# Off on a Tangent

Meditations on the derivative

Shishir Agrawal

Draft of October 6, 2019

# Introduction

These notes are intended as a guided meditation on the concept of the derivative. They start with a rigorous treatment of single variable derivatives. The next step is to generalize and consider multivariable derivatives. Then, after an interlude introducing manifolds, the notes conclude with the "ultimate" generalization of the derivative: pushforwards of tangent vectors on manifolds.

I have tried to include many pictures, because it seems to me that you don't deeply understand something until you can visualize it. I have also tried to include many exercises, because it seems to me that you really need to play with concepts yourself in order to understand them.

## How to read mathematics

The first and foremost comment I have about reading math is to remember that you'll only understand things if you do them yourself. Spend a lot of time solving exercises. It's okay if you get stuck; in fact, that's great news! That means you'll have learned something when you finally do figure it out. Don't let it bog you down, but do keep coming back to the exercises that give you trouble until you manage to pin down a solution.

Generally speaking, I think that it's useful to organize reading and learning new mathematics in the following stages.

- In the first stage, focus on the definitions, theorem statements, and examples. The examples are the most important of those. If there are exercises about explicit examples, do them. Skip the proofs of the theorems.
- (2) In the second stage, look over the proofs and try to figure out how it's structured. Do not go through line-by-line and trying to understand all of the details at this point. Try to formulate an outline of the argument by identifying the main claims that are being made. Make sure that these main claims agree with the intuition you've developed

from the examples you studied in the previous step. Try to visualize parts of the argument.

(3) In the third stage, study the details of the proofs.

Each of these three stages builds on the previous one. If your grasp on definitions, theorem statements, and examples is shaky, you're unlikely to get anything meaningful out of reading proofs. If you don't understand how a proof is broadly structured, the details of the proofs may just be an amorphous and meaningless string of logic.

I also think that each of these three stages is also less important than the previous one. If your understanding of definitions and examples is solid enough, you'll often just figure out the proofs yourself. Maybe not all of the proofs (some theorems have very hard proofs), but that's okay. Similarly, if you understand the broad outline of arguments, you'll often just be able to fill in the details yourself. Maybe not always (sometimes the details are very tricky), but that's also okay. If you're at the point where you really understand everything except perhaps the trickiest parts of the proofs of the hardest theorems, you're in a good place!

# "Do I need to prove this formally?"

If you find yourself asking this question, the answer is almost definitely "yes." You'll often learn a lot by trying to formalize arguments. It might just give you more practice structuring formal arguments, but sometimes you'll also discover that an assertion you thought was true isn't actually.

When you look at an assertion and *confidently know* that you *could* write down a formal proof, that's the point when maybe you don't actually need to write it down. I like calling this the *Bergman principle* (after George Bergman, who said something to this effect in a class I took with him at UC Berkeley in Fall 2011).

## How to use these notes

These notes assume basic familiarity with point-set topology (specifically, metric spaces, and basic topological properties of  $\mathbb{R}$ ) and with linear algebra. At a few points, there may be some ideas from point-set topology and linear algebra that you have not encountered before. A sort of "bare minimum" exposition of some of these ideas is included in chapter 0.

That said, you are advised to *not* spend time reading chapter 0 thoroughly before jumping into the main part of the text. When ideas from chapter 0 are invoked in the main part, a reference to the relevant part of chapter 0 is included. My suggestion is to only look at chapter 0 when you run into a reference to it, chasing references back as needed.

Some sections are starred (\*). This is intended to indicate one of two things: either that the section is a little more challenging than others, or that it's slightly less important for the overall development of concepts in these notes. I wouldn't say that the starred sections are all skip-able, and unstarred sections do sometimes reference results in starred sections. But, if you find yourself struggling and need help deciding what's most important to focus on, focus on the unstarred sections.

## Suggestions for improvement

These notes are still in very rough form. There are bound to be many errors, so please be on the lookout for them! If you think you've found one, please share it with me.

I'd also very much appreciate suggestions for improving the exposition. For example, I'd like to know if there are parts that are phrased confusingly, or if there are specific proofs where I could spend more time discussing the broad outline before diving into the details, or if there are places where more pictures would be useful... Any way that you think these notes could be better, please tell me!

