theta[f] = (\theta \circ f)'(0)$. Now $D\theta_u[f] = [\theta \circ f]$ and $\hat\imath[\theta \circ f] = (i \circ \theta \circ f)'(0) = (\theta \circ f)'(0)$.

The next lemma looks at maps between manfiolds. Again we leave subscripts off the coordinate functions.

Lemma 3.3. *Let M be an m-manifold and N be an n-manifold, each of class at least 1. Let $f : M \to N$ be a C^1 map and let $u \in M$ with $v = f(u)$. Let U be a coordinate patch around u with coordinate function θ and let V be a coordinate patch around v with coordinate function ϕ. To avoid restrictions, assume that $f(U) \subseteq V$ and use this to define $h = \phi \circ f \circ \theta^{-1}$. Let i and j be the inclusions* *of $\theta(U)$ and $\phi(V)$ respectively into $\mathbf{R}^m$ and $\mathbf{R}^n$. Then the following diagram commutes and the non-vertical arrows are isomorphisms.*

$$\begin{tikzcd}
T_u \arrow[rr, "\hat{\theta}"] \arrow[dr, "D\theta_u"] \arrow[ddd, "Df_u"'] & & \mathbf{R}^m \arrow[ddd, "\overline{D}h_{\theta(u)}"] \\
& T_{\theta(u)} \arrow[ur, "\hat{\imath}"'] \arrow[d, "Dh_{\theta(u)}"] & \\
& T_{\theta(v)} \arrow[dr, "\hat{\jmath}"] & \\
T_v \arrow[ur, "D\phi_v"] \arrow[rr, "\hat{\phi}"] & & \mathbf{R}^n
\end{tikzcd}$$

PROOF: The isomorphisms and the commutativity of all but the left hand trapezoid follow from the previous two lemmas. The commutativity of the left hand trapezoid follows from the chain rule.

There are three main quadrilaterals in the diagram of Lemma 3.3 — the outer square and the two trapezoids. Each can be interpreted in words. The outer square says that when h is an expression of f in local coordinates, then the isomorphisms induced by the coordinate functions used in the expression conjugate the action of Df on the tangent spaces to the action of $\overline{D}h$ as a linear map between Euclidean spaces. The two trapezoids say almost identical things in slightly different settings.

At this point the notation $\overline{D}$ ends. Even though there are two different notions of derivative that will have the same notation, the ambiguity will not be important.

4. Higher derivatives.

We give one more section that concentrates on maps between Euclidean spaces.

I'm trying as hard as I can to avoid partial derivatives. Before partial derivatives make an appearance, we have that if $f : \mathbf{R}^m \to \mathbf{R}^n$ is differentiable at x, then the derivative Df_x at x is a linear map from $\mathbf{R}^m$ to $\mathbf{R}^n$. Further if f is differentiable on all points in $\mathbf{R}^m$, then we have a function Df from $\mathbf{R}^m$ to the set of linear transformations from $\mathbf{R}^m$ to $\mathbf{R}^n$. We can call this function the derivative of f. If we stop here, then partial derivatives have not been brought in. They are brought in if we try to make the set of linear transformations from $\mathbf{R}^m$ to $\mathbf{R}^n$ look more familiar.

In order to make the set of linear transformations from $\mathbf{R}^m$ to $\mathbf{R}^n$ look more familiar, we need to choose a prefered basis for both $\mathbf{R}^m$ and $\mathbf{R}^n$. If we choose the standard bases (unit vectors in the coordinate directions), then a linear transformation from $\mathbf{R}^m$ to $\mathbf{R}^m$ is represented by an $n \times m$ matrix. At this point the partial derivatives have appeared. This is because the particular matrix that represents Df_x using the standard bases is the matrix whose entries are

$$(Df_x)_{i,j} = \frac{\partial f_i}{\partial x_j}$$

if we regard the matrix as acting on the left and we regard elements of $\mathbf{R}^m$ and $\mathbf{R}^n$ as column vectors. We drop the partial derivatives for several paragraphs to inspect the structure that we have built so far.

We have that Df is a function from $\mathbf{R}^m$ to the set of linear transformation from $\mathbf{R}^m$ to $\mathbf{R}^n$. With our choice of bases, we have a particular one to one correspondence between the set of linear transformations from $\mathbf{R}^m$ to $\mathbf{R}^n$ and the set of $n \times m$ matrices. Thus our choice of basis allows us to look at Df as a function from $\mathbf{R}^m$ to the set of $n \times m$ matrices.

We can add extra structure to the set of $n \times m$ matrices and make a topological space and a vector space out of it. This can be done by letting basis vectors for the set of $n \times m$ matrices be those $n \times m$ matrices with a one in a single position and zeros everywhere else. This (second) choice now makes Df a function from $\mathbf{R}^m$ to $\mathbf{R}^{nm}$.

Now that Df is a function between Euclidean spaces, we can discuss two things — the continuity of Df and the differentiability of Df. If Df is continuous, then f is of class C^1. If Df is differentiable, then its derivative $D^2 f$ is a function from $\mathbf{R}^m$ to $\mathbf{R}^{nm^2}$. We see that we can now discuss higher derivatives and higher classes of differentiability. In particular, we can point out that f is of class C^r if and only if Df is of class C^{r-1}.

Note that linear functions are infinitely differentiable. In fact, if f is linear, then $Df_x = f$ for all x so that Df is a constant (even though each Df_x is not the constant linear transformation). Now all higer derivatives of f are zero.

The fact that linear functions are infinitely differentiable is relevant because choices were made in setting up Df as a function from $\mathbf{R}^m$ to $\mathbf{R}^{nm}$. The correspondence depended on two choices of bases. Different choices of bases give different correspondences that can be obtained from the original by multiplying by "change of basis" matrices at appropriate places. Multiplying by matrices is linear and thus infinitley differentiable. From this it follows that if f is C^r as measured with one choice of bases, then it is as measured with another.

We now return to the partial derivatives. Our choice of bases made Df a function from $\mathbf{R}^m$ to $\mathbf{R}^{nm}$. The coordinates in $\mathbf{R}^{nm}$ are the entries in the matrices that represent the linear transformations Df_x. These entries are just the partial derivatives of f at x. Thus the coordinate functions of Df are the partial derivatives. This means that a C^1 function f has continuous partial derivatives and a C^r function f has partial derivatives of class C^{r-1}.

There are converses to this (continuous partial derivatives imply continuously differentiable) but we will not go into this. This might leave a hole a couple of sections down the way. There are proofs of this converse in various books on advanced calculus.

5. The full definition of differentiable manifold.

It is now as good a time as any to finish the definition of a differentiable manifold. In discussions that will come up sooner or later, it will be convenient to introduce more flexibility into our choice of coordinate charts. The addition to the definition will give us this flexibility. We have already seen the need for the flexibility in the statement of Lemma 3.3 where we assumed that one coordinate patch mapped into another in order to avoid having to mess up the notation with restrictions.

Our current definition of a C^r m-manifold is that it is a separable, metric space with an open cover of coordinate patches that have C^r overlap maps. We now shift our focus from coordinate patches (the domains of the coordinate functions) to coordinate charts (the domains of the coordinate functions together with the coordinate functions). (Our distinction between coordinate patches and coordinate charts is not exactly standard.) We now define a C^r m-manifold to be a separable, metric space with a collection of coordinate charts $\{(U,\theta)\}$ where θ is a homeomorphism from U to an open subset of $\mathbf{R}^m$. We drop the subscript from θ since we no longer regard θ as determined by U. In fact, there may be many coordinate functions with the same domain. We put three conditions on the collection of coordinate charts. The first two are already familiar. 1: The domains of the coordinate functions shall form an open cover of M. 2: The overlap maps shall be C^r. 3: The collection of coordinate charts shall be maximal with respect to conditions 1 and 2. The collection of coordinate charts is called the *differential structure* for the manifold.

Condition 3 seems as though it might introduce some ambiguity as to what the collection of charts should be. This is not the case. Let A be a collection of coordinate charts on M that satisfies 1 and 2 but not 3. Let B be a collection of coordinate charts on M that satisfy nothing in particular. It turns out that in order to tell if $A \cup B$ is a collection that satisfies 1 and 2, it is only necessary to check, for each chart (U,θ) in B, that all overlap maps involving (U,θ) and a chart in A are C^r. [Presenters: . . . .] Thus the "admissibility" of B as a possible addition to A depends only on the individual charts in B and not on any properties of B as a collection. Thus a maximal collection based on A is obtained by throwing in any chart whose overlap maps with the charts of A are C^r.

This has several consequences. The first consequence discusses how little information is needed to determine the structure on a manifold. Let C be a collection of coordinate charts satisfying 1 and 2. Let A and B be subcollections of C that also satisfy 1 and 2. All the charts in C are compatible with A and also with B. Thus if we start with only A and maximize to obtain 3, we will add all the charts originally in C. Similarly, if we start with only B and maximize to obtain 3, we will add all the charts originally in C. Thus, the differential structure on a manifold is determined by the class of differentiability desired and by any subcollection of charts of the differentiable structure whose domains cover the manifold.

The second consequence discusses the richness of charts available. Let M be a C^r m-manifold and let x be a point in an open set E of M and let (U,θ) be a coordinate chart with $x \in U$. But now $(U\cap E, \theta|_{U\cap E})$ is a valid coordinate chart. If it were not in the collection of charts, then its overlap maps with all existing charts would just be restrictions of existing overlap maps and would be C^r. By maximality, it must be in the collection of charts. This is the last time we will repeat this argument.

Now, instead of working with $\theta|_{U\cap E}$, we will just assume that $\theta|_{U\cap E}$ has replaced θ and that $U \subseteq E$. We will do further replacements introduced by the code words "we now assume" to improve things even more. Now $\theta(x) \in \theta(U)$ and $\theta(U)$ is an open set in $\mathbf{R}^m$. There is an open ϵ-box

$$D = \{(x_1, \ldots, x_m) | a_i < x_i < b_i,\, b_i - a_i = \epsilon,\, 1 \le i \le m\}$$

in $\theta(U)$ with $\theta(x) = ((b_1 - a_1)/2, \ldots, (b_m - a_m)/2)$ at its center. By restricting θ to $\theta^{-1}(D)$, we now assume that $\theta(U) = D$. There is a C^∞ homeomorphism taking D to $\mathbf{R}^m$. This can be done in several steps. First take D to the open ϵ-box centered at the origin by translating $\theta(x)$ to the origin. Then dilate by π/ϵ to get to $[-\pi/2, \pi/2]^m$. Now take $[-\pi/2, \pi/2]^m$ to $\mathbf{R}^m$ by taking $(x_1, \ldots, x_m)$ to $(\tan(x_1), \ldots, \tan(x_m))$. The tangent function is C^∞ and has C^∞ inverse. Thus we can now assume that the coordinate function takes U to all of $\mathbf{R}^m$. What we have shown is that every point has arbitrarily small neighborhoods that are domains of charts whose image is all of $\mathbf{R}^m$.

We can combine our two consequences and say that every differentiable structure has charts whose images are all $\mathbf{R}^m$ and whose domains contain a neighborhood base for every point in the manifold.

6. The tangent space of a manifold.

Let M be a C^r m-manifold and let TM be the union of all the T_x, for $x \in M$. We want to define a structure on TM. This means two things. We want to define a topology on TM. But the current subject is differentiable manifolds. So we also want to define a set of differentiable coordinate patches that cover TM. When we have done so, we will have defined the *tangent space* of the manifold M.

It is possible to spend an infinite amount of time on the tangent space. I want to avoid that. We will see to what extent I succeed.

Since each T_x in TM is a vector space isomorphic to $\mathbf{R}^m$, it is tempting to associate TM with $M \times \mathbf{R}^m$. However, this turns out not to be the right structure in general. For a subset $U \subseteq M$, we can define TU to be the union of all the T_x, for $x \in S$. When U is a coordinate patch, then $U \times \mathbf{R}^m$ does turn out to be the right structure for TU. From this, the right structure for TM will follow.

There are two possible approaches toward proving that the structure for TU is $U \times \mathbf{R}^m$ when U is a coordinate patch of M. One is to come up with a mathematical reason as to why this is so. The other is to simply make this a definition. The second approach is not at all unreasonable since we will show that the coordinate function induces a natural one to one correspondence between TU and $U \times \mathbf{R}^m$. This is reminiscent of our definition of the vector space structure on T_x.

The second approach above (the "just make it a definition" approach) has many advantages. The first is that it gives reasonable answers and that it is easier than the first approach. Another advantage is that many structures get defined on differentiable manifolds and they are usually defined patch by patch. The definition usually starts by declaring that the structure restricted to any single coordinate patch is a product. Often this is justified by the fact that the coordinate function induces a natural one to one correspondence between the structure over the patch and the appropriate product. It might be considered a precedent that if it is proven laboriously that the tangent space over a coordinate patch should be a product, then it should be proven that all other structures are products over coordinate patches. We will take the point of view that once it is shown that tangent spaces should be products over coordinate patches, then it will be reasonable to accept as given that other structures defined in the future should be products over coordinate patches.

We will divide our discussion of the tangent space into two parts. In this section we will assume that the tangent space over a coordinate patch is a product. (Actually, we will make it look rather reasonable because of the one to one correspondence.) In later sections we will justify this.

Now let M be a C^r m-manifold, and let (U,θ) be a coordinate chart for M. We define

$$TM = \bigcup_{x \in M} T_x$$

and

$$TU = \bigcup_{x \in U} T_x.$$

Note that these are disjoint unions since each T_x consists of classes of curves that are required (among other things) to carry 0 to x. Thus T_x and T_y have nothing in common unless $x = y$.

We have a function $\pi : TM \to M$ which takes each vector v in TM to the unique $x \in M$ for which $v \in T_x$. Note that this can be thought of as evaluation at 0. Again, this because T_x consists of classes of curves into M which carry 0 to x.

We now consider the coordinate chart (U,θ). Let $U' = \theta(U) \subseteq \mathbf{R}^m$.

Recall the isomorphism $\hat{\theta} : T_u \to \mathbf{R}^m$ for each $u \in U$ defined by $\hat{\theta}[f] = (\theta \circ f)'(0)$. This is imperfect notation since it is a different isomorphism for each $u \in U$. We recycle this notation to give a function $\hat{\theta} : TU \to \mathbf{R}^m$ defined by exactly the same formula $\hat\theta[f]=(\theta\circ f)'(0)$. It is an isomorphism when restricted to a single T_u, $u\in U$. We also invent a function $\check\theta: TU\to \mathbf{R}^m$ defined by $\check\theta[f]=\theta(\pi[f])=(\theta\circ f)(0)$. The last is well defined since all f in a class are required to take 0 to the same point.

Define a function $\bar\theta: TU\to U'\times\mathbf{R}^m$ by

$$\bar\theta(v)=\big(\check\theta(v),\hat\theta(v)\big).$$

The function $\bar\theta$ is a one to one correspondence. To show one to one, we note that if v and w come from different T_x and T_y, then $\check\theta(v)\neq\check\theta(w)$ since θ is one to one. If v and w come from one T_x but $v\neq w$, then $\hat\theta(v)\neq\hat\theta(w)$ because $\hat\theta$ is an isomorphism when restricted to T_x. The fuction is onto because $\theta: U\to U'$ is onto and each T_x, $x\in U$ is carried onto $\{\theta(x)\}\times\mathbf{R}^m$ by $\hat\theta$.

We now declare the one to one correspondence $\bar\theta$ between TU and $U'\times\mathbf{R}^m$ to be a homeomorphism by setting the open sets in TU to be the images under $\bar\theta^{-1}$ of the open sets in $U'\times\mathbf{R}^m$. Since $U'\times\mathbf{R}^m$ is an open subset of $\mathbf{R}^{2m}$, we have ourselves a coordinate chart for TM. Since the domains of the coordinate charts of M cover M, the coordinate charts that we have just defined cover TM. As mentioned in Section 1, this determines the topology on TM. We must check that the overlap maps are well behaved.

