Differential Topology

Notes for a course given in the Spring of 1999

Matthew G. Brin

Author address:

Binghamton University, Binghamton, NY 13902-6000, USA

Contents

Chapter 0. Introduction 5

Chapter 1. Calculus on Euclidean Spaces 11

1.1. Derivatives in Euclidean spaces 12

1.2. The Chain Rule 15

1.3. Derivatives and Products 17

1.4. Higher Derivatives 19

1.5. The Inverse Function Theorem 22

Chapter 2. Differentiable Manifolds 33

2.1. Definition 33

2.2. Tangent Vectors and Derivatives 40

2.3. The Tangent Space of a Differentiable Manifold 46

2.4. The Tangent Bundle 50

2.5. Inverse Function Theorem and Consequences 52

2.6. Submanifolds 58

Chapter 3. Extra topics 65

CHAPTER 0

Introduction

This course will describe a category. That is, we will describe the objects of the category and the morphisms between them. The objects will be differentiable manifolds and the morphisms will be differentiable maps. We can refer to the category as DIFF.

There are at least two reasons for the importance of this category. The first reason is that differentiable manifolds and differentiable maps appear in many other subjects. Much of physics is based on the idea that we live in a differentiable manifold. If you consider space as viewed by Newton, the number of dimensions is three. If you take time into account, the number of dimensions is four. Keep this in mind when you read the last comment in this introduction. Of course, more modern theories of physics have us living in a manifolds of even higher dimension, where the number of dimensions and the total structure are far from fully understood.

The second reason for the importance of the category is that there is a rich set of tools for analyzing its objects and morphisms. Most of the time will be spent in developing some of the tools. There will be only a small amount of time devoted to results that use the tools. Differential topology is a large subject and just learning some of the tools is a good start. Besides, the tools are useful in other subjects as well.

These notes are based on a set of notes written in 1994 for a half semester introduction to the subject. Those notes were to be presented in the class by the students. These will also be presented in class by the students, but there will be additional material to be presented by the students that is not in these notes. The extra material will be identified when it is used.

The additional material will be from the following two books

1. John W. Milnor, *Topology from the differentiable viewpoint*, Princeton, 1965.
2. Th. Bröcker and K. Jänich, *Introduction to differential topology*, Cambridge, 1982 (English Edition), out of print.

The material within these notes was gathered from several sources. These include the following.

3. Serge Lang, *Differential manifolds*, Addison Wesley, 1972.
4. Morris W. Hirsch, *Differential topology*, Springer-Verlag, 1976.
5. Michael Spivak, *Calculus on manifolds*, Benjamin, 1965.
6. James R. Munkres, Elementary differential topology, Princeton, 1966.

Another book on the subject that we did not use is

7. Andrew Wallace, *Differential topology: first steps*, Benjamin, 1968.

The only purpose of the rest of this introduction is to whet the appetite by pointing out some differences between the differentiable world and the topological world. New terms are presented on an informal basis, and definitions are given for clarity and not rigor. Formal introductions will occur later in these notes and all definitions given here will be repeated or expanded later.

A differentiable manifold M will be a topological space with extra restrictions.

1. M must be a separable metric space.
2. There must be an integer $n > 0$ so that every $x \in M$ has a neighborhood homeomorphic to $\mathbf{R}^n$.
3. There is a restriction to be given later on how the neighborhoods overlap.

The third restriction will allow us to say in a well defined way when a map between differentiable manifolds is differentiable. The differentiable maps will be continuous functions that also satisfy extra restrictions. The category DIFF will have the differentiable manifolds and objects, and the diffentiable maps as morphisms.

A topological space satisfying (1) and (2) of the previous paragraph is a topological manifold. Thus every differentiable manifold is a topological manifold. We can define a category whose objects are topological manifolds, and whose morphisms are just the continuous functions between them. We can refer to this category as TOP.

We have a forgetful functor from DIFF to TOP. That is, if M is an object in DIFF, then we can forget the third restriction in the definition of differentiable manifold and regard M as an object in TOP. If f is a morphism in DIFF, then we can forget the extra restrictions that make it differentiable and regard it as a morphism in TOP.

Questions about differentiable manifolds may have different answers when looked at in the DIFF than when looked at in the TOP. Obviously, objects that are isomorphic in DIFF will be isomorphic in TOP. However, the converse is false. We will not have time to give proofs of this in the course, but we will mention some interesting examples at the end of this introduction.

There are other examples of the differences between DIFF and TOP. Some are quite simple once the machinery is developed, and others are more delicate. These other examples do not refer to isomorphisms but to other concepts. Let $f_0 : X \to Y$ and $f_1 : X \to Y$ be two embeddings of one topological space into another. In the following definitions I denotes the unit interval $[0, 1]$. We say that f_0 is *isotopic* to f_1 if there is a map $F : X \times I \to Y$ so that if $F_t : X \to Y$ is deinfed by $F_t(x) = F(x, t)$, then $F_0 = f_0$, $F_1 = f_1$ and for all $t \in I$, F_t is

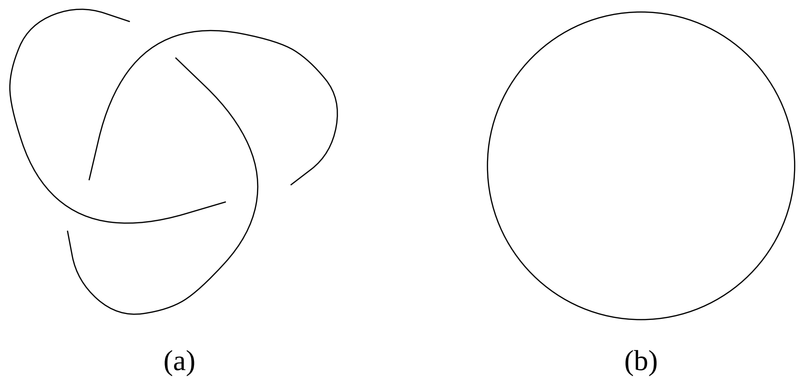

FIGURE 0.1. Two knots in 3-space

an embedding. We say that f_0 is *concordant* to f_1 if there is an embedding $F : X \times I \to Y \times I$ so that for each $x \in X$ we have $F(x, 0) = (f_0(x), 0)$ and $F(x, 1) = (f_1(x), 1)$. Note that isotopic implies concordant implies homotopic.

If all spaces and maps in the previous paragraph are in TOP, then we have the definitions of TOP isotopies and TOP concordances. If all maps and spaces are in DIFF, then we have DIFF isotopies and concordances.

Examples of maps that are TOP isotopic or TOP concordant but are not DIFF isotopic or DIFF concordant are easy to describe. Proving the claimed properties is a different matter. We give a few examples. In the examples, we use D^2 to denote the unit disk in $\mathbf{R}^2$ and S^1 to denote its boundary.

Figure 0.1 shows two knots in $\mathbf{R}^3$. That is, they show two embeddings of S^1 into $\mathbf{R}^3$, and we assume that the embeddings are in DIFF. The two embddings are TOP isotopic but not DIFF isotopic. We leave the TOP isotopy as an exercise. The lack of a DIFF isotopy is an application of a theorem in DIFF and a calculation involving the fundamental group.

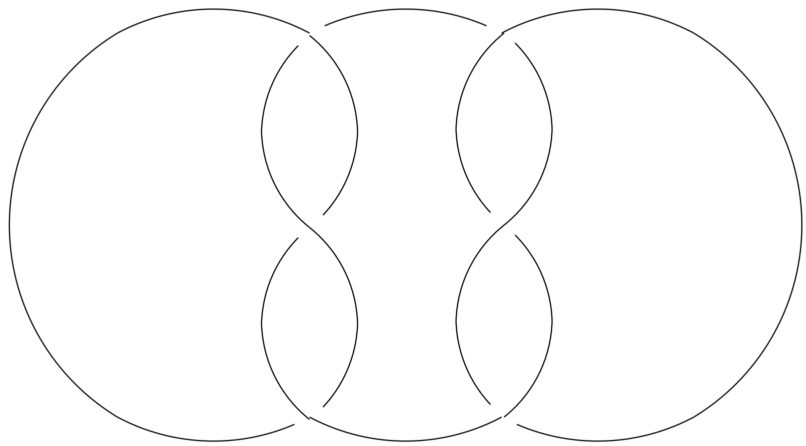

FIGURE 0.2. The square knot

Concordances can be harder to work with. Clearly the knots in Figure 0.1 are TOP concordant. The fact that they are not DIFF concordant is harder to show than the fact that they are not DIFF isotopic. We leave as an exercise the fact that the knot in Figure 0.2 is DIFF concordant to the knot in Figure 0.1(b).