# Contents

Introduction 2								
0	0 Preliminaries							
	0.1	Little-	oh notation	11				
	0.2	Produ	ct metric	13				
	0.3	Norms on vector spaces						
		0.3.A	Norms induce metrics	13				
		0.3.B	Equivalence of norms	14				
		0.3.C	Norms on finite dimensional vector spaces	14				
		0.3.D	Sup norm on real-valued functions	16				
	0.4	.4 Euclidean space		16				
		0.4.A	Linear structure on $\mathbb{R}^n$	17				
		0.4.B	Norms on $\mathbb{R}^n$	17				
	0.5	.5 Matrices		18				
		0.5.A	Matrix representations of linear maps	18				
		0.5.B	Matrix norm	20				
	0.6	6 Operator norm		21				
		0.6.A	Basics	21				
		0.6.B	Operator norm as a metric *	25				
	0.7 Topological spaces		ogical spaces	26				
		0.7.A	Basics	26				
		0.7.B	Hausdorff spaces *	30				
		0.7.C	Compact and $\sigma$ -compact spaces $\star$	30				
1	Single variable derivatives							
	1.1	.1 Two definitions of the derivative						
		1.1.A	Slope of the tangent line	31				

		1.1.B	Best linear approximation    35							
	1.2	Comp	uting derivatives							
		1.2.A	Sum and scalar multiples rule							
		1.2.B	Product and quotient rules							
		1.2.C	Chain rule							
		1.2.D	Interior extremum theorem							
		1.2.E	Non-rational functions							
	1.3	Differe	entiable functions							
		1.3.A	Mean value theorem							
		1.3.B	Monotonicity							
		1.3.C	Darboux's theorem $\star$							
		1.3.D	Inverse function theorem							
		1.3.E	Uniform limits *							
	1.4	C <sup>k</sup> hie	erarchy							
		1.4.A	Continuous differentiability							
		1.4.B	$C^k$ functions							
		1.4.C	Taylor's theorem $\star$ 66							
		1.4.D	Smooth functions   70							
2	Multivariable derivatives 7									
	2.1	Introd	uctory example							
	2.2	Defini	tion of the derivative							
	2.3	Comp	uting derivatives							
		2.3.A	Sum and scalar multiples rule							
		2.3.B	Chain rule							
		2.3.C	Differentiability by components							
		2.3.D	Product and quotient rules							
		2.3.E	Directional and partial derivatives							
		2.3.F	Jacobian matrix							
		2.3.G	Rank of a differentiable map							
	2.4	Differe	entiable functions							
		2.4.A	Vanishing derivatives							
		2.4.B	Bounded derivatives							
		2.4.C	Inverse function theorem							

		2.4.D	Proof of the inverse function theorem $\star$	109
	2.5	2.5 C <sup>k</sup> hierarchy		116
		2.5.A	Continuous differentiability	116
		2.5.B	C <sup>k</sup> functions	122
		2.5.C	Equality of mixed partials *	123
		2.5.D	Taylor's theorem $\star$ .	125
		2.5.E	Smooth functions	132
3	Man	ifolds		137
	3.1	Definition of a manifold		138
	0.12	3.1.A	Charts	138
		3.1.B	Compatibility of charts	139
		3.1.C	Atlases and manifolds	140
		3.1.D	Dimension	141
	3.2	Exam	bles	141
		3.2.A	First examples	141
		3.2.B	Möbius strip	144
		3.2.C	More quotient examples	147
		3.2.D	Submanifolds	147
		3.2.E	Submanifolds of euclidean space	151
	3.3	Smoot	h functions	154
		3.3.A	Smooth real-valued functions	154
		3.3.B	Smooth functions between manifolds	157
		3.3.C	Paths	159
		3.3.D	Diffeomorphisms	159
4	Tano	ient Sp	aces	161
1	4.1	Tanger	nt space	161
		4 1 A	Curves based at a point	161
		4.1.B	Vector space structure	162
		4.1.C	Derivations *	166
	4.2	Pushfo	prward	168
		4.2.A	Definition of the pushforward	168
		4.2.B	Rank of a smooth map	170
			r	0

Contents

Bibliography	171
Index	173

# **0** Preliminaries

The sections in this chapter cover some ideas relating to point-set topology and linear algebra that will be invoked in the main part of the text. Some sections are merely intended to establish notation; others cover topics that you may not have encountered before. I encourage skipping this chapter, and referring back to it only when you need to.