Note that $TU\cap TV\neq\emptyset$ if and only if $U\cap V\neq\emptyset$. In fact, $TU\cap TV=T(U\cap V)$. Assume that (U,θ) and (V,ϕ) are coordinate charts with $U\cap V\neq\emptyset$. Consider the homeomorphisms

$$\begin{aligned}\bar\theta&: TU\to\theta(U)\times\mathbf{R}^m,\\ \bar\phi&: TV\to\phi(V)\times\mathbf{R}^m\end{aligned}$$

and the restrictions to which we give the same names

$$\begin{aligned}\bar\theta&: T(U\cap V)\to\theta(U\cap V)\times\mathbf{R}^m,\\ \bar\phi&: T(U\cap V)\to\phi(U\cap V)\times\mathbf{R}^m.\end{aligned}$$

We now must consider

$$(\bar\phi\circ\bar\theta^{-1}):\theta(U\cap V)\times\mathbf{R}^m\to\phi(U\cap V)\times\mathbf{R}^m$$

as an overlap map. We first identify what is going on in each coordinate.

On the first coordinate, we are looking at a map that takes $\check\theta(v)$ to $\check\phi(v)$. But $\check\theta(v)$ is just $\theta(\pi(v))$ or $\theta(x)$ where $v\in T_x$. This is carried to

$$\begin{aligned}\check\phi(v)&=\phi(\pi(v))\\ &=\phi(x)\\ &=(\phi\circ\theta^{-1})(\theta(x))\\ &=(\phi\circ\theta^{-1})(\check\theta(v)).\end{aligned}$$

Thus the action on the first coordinate is just that of $(\phi \circ \theta^{-1})$ or the overlap map between the charts (U, θ) and (V, ϕ).

On the second coordinate, there is no subtlety. The map takes $\hat{\theta}(v)$ to

$$\hat{\phi}(v) = (\hat{\phi} \circ \hat{\theta}^{-1})(\hat{\theta}(v))$$

and the action on the second coordinate is that of $(\hat{\phi} \circ \hat{\theta}^{-1})$.

The action on the second coordinate can be reinterpreted with the aid of Lemma 3.3. In the setting of that lemma, let the map f be the identity. With this assumption, the lemma is discussing the identity map expressed in local coordinates under two different coordinate functions. This expression in local coordinates is just the overlap map. The conclusion of the lemma (the outer square) is that the derivative of the overlap map is the composition $(\hat{\phi} \circ \hat{\theta}^{-1})$. Of course this notation suppresses the fact that these derivatives are taken at specific points. More accurately, the map from $\{\theta(x)\} \times \mathbf{R}^m$ to $\{\phi(x)\} \times \mathbf{R}^m$ is the derivative of the overlap map $(\phi \circ \theta^{-1})$ at $\theta(x)$.

We now prepare ourselves to forget that we are looking at maps developed from an overlap map of M and use h to denote $(\phi \circ \theta^{-1})$. Let $U' = \theta(U \cap V)$ and let $V' = \phi(U \cap V)$. Our analysis above says that we are looking at a map

$$\bar{h} : U' \times \mathbf{R}^m \to V' \times \mathbf{R}^m$$

that takes (u, v) to $(h(u), Dh_u(v))$. We will analyze the differentiability of this map by representing it as a composition of several maps.

Our discussion in Section 4 gives us a map $A : U' \to \mathbf{R}^{m^2}$ that takes u to the matrix representation of Dh_u. By definition of class, this map is of class C^{r-1} if h is of class C^r. If i represents the identity on U', then we get the map

$$(i, A) : U' \to U' \times \mathbf{R}^{m^2}$$

which is of class C^{r-1} by Lemma 2.5. If j represents the identity on $\mathbf{R}^m$, then we have the map

$$((i, A) \times j) : U' \times \mathbf{R}^m \to U' \times \mathbf{R}^{m^2} \times \mathbf{R}^m$$

which is also of class C^{r-1} by Lemma 2.5. We have a map $B : \mathbf{R}^{m^2} \times \mathbf{R}^m \to \mathbf{R}^m$ which takes (Q, v) to Qv where Q is regarded as an $m \times m$ matrix and $v \in \mathbf{R}^m$ is regarded as a column vector. The formulas for matrix multiplication are infinitely differentiable, so B is C^∞. Now we have that

$$(h \times B) : U' \times \mathbf{R}^{m^2} \times \mathbf{R}^m \to V' \times \mathbf{R}^m$$

is C^r by Lemma 2.5. Now we have

$$\bar{h} = ((i, A) \times j) \circ (h \times B)$$

which is C^{r-1}. (This argument was shown to me by Erik Pedersen who said that the right approach to exercises of this type is to represent the map being analyzed as the longest possible combination of simpler maps.)

We have shown

THEOREM 6.1. *If M is a C^r manifold, then TM is a C^{r-1} manifold.*

We finish this section with a few statments about the tangent space of M.

The space TM is an example of a vector bundle. Thus it is often called the *tangent bundle* over M to distinguish it from the individual spaces T_x which are the tangent spaces over the individual $x \in M$. A vector bundle over a space is a structure over the space that includes a cover of the space and a collection of charts of the vector bundle that are made of products of the elements of the cover with a fixed vector space. A careful discussion then has to take place about overlap maps. We will not go into this.

We have the map $\pi : TM \to M$ which takes each v to the x for which $v \in T_x$. A *section* for π or a *section* of the tangent bundle is a map $\sigma : M \to TM$ which satisfies $(\pi \circ \sigma)(x) = x$ for all $x \in M$. In words, each x is carried to vector in T_x. Recall that maps are continuous, so that we have a continuous choice of a vector at x that is tangent to M at x. Another name for a section of the tangent bundle is a *vector field* on M.

Note that each T_x has a zero vector. If $\sigma : M \to TM$ is a vector field, then it is a non-zero vector field if no $\sigma(x)$ is the zero vector. We have shown previously

THEOREM 6.2. *There is no non-zero vector field on S^2.*

Note that if TM has the structure $M \times \mathbf{R}^m$, then there is a non-zero vector field. Take your favorite non-zero vector v in $\mathbf{R}^m$ and let $\sigma(x) = v$ for all $x \in M$. We thus have

COROLLARY 6.2.1. *The structure of TS^2 is not that of $S^2 \times \mathbf{R}^2$.*

7. The Inverse Function Theorem.

In this section we present the first of several theorems that derive information from the derivative of a function. The idea behind such theorems is that if the derivative is such a good approximation to a function, then properties of the derivative should be inherited to some extent by the function. The reason that this is useful is that the linearity of the derivative makes certain properties easy to detect on level of the derivative.

The main theorem of this section, the Inverse Function Theorem, is that if a C^1 function f between manifolds has Df_x a vector space isomorphism for some x, then f is locally a homeomorphism on some neighborhood of x. The continuity of the derivative is vital in reaching a conclusion about a neighborhood of x.

There are other features of this section. The first theorem that one learns in calculus that extracts information from the derivative is the Mean Value Theorem. The importance of this theorem cannot be overemphasized. One of the steps of the proof of the Inverse Function Theorem is to develop a version of the Mean Value Theorem in higher dimensions.

Another feature of this section is to introduce the phrase "by local change of coordinates, we can assume … " to the reader. This will occur several times, once as a consequence of the Inverse Function Theorem that we give as a corollary. Instead of trying to make a general lemma that states when this phrase can be invoked, we just give the examples to show how and when it is done.

A third feature of this section is that we avoid partial derivatives to a degree verging on paranoia. Our arguments lie somewhere between the specificity of direct coordinate calculations and the generality of proving these theorems on Banach spaces. (This last can be done, and is done in several texts.)

Lastly, this section unrolls the proof of the main theorem very slowly. Various intermediate results (such as the Mean Value Theorem) are stated and proven in the middle of the proof of the main theorem. To prove a homeomorphism, one must prove that a function is both one to one and onto. The proofs of these two parts are quite separate and are done in with a large interruption in between to introduce needed lemmas.

We start by stating the main theorem and giving a corollary. The theorem guarantees the existence of a homeomorphism and has something to say about the derivative of the inverse.

Theorem 7.1 (Inverse Function Theorem). *Let $f: M \to N$ be a C^r function, $r \geq 1$, between manifolds, and assume that Df_x is an isomorphism for some $x \in M$. Then there is an open set U about x so that $V = f(U)$ is open in N, so that $f|_U$ is a homeomorphism onto V and so that $(f|_U)^{-1}$ is C^r and if $(f|_U)^{-1}(z) = x$, then $D\big((f|_U)^{-1}\big)_z = (Df_x)^{-1}$.*

Corollary 7.1.1. *Let f, M, N and x be as in the theorem above with M and N of class C^r. Then there is an expression h of f in local coordinates so that h is the identity function from a Euclidean space to itself.*

Proof of corollary: Assume that M is an m-manifold. Since Df_x is an isomorphism, the dimension of $T_{f(x)}$ is m and N is an m-manifold. Assume the conclusion of the Inverse Function Theorem with the notation as in the statement. By the discussion in Section 5, we can find a coordinate chart (U_1, θ) with $U_1 \subseteq U$ in which θ is a homeomorphism onto $\mathbf{R}^m$ and so that $f(U_1)$ is contained in the domain of a chart (V_1, ϕ) for N. Thus, the expression h_1 of f in these coordinates takes $\mathbf{R}^m$ to an open subset W of $\mathbf{R}^m$. We know that h_1 and $(h_1)^{-1}$ are C^r. Let $W = f(U_1)$ and let $\zeta = (h_1)^{-1} \circ (\phi|_W)$. Now (W, ζ) is is a valid coordinate chart for N and the expression of f using coordinates (U_1, θ) and (W, ζ) is the identity from $\mathbf{R}^m$ to itself.

In the presence of the hypotheses of the Inverse Function Theorem, the corollary above is usually invoked with the words "by the Inverse Function Theorem we can assume that the function is just the identity on $\mathbf{R}^m$ in local coordinates."

We will start the proof of the Inverse Function Theorem be first showing that there is a neighborhood of x on which f is one to one. The main tool will be a technique that controls how much points move under various maps. The main tool for the control will be a Mean Value Theorem. We will start with that.

THEOREM 7.2 (MEAN VALUE THEOREM). *Let $f : \mathbf{R}^m \to \mathbf{R}^n$ be C^1 and let $a, b \in \mathbf{R}^m$. Assume that $\|Df_x\| \leq K$ for some real $K \geq 0$ and for all x on the straight line from a to b. Then $\|f(b) - f(a)\| \leq K\|b - a\|$.*

PROOF: Let x be on the line L from a to b and let ϵ be greater than 0. Consider h small enough to make the following true:

$$\|f(x+h) - f(x)\| \; - \; \|Df_x(h)\| \; \leq \; \|f(x+h) - f(x) - Df_x(h)\| \; < \; \epsilon\|h\|.$$

For such an h,

$$\begin{aligned} \|f(x+h) - f(x)\| &< \|Df_x(h)\| + \epsilon\|h\| \\ &\leq \|Df_x\| \cdot \|h\| + \epsilon\|h\| \\ &\leq (K + \epsilon)\|h\|. \end{aligned}$$

Now each $x \in L$ has a $\delta_x > 0$ so that the above holds whenever h is within δ_x of x and we get an open cover of L. Pick a Lebesgue number η for this cover and divide L into intervals of length less than η. Let the endpoints of the intervals be $a = x_0 < x_1, \cdots < x_p = b$. Now

$$\begin{aligned} \|f(b) - f(a)\| &\leq \sum \|f(x_i) - f(x_{i-1})\| \\ &< (K + \epsilon) \sum \|x_i - x_{i-1}\| \\ &= (K + \epsilon)\|b - a\|. \end{aligned}$$

This can be done for any $\epsilon > 0$ so the statement of the theorem holds.

PROOF OF THE INVERSE FUNCTION THEOREM: INJECTIVITY: Since Df_x is a linear isomorphism, the dimension of the domain and range are the same. Let this common dimension be m.

We now argue a reduction. We wish to replace the hypothesis of the Inverse Function Theorem by one which assumes more about f than is given in the statement. This will be another argument about simplifications that can be made with local change of coordinates.

Consider an expression of f in local coordinates. We can call it h now, but we will make improvements on it and still call it h. This is a function from an open set in $\mathbf{R}^m$ to $\mathbf{R}^m$ and it carries the image of x under one coordinate map to the image of $f(x)$ under another. By composing the first coordinate function with a translation we can assume that the image of x under the first coordinate function is the origin. By composing the other coordinate function with a translation, we can assume that the image of $f(x)$ under the second coordinate function is also the origin. Now we have that the expression h takes the origin to the origin, and that Dh_0 is a linear isomorphism from $\mathbf{R}^m$ to $\mathbf{R}^m$. We can compose the second coordinate function with the inverse of this linear isomorphism and we have a new expression h of f so that it carries the origin to the origin and so that Dh_0 is the identity. If the Inverse Function Theorem is proven for h, then it will be true for the f given in the statment.

We thus invoke the magic words "by a local change of coordinates ... " and we assume that f is a function from an open set U_1 in $\mathbf{R}^m$ to $\mathbf{R}^m$ that takes 0 to 0 and which has Df_0 as the identity from $\mathbf{R}^m$ to $\mathbf{R}^m$.

We now wish to show that there is a neighborhood of 0 on which f is one to one. This will follow immediately if we show that for all x, y in some neighborhood of 0, we have

$$\|f(x) - f(y)\| \;\geq\; \frac{1}{2}\|x - y\|. \tag{9}$$

To get this kind of inequality that says that f does not contract much, we apply a tranformation that reduces our task to showing that another function does not expand much. Consider the function $g(x) = x - f(x)$. Assume we can show that in some neighborhood of 0 every x and y in this neighborhood satisfies

$$\|g(x) - g(y)\| \;<\; \frac{1}{2}\|x - y\|. \tag{10}$$

So

$$\begin{aligned} \frac{1}{2}\|x - y\| &> \|g(x) - g(y)\| \\ &= \|(x - y) - (f(x) - f(y))\| \\ &\geq \|x - y\| \;-\; \|f(x) - f(y)\|. \end{aligned}$$

Thus we get (9).

Our task is now to show (10). This is now in a form that can be handled by the Mean Value Theorem. We will be done by the Mean Value Theorem if we can show that $\|Dg_x\| < 1/2$ for all x in some neighborhood of the origin. Since f is C^r, so is g. We know Df_0 is the identity, so $Dg_0 = D(x - f(x))_0 = 0$. We now need a continuity argument.

Because Dg is continuous, we have a continuous map (which we can call Dg) from U_1, the domain of g, to $\mathbf{R}^{m^2}$ which we identify with the space of linear maps from $\mathbf{R}^m$ to itself. It takes $u \in U_1$ to Dg_u. We have

$$U_1 \times \mathbf{R}^m \xrightarrow{D_g \times 1} \mathbf{R}^{m^2} \times \mathbf{R}^m \xrightarrow{\;\mu\;} \mathbf{R}^m$$

where μ represents matrix multiplication. The composition is continuous. The composition takes (x, v) to $Dg_x(v)$.

We now use this to estimate $\|Dg_x\|$ for values of x near 0. We know Dg_0 is the zero map and $\|Dg_0\| = 0$. That is, the image of the unit ball B in $\mathbf{R}^m$ is the point 0 in $\mathbf{R}^m$ under Dg_0. By the continuity of $\mu \circ (Dg \times 1)$ each (x,v) in $(\{0\} \times B) \subseteq U_1 \times \mathbf{R}^m$ has a $\delta_{(x,v)}$ so that (y,w) within $\delta_{(x,v)}$ of (x,v) implies that $Dg_y(w)$ is withing $1/2$ of 0. This gives an open cover of $(\{0\} \times B)$ with Lebesgue number η. Now for x within η of 0, we have $Dg_x(B)$ within $1/2$ of 0. Thus for x within η of 0, we have $\|Dg_x\| < 1/2$.