There are examples that are even more subtle. In Figure 0.3, we show three solid tori. Each solid torus is homeomorphic to $D^2 \times S^1$. Each solid torus has an embedded image of S^1. We can assume that the embeddings are in DIFF. We leave as an exercise the fact that embeddings (a) and (b) in Figure 0.3 are DIFF concordant. Building the concordance with a finite number of differentiable pieces will be enough to argue that the entire concordance is DIFF. Of course, this will also show that embeddings (a) and (b) are TOP concordant. It turns out that embedding (c) is not DIFF concordant to the other two. The argument is similar to the one that shows that the two knots in Figure 0.1 are not DIFF concordant. What is surprising is the fact that all three embeddings in Figure

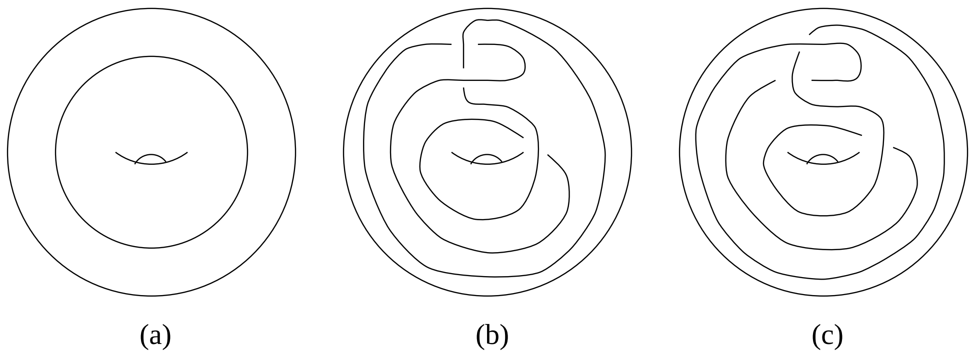

FIGURE 0.3. Three knots in solid tori

0.3 are TOP concordant. The argument is short and elementary but extremely tricky, and is due to C. Giffen. We leave it for now as a challenging exercise.

We should have time in this course to give arguments that the examples given above have the claimed properties. We will not have time to prove the following remarkable statements. (1) There are 28 objects in DIFF with no two isomorphic in DIFF, but whose images in TOP under the forgetful functor are all the 7-dimensional sphere. (2) There are uncountably many objects in DIFF with no two isomorphic in DIFF, but whose images in TOP under the forgetful functor are all $\mathbf{R}^4$.

CHAPTER 1

Calculus on Euclidean Spaces

Manifolds start as spaces based on Euclidean spaces. Every point in a manifold has a neighborhood that is homeomorphic to some Euclidean space. The notion of a differentiable manifold and the notion of a differentiable map between differentiable manifolds will be based on our understanding of differentiation in Euclidean spaces. This chapter is designed to lay out the needed material on differentiation in Euclidean spaces.

Much of the material below can be put in more general settings, such as Hilbert spaces or Banach spaces. There are technical difficulties, however. The presenter might want to see how much can be done in these setting. All of the material goes through in finite dimensional, normed vector spaces over $\mathbf{R}$. This can be exploited in Section 1.4 to avoid coordinates and partial derivatives. We will point out there how that goes, but we do not use it as the main approach. Had we done so, we would have avoided all connection to the derivatives familiar from calculus courses.

In the following, we make use of the norm $\|v\|$ of a vector in $\mathbf{R}^n$. This is length of v and is the square root of the sum of the squares of the coordinates of v. Were we to work with some arbitrary vector space V over $\mathbf{R}$, we would simply hypothesize a norm $\|\ \| : V \to \mathbf{R}$ with the usual properties that one assumes in analysis for a normed, linear space. If V is finite dimensional, then all norms induce the same topology, and the resulting metric space is complete.

1.1. Derivatives 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

$$\lim_{h \to 0} \frac{\|f(x+h) - f(x) - \lambda(h)\|}{\|h\|} = 0. \tag{1.1}$$

The quotients make sense since the denominators are real numbers. 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.

If f, x and λ satisfy (1.1), then we say that the derivative of f at x is λ and denote it Df_x. We will say that f is differentiable on all of U if f is differentiable on each $x \in U$. Note that f satisfies this vacuously if U is empty. This will be convenient later. It is important to remember that Df_x is a linear function. We can summarize the idea in (1.1) by saying that the derivative of f at x_0 is the "best linear approximation to $(f(x) - f(x_0))$."

Note that the "domain" of the limit in (1.1) 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,

$$\|f(x+h) - f(x) - \lambda(h)\| < \epsilon\|h\|. \tag{1.2}$$

Since $\lim_{h \to 0} \lambda(h) = 0$, the following is an immediate consequence of (1.2).

LEMMA 1.1.1. *If f is differentiable at x, then it is continuous at x.*

We also get the uniqueness of the derivative from (1.2).

PROPOSITION 1.1.2. *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 $\lambda_1 = \lambda_2$. ∎

We give two easily computed derivatives.

LEMMA 1.1.3. *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 1.1.4. *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 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 $D(rf)_x = rDf_x$.

1.2. The Chain Rule

The next result, the chain rule, is one of the crucial ingredients. 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 1.2.1 (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 $gf : U \to \mathbf{R}^p$ is differentiable at a and*

$$D(gf)_a = (Dg_b)(Df_a).$$

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

$$\|g\big(f(x) + \lambda(h) + E(h)\big) - g(f(x)) - \mu(\lambda(h) + E(h))\| < \frac{\epsilon}{2}\|h\| \tag{1.3}$$

and

$$\|\mu(E(h))\| < \frac{\epsilon}{2}\|h\|. \tag{1.4}$$

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

$$\|\lambda(h) + E(h)\| < \delta_1. \tag{1.5}$$

Now

$$\begin{aligned}
\|\lambda(h) + E(h)\| &\le \|\lambda(h)\| + \|E(h)\| \\
&< \|\lambda\| \cdot \|h\| + \|\epsilon_2\|h\| \qquad (1.6) \\
&= \big(\|\lambda\| + \epsilon_2\big)\|h\|
\end{aligned}$$

for

$$\|h\| < \delta_2 \tag{1.7}$$

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

$$\epsilon_1 < \frac{\epsilon}{4}, \qquad \epsilon_1 < \frac{\epsilon}{4\|\lambda\|}, \qquad \epsilon_2 < 1 \tag{1.8}$$

hold. Thus we get (1.3) if we can satisfy all of (1.8). Now

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

if

$$\epsilon_2 < \frac{\epsilon}{2\|\mu\|}. \tag{1.9}$$

Thus we get (1.4) if we can satisfy (1.9).

So given ϵ, we determine ϵ_1 and ϵ_2 from (1.8) and (1.9). This determines δ_1 and δ_2 which puts our first restriction $\delta \leq \delta_2$ on δ because of (1.7). We must deal with (1.5). But we can get this from (1.6) by putting the resriction

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

on δ. This finishes the proof. ∎

1.3. Derivatives and Products

We give a lemma that we will use to relate two notions of derivative that we will define. 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 1.3.1. *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

$$\begin{aligned} &\|(f \times g)(a + h_1, b + h_2) - (f \times g)(a, b) - (Df_a \times Dg_b)(h_1, h_2)\| \\ =&\|\big(f(a + h_1), g(b + h_2)\big) - \big(f(a), g(b)\big) - \big(Df_a(h_1), Dg_b(h_2)\big)\| \qquad (1.10)\\ =&\|\big(f(a + h_1) - f(a) - Df_a(h_1)\ ,\ g(b + h_2) - g(b) - Dg_b(h_2)\big)\|. \end{aligned}$$

The i-th coordinate, $i = 1, 2$, in (1.10) 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 (1.10) 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 $(f, h) = (f \times h)d$. Now $D(f, h) = D(f \times h)Dd = (Df \times Dh)d = (Df, Dh)$. ∎

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 $f = (f_1, \dots, f_n)$. By Lemma 1.3.1, we have $Df = (Df_1, \dots, Df_n)$. 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,

$$\begin{aligned} f'(x) &= \big(f_1'(x), \dots, f_n'(x)\big) \\ &= \big(D(f_1)_x(1), \dots, D(f_n)_x(1)\big) \\ &= Df_x(1). \end{aligned} \tag{1.11}$$

This equality will be convenient later on.

1.4. Higher Derivatives

Thus far, there has been little mention of coordinates or partial derivatives. In conjunction, there has been no mention of bases when we talk about linear maps between the vector spaces $\mathbf{R}^m$ and $\mathbf{R}^n$. We break this silence for one section and bring in partial derivatives and specific bases to discuss higher derivatives. This can be avoided, at the expense of never relating this material to more familiar material. The way to avoid partial derivatives and specific bases is given at the end of the section.

Let $f : \mathbf{R}^m \to \mathbf{R}^n$ be differentiable at x. Then the derivative Df_x at x is a linear map from $\mathbf{R}^m$ to $\mathbf{R}^n$. 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. We can bring them in to make the set of linear transformations from $\mathbf{R}^m$ to $\mathbf{R}^n$ look more familiar. We will also need to choose bases for $\mathbf{R}^m$ and $\mathbf{R}^n$.

Let the standard bases be chosen for $\mathbf{R}^m$ and $\mathbf{R}^n$ (unit vectors in the coordinate directions) and let elements of $\mathbf{R}^m$ and $\mathbf{R}^n$ be represented by column vectors. Now a linear transformation from $\mathbf{R}^m$ to $\mathbf{R}^n$ is represented as left multiplication 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 Jacobian matrix whose entries are

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

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 we say that *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 say 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 necessarily 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 if "equality of mixed partials" is satisfied) 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.