You should consult a dedicated linear algebra textbook to review the definitions of vector spaces, subspaces, linear independence, span, dimension, linear maps, matrices, matrix multiplication, determinants, and minors. You should also consult an analysis textbook to review the definitions of metric spaces, equivalence of metrics, open subsets, closed subsets, continuous maps, compactness, and connectedness.

# 0.1 Little-oh notation

Suppose X is a metric space<sup>1</sup> and  $x_0 \in X$  is a point.

**Definition 0.1.1.** A function  $g : X \setminus \{x_0\} \to \mathbb{R}$  is *positive* if g(x) > 0 for all  $x \in X \setminus \{x_0\}$ . Similarly, g is *non-negative* if  $g(x) \ge 0$  for all  $x \in X \setminus \{x_0\}$ .

Suppose g is a positive function. Definition 0.1.2 below formulates what it means for a non-negative function  $f : X \setminus \{x_0\} \to \mathbb{R}$  to be "little-oh of g as x tends to  $x_0$ ." Intuitively, this condition means that f is *much smaller* than g near  $x_0$ .

**Definition 0.1.2.** A function  $f : X \setminus \{x_0\} \to \mathbb{R}$  is *little-oh of g as x tends to x*<sub>0</sub>, written

$$f(x) = o(g(x))$$
 as  $x \to x_0$ 

if, for every  $\epsilon > 0$ , there exists an open neighborhood U of  $x_0$  such that  $|f(x)| \leq \epsilon g(x)$  for

<sup>&</sup>lt;sup>1</sup>Even more generally, X could be a topological space here. See section 0.7.

#### 0 Preliminaries

all  $x \in U \setminus \{x_0\}$ . Equivalently, this means that

$$\lim_{x\to x_0}\frac{|f(x)|}{g(x)}=0.$$

When  $x_0$  can be inferred from context, we write simply f = o(g).

**Remark 0.1.3.** It's worth noting that the use of the symbol "=" above is mathematically abusive. The left-hand side of the "=" is a function and the right-hand side is a property of functions; of course, a function cannot literally be equal to a property. Rather, the "=" is being used here to mean that the function on the left-hand side *has* the property described on the right-hand side. As annoying as it is, this abuse of notation is fairly standard, so it's probably best to just get used to it.

**Exercise 0.1.4.** Prove that the two conditions in definition 0.1.2 are in fact equivalent.

**Exercise 0.1.5.** Let  $X = \mathbb{R}$  and  $x_0 = 0$ . For each of the functions f and g described below, sketch graphs of f and g and then determine whether or not f = o(g) as  $x \to 0$ .

- (a) f(x) = x and  $g(x) = x^2$ .
- (b) f(x) = x and g(x) = |x|.
- (c)  $f(x) = x^2$  and g(x) = |x|.

**Exercise 0.1.6.** For a fixed positive function  $g : X \setminus \{x_0\} \to \mathbb{R}$ , prove that the set of functions  $f : X \setminus \{x_0\} \to \mathbb{R}$  which are o(g) as  $x \to x_0$  is a vector space under the natural operations. In other words, verify the following three facts.

- (1) ("Zero is small") The zero function is o(g(x)).
- (2) ("Scalar multiples of small are still small") If  $c \in \mathbb{R}$  is a scalar and f(x) = o(g(x)), then

$$cf(x) = o(g(x)).$$

(3) ("Sum of smalls is still small") If  $f_1(x) = o(g(x))$  and  $f_2(x) = o(g(x))$ , then

$$(f_1 + f_2)(x) = o(g(x)).$$

**Exercise 0.1.7** ("Smaller than small is still small"). Suppose  $f_1, f_2 : X \setminus \{x_0\} \to \mathbb{R}$  are functions such that  $|f_1(x)| \leq |f_2(x)|$  for all x in a punctured neighborhood of  $x_0$ , and  $f_2(x) = o(g(x))$  as  $x \to x_0$ . Then  $f_1(x) = o(g(x))$  also.