Combining this with our observations above, we have that f is one to one on the open ball E of radius η around 0.

Before we start work on the proof that f is surjective onto some open set in $\mathbf{R}^m$ that contains 0, we need some preliminaries. As a start, it becomes important at this point to mention that we are using the Euclidean metric on $\mathbf{R}^m$. That is, the square root of the sum of the squares of the differences of the coordinates. We use ρ to denote this metric. The property that we need from this metric is that straight lines give the shortest distances betweeen points. We only need this in the form of a strict triangle inequality for non-degenerate triangles which can be deduced from the law of cosines. It is used in the next chain of lemmas.

LEMMA 7.3. *Let ABC be an isosceles triangle in $\mathbf{R}^m$ with $\rho(A,B) = \rho(A,C)$ and $B \neq C$. Let D be a point in the interior of $\rho(A,B)$. Then $\rho(D,C) > \rho(D,B)$.*

PROOF: If false, then the non-degenerate triangle ADC violates the strict triangle inequality by having $\rho(A,D) + \rho(D,C)$ no greater than $\rho(A,C)$.

LEMMA 7.4. *Let B be a closed, round ball in $\mathbf{R}^m$ and let y be a point in the interior of B that is not the center. Let z be the point on the boundary of B that is the intersection of a ray from the center of B through y. Then, for any point x in $\mathbf{R}^m$ minus the interior of B, $\rho(x,y) > \rho(y,z)$.*

PROOF: If x is on the boundary of B, then x, z and the center of B form an isosceles triangle with y in the interior of one of the equal legs. The result follows from the previous lemma. If x is not on the boundary of B, then the straight line segment from y to x must hit the boundary of B in a point w interior to the segment and w will be closer to y than x. But now w is farther from y than z unless $w = z$.

LEMMA 7.5. *Let B be a closed round ball in $\mathbf{R}^m$ and let z be a point on the boundary of B. Let U be an open subset of $\mathbf{R}^n$ and let $f : \mathbf{R}^n \to \mathbf{R}^m$ be C^1 taking a point x to z. Assume that the image of f misses the interior of B. Then Df_x is not a surjection.*

PROOF: By applying a translation, we may assume that z is the origin. Let v be the center of B. We will show that the image of Df_x does not contain v. Since Df_x is linear, this is equivalent to showing that Df_x hits no multiple of

v. Assume that v is in the image. Then for some $h \in \mathbf{R}^n$ we have $Df_x(h)$ is a positive multiple of v. For real $t > 0$, consider

$$\|f(x+th) - f(x) - Df_x(th)\|. \tag{11}$$

For small values of t, the vector $Df_x(th)$ is parallel to v but shorter. Thus it represents a point y in the interior of B that is not the center and, by the previous lemma, z is the point not in the interior of B that is closest to y. Now $f(x) = z$ which is the origin, so (11) reduces to $\|f(x+th) - y\|$. Since the hypothesis says that $f(x+th)$ is not in the interior of B, we know, from the previous lemma, that $\|y\| < \|f(x+th) - y\|$ which restates as

$$\|Df_x(th)\| \ < \ \|f(x+th) - f(x) - Df_x(th)\|.$$

But for any $\epsilon > 0$, suitably small values of $t > 0$ make the right side is less than $\epsilon\|th\|$. Linearity of Df_x gives $t\|Df_x(h)\| < t\epsilon\|h\|$ or $\|Df_x(h)\| < \epsilon\|h\|$. Since this is true for any $\epsilon > 0$, we must have $Df_x(h) = 0$. But now no multiple of $Df_x(h)$ equals v.

Proof of the Inverse Function Theorem: surjectivity: We assume that we work in the open ball E about 0 on which f is one to one. Let B be the closed ball about 0 of radius half that of E. We know that f takes 0 to 0 and is one to one on B. Thus no point of S, the boundary of B, is taken to 0. Since S is compact, there is a minimum distance δ from 0 to $f(S)$. Let B' be the ball about 0 of radius $\delta/3$. We claim that B' is in the image of B. Let y be a point in B'. If y is not in the image of B, then there is a minimum distance γ from y to $f(B)$ and there is a point x in B for which $\rho(y, f(x)) = \gamma$. Now $\rho(y,0) \leq \delta/3$ and 0 is in the image of B, so $\gamma \leq \delta/3$. Since δ is the minimum distance from 0 to $f(S)$, the triangle inequality says that the distance from y to any point in $f(S)$ is at least $2\delta/3$. Thus x is not in S and is in the interior of B.

We now have the situation of the previous lemma since f is a C^r map from the interior of B to $\mathbf{R}^m$ which hits the boundary of the γ ball about y but not the interior of that ball. Thus by the previous lemma, Df_x is not surjective. In particular, it is not an isomorphism. This occured inside a given ball B, so if f is not surjective onto some open neighborhood, then it happens arbitrarily close to 0. Now if Df_x is not an isomorphism, then its matrix representation has determinant 0. Thus if f is not surjective onto some open set, then there are points x_i converging to 0 whose derivatives have determinant 0. But Df_0 is an isomorphism and has non-zero determinant. The determinant is a continuous function of the entries of a matrix. Since f is C^1, we have a contradiction.

We are not quite done. The statment of the theorem has something to say about the differentiability of the inverse function and we do not yet even know if the inverse is continuous. The next arguments finish the proof.

Proof of the Inverse Function Theorem: conclusion: We have that f is a continuous one to one correspondence from some open set U containing 0 to an open set W containing 0. By the argument just above using the continuity of Df, we can also assume that the neighborhood U has been picked so that Df_x is an isomorphism for all $x \in U$.

Let z, w be in W and let x, y in U be such that $f(x) = z$ and $f(y) = w$. Denote the inverse of f by F. From (9) we have

$$\|z - w\| \;\geq\; \frac{1}{2}\|F(z) - F(w)\|$$

or

$$\|F(z) - F(w)\| \;\leq\; 2\|z - w\|$$

which shows the continuity of F.

To validate the claim in the statement of the Inverse Function Theorem about the derivative of DF, we must look at

$$\|F(w) - F(z) - (Df_x)^{-1}(w - z)\| \;=\; \|y - x - (Df_x)^{-1}(f(y) - f(x)\|. \tag{12}$$

The expression inside the norm in (12) is obtained from the expression inside the norm of the next expression by applying $(Df_x)^{-1}$. Thus if $K = \|(Df_x)^{-1}\|$, then (12) is no greater than

$$K\|Df_x(y - x) - f(y) + f(x)\| \;=\; K\|f(y) - f(x) - Df_x(y - x)\|. \tag{13}$$

Now (13) can be kept less than $(\epsilon/2)\|y - x\|$ for a given $\epsilon > 0$ by keeping $\|y - x\|$ suitably small. We want our original (12) (which is no greater than (13)) smaller than $\epsilon\|w - z\|$. But another application of (9) gives us

$$(\epsilon/2)\|y - x\| \leq \epsilon\|f(y) - f(x)\| = \epsilon\|w - z\|.$$

We obtain this by controling $\|y - x\| = \|F(w) - F(z)\|$. We want to do it by controlling $\|w - z\|$. But by (9) again, $\|F(w) - F(z)\| \leq 2\|w - z\|$ so keeping $\|w - z\|$ half the size required for $\|y - x\| = \|F(w) - F(z)\|$ will do the job. This shows that F is differentiable and that its derivative is as claimed in the statement of the theorem.

We now show that F is C^r. We have $DF_z = (Df_{F(z)})^{-1}$. We can regard $z \mapsto DF_z$ as a composition of three functions $i \circ Df \circ F$ where $i : \mathbf{R}^{m^2} \to \mathbf{R}^{m^2}$ is the operation of matrix inverse. Cramer's rule (a formula for matrix inversion involving determinants) shows that i is C^∞. Since f is C^1, the function $x \mapsto Df_x$ is continuous. Thus

$$DF = i \circ Df \circ F \tag{14}$$

is continuous and F is C^1. But now if f is C^2, then all the functions on the right side of (14) have continuous derivatives and F is C^2. Further, the derivative of both sides of (14) and the chain rule give D^2F as a composition involving DF, Di and D^2f. But (14) can be used again to replace DF in the composition with the right side of (14) in which only F and not DF appears. Since i is infinitely differentiable, the only thing to stop this process is the limit on the differentiability of f. Inductively, we get that if f is C^r, then so is F.

[The proof of surjectivity above can be short circuited significantly by replacing the geometric argument about the derivative at the point of closest approach to a point in the range by a more algebraic one. The right way to measure to detect the closest approach is to use the square of the distance. This has the double advantage that the square of the distance has a simple formula that is differentiable and that it can be represented by a dot product. It turns out that formulas involving the dot product are easy to differentiate. In fact, the dot product is an example of a bilinear map and these are easy to differentiate. Let $f : A \times B \to C$ be a bilinear map between vector spaces. That means that $f(a, b_1 + b_2) = f(a, b_1) + f(a, b_2)$, $f(a_1+a_2, b) = f(a_1, b)+f(a_2, b)$, and $rf(a, b) = f(ra, b) = f(a, rb)$. Unfortunately, it also means that f is not linear unless one of A or B is trivial so we cannot say that $Df = f$. Consider the inclusions $i_v : A \to A \times B$ defined by $i_v(u) = (u, v)$ and $j_u : B \to A \times B$ defined by $j_u(v) = (u, v)$. Each is a constant plus a linear map. For example $i_v(u) = (0, v) + i_0(u)$ and i_0 is linear. Thus $D(i_v)_u = i_0$ for all u and v, and $D(j_u)_v = j_0$ for all u and v. Now the compositions $(f \circ i_v)$ and $(f \circ j_u)$ are basically the restrictions of f to $A \times \{v\}$ and to $\{u\} \times B$ respectively and are also linear (since f is bilinear) and are their own derivatives.

This observation and the chain rule give

$$\begin{aligned}(f \circ i_v) &= D(f \circ i_v)_u \\ &= (Df_{i_v(u)}) \circ i_0 \\ &= (Df_{(u,v)} \circ i_0),\end{aligned}$$

and

$$\begin{aligned}(f \circ j_u) &= D(f \circ j_u)_v \\ &= (Df_{j_u(v)}) \circ j_0 \\ &= (Df_{(u,v)} \circ j_0).\end{aligned}$$

These can be applied to $a \in A$ and $b \in B$ as appropriate to give

$$\begin{aligned}(f \circ i_v)(a) = (Df_{(u,v)} \circ i_0)(a), \ \text{or} \\ f(a, v) = Df_{(u,v)}(a, 0),\end{aligned}$$

and

$$\begin{aligned}(f \circ j_u)(b) = (Df_{(u,v)} \circ j_0)(b), \ \text{or} \\ f(u, b) = Df_{(u,v)}(0, b).\end{aligned}$$

Since $Df_{(u,v)}$ is a linear map, we have

$$Df_{(u,v)}(a,b) = f(a,v) + f(u,b).$$

We can now apply this to dot products. Consider $d : \mathbf{R}^m \times \mathbf{R}^m \to \mathbf{R}$ where $d(u,v)$ is the dot product of u and v. This is bilinear so the above applies. Consider $f : X \to \mathbf{R}^m$ and $g : Y \to \mathbf{R}^m$. We have $(f \cdot g) = d \circ (f \times g)$. Now $D(f \cdot g) = Dd \circ (Df \times Dg)$. More specifically

$$\begin{aligned} D(f \cdot g)_{(x,y)}(a,b) &= Dd_{(f(x),g(y))} \circ (Df_x \times Dg_y)(a,b) \\ &= Dd_{(f(x),g(y))}(Df_x(a), Dg_y(b)) \\ &= f(x) \cdot Dg_y(b) + g(y) \cdot Df_x(a). \end{aligned}$$

This is often referred to as a product formula.

Going back to the proof of surjectivity, it is now possible to use this to show that if x has $f(x)$ the closest point to y, then all vectors in the image of Df_x are perpendicular to the vector from $f(x)$ to y.]

8. The C^r category and diffeomorphisms.

There is a category whose objects are C^r manifolds and whose morphisms are C^r functions. The categorical isomorphisms are called C^r *diffeomorphisms.* They are the morphisms in the category that have inverses in the category. This is a stronger requirement than just requiring that the morphism have an inverse as a function.

Consider the function $f(x) = x^3$ from $\mathbf{R}$ to $\mathbf{R}$. The function f is C^∞ and is a homeomorphism. However it is not even a C^1 diffeomorphism since its inverse has no derivative at 0. However it is a consequence of the Inverse Function Theorem that if f is a C^r homeomorphism (that is, a homeomorphism that happens to be C^r) and Df_x is non-singular for each x, then f is a C^r diffeomorphism. Note how this does not apply to $f(x) = x^3$.

Two diffeomorphic manifolds “behave the same” with respect questions about differential maps. Every diffeomorphism is a homeomorphism so diffeomorphic manifolds are homeomorphic. The converse is not true. There are eight manifolds that are not C^∞ diffeomorphic, but they are all homeomorphic to S^7. There is an uncountable collection of manifolds, no two of which are C^∞ diffeomorphic, but which are all homeomorphic to $\mathbf{R}^4$. The class of differentiability is uninteresting in these questions once C^1 is reached. The following is one version of this.

THEOREM 8.1.

(1) *Let $1 \le r < \infty$. Every C^r manifold is C^r diffeomorphic to a C^∞ manifold.*
(2) *Let $1 \le r < s \le \infty$. If two C^s manifolds are C^r diffeomorphic, then they are C^s diffeomorphic.*

The above theorem can be found in *Differential topology* by Morris W. Hirsch, Page 52.

Consider $f : M \to N$ a C^r map between C^r manifolds. Let the dimensions of M and N be m and n respectively. We have that $Df_x : T_x \to T_{f(x)}$ is a linear map. This allows us to define $Tf : TM \to TN$ by $Tf(v) = Df_{\pi(v)}(v) \in T_{f(x)}$. This gives a nice well defined function, but it tells us little about how it cooperates with the structures on TM and TN as C^{r-1} manifolds. If (U, θ) is a chart with $x \in U$ and (V, ϕ) is a chart with $f(x) \in V$, then we can express f in local coordinates as $h = \phi \circ f \circ \theta^{-1}$. We also get coordinate charts $(TU, \bar{\theta})$ and $(TV, \bar{\phi})$ for TM and TN that contain the relevant points. The images of these coordinate functions are $\theta(U) \times \mathbf{R}^m$ and $\phi(V) \times \mathbf{R}^n$ respectively. The expression of Tf in these local coordinates from $\theta(U) \times \mathbf{R}^m$ to $\phi(V) \times \mathbf{R}^n$ takes $(\theta(x), v)$ to $(\phi(x), (\hat{\phi} \circ Df_x \circ \hat{\theta}^{-1})(v))$ which by Lemma 3.3 means that (p, v) is taken to $(h(p), Dh_p(v))$. As discussed in Section 6, this is a C^{r-1} map. Since Tf behaves functorially on each T_x and it carries each T_x into $T_{f(x)}$, it is easy to show that Tf behaves functorially in general. Specifically, $T(f \circ g) = Tf \circ Tg$ and if f is the identity on M, then Tf is the identity on TM. We thus have

THEOREM 8.2. *The operator T is a functor from the category of C^r manifolds and C^r maps, $r \geq 1$, to the category of C^{r-1} manifolds and C^{r-1} maps.*

9. Vector fields and flows.

This section is about differential equations and their solutions. Rather than start this section with a diffential equation and look for a solution, we look at a function and see what differential equation it solves. Then we can discuss general differential equations and their solutions.

Let $f : \mathbf{R} \to M$ be a C^1 function into a C^r manifold. We regard $\mathbf{R}$ as a C^1 manifold and we assume a C^1 differential structure on it that contains the coordinate chart $(\mathbf{R}, i)$ where i is the identity map from $\mathbf{R}$ to itself.