It is an elementary exercise that the composition of C^r functions is C^r for a fixed r. The (local) inverses of C^r functions are discussed in the next section.

We say that a function is of class C^∞ (or *smooth*) if it is of class C^r for all r. The composition of smooth functions is also smooth. Since smooth functions tend to be as well behaved as C^r functions, we will often make statements about C^r functions for $1 \leq r \leq \infty$.

[To avoid coordinates, and partial derivatives, we can use the norm of a linear function between normed linear spaces as defined in Section 1.2. Let V and W be normed linear spaces. Now for a differentiable $f : V \to W$, we have that Df is a function from V to $\mathrm{Hom}(V, W)$, the linear functions from V to W. Since $\mathrm{Hom}(V, W)$ is also a normed linear space and thus a metric space, the continuity of Df can be discussed. Further, the differentiability of $Df : V \to \mathrm{Hom}(V, W)$

can be discussed, and $D^2 f$ can be defined. Now the classes C^r, $1, \leq r \leq \infty$, can be defined.]

1.5. The Inverse Function Theorem

Hirsh makes the following comments. A derivative gives information about infinitesimal behavior. It tells about the limiting behavior near a point. The next step up from limiting behavior is local behavior—behavior on all of some open set about a point. The theorem in this section is the basis of several results that make the passage from infinitesimal to local. That is, they take information about infinitesimal behavior and make conclusions about local behavior. This transition is one of the key tools in differential topology. The transition from local to global is another goal of all branches of topology and will be illustrated elsewhere.

A rule of thumb is that differentiable functions with invertible derivatives are locally invertible. We will want a statement to that effect for manifolds, but that will follow with very few words from such a statement for Euclidean spaces. We start with the statement. The statement for manifolds will be identical except that M and N will represent differentiable manifolds.

THEOREM 1.5.1 (Inverse Function Theorem on Euclidean spaces). *Assume that $f : M \to N$ is a C^r function, $1 \leq r \leq \infty$, between open subsets of Euclidean spaces, 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 $D\big((f|_U)^{-1}\big)_z = (Df_x)^{-1}$.*

We will unroll the proof of the Inverse Function Theorem very slowly. Various intermediate results are stated and proven in the middle of the proof of the Inverse Function Theorem 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.

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. We give a version of the Mean Value Theorem in higher dimensions.

THEOREM 1.5.2 (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. ∎

We will start the proof of the Inverse Function Theorem by 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 the Mean Value Theorem.

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.

We have that f is a function from an open set in $\mathbf{R}^m$ to $\mathbf{R}^m$. If we compose f on the right or left with linear isomorphisms, then the point of interest and its image will shift, but the hypotheses will remain the same. This is because linear isomorphisms are infinitely differentiable (so the class of the function is not altered), and because the derivative of a linear isomorphism is a linear isomorphism (so the invertibility of the derivative is not lost).

By composing on the left and right with translations, we can assume that f takes 0 to 0 and that we are concerned with the behavior of f at 0. We can do this since translations are linear isomorphisms. Now we have that Df_0 is a linear isomorphism from $\mathbf{R}^m$ to $\mathbf{R}^m$. We can compose f with the inverse of this linear isomorphism and we now have that f carries the origin to the origin and Df_0 is the identity.

Thus 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

$$\|f(x) - f(y)\| \geq \frac{1}{2}\|x - y\|. \tag{1.12}$$

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

$$\|g(x) - g(y)\| < \frac{1}{2}\|x - y\|. \tag{1.13}$$

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 (1.12).

Our task is now to show (1.13). 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 $\mu(Dg \times 1)$ 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 1.5.3. *Let ABC be an isosceles triangle in $\mathbf{R}^m$ so that $\rho(A,B)$ equals $\rho(A,C)$ and so that $B \neq C$. 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 1.5.4. *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 1.5.5. *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

$$\|f(x+th) - f(x) - Df_x(th)\|. \tag{1.14}$$

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 (1.14) 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 (1.12) 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 DF, we must look at

$$\text{(1.15)} \quad \|F(w) - F(z) - (Df_x)^{-1}(w - z)\| \;=\; \|y - x - (Df_x)^{-1}(f(y) - f(x)\|.$$

The expression inside the norm in (1.15) is obtained from the expression inside the norm of (1.16) by applying $(Df_x)^{-1}$. Thus if $K = \|(Df_x)^{-1}\|$, then (1.15) is no greater than

$$\text{(1.16)} \qquad K\|Df_x(y - x) - f(y) + f(x)\| \;=\; K\|f(y) - f(x) - Df_x(y - x)\|.$$

Now (1.16) can be kept less than $(\epsilon/2)\|y - x\|$ for a given $\epsilon > 0$ by keeping $\|y - x\|$ suitably small. We want our original (1.15) (which is no greater than (1.16)) smaller than $\epsilon\|w - z\|$. But another application of (1.12) 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 (1.12) 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 $i(Df)F$ 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

$$DF = i(Df)F \tag{1.17}$$

is continuous and F is C^1. But now if f is C^2, then all the functions on the right side of (1.17) have continuous derivatives and F is C^2. Further, the derivative of both sides of (1.17) and the chain rule give D^2F as a composition involving DF, Di and D^2f. But (1.17) can be used again to replace DF in the composition with the right side of (1.17) 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 (fi_v) and (fj_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

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

and

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

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

$$(fi_v)(a) = (Df_{(u,v)})(i_0)(a), \text{ or}$$
$$f(a,v) = Df_{(u,v)}(a,0),$$

and

$$(fj_u)(b) = (Df_{(u,v)})(j_0)(b), \text{ or}$$
$$f(u,b) = Df_{(u,v)}(0,b).$$

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 $(f \cdot g) = d(f \times g)$. Now $D(f \cdot g) = Dd(Df \times Dg)$. More specifically

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

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.]

There are three standard consequences of the Inverse Function Theorem: the Immersion Theorem, the Submersion Theorem and the Implicit Function Theorem. Their proofs from the Inverse Function Theorem are not long. Further, it is actually a disadvantage to have to state them separately for Euclidean spaces before stating them for differentiable manifolds. Thus we will put off these three theorems until we define differentiable manifolds.

CHAPTER 2

Differentiable Manifolds

2.1. Definition

A differentiable manifold M of class C^r and dimension n (or a C^r n-manifold), with $n \geq 1$, will be a separable metric space together with a set A (called an *atlas*) of homeomorphisms (called *charts*) that satisfy the following properties.

1. Each θ in A has domain an open set in M and image an open set in $\mathbf{R}^n$.
2. The domains of the elements of A form an open cover of M.
3. For each pair (θ, ϕ) of elements in A with domains U and V respectively, the map

   $$\phi\theta^{-1} : \theta(U \cap V) \to \phi(U \cap V)$$

   is differentiable of class C^r.
4. The set A is maximal with respect to the first three properties.

We say that A gives M a C^r *differential structure*. The definition needs some discussion.

If one ignores Property 3, then one has a definition of a manifold. The words "separable metric" are always thrown in to avoid certain examples that no one wants to call manifolds. The property that every point must have a neighborhood that is homeomorphic to an open subset of some Euclidean space easily implies that each point has a neighborhood that is homeomorphic to all of some Euclidean space. This is often referred to as "locally Euclidean." Thus a manifold is a separable, metric, locally Euclidean space. A locally Euclidean space that is not separable is the "long line." Locally Euclidean spaces that are not Hausdorff are easy to construct. (For example, this can be done by adding an extra point to the real line and putting in the right set of neighborhoods for the new point.) These examples are certainly not metric.

It turns out that Hausdorff is the main obstruction to being nice. Once a locally Euclidean, Hausdorff space is assumed, then there is a long list of properties that turn out to be equivalent. This list includes: (a) metrisable, (b) paracompact, (c) Lindelöf, (d) second countable and (e) sigma compact. However, the list does not include separable which is a consequence of (a)–(e) for locally Euclidean, Hausdorff spaces, but which does not imply (a)–(e).

We will see later, that paracompact is an extremely relevant property. So "paracompact, Hausdorff" might be a good pair of words to use with locally Euclidean if the properties that are useful are to be emphasized. Many texts include "Hausdorff" in the definition of "paracompact." Thus in theory, "paracompact and locally Euclidean" is enough to define a manifold. However, it is not good to rely on defintions that might not be assumed as standard.

Metrisable is equivalent to paracompact for locally Euclidean, Hausdorff spaces, so "metric, Hausdorff" might also be a good pair from the point of view of useful properties. Since metric implies Hausdorff (no assumed definitions needed here), this is also redundant and metric and locally Euclidean is enough. This makes the pair of properties "separable, metric" an odd choice to go with locally Euclidean. "Separable, metric" is used since each one of the pair is used to rule out a specific example.