## 0.2 Product metric

**Exercise 0.2.1.** Suppose X and Y are metric spaces with metrics  $d_X$  and  $d_Y$ , respectively. Define a function  $d : (X \times Y) \times (X \times Y) \rightarrow \mathbb{R}$  by

$$d((x_1, y_1), (x_2, y_2)) = \max\{d_X(x_1, x_2), d_Y(y_1, y_2)\}.$$

Show that d is a metric on  $X \times Y$ .

**Exercise 0.2.2.** Suppose X and Y are metric spaces with metrics  $d_X$  and  $d_Y$ , respectively. Define a function  $d : (X \times Y) \times (X \times Y) \rightarrow \mathbb{R}$  by

$$d((x_1, y_1), (x_2, y_2)) = \sqrt{d_X(x_1, x_2)^2 + d_Y(y_1, y_2)^2}.$$

Show that d is a metric on  $X \times Y$ , and that it is equivalent to the metric from section 0.2.

## 0.3 Norms on vector spaces

**Definition 0.3.1.** Let V be a vector space. A *norm*  $\| - \|$  on V is a function  $V \to \mathbb{R}$  satisfying the following two axioms.

- (N1)  $|v| \ge 0$ .
- (N2) |v| = 0 if and only if v = 0.
- (N3)  $|\lambda \nu| = |\lambda| |\nu|$  for all  $\lambda \in \mathbb{R}$  and  $\nu \in V$ .
- (N4)  $|v + w| \leq |v| + |w|$  for all  $v, w \in V$ .

#### 0.3.A Norms induce metrics

**Exercise 0.3.2.** Suppose  $\| - \|$  is a norm on a vector space V. Show that the function  $d : V \times V \to \mathbb{R}$  given by d(v, w) = |v - w| is a metric.

#### 0 Preliminaries

**Exercise 0.3.3.** Suppose |-| is a norm on a vector space V. Show that the function  $V \times V \rightarrow V$  given by  $(v, w) \mapsto v + w$  is continuous, where  $V \times V$  is regarded as a metric space via one of the metrics defined in section 0.2.

**Exercise 0.3.4.** Suppose |-| is a norm on a vector space V. Show that the function  $\mathbb{R} \times V \to V$  given by  $(\lambda, \nu) \mapsto \lambda \nu$  is continuous, where  $\mathbb{R} \times V$  is regarded as a metric space via one of the metrics defined in section 0.2.

**Exercise 0.3.5.** Suppose |-| is a norm on a vector space V and |-|' is a norm on a vector space W and  $\ell : V \to W$  is a linear map. Then the following are equivalent.

(a)  $\ell$  is continuous.

(b) There exists a constant M > 0 such that  $|\ell(v)|' \leq M|v|$  for all  $v \in V$ .

*Possible hint*. The harder direction is (a) implies (b). For this direction, since  $\ell$  is continuous at 0, there exists  $\delta > 0$  such that  $|\ell(v) - \ell(0)|' \leq 1$  whenever  $|v - 0| \leq \delta$ . Then note that for any nonzero vector  $v \in V$ , the vector  $\delta v/|v|$  is within  $\delta$  of 0, so the above inequality applies.

### 0.3.B Equivalence of norms

**Definition 0.3.6.** Two norms |-| and |-|' on a vector space V are *equivalent* if there exist nonzero constants  $C_1$  and  $C_2$  such that

$$C_1|\nu| \leqslant |\nu|' \leqslant C_2|\nu|$$

for all  $\nu \in V$ .

**Exercise 0.3.7.** If V is a vector space, show that equivalence of norms is an equivalence relation on the set of all norms on V.

**Exercise 0.3.8.** Suppose V is a vector space and |-| and |-|' are two equivalent norms on V. Show that the metrics d and d' corresponding to |-| and |-|' (cf. exercise 0.3.2) are equivalent.

### 0.3.C Norms on finite dimensional vector spaces

If V is a finite dimensional vector space, we can construct a number of norms on V by choosing a basis. Of these, two are especially important: the max norm (also called the

 $L^{\infty}$  norm, which is often the most convenient), and the euclidean norm (also called the  $L^2$  norm, which is the most common).