Since $i : \mathbf{R} \to \mathbf{R}$ is the identity map, $[i]$ represents an element of $T_0 \subseteq T\mathbf{R}$. Note that 0 (the additive identity) in the vector space T_0 is $[0]$, the class of the constant map taking all of $\mathbf{R}$ to 0. This is because the isomorphism $\hat{\imath} : T_0 \to \mathbf{R}$ of Lemma 1.1 has $\hat{\imath}[0] = (i \circ 0)'(0) = 0$. We also have $\hat{\imath}[i] = (i \circ i)'(0) = 1$ so $[i] \neq 0$ in T_0. (Because $\hat{\imath}[i] = (i \circ i)'(0) = 1$, we could try to identify $[i]$ with 1 in T_0, but this is dependent on our choice of coordinate function and we will content ourselves with the fact that $[i]$ is not 0 in T_0.)

From the definition of tangent spaces, $[f]$ is an element of $T_{f(0)}$. We have $Df_0[i] = [f \circ i] = [f]$. We thus have an interpretation of the vector that f represents at $f(0)$.

It should also be possible for f to represent vectors at other points of its image. Note that $[f]$ is the set of curves that take 0 to $f(0)$ and that have derivatives at 0 the same as $f'(0)$ (as measured in any coordinate chart). It is reasonable to define, for any $t \in \mathbf{R}$, that f represents a vector at $f(t)$ which is the class of curves that take 0 to $f(t)$ and that have the same derivatives at 0 as $f'(t)$ (as measured in any coordinate chart) so we make this a definition. Note that one curve in this class is the curve defined by $f_t(x) = f(x+t) = (f \circ \theta_t)(x)$ (where $\theta_t(x) = x + t$ is the translation of $\mathbf{R}$ that takes 0 to t) since $f_t(0) = f(t)$ and $f_t'(0) = f'(t)$. Also note that $Df_t[i] = (Df \circ D\theta_t)[i] = Df(D\theta_t[i])$ where $D\theta_t[i]$ is an element of T_t in $T\mathbf{R}$. Thus we are using the translations to give preferred isomorphisms from T_0 to the various T_t in $T\mathbf{R}$. We can use $[f_t]$ as the tangent to the curve f at $f(t)$ and, tempting danger, we recycle the prime notation for derivative and let $f'(t)$ denote this tangent $[f_t]$. [Note also that $[\theta_t] \in T_t \subseteq T\mathbf{R}$ since $\theta_t(0) = t$. Thus $Df_t[\theta_t]$ makes sense and $Df_t[\theta_t] = [f \circ \theta_t] = [f_t] = f'(t)$ in our new notation, so we have another view of $f'(t)$.]

From the above discussion, a curve $f : \mathbf{R} \to M$ defines a set of vectors $f'(t) = [f_t]$ that are tangent to the curve at the various points of its image. These tangents give derivative information about the curve at each of its points. A differential equation will go the other way. We will start with vectors and try to find curves that the vectors are tangent to.

One way to start with vectors is to start with a vector field. In deference to customary notation, we will usually use capital letters from the end of the Roman alphabet to denote vector fields. Thus, let $X : M \to TM$ be a vector field. Specifically, X is a section of the tangent bundle. A curve $f : \mathbf{R} \to M$ is an *integral curve* for X, if for each $t \in \mathbf{R}$ we have $f'(t) = X(f(t))$. If $x \in M$, then we say that the integral curve *starts at* x if $x = f(0)$. An *initial value problem* is a vector field X on M and a point $x \in M$. A solution of the initial value problem is an integral curve for X starting at x. We will relate the solutions of initial value problems with the standard existence and uniqueness theorems for differential equations of functions of a real variable.

The following was proven in class in the Fall semester.

THEOREM 9.1. *Let $f(t,x)$ be a function of two real variables defined on some open set U of $\mathbf{R}^2$. Assume that f is continuous, and that (t_0, x_0) is given in U. Then there is an open interval J in $\mathbf{R}$ containing t_0 and a C^1 function $\phi : J \to \mathbf{R}$ so that $\phi(t_0) = x_0$ and so that for all $t \in J$, $(t, \phi(t))$ is in U and $\phi'(t) = f(t, \phi(t))$. Further, if f satisfies a Lipschitz condition with respect to the second variable, and $\theta : K \to \mathbf{R}$ for an open interval $K \subseteq J$ satisfies all the same requirements as ϕ, then $\theta = \phi|_K$.*

This is the standard theorem that guarantees for each initial value problem

$$x' = f(t,x), \qquad x(t_0) = x_0, \tag{15}$$

there exists locally a unique solution. We must make a comment about the solutions. Consider $x(t) = \tan(t)$. This cannot be defined continuously on any open interval containing $\pm\pi/2$. Thus the maximal open interval continaing 0 that this function can be defined on is $(-\pi/2, \pi/2)$. Note that $x'(t) = \sec^2(t) = 1+\tan^2(t) = 1+x^2(t)$ so that x satisfies the initial value problem

$$x' = 1 + x^2, \qquad\qquad x(0) = 0.$$

Thus it may be impossible for the solutions guaranteed in Theorem 9.1 to be defined on all of $\mathbf{R}$. This will have some effect later in this section. We will mention later how this is sometimes prevented.

We would like to apply a theorem like Theorem 9.1 to a manifold setting. We will comment on some aspects of this theorem that need modification before we make the application.

Theorem 9.1 has the derivative conditions given by f varying with both time and position. This is reflected in the notation $f(t,x)$. The setting to which we would like to apply the theorem has a fixed vector field which gives derivative (tangent) conditions at each point, but which does not depend on time (does not depend on the time of arrival of the curve). Extracting less information from Theorem 9.1 is no problem. We can restrict ourselves to time independent systems (the adjective is *autonomous*) which we disguise as time dependent ones by taking an autonomous $f(x)$ and rewriting it as an apparently time dependent $F(t,x)$ defined by $F(t,x) = f(x)$. At this point we can apply standard existence and uniqueness theorems as if time were a factor. Note that autonomous systems are ones where the function giving the derivative information does not depend on time, however the parameter for any solution is still time. Thus $x' = f(x)$ still has x as a function of t and x' still means dx/dt.

[If the entire theory were developed for autonomous systems, then the theory for time dependent systems could actually be recovered. Given a time dependent system, we can regard it as an autonomous system on a domain that has one more dimension than the original. The derivative information in the new system will have vector components the same as they were in the original dimensions and vector component 1 in the new dimension (which may as well be regarded as the time dimension). This will force solution curves to move along in the extra dimension at unit speed and thus pass through points in the other dimensions with the right derivative information for each time t.]

The result of the previous two paragraphs' discussion is that vector fields and differential equations will be assumed autonomous.

The next modification is to introduce extra space dimensions into the theorem. We can use the same notation (taking into account the removal of the dependence on time) and write problems as $x' = f(x)$. However, we now regard x as an element of $\mathbf{R}^m$ instead of $\mathbf{R}$ and the derivative x' will be also be an element of $\mathbf{R}^m$. Thus $f(x)$ has to be an element of $\mathbf{R}^m$ and f is a function from $\mathbf{R}^m$ to $\mathbf{R}^m$. This change turns out to be very minor. The proof of Theorem 9.1 from last semester goes through almost without change to prove a version of Theorem 9.1 in dimensions above 1.

At this point we can sketch how a modified version of Theorem 9.1 can be applied to vector fields on a manifold. Let $X : M \to TM$ be a vector field on a C^r m-manifold M. If we wish some uniqueness in our discussion (and we do), we will need a Lipschitz condition at the appropriate place. One easy way to get a Lipschitz condition for a function is to assume that it is differentiable. This follows from the Mean Value Theorem (exercise). The Lipshitz condition is to be applied to the function giving the derivative information as a function of the spatial coordinates. In our setting this is the vector field X. Thus, we want to assume that X is C^1. This means that TM must have at least a C^1 structure. From Section 6, we know that M must have at least a C^2 structure. We thus assume that $r \geq 2$.

Let (U, ζ) be a coordinate chart for M. We have available the homeomorphism $\bar{\zeta} : TU \to \zeta(U) \times \mathbf{R}^m$ where $\bar{\zeta}(v) = (\check{\zeta}(v), \hat{\zeta}(v))$. We can set up an autonomous differential equation $x' = \hat{\zeta}(X(\zeta^{-1}(x)))$ on $\zeta(U)$. Let ϕ be a solution satisfying an initial condition $\phi(0) = x_0 \in \zeta(U)$. Consider $f = \zeta^{-1}(\phi)$ as a curve in M. We have $f'(t) = [f_t] = [f \circ \theta_t]$ where θ_t is translation by t. But $[f \circ \theta_t]$ is understood by looking at its image under $\hat{\zeta}$. Namely, at the derivative of $\zeta \circ f \circ \theta_t$ at 0. This is

$$\begin{aligned}(\phi \circ \theta_t)'(0) &= \phi'(t) \\ &= \hat{\zeta}(X(\zeta^{-1}(\phi(t)))) \\ &= \hat{\zeta}(X(f(t))).\end{aligned}$$

But this just says that the image under $\hat{\zeta}$ of $f'(t)$ is just the image of $X(f(t))$ under $\hat{\zeta}$. Thus $f'(t) = X(f(t))$ and f is an integral curve for X. It starts at $\zeta^{-1}(\phi(0)) = \zeta^{-1}(x_0)$. It is an exercise to show that another coordinate chart containing $\zeta^{-1}(x_0)$ gives an integral curve starting there that must agree on overlapping parts of the domains. The exercise would use the overlap maps to relate one solution to the other and then quote uniqueness to show that they must agree as maps into M.

The above sketch gives support to the following.

THEOREM 9.2. *Let M be a C^r manifold with $r \geq 2$. Let X be a C^s vector field on M with $s \geq 1$. Then for any $x \in M$, there is a unique integral curve for X that starts at x and that is defined on some open interval in $\mathbf{R}$ containing 0.*

We want more. This will require another modification to the existence and uniqueness theorems above. Because of the techniques that allow results on Euclidean spaces to be applied to manifolds and vice versa, we will not distinguish much from now on between Theorems 9.1 and 9.2.

The last modification is far from minor. We introduce a new concept to discuss it. Let $\phi : J \to M$, be a curve where J is an open interval in $\mathbf{R}$. Assume for the moment that ϕ is one to one. We can talk about a flow that is defined along the image of the curve. The flow will involve a motion of the points on the image of the curve. If $x = \phi(t_0)$ then we can define $\Phi_t(x) = \phi(t_0 + t)$. Note that $\Phi_0(x) = x$. We can think of Φ_t as a function that pushes points t units along the curve with t measured in the domain of ϕ. We have to be careful if J is not all of $\mathbf{R}$. If this is the case, then Φ_t is only defined on those x with a $t_0 \in J$ for which $\phi(t_0) = x$ and $t_0 + t \in J$. The domain of a given Φ_t can easily turn out to be empty. We have actually defined a family of functions and we will refer to the entire family as a *flow*. One relation that the maps Φ_t satisfy, for any x in the image of ϕ, is

$$\begin{aligned}(\Phi_t \circ \Phi_s)(x) &= \Phi_t(\phi(s + t_0)) \\ &= \phi(t + s + t_0) \\ &= \Phi_{s+t}(x)\end{aligned}$$

using the fact that x in the image of ϕ has a unique t_0 satisfying $x = \phi(t_0)$. The above relation must be treated with care in those situations where the domain of ϕ is not all of $\mathbf{R}$.

If ϕ is not one to one, then we get into potential problems of well definedness. These problems go away if the curve is an integral for an autonomous system for which uniqueness holds.

Now assume that ϕ is an integral curve for a vector field X in that $\phi'(t) = X(\phi(t))$. (It will be very important for what we want to say that we are in the autonomous case.) Assume that ϕ is not one to one and assume that the differential equation satifies hypotheses that make solutions to the initial value problems unique. Let $x_0 = \phi(t_0) = \phi(t_1)$ with $t_0 \neq t_1$. Now $\phi(t)$ is a solution to the initial value problem

$$x' = X(x), \qquad\qquad x(t_0) = x_0.$$

Consider

$$\begin{aligned}\phi_1(t) &= \phi(t + (t_1 - t_0)) \\ &= (\phi \circ \theta_{t_1 - t_0})(t)\end{aligned}$$

where $\theta_{t_1 - t_0}$ is translation in $\mathbf{R}$ by $t_1 - t_0$. We have

$$\begin{aligned}\phi_1'(t) &= \phi'(\theta_{t_1 - t_0}(t)) \\ &= \phi'(t + (t_1 - t_0)) \\ &= X(\phi(t + (t_1 - t_0))) \\ &= X(\phi_1(t))\end{aligned}$$

and

$$\phi_1(t_0) = \phi(t_1) = x_0$$

so ϕ_1 is also a solution to the same initial value problem. Thus by uniqueness $\phi_1 = \phi$ and for all t, $\phi(t) = \phi(t + (t_1 - t_0))$. This makes ϕ periodic. It also makes the flow well defined. If $\phi(t_0) = \phi(t_1) = x$ then $\Phi_t(x)$ written as $\phi(t+t_0)$ or $\phi(t+t_1) = \phi(t+t_0+(t_1-t_0))$ specifies only one point.

We claim that there are two possibilities in the above situation (non-injective integral curve for autonomous system) — either ϕ is a constant map or there is a $\delta > 0$ so that

$$\phi(t+\delta) = \phi(t) \tag{16}$$

for all t and δ is the minimum positive real for which (16) holds. If (16) holds for a given δ, then $\phi(t+n\delta) = \phi(t)$ for all $n \in \mathbf{Z}$. If there are arbitrarily small, positive δ for which (16) holds, then the set of points in $\mathbf{R}$ which map to $\phi(t)$ is dense in $\mathbf{R}$. But this is the set $\phi^{-1}(t)$ which must be closed and therefor all of $\mathbf{R}$. Note that a flow using a constant curve makes sense. It is just the constant flow.

Now we note that the existence and uniqeness theorem guarantees solution curves through all points in M. Thus we can define a flow at every point in M. Specifically, $\Phi_t(x) = \phi(t+t_0)$ where ϕ is a solution curve that passes through x, and t_0 is a real number for which $\phi(t_0) = x$. The collection of the Φ_t will be called a *flow* on M determined by X. Since $\Phi_t \circ \Phi_s = \Phi_{t+s}$ holds at each point, it holds in general (whenver the composition makes sense). We can prove more.

Suppose $\Phi_t(x) = \Phi_t(y) = z$. This means that the integral curve passing through x and the integral curve passing through y meet at z. Say $\phi_1(t_1) = x$, $\phi_2(t_2) = y$ and $\phi_1(t_3) = \phi_2(t_4) = z$. Now $\phi_3(t) = \phi_2(t+(t_4-t_3))$ solves the same initial value problem as ϕ_1 (repeat the analysis several paragrpahs above), so $\phi_3 = \phi_1$ and $\phi_1(t) = \phi_2(t+(t_4-t_3))$. So $x = \phi_1(t_1) = \phi_2(t_1+(t_4-t_3))$. Now $z = \Phi_t(x) = \phi_2(t+t_1+(t_4-t_3))$ and $z = \Phi_t(y) = \phi_2(t+t_2)$. Thus ϕ_2 is periodic and $\phi_2(t) = \phi_2(t+(t_1-t_2)+(t_4-t_3))$ for all t. But $y = \phi_2(t_2) = \phi_2(t_1+(t_4-t_3)) = x$. We have shown that each Φ_t is one to one.

Showing that Φ_t is onto requires an assumption. We now assume that the domains of each integral curve is all of $\mathbf{R}$. Let x be in the domain of the system. Then Φ_{-t} is defined as well as Φ_t. We have $\Phi_t \circ \Phi_{-t} = \Phi_0$ which is the identity. Thus $x = \Phi_t(\Phi_{-t}(x))$ and Φ_t is onto. Note that consideration of Φ_{-t} also shows that Φ_t is one to one, but the paragraph above shows that Φ_t is one to one without the assumption that integral curves are defined on all of $\mathbf{R}$.