All of that being said, we point out that proving that a space is a manifold does not always come down to proving the defining properties directly. Proving metrisability directly, or proving paracompact directly is not easy. The Urysohn metrisation theorem says that a regular (T_3) space that is second countable is metrisable. One can show that locally Euclidean and Hausdorff imply regular. From the point of view of easily provable properties that imply that a space is a manifold, the pair of properties "Hausdorff, second countable" could be used with locally Euclidean. In spite of this, few say it that way. (There is at least one text that does.) The pair "separable, metric" is there by way of popularity.

We now look at the property that makes a manifold differentiable. Property 3 holds for all pairs of charts. Since (ϕ, θ) is as valid a pair as (θ, ϕ), it follows that

$$\theta\phi^{-1} : \phi(U \cap V) \to \theta(U \cap V)$$

must also be differentiable of class C^r. The functions $\theta\phi^{-1}$ and $\phi\theta^{-1}$ are called *coordinate change functions* or sometimes *overlap maps*. Property 3 demands that all overlap maps are C^r. All overlap maps are homeomorphisms from their domains to their images.

Since each $\theta \in A$ is a homeomorphism, the open subsets of its domain are known once θ is understood. Also, since the domains of the elements of A form an open cover of M, the topologies on the domains of elements of A determine the topology of M. Thus the topology of M is fixed once the elements of A are understood. We will exploit this later in the following fashion. If M is given as only a set and not as a topological space, and a set of charts A is given whose domains cover M and whose overlap maps are at least continuous, then the collection of sets

$$\{\theta^{-1}(U) \mid \theta \in A,\ U \text{ open in image } \theta\}$$

is a basis for a topology on M.

We take the point of view that part of the data of a function is its domain. Thus it is superfluous to specify the domain of a function. However, it is sometimes very useful to have notation for the domain of a function introduced simultaneously with the notation for a function. Thus we will often (and many authors always) give a chart as an ordered pair (ϕ, U) where ϕ is the homeomorphism and U is its domain.

Property 4 raises the question of how much information is needed to specify an atlas, and even whether an atlas can be completely specified in the first place. Also, the entire definition needs motivation. We will deal with the motiviation before dealing with Property 4.

Let $f : M \to N$ be a function between C^r manifolds. The dimensions are not important, but we will assume that M and N have dimensions m and n, respectively. Let x be in M and let (θ_1, U_1) and (θ_2, U_2) be two charts in the atlas on M whose domains contain x. Let (ϕ_1, V_1) and (ϕ_2, V_2) be two charts in the atlas on N whose domains contain $f(x)$. An easy combination of the chain rule and Property 3 shows that if

$$\phi_1 f \theta_1^{-1} : \theta_1(U_1) \to \phi_1(V_1)$$

is C^r at $\theta_1(x)$, then so is

$$\phi_2 f \theta_2^{-1} : \theta_2(U_2) \to \phi_2(V_2).$$

Thus Property 3 makes the following well defined. We say that an f as given above is differentiable of class C^r at x if, for some chart (θ, U) on M with $x \in U$ and some chart (ϕ, V) on N with $f(x) \in V$, we have that $\phi f \theta^{-1}$ is differentiable of class C^r at $\theta(x)$. We now declare that f is C^r (or differentiable of class C^r) if it is C^r at every $x \in M$.

The compositions of the form $\phi f \theta^{-1}$ are very important. The homeomorphism θ associates a unique m-tuple of real coordinates to each point in U and the homeomorphim ϕ associates a unique n-tuple of real coordinates to each point in V. The composition $\phi f \theta^{-1}$ replaces f in a neighborhood of x by a function from the coordinate system given by θ to U to the coordinate system given by ϕ to V. We say that the composition $\phi f \theta^{-1}$ is an *expression (or representation)* *of f near (or about) x in local coordinates.* Associated to this expression of f in local coordinates is the diagram

$$\begin{array}{ccc} U & \xrightarrow{f} & V \\ {\scriptstyle\theta}\downarrow & & \downarrow{\scriptstyle\phi} \\ \theta(U) & \xrightarrow{\phi f \theta^{-1}} & \phi(V) \end{array} \tag{2.1}$$

which commutes by definition. As trivial as this is, it is useful to keep in mind when derivatives are discussed later.

Note that there are many expressions in local coordinates for a given f and x. In particular, even the identity map from a manifold M to itself has many expressions in local coordinates. If two charts θ and ϕ have x in their domains, then each of $\phi\theta^{-1}$ and $\theta\phi^{-1}$ is an expression of f in local coordinates near x. Note that these are just overlap maps. Thus expressions in local coordinates for identity maps are just overlap maps. Any two expression of a given f at one x in local coordinates can be gotten from one another by composing on the right and left by overlap maps.

This leaves Property 4. Let us call a set A of homeomorphisms satisfying Properties 1, 2 and 3 a *partial atlas* on M. If ρ is a homeomorphism from some open set W in M to some open subset in R^n, then we say that ρ is *compatible* with A if $A \cup \{\rho\}$ is also a partial atlas on M. The following are easy consequences of the chain rule. (1) If $A_1 \subseteq A_2$ are two partial atlases on M, then ρ is compatible with A_1 if and only if it is compatible with A_2. (2) If A_1 and A_2 are two atlases containing a partial atlas in common, then they are equal. (3) Every partial atlas A is contained in a unique atlas, and this atlas consists of all charts that are compatible with A.

The last observation makes Property 4 seem superfluous. However, it is extremently convenient, and allows the elimination of much verbiage in arguments. We give an example of this.

Let M be a C^r n-manifold. Let x be a point in M. Let (θ, U) be a chart with $x \in U$. There is a subset W of U that maps to a round ball V of radius $\epsilon > 0$ centered at $\theta(x)$ in $\mathbf{R}^n$. There is a C^∞ homeomorphism from V to to $\mathbf{R}^n$ that takes $\theta(x)$ to 0. This can be done by a translation and a combination of tangent functions. Now a restriction of θ to W composed with this homeomorphism must also be in the atlas on M by Property 4. Thus for any $x \in M$ we can assume a chart that takes x to 0 and whose image is all of $\mathbf{R}^n$.

The unique atlas that contains a partial atlas can be referred to as the atlas generated by (or determined by) the partial atlas. If two partial atlases determine the same atlas, we can think of them as equivalent. Two partial atlases are equivalent if and only if every chart in one is compatible with the other and vice versa. This is equivlent to asking that the union of the two partial atlases be a partial atlas.

The chain rule gives that compositions of C^r maps between C^r manifolds are C^r. Further, the identity map from a C^r manifold to itself is C^r. (In that last sentence domain and range not only have the same underlying set, but the same differentiable structure. An example later in this section shows why that point is important.) Thus for each $r \geq 1$ there is a category whose objects are C^r manifolds and whose morphisms are C^r maps. An isomorphism in such a category is called a *diffeomorphism*. Note that a diffeomorphism is not just a C^r map that is invertible as a function. The inverse of a diffeomorphism must also be C^r. An example relevant to this point will be given in the next paragraph.

We give some standard examples of differentiable manifolds. Let $\mathbf{R}^n$ have a partial atlas consisting of the identity map from $\mathbf{R}^n$ to itself. The C^∞ atlas determined by this partial atlas gives $\mathbf{R}^n$ a differential structure that we can refer to as the usual C^∞ structure on $\mathbf{R}^n$. If $\mathbf{R} = \mathbf{R}^1$ is given the usual structure, then $x \mapsto x$ is a C^∞ diffeomorphism but $x \mapsto x^3$ is a C^∞ homeomorphism but not even a C^1 diffeomorphism.

Let S^{n-1} be the set of points in $\mathbf{R}^n$ for which the squares of the coordinates sum to one. There are $2n$ open hemispheres $S^{n-1}_{\pm i}$ in S^{n-1} where S^{n-1}_{+i} consists of all points in S^{n-1} with strictly positive i-th coordinate, and S^{n-1}_{-i} consists of all points in S^{n-1} with strictly negative i-th coordinate. There is an obvious projection from any $S^{n-1}_{\pm i}$ to the open unit disk in $\mathbf{R}^{n-1}$. This projection is a homeomorphism and it is not hard to show that this gives a C^∞ partial atlas on S^{n-1}. We will refer to the resulting structure as the usual C^∞ structure on S^{n-1}.

Let us now give a less standard example. Let $M = \mathbf{R}$ with the usual C^∞ structure on it determined by α, the identity on $M = \mathbf{R}$. Let $N = \mathbf{R}$ and let the only element of a partial atlas on N be the function $\beta : \mathbf{R} \to \mathbf{R}$ given by

$$\beta(x) = \begin{cases} x & x \leq 0 \\ 2x & x \geq 0. \end{cases}$$

Note that α and β cannot be in the same atlas.

Now α is a C^∞ function from M to M and from N to N, but is not C^1 from M to N nor from N to M. The function β is a C^∞ function from N to M, but not from M to N. The function β^{-1} is a C^∞ function from M to N but not from N to M.

Now that we have differential structures and differentiable maps between them, we should start differentiating these differentiable maps. However, this takes some doing. The following example shows why.