From now on, we assume that integral curves are defined on all of $\mathbf{R}$. This gives us one to one correspondences Φ_t. Because of the fact that Φ_0 is the identity one to one correspondence and $\Phi_t \circ \Phi_s = \Phi_{s+t}$, we have a group of one to one correspondences and the function $t \mapsto \Phi_t$ is a homomorphism. This situation is almost never referred to as a *one parameter family of one to one correspondences.* There is such a thing as a *one parameter family of homeomorphisms*, but we don't know yet that the functions Φ_t are homeomorphisms. It remains to discuss what kind of one to one correspondences the Φ_t are.

The following can be proven, but will not be proven here. To simplify the statment, we use Φ to represent the flow Φ_t on M and regard the domain of Φ to be

$\mathbf{R} \times M$. Here $\Phi(t,x) = \Phi_t(x)$.

THEOREM 9.3. *Let M be a C^{r+1} manifold with $r \geq 1$. Let X be a C^r vector field on M. Then the flow Φ on M determined by X is C^r on its domain. In particular, each Φ_t is a C^r homeomorphism from M to itself.*

Of course the above statment is limited by the fact that the integral curves for X may have limited domains of definition. The following gives a condition that avoids this problem. We will not prove it here.

THEOREM 9.4. *Let M in Theorem 9.3 be compact. Then the domain of the flow Φ determined by the vector field X is all of $\mathbf{R} \times M$ and each Φ_t is a C^r diffeomorphism.*

10. Consequences of the Inverse Function Theorem.

In this section we present more theorems that obtain information from the derivative of a function. They are all based on the Inverse Function Theorem.

To make the statements simpler we invent some notation. Let $f : M \to N$ be a C^r map, $r \geq 1$, from an m-manifold to an n-manifold and let $x \in M$. If (U,θ) and (V,ϕ) are coordinate charts of M and N respectively with $x \in U$ and $f(x) \in V$ so that $\theta(x) = 0$ and $\phi(f(x)) = 0$, then we say that $h = \phi \circ f \circ \theta^{-1}$ is an expression of f in local coordinates *centered about* x.

THEOREM 10.1 (IMMERSION THEOREM). *Let $f : M \to N$ be a C^r map, $r \geq 1$, from an m-manifold to an n-manifold. Let Df_x be a monomorphism for some $x \in M$. Then there is an expression $h : \mathbf{R}^m \to \mathbf{R}^n$ of f in local coordinates centered about x for which $h(x_1,\ldots,x_m) = (x_1,\ldots,x_m,0,\ldots,0)$.*

PROOF: As in the beginning of the proof of the Inverse Function Theorem, a local change of coordinates allows us to assume that f is a function from an open set U_1 in $\mathbf{R}^m$ into $\mathbf{R}^n$ that takes 0 to 0 and which has $Df_0 : \mathbf{R}^m \to \mathbf{R}^n$ act by taking $(x_1,\ldots,x_m)$ to $(x_1,\ldots,x_m,0,\ldots,0)$.

Let $j : \mathbf{R}^{n-m} \to \mathbf{R}^n$ act by taking $(x_1,\ldots,x_{n-m})$ to $(0,\ldots,0,x_1,\ldots,x_{n-m})$. We define $\bar{f} : U_1 \times \mathbf{R}^{n-m} \to \mathbf{R}^n$ by $\bar{f}(u,v) = f(u) + j(v)$. The domains of $\bar{f}$, f and j do not agree, but we can fix this up by introducing π_1 and π_2 which project $U_1 \times \mathbf{R}^{n-m}$ onto its first and second factors respectively. Now we have

$$\bar{f}(u,v) = (f \circ \pi_1)(u,v) + (j \circ \pi_2)(u,v).$$

Each of j, π_1 and π_2 is linear and its own derivative. We have

$$\begin{aligned} D\bar{f}_{(0,0)}(a,b) &= D(f \circ \pi_1)_{(0,0)}(a,b) + D(j \circ \pi_2)_{(0,0)}(a,b) \\ &= Df_0(a) + j(b) \\ &= (a,b) \end{aligned}$$

by our assumptions about Df_0.

By the the Inverse Function Theorem, there is an open set U_2 in $U_1 \times \mathbf{R}^{n-m}$ containing $(0,0)$ on which $\bar{f}$ is a C^r diffeomorphism onto an open set in $\mathbf{R}^n$. By the discussion in Section 5, there is a coordinate chart (U_3, ρ) in U_2 taking U_3 to $\mathbf{R}^n$ in a way that takes $U_1 \cap U_3$ to $\mathbf{R}^m \times \{(0,\ldots,0)\}$. (The functions discussed in Section 5 "respect" the coordinates.) Now the last few lines in the proof of the corollary to the Inverse Function Theorem can be duplicated.

Theorem 10.2 (Submersion Theorem). *Let $f : M \to N$ be a C^r map, $r \geq 1$, from an m-manifold to an n-manifold. Let Df_x be an epimorphism for some $x \in M$. Then there is an expression $h : \mathbf{R}^m \to \mathbf{R}^n$ of f in local coordinates centered about x for which $h(x_1, \ldots, x_n, x_{n+1}, \ldots, x_m) = (x_1, \ldots, x_n)$.*

Proof: Again, a local change of coordinates allows us to assume that f is a function from an open set U_1 in $\mathbf{R}^m$ into $\mathbf{R}^n$ that takes 0 to 0 and which has $Df_0 : \mathbf{R}^m \to \mathbf{R}^n$ act by taking $(x_1, \ldots, x_n, x_{n+1}, \ldots, x_m)$ to $(x_1, \ldots, x_n)$.

Let $\pi : \mathbf{R}^m \to \mathbf{R}^{m-n}$ take $(x_1, \ldots, x_n, x_{n+1}, \ldots, x_m)$ to $(x_{n+1}, \ldots, x_m)$. Define $\bar{f} : U_1 \to \mathbf{R}^n \times \mathbf{R}^{m-n}$ by setting $\bar{f}(u) = (f(u), \pi(u))$. Since π is linear, we have

$$D\bar{f}_0(a) = (Df_0(a), \pi(a)) = a$$

by our assumption on Df_0. The rest of the argument proceeds as in the proof of the Immersion Theorem.

A function is called an immersion (submersion) at an x in its domain, if the Immersion (Submersion) Theorem applies to the function at x. A function is called an immersion (submersion) if it is an immersion (submersion) at each point in its domain.

This leads to more terminology. A point in the domain of a function is a *regular point* of the function if the function is a submersion there. A point in the domain of a function is a *critical point* of the function if it is not a regular point of the function. A point in the range of a function is a *critical value* of the function if it is the image of a dritical point of the function. A point in the range of a function is a *regular value* of the function if it is not a critical value of the function. This chain of positive and negative definitions leads to conclusions that are worth getting used to. A point that is in the range but not the image of a function must be a regular value of the function since it cannot be a critical value. If $f : M \to N$ is a function from an m-manifold to an n-manifold with $m < n$, then all points in M are critical points and all points in the image of f are critical values since it is impossible for f to be a submersion anywhere. If a function is a submersion, then all points in the domain are regular points and all points in the range (whether in the image or not) are regular values. Lastly, the image of a regular point might still be a critical value if it is also the image of a critical point. That is, a regular value has the property that no point in its preimage is a critical point.

The "subimmersion theorem" fails. The function $x \mapsto x^2$ from $\mathbf{R}$ to $\mathbf{R}$ has derivative at 0 that is neither one to one nor onto. There is also no expression of the function in local coordinates centered at 0 that is linear. It is interesting to see how far a combined proof of the Immersion and Submersion Theorems can be pushed before it fails.

If k is a constant and x is a vector of several components, then under some conditions a formula such as $f(x) = k$ can define some of the coordinates as functions of some of the others. The Implicit Function Theorem says when and to what extent. The standard example of $x^2 + y^2 = 1$ shows that the hypotheses and conclusions are reasonable.

To help with the statement of the theorem, we need a reasonable way to refer to a partial derivative with respect to one variable. Let $f : U \times V \to W$ be given and let $j_u : V \to U \times V$ be defined by $j_u(v) = (u, v)$. As in the remarks at the end of Section 7, j_u is not linear but a constant plus a linear. It derivative is the linear part and we have $D(j_u)_v = j_0$ for any v. (We have to keep careful track of the meaning of the subscripts.) We define $D_2 f_{(u,v)}$ to be $D(f \circ j_u)_v = (Df_{(u,v)} \circ j_0)$.

Theorem 10.3 (Implicit Function Theorem). *Let $f : U \times V \to N$ be a C^r function, $r \geq 1$, between manifolds. Assume that $D_2 f_{(u,v)}$ is an isomorphism for some (u, v) and let $k = f(u, v)$. Then there is an open set U_1 about u in U, an open set V_1 about v in V and a C^r function $g : U_1 \to V_1$ so that for every $(x, y) \in U_1 \times V$, we have $f(x, y) = k$ if and only if $y = g(x)$. Further, if $U_2 \subseteq U_1$ is open and connected about u, then any continuous $g_0 : U_2 \to V$ with $g_0(u) = v$ and satisfying $f(x, g_0(x)) = k$ for every $x \in U_2$ must agree with g on U_2.*

Remark: The function g is the function that is being "implicitly" defined by the equation $f(u, v) = k$.

Proof: By local change of coordinates, we can assume that U and V are open subsets of $\mathbf{R}^m$ and $\mathbf{R}^n$ respectively, that $(u, v) = (0, 0)$, that N is $\mathbf{R}^n$ (the dimension is fixed by the isomorphism $D_2 f_{(0,0)}$), that $f(0, 0) = 0$, and that

$$D_2 f_{(0,0)}(b) = D(f \circ j_0)_0(b) = (Df_{(0,0)} \circ j_0)(b) = Df_{(0,0)}(0, b) = b.$$

We now use u and v as arbitrary elements of U and V and not as reference to items in the statement.

Let $\bar{f} : U \times V \to \mathbf{R}^m \times \mathbf{R}^n$ be defined by

$$\bar{f}(u, v) = (u, f(u, v)) = (\pi(u, v), f(u, v))$$

where $\pi : U \times V \to U$ is projection. Now

$$D\bar{f}_{(0,0)}(a, b) = (\pi(a, b), Df_{(0,0)}(a, b)) = (a, b).$$

So $\bar{f}$ is a C^r diffeomorphism from some open set about $(0,0)$ to an open set about 0. Thus on some open set of the form $U_1 \times V_1$, we have a C^r inverse h of $\bar{f}$ from an open set W about $(0,0) \in \mathbf{R}^m \times \mathbf{R}^n$ onto $U_1 \times V_1$. Every $(x,y) \in W$ has

$$h(x,y) = (h_1(x,y), h_2(x,y))$$

where, by Lemma 2.5, both h_1 and h_2 are C^r. Now

$$\begin{aligned}(x,y) &= \bar{f}(h(x,y)) \\ &= \bar{f}(h_1(x,y), h_2(x,y)) \\ &= (h_1(x,y), f(h_1(x,y), h_2(x,y))),\end{aligned}$$

so $h_1(x,y) = x$ for all (x,y) in W. So $h(x,y) = (x, h_2(x,y))$ and

$$\begin{aligned}(x,y) &= \bar{f}(h(x,y)) \\ &= \bar{f}(x, h_2(x,y)) \\ &= (x, f(x, h_2(x,y))).\end{aligned}$$

This gives that $f(x, h_2(x,y)) = 0$ if and only if $y = 0$. Let $g(x) = h_2(x,0)$. Now $f(x,z) = 0$ if and only if $z = h_2(x,0) = g(x)$. This holds for all $(x,z) \in U_1 \times V_1$ since every such (x,z) is of the form $(x, h_2(x,y))$ for an $(x,y) \in W$.

Now assume U_2 is a connected, open subset of U_1 about 0 and assume there is a continuous function $g_0 : U_2 \to V$ for which has $g_0(0) = 0$ and $f(x, g_0(x)) = 0$ for every $x \in U_2$. Consider the subset A of U_2 on which $g_0 = g$. We know $0 \in A$. Let x_0 be in A. By the continuity of g_0, there is an open $U_3 \subseteq U_2$ about x_0 so that $g_0(U_3) \subseteq V_1$. But for $x \in U_3$, we have $(x, g_0(x)) \in U_3 \times V_1 \subseteq U_1 \times V_1$ and here $f(x, g_0(x)) = 0$ if and only if $g_0(x) = g(x)$. Thus A is open in U_2. Now A is the inverse image of 0 under the continuous $g - g_0$. Thus A is also closed in U_2. Since U_2 is connected, A is all of U_2.

11. Submanifolds.

Let A be a subset of a C^r m-manifold M. We say that A is a C^r *submanifold* of M of dimension k if each point a of A lies in the domain of a chart (U,θ) of M so that if $\mathbf{R}^k \subseteq \mathbf{R}^m$ is the set of points in $\mathbf{R}^m$ whose last $m-k$ coordinates are 0, then

$$U \cap A = \theta^{-1}(\mathbf{R}^k).$$

The chart (U,θ) is called a submanifold chart for A in M. Note that all the charts $(U \cap A, \theta|_{U\cap A})$ where (U,θ) is a submanifold chart for A in M define a C^r differentiable structure for A.

The inclusion of the submanifold A into M is an immersion. That is because a non-zero tangent vector in A cannot become zero in M since a coordinate function to test the tangent vector in A is the restriction of a coordinate function that tests it in M. The inclusion is also more than that. A basic open set in A (say the domain of a coordinate chart) is also open in A in the subspace topology that A gets from M. Thus the inclusion map is open and is a homeomorphism onto A. That this obvious fact is worth pointing out is seen from the next two examples example. We give the more complicated one first.

Let $S^1 \times S^1$ be covered by $\mathbf{R}^2$ in the usual way so that two points in $\mathbf{R}^2$ project to the same point in $S^1 \times S^1$ if and only if their coordinates differ by integers. Let L be a straight line in $\mathbf{R}^2$ of irrational slope. It is impossible for two points on L to have coordinates that differ by integers, so the covering projection restricted to L is one to one. It is also an immersion. (Covering projections are immersions under the reasonable assumption that the charts of the base space and the charts of the covering space are chosen compatibly.) However it is not a homeomorphism onto its image in $S^1 \times S^1$ and its image is not a submanifold of $S^1 \times S^1$. To argue that these statements are true, we argue that the image is dense in $S^1 \times S^1$. First we need a lemma.

Lemma 11.1. *Let r be a positive irrational number, let x and $\epsilon > 0$ be real, and let k be a positive integer. Then there are integers m and n with $|m| \geq k$ so that $mr - n$ is within ϵ of x.*

Proof: Consider the half open interval $[0,1)$ as representative of the real numbers modulo 1. Then the function from $k\mathbf{Z}$ to $[0,1)$ taking km to $kmr \mod 1$ is one to one since $km_1r - km_2r \in \mathbf{Z}$ implies that r is rational. Thus there are infinitely many different numbers in $[0,1)$ of the form $kmr - kn$ for integers km and kn. There must be two $(km_1r - kn_1) < (km_2r - kn_2)$ in $[0,1)$ that differ by less than ϵ. Let $\delta = k(m_2 - m_1)r - k(n_2 - n_1)$. Now $0 < \delta$ and δ is smaller than both 1 and ϵ. If $m_2 = m_1$, then δ is an integer and cannot be greater than 0 and less than 1. Now the integral multiples of δ divide the real line into intervals of length δ so x is within δ (which is less than ϵ) of at least two consecutive integral multiples of δ. We can thus choose one integral multiple of δ that is not 0 and is within ϵ of x. We now have that x is within ϵ of a number of the form $kpr - kq$ where p and q are integers and p is not 0. This completes the lemma.

Now back to the line L in $\mathbf{R}^2$ of irrational slope r. Let its equation be $y = rx + c$. The distance from a point (a,b) in $\mathbf{R}^2$ to L is no more than $b-(ra+c)$ since this is the vertical distance from L to (a,b). If m and n are integers, then $(a+m, b+n)$ projects to the same point in $S^1 \times S^1$ as (a,b) does. The distance from such a point to L is less than $b + n - (ra + rm + c) = (b - ra - c) - (rm - n)$. From the lemma above, we know that we can make $(rm - n)$ as close to $(b - ra - c)$ as we like and we can do it with arbitrarily large values of $|m|$. It is now easy to create a sequence of points in L that is discrete in L but whose images under projection to $S^1 \times S^1$ converge to the image of (a,b). This allows us to make two conclusions. The first is that the image of L is dense in $S^1 \times S^1$. The second is that the projection restricted to L does not carry L homeomorphically onto its image. For let x be a point of L and let x_i be a sequence of discrete points in L whose image converges in $S^1 \times S^1$ to the image of x. The inverse map from the image of L to L cannot be continuous since it will not preserve the limit of the convergent sequence. The problem with the projection restricted to L is that while it is a one to one continuous map, it is not open.

To argue that the image of L is not a submanifold of $S^1 \times S^1$ we note that any open set around a point in the image has its intersection with the image dense in the open set. But the definition of submanifold would demand a coordinate chart (U, θ) in which the intersection of the image of L with U would definitely not be dense in U.

We have constructed an example of an injective immersion that is not a homeomorphism onto its image and whose image is not a submanifold. A much easier example is an injective immersion of the open unit interval into the open unit disk in $\mathbf{R}^2$ so that its image is homeomorphic to the numeral "6." These examples lead to a definition and a lemma. We say that an immersion that is a homeomorphism onto its image is an *embedding*.

LEMMA 11.2. *Let N be a C^r manifold, $r \geq 1$. A subset A of N is a C^r submanifold if and only if A is the image of a C^r embedding.*

PROOF: The forward direction has been argued above. We consider the reverse direction. Let A be the image of the C^r embedding $f : M \to N$. A point x in A has an open neighborhood U which is the image of an open V in M. The set U is of the form $U' \cap A$ where U' is open in N. From the Immersion Theorem, there is an expression of f in local coordinates based on charts contained in U' and V that gives exactly the structure needed for a submanifold chart around x.

In the above, we exploited the fact that the expression in local coordinates guaranteed by the Immersion Theorem gives a structure that fits the definition of a submanifold chart. We can also look at the expression in local coordinates that is guaranteed by the Submersion Theorem. Here we are looking at the projection of $\mathbf{R}^n$ onto the subspace spanned by a subset of its coordiante axes. The preimage of 0 under this projection (the kernel) lies in $\mathbf{R}^n$ exactly as required by the definition of a submanifold chart. That makes the next lemma an easy exercise.

LEMMA 11.3. *Let $f : M \to N$ be a C^r map, $r \geq 1$. If $y \in f(M)$ is a regular value, then $f^{-1}(y)$ is a C^r submanifold of M.*

There is no "only if" in the above. There are submanifolds that are not the inverse images of regular values under any map. The center line L of the Möbius band M does not separate any neighborhood of itself in M. (We have not dealt with manifolds with boundary, so we consider M to be the open Möbius band.) For L to be the inverse image of a regular value, there has to be a submersion to a manifold of dimension 1. But every point in a manifold of dimension 1 separates some neighborhood of itself. [Exercise: the centerline L of the Möbius band M is the inverse image of a critical value of a function $f : M \to \mathbf{R}$.]

It should be noted that there is nothing in the definition of a submanifold that requires it be a closed subset of the manifold that contains it. Some like to include a requirement that submanifolds be closed subsets. Exercise: find an example of a submanifold of $\mathbf{R}^2$ that is not a closed subset.

We end this section with some notation. We have been using T_x to denote the tangent space to a manifold at x. Until now this has offered no opprotunity for ambiguity since the manifold in question was always the unique manifold containing x. Now that one manifold can be a submanifold of another, the notation is not specific enough. We will continue to use it when there is no problem. There are two notations that are standard to resolve the ambiguity. One is to use M_x to denote the tangent space to M at x and the other is to use T_xM to denote the same thing. We will use the first when needed because it is one less character to type.

It is important to note that if M is a C^r submanifold of N and $x \in M$, then M_x is a vector subspace of N_x and that if $i : M \to N$ is the inclusion map, then Di_x is the linear inclusion of M_x into N_x. This is straightforward from the definitions of "submanifold", M_x, N_x, and Di_x.

12. Bump functions and partitions of unity.

This section introduces two very powerful tools available when working with differentiable functions. One typical way that they are used is to deduce global information from local information. Before we give sample applications, we have to develop the techniques.

Consider the function

$$f(x) = \begin{cases} e^{-\frac{1}{t}}, & t > 0 \\ 0, & t \le 0. \end{cases}$$

Before we look at properties of f, we show

$$\lim_{t \to 0_+} \frac{e^{-\frac{1}{t}}}{t^n} = 0. \tag{17}$$

Replacing t^{-1} by x lets us rewrite (17) as

$$\lim_{x \to \infty} \frac{e^{-x}}{x^{-n}} = \lim_{x \to \infty} \frac{x^n}{e^x}$$

which is shown to be 0 by L'Hôpital's rule. The first consequence of (17) is that f is continuous.

We note that $f'(t) = 0$ for negative t. We now discuss $f'(x)$ for positive t and assume that $t > 0$ for the rest of the paragraph. The function f has the form e^g

where g is the function $g(t) = t^{-1}$. It is the case that higher derivatives $f^{(n)}(t)$ have the form $(e^g)(P(g))$ where $P(g)$ is a polynomial combination of derivatives of g. This is easily shown by induction and the chain rule. It is also proven by induction that derivatives of g are polynomial combinations of negative powers of t. Thus $f^{(n)}(t)$ is of the form $(e^g)(Q(t))$ where $Q(t)$ is a polynomial in negative powers of t. By (17) we now have

$$\lim_{t\to 0_+} f^{(n)}(t) = 0.$$

Thus if we show that $f^{(n)}(0) = 0$ for all n, then f is C^∞. But to show that $f^{(n)}(0) = 0$ inductively from the definition of the first derivative, we are reduced to showing that

$$\lim_{t\to 0_+} \frac{f^{(n-1)}(t)}{t} = 0$$

which follows from (17).

Note that while f is C^∞, it is not analytic at 0. No power series can give the constant function 0 to the left of 0 and simultaneously the non-constant function $e^{-1/t}$ to the right of 0. There is a notion of an analytic manifold based on coordinate charts with analytic overlap maps. They are harder to work with since the techniques of this section are not available with these spaces.

We can build various interesting functions from f.

Let

$$g_1(t) = \frac{f(t)}{f(t) + f(1-t)}.$$

The denominator is never 0 since t and $1-t$ are never simultaneously negative. Thus g_1 is C^∞. Now $g_1(t) = 0$ for $t \le 0$, $0 < g_1(t) \le 1$ for $t > 0$ and $g_1(t) = 1$ for $t \ge 1$. Setting $g_2(t) = g_1(t-1)$ and $g_3(t) = g_2(-t)$ give C^∞ functions where g_2 is 0 on $(-\infty, 1]$ and 1 on $[2,\infty)$ and g_3 is 1 on $(-\infty,-2]$ and 0 on $[-1,\infty)$. Thus if $h(t) = 1 - (g_2(t) + g_3(t))$, then $0 \le h(t) \le 1$ for all t, and $h(t)$ is 1 when $|t| \le 1$ and 0 when $|t| \ge 2$. The function h is typically called a bump function.

Higher dimensional versions can be constructed. Consider the function $\phi : \mathbf{R}^m \to \mathbf{R}$ defined by

$$\phi(x_1, \ldots, x_m) = h(x_1)h(x_2)\cdots h(x_m).$$

The function ϕ is C^∞, has its values in $[0,1]$, takes on the value 1 on $[-1,1]^m$ and takes on the value 0 off $(-2,2)^m$. Clearly ϕ can be adjusted so that given an $\epsilon > 0$, the boxes $[-\epsilon,\epsilon]^m$ and $(-2\epsilon, 2\epsilon)^m$ replace $[-1,1]^m$ and $(-2,2)^m$. Also, these boxes can be centered at points other than the origin. This is worth noting as a lemma. We introduce some notation to make this lemma and later lemmas easier to state.

Let $C \subseteq U$ be a closed set in an open set in a C^r manifold M. We say that a C^r function $\phi : M \to \mathbf{R}$ is a bump function for the pair (U, C) if $f(M) \subseteq [0,1]$, $f(C) = \{1\}$, and $f(M - U) = \{0\}$. So far we have shown:

Lemma 12.1. *Let $\epsilon > 0$ be real. Let $x = (x_1, \ldots, x_m) \in \mathbf{R}^m$. Let*

$$C = \{(y_1, \ldots, y_m) \in \mathbf{R}^m | x_i - \epsilon \le y_i \le x_i + \epsilon, \ 1 \le i \le n\}$$

and let

$$U = \{(y_1, \ldots, y_m) \in \mathbf{R}^m | x_i - 2\epsilon < y_i < x_i + 2\epsilon, \ 1 \le i \le n\}.$$

Then there is a C^∞ bump fucntion for (U, C).

Now let $K \subseteq U$ be a compact set in an open set in a C^r m-manifold M. Let x lie in the domain of a coordinate function θ. Then in the domain of θ we can arrange $x \in C_x \in U_x$ where $\overline{U_x}$ lies in the domain of θ, where $\theta(C_x)$ is a box of diameter δ_x centered at $\theta(x)$, and where $\theta(\overline{U_x})$ is a box of diameter $2\delta_x$ centered at $\theta(x)$. Note that this forces x to be in the interior of C_x. By composing θ with a C^∞ bump function for the pair $(\theta(U_x), \theta(C_x))$ we get a C^r bump function for (U_x, C_x) that is defined on the domain of the coordinate function. We extend the bump function to a function ϕ_x defined on all of M be letting ϕ_x be 0 off the domain of the coordinate function. This extends all the relevant derivatives continuously since they all vanish off U_x. The interiors of the C_x form an open cover of K from which a finite subcover can be extracted. Let the corresponding "centers" be $\{x_1, \ldots, x_s\}$ and let the corresponding (U, C) pairs be denoted (U_i, C_i), $1 \le i \le s$. For each i, let ϕ_i be the bump function above for (U_i, C_i). Now if we define $\Phi : M \to \mathbf{R}$ by

$$\Phi(x) = \sum \phi_i(x),$$

then Φ is non-negative and C^r and $\Phi(x)$ has strictly positive values on K and is 0 off U. This is not exactly a bump function because we have no control on the exact values of Φ on K. We can improve on this if desired. We will need what we have just proven in order to get to the improvements so we state it as a lemma.

Lemma 12.2. *Let $K \subseteq U \subseteq M$ where K is compact and U is open and M is a C^r manifold. Then there is a C^r function from M to $\mathbf{R}$ taking values in $[0, \infty)$, taking the value 0 off U and strictly positive values on K.*

In order to get more, we need the notions of paracompact and partition of unity. A topological space is *paracompact* if every open cover of the space has a locally finite open refinement. A *refinement* of a cover is another cover so that every element of the refinement is contained in some element of the original. A cover is *locally finite* if every point of the space has a neighborhood that intersects only finitely many elements of the cover. The following are proven in Section 6-4 of Munkres:

Theorem 12.3 (Stone's theorem). *Every metric space is paracompact.*

THEOREM 12.4. *Every paracompact space is normal.*

The first result applies here because we are only looking at metric spaces. The second result applies as well, but a direct proof that metric spaces are normal is much easier than going through Stone's theorem.

Let $f : X \to [0, \infty)$ be a map. The *support* of f is the closure of the pre-image of $(0, \infty)$. If O is an open cover and $\{\phi_\alpha\}$ is a collection of functions from X to $[0, \infty)$, then the collection of functions is a *partition of unity subordinate to* the cover O if the collection of supports of the ϕ_α is a refinement of O, if for all x

$$\sum \phi_\alpha(x) = 1$$

and if the sum involves only finitely many non-zero terms for each x. Since the values of the functions are never negative, they can never exceed 1. Note that even if O is locally finite, there might be infinitely many non-zero terms in the sum without the extra assumption that this does not happen. The following modification of the definition of partition of unity is used to make the finiteness automatic if O is locally finite. If $O = \{U_\alpha\}_{\alpha \in J}$ is the open cover, then the partition of unity $\{\phi_\alpha\}_{\alpha \in J}$ is *dominated by* O if the support of ϕ_α lies in U_α for each $\alpha \in J$.

We will not prove Stone's Theorem. There is a perfectly good proof in Munkres. It takes about three pages there. We will look at some consequences. We will show:

THEOREM 12.5. *Every open cover of a C^r manifold dominates a C^r partition of unity.*

This will take several steps. We will need various technical lemmas along the way, as well as partial results.

LEMMA 12.6. *A locally finite open cover of a separable space has countably many non-empty sets.*

PROOF: The wording of the statment is to allow a given indexing set to be used for a cover even if some (or most) of the index values refer to empty sets.

Pick a countable dense subset S. Locally finite implies the weaker *point finite*, that every point in $\mathbf{R}^m$ lies in a finite number of elements of the cover. Since every non-empty open set contains a point in S, a list of the elements of the cover that contain each point in S will list all the non-empty elements of the cover. But each point in S lies in finitely many elements of the cover, so the list is countable.

LEMMA 12.7. *A point finite, countable open cover $\{U_i\}$ of a normal space X has a refinement $\{C_i\}$ of closed sets whose interiors cover X and with each $C_i \subseteq U_i$.*

PROOF: Assume that $\{C_1, \ldots, C_n\}$ have been found so that each C_i is closed and in U_i and so that the interiors of the C_i and the U_j for $j > n$ cover X. Let C'_{n+1} be X minus the interiors of all the C_i, $i \leq n$, and minus all the U_j, $j > (n+1)$.

This is a closed set. Since the only set not removed is U_{n+1} and removing U_{n+1} would yield the empty set, we have $C'_{n+1} \subseteq U_{n+1}$. Now because X is normal, there is a closed set C_{n+1} in U_{n+1} whose interior contains C'_{n+1}. We now have our assumption with n replaced by $n+1$. In this way we inductively end up with a collection $\{C_i\}$. To argue that the interiors cover, we note that every $x \in X$ lies in finitely many U_i. After a finite number of steps, these U_i will have been replaced by C_i. By our assumption, x must lie in one or more of the interiors of the C_i.

LEMMA 12.8. *Every open cover $\{U_\alpha\}_{\alpha\in J}$ of a paracompact X has a locally finite open refinement $\{W_\alpha\}_{\alpha\in J}$ where each $W_\alpha \subseteq U_\alpha$.*

PROOF: Note that various W_α may be empty. Let $\{V_\beta\}_{\beta\in K}$ be a locally finite open refinement. Chose a function $f : K \to J$ so that each $V_\beta \subseteq U_{f(\beta)}$. Now form $\{W_\alpha\}_{\alpha\in J}$ by setting W_α to be the union of those V_β for which $f(\beta) = \alpha$. This is an open refinement since each W_α is a union of open subsets of U_α and since each V_β is used in some W_α. Since each V_β is used in only one W_α any neighborhood hitting only finitely many V_β hits only finitely many W_α. Thus $\{W_\alpha\}_{\alpha\in J}$ is locally finite.

LEMMA 12.9. *Every open cover of a C^r manifold M by sets with compact closure dominates a C^r partition of unity.*

PROOF: We can replace the given cover by a locally finite open refinement using the same indexing set as the original. A partition of unity dominated by the new cover will be dominated by the original. The new cover has countably many non-empty sets. Since it is a refinement of the original the elements have compact closure. Let the non-empty sets in the cover that we are working with be $\{V_i\}$. We can extract a closed refinement $\{C_i\}$ whose interiors cover. Since each C_i is closed in a compact set, it is compact. By Lemma 12.2, we now have C^r non-negative functions ϕ_i from M to $\mathbf{R}$ with each ϕ_i strictily positive on C_i and zero off V_i. Thus the supports of the ϕ_i are locally finite and the sum $\sum \phi_i(x)$ is defined for each x. Since the interiors of the C_i cover M, the sum $\sum \phi_i(x)$ is never 0. Now we let

$$\Phi_i(x) = \frac{\phi_i(x)}{\sum \phi_j(x)}.$$

The collection of the Φ_i is now a partition of unity dominated by the $\{V_i\}$. To get a partition of unity for the original indexing set, let the function for those indexes of empty sets be the constant function to 0.