If we let $\gamma : \mathbf{R} \to \mathbf{R}$ be defined by $\gamma(x) = 2x$, then we have γ in the unique atlas on M that contains the partial atlas $\{\alpha\}$. The presence of γ makes it difficult to say what the derivative of a particular differentiable function is. The compositions $\alpha\alpha\alpha^{-1}$, $\alpha\alpha\gamma^{-1}$, $\gamma\alpha\alpha^{-1}$ and $\gamma\alpha\gamma^{-1}$ are all expressions of $\alpha : M \to M$ in local coordinates. However, they do not all have the same derivatives at a given point (say 0). We got a consistent criterion for saying whether a function was differentiable by making suitable restrictions (Property 3) on the charts of an atlas. We might get consistent values for the calculation of the derivative if we made more restrictions. It turns out that restrictions that are strong enough to work make it harder to build examples of manifolds. Thus it seems that even though we know what functions are differentiable, we don't know what their derivatives are. It turns out that there is a way to build the derivative of a function with the structures that we have defined. However, that takes another section.

2.2. Tangent Vectors and Derivatives

The problems raised at the end of the last section come from the fact that we are discussing a topic in topology and not a topic in geometry. The driving fact behind the examples given is that we have no absolute measure of length. Since a derivative is supposed to measure the stretch involved in going from domain to range, we find ourselves without the machinery to measure that stretch.

However, even though we cannot say much about one function, we can compare two functions. It turns out that we can say when two functions have the same local stretch coefficients at a point (have the same derivatives) without saying exactly what those coefficients are. Even further, we can say that the stretch of one function is twice that of the other. Even more importantly, we can say that the stretch of one function is zero times that of another. The last implies that we can tell when a derivative is zero or not, even while being unable to say what any of the non-zero derivates are.

The negatives of the last paragraph are exaggerated. The method of comparing derivatives leads to a very powerful machine that arranges derivatives into a linear space. We now give the details.

We will be comparing all functions against classes of reference functions. The reference functions will be functions with a one-dimensional domain. This will pick out one dimensional information about our manifold. The functions will be divided into classes since functions with the same derivative at a point will not be distinguishable in what we do.

We start with the real line $\mathbf{R}$ with its usual C^∞ structure. Let M be a C^r m-manifold. Let f and g be two C^r maps from $\mathbf{R}$ to M. We can write $f \sim g$ if $f(0) = g(0)$ and for some chart (θ, U) with $f(0) \in U$, we have $D(\theta f)_0 = D(\theta g)_0$. The chain rule implies that the second test is independent of the choice of chart. Since only equalities are used in the definition, the relation $\sim$ is clearly an equivalence relation. We let $[f]$ denote the equivalence class of f under this equivalence relation.

Note that $D(\theta f)_0$ is an element of $\mathbf{R}^m_* = \mathrm{Hom}(\mathbf{R}, \mathbf{R}^m)$, the vector space of linear maps from $\mathbf{R}$ to $\mathbf{R}^m$. For each $r \neq 0$ in $\mathbf{R}$, there is an evaluation isomorphism $e_r : \mathbf{R}^m_* \to \mathbf{R}^m$ given by $e_r(\lambda) = \lambda(r)$. It is convenient to let $r = 1$ and look at $D(\theta f)_0(1) \in \mathbf{R}^m$. By the remarks at the end of Section 1.3, this is just $(\theta f)'(0)$. Since e_1 is an isomorphism, we can define the equivalence relation $\sim$ by saying $f \sim g$ if and only if $f(0) = g(0)$ and $D(\theta f)_0(1) = D(\theta g)_0(1)$.

We use TU to denote the set of equivalence classes under $\sim$. We call the elements of TU *tangent vectors* or *tangent vectors of* U or *vectors tangent to* U. It is not convenient to carry around the brackets [ ] at all times, so we will use single letters for tangent vectors. This will result in abuse of notation when we write $v(0)$ and $D(\theta v)_0(1)$ for some $v \in TU$. However, the abused notation does not result in ambiguity of these values because of the definition of the equivalence relation.

For $x \in U$, we use TM_x to denote those tangent vectors $v \in TU$ for which $v(0) = x$. Note that the set TM_x is independent of the choice of chart that contains x in its domain. We call the elements of TM_x the *tangent vectors to* M *at* x. We let TM denote the union of all the TM_x as x ranges over M. We would like to understand more of the structure of TM. That will be done in the next section. Here we work to gain an understanding of TM_x and TU.

Since the equivalence relation $\sim$ is detected by looking at values, it is easy to relate the set of equivalence classes to the set of values. For $v \in TU$, let $\bar{\theta}v = ((\theta v)(0), D(\theta v)_0(1))$. We use $(\theta v)(0)$ instead of $v(0)$ since it is useful to have our values in standard spaces such as $\mathbf{R}^m$. If we define $\check{\theta}(v) = (\theta v)(0)$ and $\hat{\theta}(v) = D(\theta v)_0(1)$, then we have $\bar{\theta}(v) = (\check{\theta}(v), \hat{\theta}(v))$. That is, $\check{\theta}$ and $\hat{\theta}$ pick out the two values that define $\sim$. In particular, $TM_x = \check{\theta}^{-1}(\theta(x))$.

From the definitions, we have that $\bar{\theta}$ is a one-to-one correspondence onto its image. However, the image is in the set $\theta(U) \times \mathbf{R}^m$. We claim that $\bar{\theta}$ is a one-to-one correspondence between TU and all of $\theta(U) \times \mathbf{R}^m$. This will follow when we show that for any $x \in U$, the map $\hat{\theta} : TM_x \to \mathbf{R}^m$ is a surjection.

Given $u \in \mathbf{R}^m$, let $\alpha_u : \mathbf{R} \to \mathbf{R}^m$ be defined by $\alpha_u(t) = \theta(x) + tu$ and form $g_u = \theta^{-1}\alpha_u$. Now $g_u(0) = x$ and $\hat{\theta}[g_u] = D(\theta g_u)_0(1) = D(\alpha_u)_0(1) = u$. This makes $\hat{\theta}$ a surjection and thus also $\bar{\theta}$ is a surjection.

We can use $\hat{\theta}$ to induce a vector space structure on TM_x by setting

$$u + v = \hat{\theta}^{-1}(\hat{\theta}(u) + \hat{\theta}(v)), \quad \text{and}$$
$$rv = \hat{\theta}^{-1}(r\hat{\theta}(v)).$$

However the bijection $\hat{\theta}$ from TM_x to $\mathbf{R}^m$ depends heavily on θ. Thus if a chart ϕ with x in its domain is used, we might end up with a different structure. It is easy to show that we end up with the same structure if $\hat{\phi}\hat{\theta}^{-1} : \mathbf{R}^m \to \mathbf{R}^m$ is a linear map. Given $u \in \mathbf{R}^m$, we have already noted, that $\hat{\theta}[g_u] = u$ where $g_u = \theta^{-1}\alpha_u$ and $\alpha_u(t) = \theta(x) + tu$. Thus $[g_u] = \hat{\theta}^{-1}(u)$ and

$$\begin{aligned} \hat{\phi}\hat{\theta}^{-1}(u) &= \hat{\phi}(g_u) \\ &= D(\phi g_u)_0(1) \\ &= D(\phi\theta^{-1}\alpha_u)_0(1) \\ &= D(\phi\theta^{-1})_{\theta(x)} D(\alpha_u)_0(1) \\ &= D(\phi\theta^{-1})_{\theta(x)}(u) \end{aligned} \tag{2.2}$$

and $\hat{\phi}\hat{\theta}^{-1}$ is just the derivative of the overlap map $\phi\theta^{-1}$. However, the derivative is linear.

Thus we can declare TM_x to be a vector space. Every chart θ with x in its domain gives an isomorphism $\hat{\theta} : TM_x \to \mathbf{R}^m$ given by $\hat{\theta}(v) = D(\theta v)_0(1)$.

We now let N be a C^r n-manifold, and let $f : M \to N$ be a C^r map. For $v \in TM_x$ with $v = [\alpha]$ and $\alpha : \mathbf{R} \to M$ taking 0 to x, we let $\hat{f}(v) = [f\alpha] \in TN_{f(x)}$. Well definedness is covered in the next paragraph. Letting x run over all of M and v over all of TM gives a map from TM to TN. For any $x \in M$, we will use $\hat{f}_x$ to denote the restriction of $\hat{f}$ to TM_x.

The well definedness check will also give the tool needed to show that $\hat{f}_x : TM_x \to TN_{f(x)}$ is a homomorphism. Let θ and ϕ be charts for M and N, respectively, with x in domain θ and $f(x)$ in domain ϕ. If $\beta \in [\alpha]$, then $D(\theta\alpha)_0(1) = D(\theta\beta)_0(1)$ and

$$\begin{aligned} D(\phi f\alpha)_0(1) &= D(\phi f\theta^{-1})D(\theta\alpha)_0(1) \\ &= D(\phi f\theta^{-1})D(\theta\beta)_0(1) \qquad (2.3) \\ &= D(\phi f\beta)_0(1) \end{aligned}$$

which implies that $[f\alpha] = [f\beta]$. We now have a well defined function $\hat{f} : TM_x \to TN_{f(x)}$.