The next lemma gives the promised improvement to Lemma 12.2. It also leads to a proof of Theorem 12.5.

LEMMA 12.10. *Let $C \subseteq V \subseteq M$ where C is closed and V is open and M is a C^r manifold. Then there is a C^r bump function for (V, C).*

Proof: By using coordinate charts, we can cover C by open subsets of V with compact closure. Let $U = M - C$. We can also cover U by open subsets of U which also have compact closure. These two covers together will cover M. Let Φ_α be a C^r partition of unity dominated by the cover. The sum of all the elements of the partition that satisfy the restriction that they correspond to open sets that intersect C gives us a C^r function. It is the function we want since all the supports are in V and since all the functions omitted by the restriction have their supports in U and are not contributors to the fact that the sum is 1 on C.

Proof of Theorem 12.5: The proof is exactly the same as the proof of Lemma 12.9 except that Lemma 12.10 is used instead of Lemma 12.2.

We now give two applications. The first is an example of the use of bump functions, and the second is an example of the use of partitions of unity. They both deduce global information from local information.

The definition of a C^r manifold states that locally the manifold has C^r embeddings into a Euclidean space. If the manifold is compact, then we can use partitions of unity to guarantee the existence of a C^r embedding of the entire manifold into a Euclidean space.

Lemma 12.11. *Let M be a compact C^r m-manifold, $r \geq 1$. Then there is an integer n and an embedding $f : M \to \mathbf{R}^n$.*

Proof: Since M is compact, there is a finite cover of M by coordinate charts (U_i, θ_i), $1 \leq i \leq k$. We can extract a closed cover $\{C_i\}$ with each $C_i \subseteq U_i$ and with the interiors of the C_i covering M. For each i, let $\phi_i : M \to \mathbf{R}$ be a bump function for the pair (U_i, C_i). Each $\theta_i : U_i \to \mathbf{R}^m$ is an embedding. Define

$$g_i : M \to \mathbf{R}^m \times \mathbf{R} = \mathbf{R}^{m+1} \text{ by } g_i(x) = (\phi_i(x)\theta_i(x), \phi_i(x)).$$

Now let

$$g = (g_1, \ldots, g_k) : M \to \mathbf{R}^{m+1} \times \cdots \times \mathbf{R}^{m+1} = \mathbf{R}^{k(m+1)}.$$

Now g is C^r. If $x \in C_i$, then g_i is an immersion at x since the first coordinate of g_i on C_i is θ_i. Thus no tangent vector at x is taken to zero by Dg_i and thus not by Dg since the Dg_i go into independent subspaces of $T\mathbf{R}^{k(m+1)}$. To see that g is an injection, consider $x \neq y$. If x and y lie in one C_i, then $g(x) \neq g(y)$ again since the first coordinate of g_i is θ_i which is injective on C_i. If $x \in C_i$ and $y \notin C_i$ then the second coordinate of g_i disagrees on x and y and $g(x) \neq g(y)$. So g is an injective immersion and thus an embedding.

Remark: The result above gives no where close to the best estimate on the dimension of the Euclidean space needed to receive the embedding. There is an argument that shows that the embedding can take place in $\mathbf{R}^{2m+1}$. A much more difficult argument shows that the embedding can take place in $\mathbf{R}^{2m}$.

Now for the second example. Let M and N be C^r manifolds and let C be a closed set in M. Let $f : C \to N$ be a function. We say that f is C^r if for every x in C, there is an open set U in M about x and a C^r function $f_U : U \to N$ so that $f|_{U\cap C} = f_U|_{U\cap C}$.

LEMMA 12.12. *A function $f : C \to \mathbf{R}^n$ where C is a closed subset of a C^r manifold M is C^r if and only if there is an open set U in M about C and a C^r function $f_U : U \to \mathbf{R}^n$ so that $f = f_U|_C$.*

PROOF: For the "if" direction, use U for every x.

Now if f is C^r, then there is a cover $\{U_x\}_{x\in C}$ of C by open sets of M and C^r functions f_x that extend the various $f|_{C\cap U_x}$. Let $V = M - C$ and let a partition of unity dominated by the open cover $\{U_x\}_{x\in C} \cup \{V\}$ of M consist of functions denoted ϕ_x and ϕ_V. Now

$$\sum_{x\in C} \phi_x f_x$$

is C^r, is defined on all of M, and equals f on C.

13. The C^1 metric.

The tangent vectors to a manifold M are defined as equivalence classes of curves. Curves are maps from subsets of $\mathbf{R}$ to M. The set of curves can be formed into a topological space (function space) in many ways. We are familiar with some. Once the set of curves is formed into a function space, we can use a quotient topology on the set of tangent vectors. It turns out that the function space topologies that we are familiar with (e.g., uniform topology, uniform convergence on compact sets, etc.) will give bad topologies on the set of tangent vectors. In particular the quotient topologies are not Hausdorff. This is not hard to see, so we will go into some detail.

The function space topologies that we know give some control on the values of a function. An open set of functions can be defined that will force any function in this open set to have its values on some restricted part of the domain to be near a given value in the range. For example, the compact open topology can be used to build an open set O of functions where the values on a compact subset in the domain are constrained to lie in a neighborhood of a given value in the range. But this will not control the derivative. One can build functions in O that race around the range neighborhood like mad giving arbitrarily large values for the derivatives at given points, and there will be functions in O that will stall at various points (see, for example, the bump functions of Section 12) giving low values of the derivative (even 0) at those points.

A curve identifies a particular tangent vector in TM by seeing what the value of the curve is at 0 (this identifies which T_x we are in) and what its derivative is at 0 (which identifies which v in T_x we are looking at). The topologies that we know build open sets of curves in which the values of the curves at 0 are near a certain point. For such an open set O of curves, the set of tangent vectors defined will lie in a set of tangent spaces T_x where the points x are confined to some neighborhood W in M. However, the derivatives of the curves in O will take on all possible values at 0. The set of tangent vectors defined by the curves will thus be the union of all the T_x for $x \in W$. Taking unions and intersections of these sets of curves will still give sets that represent entire copies of the tangent spaces T_x. Thus the topologies that we know on the set of curves will allow us to separate points in M by open sets but not vectors in any one T_x.

We now discuss how to control the derivative. The problem that we are working on is the structure of TU where (U, θ) is a chart of a C^r m-manifold M. We will use the coordinate function as a tool. This is reasonable since it is the coordinate function that sets up the one to one correspondence between TU and $\theta(U) \times \mathbf{R}^m$ in the first place. Also, a curve $f : J \to U$, where J is an open interval about 0 in $\mathbf{R}$, can be composed with θ so that both its values and its derivatives are elements of $\mathbf{R}^m$.

We will use the metric on $\mathbf{R}^m$ to imitate the construction of the uniform metric. The easiest way to make use of the metric is to take supremums. If we have a compact domain, then our formulas are a little simpler since we don't have to bound distances by 1 all the time. Thus we restrict ourselves to the "unit disk" $[-1, 1]$ in $\mathbf{R}$ and use this for our domain for all curves. Since the relevant information about a curve is its value and derivative at 0, this will suffice. For the rest of this section, let I deonte the interval $[-1, 1]$ in $\mathbf{R}$. When we discuss the derivative of a function defined on I, we will use the right hand derivative at -1 and the left hand derivative at $+1$.

Let d be the metric on $\mathbf{R}^m$. Let $C^1(I, U)$ be the set of C^1 functions from I to U. Let f be an element of $C^1(I, U)$. To simplify notation, we let $\bar{f}$ denote $\theta \circ f$. This is a curve into $\mathbf{R}^m$. For f and g in $C^1(I, U)$ define

$$\begin{aligned}\rho(f, g) = \max[&\sup\{d(\bar{f}(x), \bar{g}(x)) | x \in I\},\\ &\sup\{d(\bar{f}'(x), \bar{g}'(x)) | x \in I\}].\end{aligned}$$

This can be compared with the uniform metric defined near the top of page 266 of Munkres.

Certain calculations go through exactly as they do for the uniform metric.

LEMMA 13.1. *The function ρ is a metric.*

Call this the C^1 metric on $C^1(I, M)$.

LEMMA 13.2. *A sequence $f_n : I \to U$ of C^1 functions converges to the C^1 function f in the C^1 metric if and only if the sequences f_n and f_n' converge uniformly to f and f' respectively.*

In the next section, we will discuss the quotient topology that the C^1 metric induces on TU, and show that with this topology, the one to one correspondence $\bar{\theta} : TU \to \theta(U) \times \mathbf{R}^m$ of Section 6 is a homeomorphism.

Before we end this section, we want to show that the C^1 metric has reasonable properties. The lemma above tells only what happens if convergence in the C^1 metric takes place. It says nothing about how often it happens. It may be rare for a sequence of functions with limit f to have the corresponding sequence of derivatives converge to f'. In fact, it is not rare. If U is complete, then $C^1(I, U)$ is complete. For simplicity, we will show this in the case that U is $\mathbf{R}^m$.

Much of the argument is familiar. If f_n is a Cauchy sequence in $C^1(I, \mathbf{R}^m)$, then for each x in I, $f_n(x)$ is Cauchy and $f_n'(x)$ is Cauchy. Since $\mathbf{R}^m$ is complete, there is a limit for each $f_n(x)$ which we can call $f(x)$ and there is a limit for each $f_n'(x)$ which we can call $g(x)$. It would be a little premature to call g the derivative of f. Since the definition of C^1 demands continuous derivative, the f_n and the f_n' are all continuous. A uniform limit of continuous functions is continuous, so f and g are continuous. Since the convergence $f_n' \to g$ is uniform, there is a tail of the sequence that is within ϵ of g. So every member in this tail satisfies

$$(g(x) - \epsilon) < f_n'(x) < (g(x) + \epsilon)$$

for each x in I. If K is the maximum of g on I, then on this tail $|f_n'(x)| < K + \epsilon$ for all x in I. Thus the tail satisfies the hypotheses of the dominated convergence theorem for integrals. (Our functions are integrable since they are continuous.) We get

$$\int_{-1}^{x} g = \lim \int_{-1}^{x} f_n' = \lim\bigl(f_n(x) - f_n(-1)\bigr) = f(x) - f(-1)$$

for all x in I which demonstrates that $f' = g$. This finishes the argument.

[There is another argument that shows that $f' = g$ based on the Mean Value Theorem and direct computation of the derivative. We give it here for those uncompfortable with the use of the dominated convergence theorem. It is nice in that it can be applied when the defintion of the C^1 metric is generalized to functions from $\mathbf{R}^m$ to $\mathbf{R}^n$ instead of just functions defined on I.

Given $\epsilon > 0$, we wish to find a $\delta > 0$ so that $\|h\| < \delta$ implies

$$\|f(x + h) - f(x) - g(x)h\| < \epsilon\|h\|.$$

Now

$$\begin{aligned}\|f(x+h) - f(x) - g(x)h\| \le \|f(x+h) - f_n(x+h)\| \\ + \|f_n(x+h) - f_n(x) - f_n'(x)h\| \\ + \|f_n(x) - f(x)\| \\ + \|f_n'(x)h - g(x)h\|.\end{aligned}$$

The fourth term on the right is the difference of two linear functions to $\mathbf{R}^m$ evaluated at the same point. (Actually in our setting it is the difference of two function values multiplied by the same displacement.) Thus for a fixed value of h, we can make the first, third and fourth terms on the right as small as we like, say less than $\eta/3$, by using the uniform convergence of f_n to f and f_n' to g by keeping n large enough. Thus if the second term is shown to be less than $\epsilon\|h\|$, then we will have

$$\|f(x+h) - f(x) - g(x)h\| \le \epsilon\|h\| + \eta$$

which can be made to hold for any η by chosing n large enough. Thus we will have shown

$$\|f(x+h) - f(x) - g(x)h\| \le \epsilon\|h\|.$$

We now concentrate on how to show

$$\|f_n(x+h) - f_n(x) - f_n'(x)h\| < \epsilon\|h\|. \tag{18}$$

Note that (18) can be made true for each n by restricting h differently for each n. However, we need to show once h has been chosen sufficiently small, that (18) is true for all sufficiently large n.

We note that as a function of h, the expression $f_n(x+h) - f_n(x) - f_n'(x)h$ is equal to 0 when $h = 0$. Thus we are asking how much $f_n(x+h) - f_n(x) - f_n'(x)h$ varies from its value at $h = 0$ for a given value of h. This is where we apply the Mean Value Theorem.

Let

$$\phi(t) = f_n(x+th) - f_n(x) - f_n'(x)(ht).$$

We have

$$\|f_n(x+h) - f_n(x) - f_n'(x)h\| = \|\phi(1) - \phi(0)\|. \tag{19}$$

We can estimate this by using the Mean Value Theorem.

We will have to take some derivatives. We are already mixing them up pretty well ($f'(x)$ versus Df_x), so we will stick to the "prime" notation and regard the expression $f_n'(x)(ht)$ as the constant $f_n'(x)$ (it does not depend on t) multiplied by ht. Now we have

$$\phi'(t) = f_n'(x+th)(h) - f_n'(x)(h) = (f_n'(x+th) - f_n'(x))(h)$$

by the chain rule. Now

$$\begin{aligned}\|f_n'(x+th) - f_n'(x)\| \le \|f_n'(x+th) &- g(x+th)\| \\ &+ \|g(x+th) - g(x)\| \\ &\quad + \|g(x) - f_n'(x)\|.\end{aligned}$$

The first and third terms can be kept less than $\epsilon/3$ by unifom convergence and keeping n sufficiently large. The middle term is where we get our δ. We chose δ to keep the middle term less than $\epsilon/3$ whenever $\|h\| < \delta$ which can be done by the continuity of g and the fact that t is restricted to lie in $[0,1]$. Now we have $\|\phi'(t)\| \leq \epsilon\|h\|$ for t in $[0,1]$. By the Mean Value Theorem, the right side of (19) is less than $\epsilon\|h\|(1-0)$ and we have shown that (18) holds.]

14. The tangent space over a coordinate patch.

We continue the discussion of the previous section. We have a C^r m-manifold M with a coordinate chart (U,θ). We have the one to one correspondence $\bar{\theta} : TU \to \theta(U) \times \mathbf{R}^m$ as defined in Section 6. We have that TU is a quotient of $C^1(I,U)$ and we have the C^1 metric ρ on $C^1(I,U)$. This gives the quotient topology on TU. We wish to show that $\bar{\theta}$ is a homeomorphism under this topology.

First we show that $\bar{\theta}$ is continuous. Let $\epsilon > 0$ be real. We want a $\delta > 0$ so that if $f, g \in C^1(I,U)$ have $\rho(f,g) < \delta$, then $d(\bar{\theta}[f], \bar{\theta}[g]) < \epsilon$. Here we need to decide on the metric on $\theta(U) \times \mathbf{R}^m$. We decide on the metric $d((a,b),(c,d)) = \max\{d_1(a,c), d_1(b,d)\}$ where d_1 is the metric on $\theta(U) \subseteq \mathbf{R}^m$ and on $\mathbf{R}^m$. We make this choice because it makes the next argument a triviality.

Now $\rho(f,g) < \delta$ implies that $(\theta \circ f)(0)$ and $(\theta \circ g)(0)$ differ by less than δ and $(\theta \circ f)'(0)$ and $(\theta \circ g)'(0)$ differ by less than δ. So $d(\bar{\theta}[f], \bar{\theta}[g]) < \delta$. We now let $\delta = \epsilon$ and are done.