Note the use of the expression $\phi f\theta^{-1}$ of f in local coordinates. In fact, the first equality in (2.3) is a verification that

$$\begin{CD} TM_x @>{\hat{f}}>> TN_{f(x)} \\ @V{\hat{\theta}}VV @VV{\hat{\phi}}V \\ \mathbf{R}^m @>{D(\phi f\theta^{-1})_{\theta(x)}}>> \mathbf{R}^n \end{CD} \qquad (2.4)$$

commutes. Since the vertical arrows are isomorphisms and the bottom arrow is a homomorphism, we get that $\hat{f}$ is a "conjugate" of the linear $D(\phi f\theta^{-1})$ and is a homomorphism. Diagram (2.4) is a companion to Diagram (2.1). The two will be combined later.

In the rather special case where M and N represent open subsets of $\mathbf{R}^m$ and $\mathbf{R}^n$, respectively, and θ and ϕ are just inclusion maps, then (2.4) reduces to

$$\begin{array}{ccc} TM_x & \xrightarrow{\quad \hat{f} \quad} & TN_{f(x)} \\ \big\downarrow{\scriptstyle\prime} & & \big\downarrow{\scriptstyle\prime} \\ \mathbf{R}^m & \xrightarrow{\quad Df_x \quad} & \mathbf{R}^n \end{array}$$

where the vertical arrows are just the calculus derivatives of curves evaluated at 0. This makes $\hat{f}$ hard to distinguish from Df in this case.

In fact, what we are (temporarily) calling $\hat{f}$ is usually called Df and we will use that notation for it starting now. Even in the case where M and N are not open subsets of Euclidean spaces, (2.4) shows that the $\hat{f} = Df$ above is intimately tied to derivatives of functions between Euclidean spaces.

The chain rule now resembles the chain rule for derivatives of functions between Euclidean spaces. If $f : M \to N$ and $g : N \to Q$ are C^r, then we already know that gf is C^r. It is a triviality to verify that the chain rule now takes the usual form $D(gf)_x = (Dg)_{f(x)} Df_x$.

Now that we have $Df : TM \to TN$ for a differentiable $f : M \to N$, we would like to know more about the structure of TM and TN. This will be done in the next section.

The isomorphism $\hat{\theta}$ where θ is a chart, and Diagrams (2.1) and (2.4) are quite helpful in understanding some of the structures that we have defined. We leave as an exercise that if M and N are C^r manifolds with atlases A and B, respectively, then the set of all $(\theta \times \phi, U \times V)$, $(\theta, U) \in A$ and $(\phi, V) \in B$, forms a valid partial C^r atlas for $M \times N$. We can ask about the structure of $T(M \times N)_{(x,y)}$ for an $(x, y) \in M \times N$.

Let m and n be the dimensions of M and N, respectively. Consider the map $\langle x \rangle : N \to M \times N$ given by $\langle x \rangle(y) = (x, y)$. Let (θ, U) be a chart for M about x with $\theta(x) = 0$ and $\theta(U) = \mathbf{R}^m$ and (ϕ, V) be a chart for V about y with $\phi(y) = 0$

adn $\phi(V) = \mathbf{R}^n$. We get a chart $(\theta \times \phi, U \times V)$ for $M \times N$ about (x, y) and the following commutative diagrams from (2.1) and (2.4)

$$\begin{array}{ccc} N & \xrightarrow{\langle x \rangle} & M \times N \\ {\scriptstyle \phi}\downarrow & & \downarrow{\scriptstyle \theta\times\phi} \\ \mathbf{R}^n & \xrightarrow{h} & \mathbf{R}^m \times \mathbf{R}^n \end{array} \qquad \begin{array}{ccc} TN_y & \xrightarrow{D(\langle x \rangle)_y} & T(M \times N)_{(x,y)} \\ {\scriptstyle \hat{\phi}}\downarrow & & \downarrow{\scriptstyle \widehat{\theta\times\phi}} \\ \mathbf{R}^n & \xrightarrow{Dh_{\phi(y)}} & \mathbf{R}^m \times \mathbf{R}^n \end{array}$$

where h is the expression of $\langle x \rangle$ in local coordinates. Our choices have made the map h the inclusion of $\mathbf{R}^n$ into the second factor of $\mathbf{R}^m \times \mathbf{R}^n$. This is linear and its own derivative. Thus $Dh_{\phi(y)}$ has the same description and $Dh_{\phi(y)}\hat{\phi}$ is an isomorphism from TN_y to $\{0\} \times \mathbf{R}^n$.

Now $(\widehat{\theta \times \phi})D(\langle x \rangle)_y$ is an isomorphism from TN_y to $\{0\} \times \mathbf{R}^n$, so $\widehat{\theta \times \phi}$ must restrict to an isomorphism from the image of $D(\langle x \rangle)_y$ to $\{0\} \times \mathbf{R}^n$. The image of $D(\langle x \rangle)_y$ is clearly equal to the subspace of $T(M \times N)_{(x,y)}$ consisting of those $v \in T(M \times N)_{(x,y)}$ with a representative curve with image entirely in $\{x\} \times N$. Apologizing in advance, we will denote this subspace by $T(M \times N)_{N(x,y)}$. Thus $D(\langle x \rangle)_y$ is an isomorphism from TN_y to $T(M \times N)_{N(x,y)}$ and $\widehat{\theta \times \phi}$ restricts to an isomorphism from $T(M \times N)_{N(x,y)}$ to $\{0\} \times \mathbf{R}^n$.

We can invent parallel notations for the other factors in the products and without defining things, we have that $D(\langle y \rangle)_x$ is an isomorphism from TM_x to $T(M \times N)_{M(x,y)}$ and $\widehat{\theta \times \phi}$ restricts to an isomorphism from $T(M \times N)_{M(x,y)}$ to $\mathbf{R}^m \times \{0\}$.

We have proven the perhaps obvious fact that $T(M \times N)_{(x,y)}$ is the direct sum of $T(M \times N)_{N(x,y)}$ and $T(M \times N)_{M(x,y)}$ and that these subspaces are the images of $D(\langle x \rangle)_y$ and of $D(\langle y \rangle)_x$, respectively.

This will be used later when we need a notion of a partial derivative. Let $f : (M \times N) \to Q$ be a C^r map with Q of dimension q. We define $D_2 f_{(x,y)}$ by

$$D_2 f_{(x,y)} = D((f)\langle x \rangle)_y = (Df)_{(x,y)} D(\langle x \rangle)_y. \tag{2.5}$$

If g is an expression of f with respect to $(\theta \times \phi, U \times V)$ and some chart ρ with domain containing $z = f(x, y)$, then we get the diagram

$$\begin{CD} TN_y @>{D(\langle x \rangle)_y}>> T(M \times N)_{(x,y)} @>{Df_{(x,y)}}>> TQ_z \\ @V{\hat{\phi}}VV @VV{\widehat{\theta\times\phi}}V @VV{\hat{\rho}}V \\ \mathbf{R}^n @>{Dh_{\phi(y)}}>> \mathbf{R}^m \times \mathbf{R}^n @>{Dg_{(\theta(x),\phi(y))}}>> \mathbf{R}^q \end{CD} \tag{2.6}$$

from (2.4). A similar definition can be made for $D_1 f_{(x,y)}$.

2.3. The Tangent Space of a Differentiable Manifold

For a C^r m-manifold M, we want TM to be a differentiable manifold as well. It turns out that TM is only a C^{r-1} manifold, but we won't discover that until after we try to put some structure on it.

For each chart (θ, U) of M, we have a one-to-one correspondence $\bar{\theta} : TU \to \theta(U) \times \mathbf{R}^m$. Since M is the union of the TU as U varies over the domains of the charts, we have that the sets TU cover TM. The topology of $\theta(U) \times \mathbf{R}^m$ induces a topology on TU under the one-to-one correspondence $\bar{\theta}$. Also, each $\theta(U) \times \mathbf{R}^m$ is a subset of $\mathbf{R}^m \times \mathbf{R}^m$ since $\theta(U) \subseteq \mathbf{R}^m$. Thus it makes sense to declare each $\bar{\theta}$ a chart and see if we get a differentiable structure that way.

Thus given a chart (θ, U) we identify TU with $\theta(U) \times \mathbf{R}^m$ in both structure and topology. The notation is lacking since another chart may have U as domain. However, this lack will not be crippling in what follows.

As mentioned in Section 2.1, the declaration that $\bar{\theta}$ is a chart for each chart θ of M determines a topology on TM. Since a differentiable manifold must start as a separable metric space, we should look at the topology that we get first. However, there is useful information in the overlap maps, so we look at them first instead.

Let (θ, U) and (ϕ, V) be two charts on M. From the definitions of TU and TV, we have that $(TU) \cap (TV) = T(U \cap V)$. Thus the overlap map that we wish to consider is

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

For an $x \in U \cap V$ and $v \in TM_x$, we have $\bar{\theta}(v) = (\check{\theta}(v), \hat{\theta}(v))$ and $\bar{\phi}(v) = (\check{\phi}(v), \hat{\phi}(v))$. Thus the action of $\bar{\phi}\bar{\theta}^{-1}$ is that of $\check{\phi}\check{\theta}^{-1}$ on the first coordinate and is that of $\hat{\phi}\hat{\theta}^{-1}$ on the second. However, $(\bar{\phi}\bar{\theta})(\theta(v(0))) = \phi(v(0))$ and the action on the first coordinate is just that of the overlap map $\phi\theta^{-1}$. The calculations in (2.2) show that $\hat{\phi}\hat{\theta}^{-1}$ is just the derivative of the overlap map $\phi\theta^{-1}$. Thus we have

$$\bar{\phi}\bar{\theta}^{-1}(x, y) = \left((\phi\theta^{-1})(x), D(\phi\theta^{-1})_x(y)\right).$$

To simplify notation, let $h = \phi\theta^{-1}$ and write $\bar{h}(x, y) = (h(x), Dh_x(y))$.

Since M is C^r with $r \geq 1$, we have that Dh_x varies continuously with x in $\operatorname{Hom}(\mathbf{R}^m, \mathbf{R}^m)$. This makes $\bar{h}(x, y)$ vary continuously with x and y. As mentioned in Section 2.1, the continuity of the overlap maps shows that the collection of open sets in all the TU with (θ, U) a chart of M is a basis for a topology on TM. We now look at this topology.

Since M is separable, there is a countable dense subset S of M. For each $x \in S$ we can take a countable dense subset D_x of TM_x. The union D of the D_x as x ranges over S is countable. Each open set in TM contains an open set V in some TU whose projection to U is open and contains some $x \in S$. Now the intersection of V with TM_x is open in TM_x and contains some point in D_x. Thus TM is separable.

There are many ways to argue that TM is metrizable. We will show that TM is second countable and regular. Then from the Urysohn metrization theorem, we will have that TM is metrizable.

Separable and metric implies second countable. We do not know that TM is metric, but it is locally metric. Each chart $(\bar{\theta}, TU)$ induces a metric on TU.

Since M is second countable, there are countably many charts whose domains cover M. Each x in our countable dense set D has a countable neighborhood base in each TU for which $x \in U$ where U is in our collection of countably many domains that cover M. We will show that the union of all these neighborhoods bases is a countable basis for the topology on TM. Note that one $x \in D$ may be used many times as a contributor of neighborhood bases. It will contribute one neighborhood base for each U in our collection that contains x.

For an open V in TM and $y \in V$, there is a U in our countable collection of chart domains with $y \in U$. With the metric on TU induced by the chart on U, it is now a standard exercise to show that there is a member of our candidate basis that contains y and is contained in V. This shows that we have a true basis.

To show regular, we need to separate a point y from a closed set C. Equivalently, we need to find an open V about y whose closure misses C. There is a chart (θ, U) so that y is in TU. There is an open V in TU with $y \in V$ and with the closure of V in TU disjoint from C. Since TU is the union of the TM_x with $x \in U$, we have a projection of V on M that takes $v \in TU$ to $v(0) \in M$. Since M is metric, we can also assume that V is so small that its projection in M has its closure in U.

Now assume that V has a limit point p in TM outside of TU. The point p lies in some TW for some chart (ϕ, W). Now some sequence v_i in $V \cap TW$ converges to p. The structure of TW is gotten from $\bar{\phi} : TW \to \phi(W) \times \mathbf{R}^m$. Looking at the first coordinates, this means that the sequence $\bar{\phi}(v_i)$ converges to $\bar{\phi}(p)$. Thus the $\phi(v_i(0))$ converge to $\phi(p(0))$. Since the charts on M determine the topology on M, we have that the $v_i(0)$ in the projection of V on M converge to $p(0)$ in M. But since p is outside fo TU, we have $p(0)$ outside of U, a contradiction. Thus the closure of V in TM is the closure of V in TU and TM is regular.

The last thing we wish to show is that the overlap maps are of class C^{r-1}. We noted above that the overlap maps take the form

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

where U' and V' are open subsets of $\mathbf{R}^m$ (they are the images of charts of M), and $\bar{h}(x, y) = (h(x), Dh_x(y))$. We assume that h is C^r, so $Dh : U' \to \hom(\mathbf{R}^m, \mathbf{R}^m)$ is C^{r-1}. To simplify notation, let $H\mathbf{R}^m$ denote $\mathrm{Hom}(\mathbf{R}^m, \mathbf{R}^m)$. To break $\bar{h}$ into simpler pieces, we let i be the identity on U', we let j be the identity on $\mathbf{R}^m$ and let E be the evaluation map

$$E : H\mathbf{R}^m \times \mathbf{R}^m \to \mathbf{R}^m$$

given by $E(\alpha, u) = \alpha(u)$.

Both i and j are C^∞. We wish to look at the bilinear E. From our calculations at the end of Section 1.5, we have

$$DE : H\mathbf{R}^m \times \mathbf{R}^m \to \mathrm{Hom}(H\mathbf{R}^m \times \mathbf{R}^m, \mathbf{R}^m)$$

given by

$$DE_{(\alpha,u)}(\beta, v) = \alpha(v) + \beta(u).$$

A straightforward check using this formula shows that DE is a linear map (which is a different claim than the true fact that the values of DE are linear maps). Thus DE is C^∞ and so is E.

Now Lemma 1.3.1 implies that

$$(i, Dh) : U' \to U' \times H\mathbf{R}^m, \quad \text{and,}$$
$$((i, Dh) \times j) : U' \times \mathbf{R}^m \to U' \times H\mathbf{R}^m \times \mathbf{R}^m$$

are both C^{r-1} and

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

is C^r. Thus we get that

$$\bar{h} = ((i, Dh) \times j) \circ (h \times E)$$

is C^{r-1} as desired. The above argument about the differentiability of $\bar{h}$ follows the advice of Erik Pedersen to represent the map being analyzed as the longest possible combination of simpler maps.

The collection of charts that we have introduced now form a partial atlas for TM and give TM the structure of a C^{r-1} $2m$-manifold.

It now remains to be shown that if $f: M \to N$ is a C^r map between C^r manifolds, then $Df: TM \to TN$ is a C^{r-1} map. The argument for this resembles the argument above, and is left as an exercise.

The question of what TM is as a topological space when M is given is not a trivial one. We introduce some relevant terminology in the next section.

2.4. The Tangent Bundle

The tangent space of a differentiable manifold is an example of a vector bundle. We start with a definition.

An *n-dimensional vector bundle* is a triple (E, π, B) where E and B are topological spaces, where $\pi : E \to B$ is a surjective map and every $\pi^{-1}(b)$ has the structure of an n-dimensional vector space. Further every $b \in B$ has a neighborhood U in B and a homeomorphism $h_U : \pi^{-1}(U) \to U \times \mathbf{R}^n$ so that for every $x \in U$, the restriction of h_U to $\pi^{-1}(x)$ is a vector space isomorphism to $\{x\} \times \mathbf{R}^n$.

We refer to E as the *total space* of the bundle, we refer to B as the *base space* of the bundle, we refer to π as a projection, and we refer to each $\pi^{-1}(b)$ as a *fiber* of the bundle. The bundle is usually said to be *over* the base. Note that the dimension mentioned in the term defined refers to the dimension of the fibers and not to the dimension of either the base or the total space.

Sufficient information is given in the previous section to show that if M is a C^r n-manifold, $r \geq 1$, then (TM, π, M) is a vector bundle where π takes each $v \in TM$ to $v(0)$. Because of this TM is usually called the *tangent bundle* of M.

Vector bundles are very easy to find. If B is any space and V is a finite dimensional vector space, then $(B \times V, \pi, B)$ is a vector bundle where π is projection onto the first factor. A bundle such as this is called a *trivial vector bundle.*

Vector bundles can be put into a category where the objects are vector bundles. Given two vector bundles, (E, π, B) and (F, ρ, X), a morphism (called a homomorphism) from the firstbundle to the second is a commutative diagram

$$\begin{array}{ccc} E & \xrightarrow{f_1} & F \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\rho} \\ B & \xrightarrow{f_2} & X \end{array}$$

so that for every $b \in B$ the restriction of f_1 to $\pi^{-1}(b)$ is a linear map to $\rho^{-1}(f_2(b))$.

Enough information is given in the previous section to show that if $f : M \to N$ is a differentiable map between differentiable manifolds, then $Df : TM \to TN$ is a bundle homomorphism. The previous two sections combine to show that we now have a functor from the category of C^r manifolds and C^r maps to vector bundles where an object M in the first category is taken to TM in the second, and a morphism $f : M \to N$ in the first category is taken to $Df : TM \to TN$ in the second.

The usual notion of isomorphism comes in at this point. We have that diffeomorphic manifolds have isomorphic tangent bundles. Any vector bundle isomorphic to a trivial vector bundle is also called a trivial vector bundle. An obvious question is whether the tangent bundle of a differentiable manifold is trivial. It is not obvious that the answer is often no. It is quite immediate that if M is an open subset of some $\mathbf{R}^m$ with partial atlas including the inclusion from M to $\mathbf{R}^m$, then TM is a trivial vector bundle over M. We will see in Milnor's book that TS^2 is not trivial.