Now we show that $\bar{\theta}$ is open. Suppose that $S \subseteq TU$ is open. We want to show that $\theta(S)$ is open in $\theta(U) \times \mathbf{R}^m$. Let $[f] \in S$. We want a $\delta > 0$ so that if (x,y) is within δ of $\theta[f]$, then there is a $[g]$ in S go that $\bar{\theta}[g] = (x,y)$. Since S is open in TU, it is the image of an open set in $C^1(I,U)$. Thus there is an ϵ so that if $\rho(f,h) < \epsilon$, then $[h]$ is in S. We argue that letting $\delta \leq \epsilon/2$ will work.

Let (x,y) be within δ of $\bar{\theta}[f]$. The notation is easier with displacements, so let $u = x - (\theta \circ f)(0)$ and let $v = y - (\theta \circ f)'(0)$. Consider

$$g_1(t) = u + (\theta \circ f)(t) + tv$$

defined on I. We ignore for a minute that the range of g_1 might not be in $\theta(U)$. We have $g_1(0) = u + (\theta \circ f)(0) = x$ and $g_1'(0) = (\theta \circ f)'(0) + v = y$. So if the range of g_1 is in $\theta(U)$ we are done by letting $g(t) = \theta^{-1} \circ g_1$ so that $\bar{\theta}[g] = (x,y)$. It is easy to show that $\rho(f,g) < \epsilon$ so that $[g]$ is in S. We now modify g_1 to get a g_2 with similar properties but whose range is in $\theta(U)$.

We first take δ smaller if necessary so that the δ ball B around $(\theta \circ f)(0)$ lies in $\theta(U)$. There is a straight line homotopy from $(\theta \circ f)$ to g_1 defined by

$$F(t,s) = su + (\theta \circ f)(t) + stv$$

where $s \in [0,1]$. The homotopy goes into $\mathbf{R}^m$ but not necessarily into $\theta(U)$. Now $F(0,0) = (\theta \circ f)(0)$ which is in the center of the ball B. Also $F(0,1) = g_1(0) = x$

which is within δ of $(\theta \circ f)(0)$ and so is also in B. Since the homotopy is the straight line homotopy, the straight line $\{F(0,s) | s \in [0,1]\}$ is also in B. By the continuity of F and the compactness of $[0,1]$, there is an η so that $F(t,s)$ lies in B for $s \in [0,1]$ and $t \in [-\eta, \eta]$. Let $\phi : I \to [0,1]$ be a bump function which is 1 on $[-\eta/2, \eta/2]$ and 0 off $[-\eta, \eta]$. Now let

$$g_2(t) = \phi(t)u + (\theta \circ f)(t) + \phi(t)v.$$

On $[-\eta/2, \eta/2]$ we have $g_2 = g_1$. This guarantees that $g_2(0) = g_1(0)$ and $g_2'(0) = g_1'(0)$ so that $g(t) = \theta^{-1} \circ g_2$ also has $\bar{\theta}[g] = (x,y)$. It is again easy to show that $\rho(f,g) < \epsilon$ so that $[g]$ is in S. Off $[-\eta, \eta]$ we have $g_2 = (\theta \circ f)$. This guarantees that the image of g_2 off $[-\eta, \eta]$ lies in $\theta(U)$. On $[-\eta, \eta]$ we have that the image of g_2 lies in the image of F on $[-\eta, \eta] \times [0,1]$ which lies in B. This completes the argument.

15. Approximations.

None of the statements in this section will be proven.

Just as one can define the C^1 metric, one can define the C^r metric for any $r > 1$ and also a C^∞ metric. These are for functions with range in some Euclidean space. For maps to an arbitrary manifold, it is harder to make well defined measurements, so one defines C^r topologies and C^∞ topologies instead of metrics. Once a topology is established, then questions about open, closed, compact and dense sets can be discussed. A statment that a set of functions is an open set in a topology says that if a function has the defining property of the set, then all nearby functions have the property. A statement that a set of functions is dense says that any function can be approximated by a function in the set.

There is more than one C^r topology to chose from. There is the "weak" topology and the "strong" topology and there are perhaps others. The weak and strong coincide for a compact domain. We do not provide definitions. The results below leave out which of the C^r topologies are being used on the function spaces.

Many of the approximation results are proven locally first and then extended to global results using bump functions or partitions of unity. As an exercise, one can show that C^∞ functions are dense in the continuous functions using the uniform metric by approximating a continuous function by constant functions on small sets and then using partitions of unity to smooth things out.

Consider the next two results.

Lemma 15.1. *Let M be a C^r m-manifold, $2 \le r \le \infty$. Then, in the space of C^r functions from M to $\mathbf{R}^n$ with the C^r topology, the embeddings are dense if $n > 2m$ and the immersions are dense if $n \ge 2m$.*

Theorem 15.2. *Let M and N be C^r manifolds of dimension m and n repsectively with $2 \le r \le \infty$. If $n \ge 2m$, then the immersions of M into N are dense in the C^r maps from M to N with the C^r topology.*

The proof of the second result will use the first to get approximations on charts. Then bump functions will be used to piece together an apparently incompatible collection of pieces of aproximations.

An openness result is:

LEMMA 15.3. *In the space of C^r maps with the C^r topology, $r \geq 1$, between manifolds, the immersions, the submersions and the embeddings each form an open set.*

A main approximation theorem is:

THEOREM 15.4. *Let M and N be C^s manifolds, $1 \leq s \leq \infty$. Then the C^s functions from M to N are dense in the C^r topology on the C^r functions from M to N for $0 \leq r < s$.*

Approximations are also used to increase the differentiability of a differentiable structure on a manifold. A typical result in this direction is quoted above as Theorem 8.1.

16. Sard's theorem.

Regular values of C^r maps are nicer than critical values. Recall Lemma 11.3 which says that the inverse image of a regular value is a submanifold. It turns out that regular values are dense in the range. The idea behind this is that critical points are places where the map is squashing the domain more than required to fit into the range. The image of such squashing cannot occupy much of the range. This is the content of Sard's theorem. It turns out to have many applications. It also turns out to be rather delicate to prove. We will prove a very special case to illustrate some of the ideas. We will mention an application of the full theorem in the next section.

The fact that it is delicate to prove is supported by the fact that it is false without the proper restrictions. There is a C^1 map from $\mathbf{R}^2$ to $\mathbf{R}$ whose set of critical values includes an interval. Thus the regular values cannot be dense in the range. In fact the map is quite strange. A critical point in a map from $\mathbf{R}^2$ to $\mathbf{R}^1$ can only be one at which the derivative is the zero linear map. That means that the tangent plane to the graph is horizontal. The map has the property that there is an arc of critical points in $\mathbf{R}^2$ whose image in $\mathbf{R}^1$ is an interval. Thus there is a path in the graph which rises in spite of the fact that there is a horizontal tangent to the graph at every point along the path.

To properly state Sard's theorem, we need some defintions. A cube of side a in $\mathbf{R}^n$ is a translate of $[0,a]^n = \{(x_1, \ldots, x_n) | 0 \leq x_i \leq a\}$. The volume of a cube of side a in $\mathbf{R}^n$ is defined to be a^n. We denote the volume of the cube C by $\mu(C)$. One can similarly define the volume of a rectangular solid. A set A in $\mathbf{R}^n$ is said to have measure 0 if, for every $\epsilon > 0$, it can be covered by a countable collection of cubes whose volumes sum to less than ϵ. Countable unions of sets of measure

0 have measure 0. Thus checking that a set has measure 0 can be done on small open sets. It is provable that an open set cannot have measure 0. Thus a set of measure 0 can contain no open set and thus has dense complement. It turns out that the regular values are more than just dense. A set is called *residual* if it is the intersection of a countable collection of dense open sets. The Baire category theorem (which applies to $\mathbf{R}^n$ since it is a complete metric space) says that a residual set is dense. However, there are dense sets (e.g., the rationals in $\mathbf{R}$) that are not residual.

We have only defined sets of measure 0 in $\mathbf{R}^n$. We define a set to have measure 0 in a manifold M if the intersection of the set with the domain of each coordinate map has its image under the coordinate map a set of measure 0. That this defintion makes some sense is supported by the next lemma.

LEMMA 16.1. *Let U be an open set in $\mathbf{R}^n$ and let $f: U \to \mathbf{R}^n$ be a C^1 map. If $X \subseteq U$ has measure 0, then so does $f(X)$.*

PROOF: Because f is C^1, $\|Df_x\|$ is bounded on compact sets. Thus on a ball B, we have a bound K for $\|Df_x\|$ and

$$\|f(x) - f(y)\| \leq K\|x - y\|$$

for any x and y in B. In a cube C of side a, the distances are bounded by $a\sqrt{n}$. Thus the distances in $f(C)$ are bounded by $aK\sqrt{n}$. Let $L = K\sqrt{n}$. We have that $f(C)$ is contained in a cube of side no more than aL with volume no more than $a^n L^n = L^n \mu(C)$.

Since X can be covered by countably many balls and contable unions of sets of measure 0 have measure 0, we need only prove the lemma for $X \cap B$. Now given $\epsilon > 0$, we can cover $X \cap B$ by cubes whose volumes add up to less than ϵ. Thus $f(X \cap B)$ can be covered by cubes whose volumes add up to less than $L^n \epsilon$. But L^n is fixed for this B and we can make the image sum as small as we like. This completes the proof.

The full statement if Sard's theorem is:

THEOREM 16.2 (SARD'S THEOREM). *Let M and N be manifolds of dimensions m and n repsectively and let $f: M \to N$ be a C^r map. If*

$$r > \max\{0, m - n\},$$

then the critical values have measure 0 in N and the regular values are residual in N.

Note that the example claimed above has $m = 2$, $n = 1$ and $r = 1$ which just misses the hypotheses of the theorem. There is no such example of a C^2 map from $\mathbf{R}^2$ to $\mathbf{R}$. The case where $r = \infty$ is easier than the full theorem and the proof in this case is found in many textbooks. It is also sufficient for most applications because approximation theorems (see Section 15) usually allow the assumption that all maps are C^∞. We will prove even less than the full C^∞ case. We will prove:

THEOREM 16.3 (VERY BABY SARD'S THEOREM). *Let $f : M \to N$ be a C^1 map between m-manifolds. Then the set of critical points has measure 0 in N.*

PROOF: A countable union of sets of measure 0 has measure 0 and both domain and range can be covered by countable collections of coordinate charts. Thus we assume that we are looking at a piece from a coordinate chart to a coordinate chart. From the lemma and the defintion, we can assume that we are looking at the map expressed in local coordinates. Thus we will assume that f is a C^1 map from an open set U of $\mathbf{R}^m$ into $\mathbf{R}^m$.

Let C be a cube of side a in U. Again by countable unions, it suffices to consider only the image of the critical points that lie in C.

We can divide C up into n^m cubes of side a/n. The idea of the proof is this. With a/n very small, a constant plus Df will be a very good approximation of f. But at a critical point, the image of Df will be a linear subspace of dimension no more than $m-1$. Thus a small cube of side a/n will have extent in the direction of this linear subspace that will be approximated by a/n and extent in the direction perpendicular to the subspace that will be approximated by $\epsilon a/n$ for very small ϵ. This will give that the image of the cube has a very small volume.

Let S be one of the small cubes of side a/n. We have $\|y-x\| \leq \sqrt{m}(a/n)$ for x, y in S. For n large enough, we can get

$$\|f(y) - f(x) - Df_x(y-x)\| < \epsilon\|y-x\| \leq \epsilon\sqrt{m}(a/n).$$

If S contains a critical point we can choose x to be a critical point. This makes the set of points $\{Df_x(y-x) | y \in S\}$ lie in a linear subspace V of dimension no more than $m-1$ in $\mathbf{R}^m$. Thus the set $\{f(y) - f(x) | y \in S\}$ lies within $\epsilon\sqrt{m}(a/n)$ of V so that $\{f(y) | y \in S\}$ lies within $\epsilon\sqrt{m}(a/n)$ of the translate $W = f(x) + V$. Now $\|Df\|$ is bounded by some K on the cube C. Thus

$$\|f(y) - f(x)\| \leq K\|y-x\| \leq K\sqrt{m}(a/n)$$

and we have that $f(y)$ lies within $K\sqrt{m}(a/n)$ of $f(x)$ and withing $\epsilon\sqrt{m}(a/n)$ of W. Thus $f(S)$ lies in a rectangular solid where $m-1$ of its dimensions are $2K\sqrt{m}(a/n)$ and one of its dimensions is $2\epsilon\sqrt{m}(a/n)$. The volume of S is $\mu(S) = (a/n)^m$ and the volume of $f(S)$ is no more than $\epsilon K^{m-1}(2\sqrt{m})^m(a/n)^m$ or $\epsilon K'\mu(S)$. Here K' depends on C and not on S. The sum of all $\mu(S)$ for the n^m small cubes in C is $\mu(C)$. The sum of the volumes of the $f(S)$ for those S that contain a critical point is thus no more than $\epsilon K'\mu(C)$. We can make ϵ as small as we like by increasing n. Thus the image of the critical points in C has measure 0.

17. Transversality.

None of the statements in this section will be proven.

Let $f : M \to N$ be a C^1 map and let $A \subseteq N$ be a submanifold. We say that f is *transverse* to A if for every x with $y = f(x) \in A$, the tangent space N_y of N

at y is spanned by A_y and $Df_x(M_x)$. In other words, $N_y = A_y + Df_x(M_x)$. This is written $f \pitchfork A$. We define the *codimension* of A in N to be the dimension of N minus the dimension of A.

Transversality generalizes the notion of submersion. In a submersion at a point, the tangent space in the domain must map to cover the tangent space in the range. In a transverse map, the tangent space from the domain may not cover that in the range, but it does so with the help of the submanifold that it is transverse to. Note that transversality cannot take place if the dimensions of domain and submanifold are too small to add up to the dimension of the range. If they are big enough to add up, then transversality fails if the image is too "tangent" to the submanifold. Transversality says that this degree of tangency does not take place. The map $x \mapsto x^2$ is not transverse to the x-axis but it is transverse to the y-axis.

That transversality is a nice condition is seen by the following.

THEOREM 17.1. *Let $f : M \to N$ be a C^r map, $r \geq 1$, and $A \subseteq N$ a C^r submanifold. If f is transverse to A, then $f^{-1}(A)$ is a C^r submanifold of M and the codimension of $f^{-1}(A)$ in M is that of A in N.*

This is not hard to show by reducing the theorem locally to a question about regular values.

Niceness is nice and availability is better. The following is a version of the main result about transversality. As in previous sections we are not careful about exactly which C^r topology is being used on the space of functions.

THEOREM 17.2. *Let M and N be C^r manifolds and A a C^r submanifold of N, $r \geq 1$. Let $C^r(M,N)$ be the space of C^r maps from M to N with the C^r topology.*

(1) *The maps that are transverse to A are residual in $C^r(M,N)$.*
(2) *If M is compact and A is a closed subset of N, then the maps that are transverse to A are also open in $C^r(M,N)$.*

The theorem is proven with the help of Sard's theorem and various of the techniques discussed in the other sections.

18. Manifolds with boundary.

This section is even sketchier. We prove nothing and define nothing.

The manifolds that we have considered have been modeled on Euclidean spaces. The manifolds have had no boundary since each point has to have a neighborhood homeomorphic to an open subset of some $\mathbf{R}^m$. To achieve boundary we have to allow homeomorphisms to open subsets of $\mathbf{R}^m_+$ the upper half space

$$\{(x_1, \ldots, x_m) | x_m \geq 0\}.$$

Various notions have to be redifined to take the new structures into account. Submanifolds with boundary of a given manifold will intersect (if their boundaries are transverse) in subspaces that are not even modeled on $\mathbf{R}^m_+$. They will have corners. A technique for rounding corners can be developed so as to avoid building up even more variety into the structures.