If (E, π, B) is a vector bundle, then a *section* for the bundle is a map $\sigma : B \to E$ so that $\pi\sigma(x) = x$ for all $x \in B$. The section is said to be *non-zero* if $\sigma(x) \neq 0$

for all $x \in B$. A trivial bundle $B \times V$ always has a non-zero section. Take your favorite non-zero $v \in V$ and let $\sigma(x) = v$ for all $x \in B$. Thus one way to show that a bundle is not trivial is to show that there are no non-zero sections.

The open Möbius band is a 1-dimensional vector bundle over S^1. It is not trivial. These statements are left as exercises.

Vector bundles admit various operations that derive from operations on vector spaces. Direct sums, quotients, various products and other constructions on vector spaces can be adapted to the vector bundle setting. We will probably see some of these in action later in the course.

2.5. Inverse Function Theorem and Consequences

In this section we take the basic infinitesimal to local theorem—the Inverse Function Theorem for Euclidean spaces—and deduce several theorems of the same flavor for manifolds.

We start with the following.

THEOREM 2.5.1 (Inverse Function Theorem on Manifolds). *Assume that* $f : M \to N$ *is a* C^r *function,* $1 \leq r \leq \infty$*, between* C^r *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* $D\big((f|_U)^{-1}\big)_z = (Df_x)^{-1}$.

PROOF. Let m be the dimension of M. Since Df_x is an isomorphism, the dimensions of TM_x and TM_y are the same and m is the dimension of N as well. There are charts (θ, U) and (ϕ, V) expressing f in local coordinates about x with the following commutative diagrams.

$$\begin{array}{ccc} U & \xrightarrow{f} & V \\ {\scriptstyle\theta}\downarrow & & \downarrow{\scriptstyle\phi} \\ \theta(U) & \xrightarrow{\phi f \theta^{-1}} & \phi(V) \end{array} \qquad \begin{array}{ccc} TM_x & \xrightarrow{Df_x} & TN_y \\ {\scriptstyle\hat{\theta}}\downarrow & & \downarrow{\scriptstyle\hat{\phi}} \\ \mathbf{R}^m & \xrightarrow{D(\phi f \theta^{-1})_{\theta(x)}} & \mathbf{R}^m \end{array} \tag{2.7}$$

The diagrams in (2.7) show show that the hypotheses and thus the conclusions of Theorem 1.5.1 hold, and that the conclusion that we are trying to prove must hold. □

We will often use a single letter for an expression of a function in local coordinates in order to decrease complexity of notation.

COROLLARY 2.5.1.1. *Let f, M, N and x be as in the statement above. Then there is an expression h of f about x in local coordinates so that h is the identity function from a Euclidean space to itself.*

PROOF. We know that M and N have a common dimension that we take to be m. Assume the conclusion of the Inverse Function Theorem with the notation as in the statement. We can find a coordinate chart (θ, U_1) 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 (ϕ, V_1) for N. Thus, the expression h_1 of f in these coordinates takes $\mathbf{R}^m$ to an open subset of $\mathbf{R}^m$. We know that h_1 and $(h_1)^{-1}$ are C^r. Let $W = f(U_1)$ and let $\zeta = (h_1)^{-1}(\phi|_W)$. Now (ζ, W) is is a valid coordinate chart for N and the expression of f using coordinates (θ, U_1) and (ζ, W) is the identity from $\mathbf{R}^m$ to itself. □

In each of the following arguments, we will use the phrase "by suitable choice of local coordinates we can assume ..." with the dots replaced by what we can assume. Rather than make this rigorous, we point to paragraphs 2–5 in the proof of the injectivity part of the Inverse Function Theorem for Euclidean spaces in Section 1.5 as an example of what it takes to make the phrase rigorous. We also will refer to expressions of an f in local coordinates that are "centered" about a point x in the domain. This will refer to a choice of charts that take both x and $f(x)$ to 0 in appropriate Euclidean spaces. We have noted that this can always be done. This use of the word "centered" is probably not all that standard.

We now give three standard consequences of the Inverse Function Theorem.

THEOREM 2.5.2 (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. By suitable choice of local coordinates we 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 $\mathbf{R}^n$. 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 $\mathbf{R}^m \times \{(0, \ldots, 0)\}$. (It is easy to preserve the various coordinates.) Now the last few lines in the proof of the corollary to the Inverse Function Theorem can be duplicated. □

THEOREM 2.5.3 (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, suitable choice of local coordinates lets 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. Later we will see the important theorem that regular values greatly outnumber the critical values.

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.

In the statement of the next theorem, we use the notion of a partial derivative as defined in (2.5). The proof will refer to the associated diagram (2.6).

THEOREM 2.5.4 (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 $f(u, v) = k$. It is worth finiding examples that illustrate the need for the various assumptions in the statement of the theorem.

PROOF. We get the diagram

$$\begin{array}{ccccc}
TV_v & \xrightarrow{D(\langle u\rangle)_v} & T(U\times U)_{(u,v)} & \xrightarrow{Df_{(u,v)}} & TN_k \\
\Big\downarrow{\scriptstyle\hat{\phi}} & & \Big\downarrow{\scriptstyle\widehat{\theta\times\phi}} & & \Big\downarrow{\scriptstyle\hat{\rho}} \\
\mathbf{R}^n & \xrightarrow{Dh_{\phi(v)}} & \mathbf{R}^m\times\mathbf{R}^n & \xrightarrow{Dg_{(\theta(u),\phi(v))}} & \mathbf{R}^n
\end{array}$$

where m and n are the dimensions of U and V, respectively, and n must be the dimension of N since $D_2f_{(u,v)}$ is an isomorphism. By replacing ρ with the composition of ρ with the inverse of the composition of the bottom two arrows, we can assume that the expression of $D_2f_{(u,v)}$ in local coordinates is the identity map from $\mathbf{R}^n$ to itself.

We now replace the original data by the expression in local coordinates and 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 $\mathbf{R}^n$, that $f(0,0) = 0$, and that $D_2f_{(0,0)}(b)$ is the identity map. We will now use u and v as arbitrary elements of U and V and not as references to items in the statement.

We get information about $Df_{(0,0)}$ because for any $b \in \mathbf{R}^n$, we have

$$Df_{(0,0)}(0,b) = (Df_{(0,0)})\langle 0\rangle(b) = D((f)\langle 0\rangle)_0(b) = D_2f_{(0,0)}(b) = b.$$

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

$$\begin{aligned}
D\bar{f}_{(0,0)}(a,b) &= (\pi(a,b), Df_{(0,0)}(a,b)) \\
&= (a, Df_{(0,0)}(a,0) + Df_{(0,0)}(0,b)) \\
&= (a, b + \lambda(a))
\end{aligned}$$

where λ is a linear function depending only on a. Now $D\bar{f}_{(0,0)}(a, b-\lambda(a)) = (a,b)$ and $D\bar{f}_{(0,0)}$ is invertible. (We do not even need the finite dimensionality of the domain and range since $D\bar{f}_{(0,0)}(a,b) = (0,0)$ easily implies that $(a,b) = (0,0)$.) 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 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 and $(x,y) = (x, f(x, h_2(x,y)))$. 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. □

2.6. 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 (θ, U) 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 (θ, U) is called a submanifold chart for A in M. Note that all the charts $(\theta|_{U\cap A}, U\cap A)$ where (θ, U) 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 chart to test the tangent vector in A is the restriction of a chart that tests it in M. The inclusion is also more than that. A basic open set in A (say the domain of a 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 2.6.1. *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 chart (θ, U) 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 2.6.2. *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 2.6.3. *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.

Note that an immersion $i : M \to N$ takes each TM_x isomorphically onto a linear subspace of $TN_{i(x)}$. When the immersion is just the inclusion of a submanifold $M \subseteq N$, then $i(x) = x$ and we have a linear injection $TM_x \to TN_x$. In fact, the sets TM_x and $Di_x(TM_x) \subseteq TN_x$ are rather hard to distinguish since they are equivalence classes of functions that are equal from the set theory point of view. (A function from A to B is a subset of $A \times B$, and from this point of view, a function does not change when its range changes.) The equivalence relation on the functions is the same when viewed in TM_x or TN_x (a consequence of the immersion), and so in some sense it is correct to write $TM_x = Di_x(TM_x)$. Not only is it technically correct, but it is notationally painful to distinguish them. If it ever becomes important to make the distinction, we will take the pains to do so. However whenever we have a submanifold $M \subseteq N$, it is usually convenient to think of the tangent bundle TM of M as a "subbundle" of the tangent bundle TN of N.

CHAPTER 3

Extra topics

Possible topics for inclusion after going through Milnor's book.

1. vector fields, differential equations, flows,
2. Partitions of unity and some applications,
3. Normal bundles,
4. Exponential maps and tubular neighborhoods (sprays),
5. Isotopy extension theorem,
6. Approximations, increasing differentiability,
7. Approximations, transversality,
8. Manfiolds with boundary,
9. Connected sums,
10. Embedding theorems (easy and hard Whitney theorems).