**Example 0.3.9.** Suppose V is a finite dimensional vector space, and  $v_1, \ldots, v_n$  is a basis for V. The L<sup> $\infty$ </sup>-*norm*, also called the *max norm*, on V with respect to this basis is defined by

 $|a_1v_1+\cdots+a_nv_n|_{\infty}=\max\{|a_1|,\ldots,|a_n|\}.$ 

Axioms (N1) through (N3) are straightforward to verify. For the triangle inequality (N4), suppose  $v = a_1v_1 + \cdots + a_nv_n$  and  $w = b_1v_1 + \cdots + b_nv_n$ . Then

$$|v + w|_{\infty} = \max\{|a_1 + b_1| + \dots + |a_n + b_n|\}$$
  

$$\leq \max\{|a_1| + |b_1|, \dots, |a_n| + |b_n|\}$$
  

$$\leq \max\{|a_1|, \dots, |a_n|\} + \max\{|b_1|, \dots, |b_n|\}$$
  

$$= |v|_{\infty} + |w|_{\infty}$$

where we used the triangle inequality for real numbers for the second step.

**Example 0.3.10.** Suppose V is a finite dimensional vector space, and  $v_1, \ldots, v_n$  is a basis for V. The L<sup>2</sup> *norm*, also called the *euclidean norm*, on V with respect to this basis is defined by

$$|\mathfrak{a}_1\mathfrak{v}_1+\cdots+\mathfrak{a}_n\mathfrak{v}_n|_2=\sqrt{\mathfrak{a}_1^2+\cdots+\mathfrak{a}_n^2}.$$

Axioms (N1) through (N3) are straightforward to verify. The triangle inequality is harder: see [PM91, theorem 6.2], for instance.

The following states that these two norms are equivalent.

**Exercise 0.3.11.** Suppose V is a finite dimensional vector space and  $v_1, \ldots, v_n$  is a basis for V. Show that, for any  $v \in V$ , we have

$$|v|_{\infty} \leq |v|_2 \leq \sqrt{n} |v|_{\infty}.$$

In fact, here is vast generalization of exercise 0.3.11.

**Theorem 0.3.12.** If V is a finite dimensional vector space, all norms on V are equivalent.

I'll add a proof of this eventually...

#### 0.3.D Sup norm on real-valued functions

Let X be a set and consider the set of functions  $X \to \mathbb{R}$ . This set is naturally a vector space under pointwise operations: if f, g :  $X \to \mathbb{R}$ , then

$$(f+g)(x) = f(x) + g(x)$$

and if  $\lambda \in \mathbb{R}$ , then

$$(\lambda f)(x) = \lambda f(x).$$

**Definition 0.3.13.** If X is a set, we define the *sup norm* of a function  $f : X \to \mathbb{R}$ , denoted either  $||f||_{\sup,X}$  or just  $||f||_{\sup}$  when X can be inferred from context, by

$$\|f\|_{\sup} = \sup_{x \in X} |f(x)|.$$

We say that f is *bounded* if  $||f||_{sup} < \infty$ .

**Exercise 0.3.14.** Show that  $\| - \|_{sup}$  is a norm on the vector space of bounded functions  $X \to \mathbb{R}$ .

**Definition 0.3.15** (Uniform convergence). A sequence of functions  $f_n : X \to \mathbb{R}$  converges uniformly to a function  $f : X \to \mathbb{R}$  if, for every  $\varepsilon > 0$ , there exists N such that  $\|f_n - f\|_{sup} < \varepsilon$  for all  $n \ge N$ .

**Definition 0.3.16** (Uniformly Cauchy sequences). A sequence of functions  $f_n : X \to \mathbb{R}$  is *uniformly Cauchy* if, for every  $\varepsilon > 0$ , there exists N such that  $\|f_m - f_n\|_{sup} < \varepsilon$  for all  $m, n \ge N$ .

# 0.4 Euclidean space

For any non-negative integer n, we write  $\mathbb{R}^n$  to denote the set of lists of n real numbers. We will sometimes write its elements as horizontal lists of numbers separated by commas, as in

$$(h_1,\ldots,h_n),$$

and sometimes as vertical column of numbers, as in

$$\begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix}$$

The set  $\mathbb{R}^n$  is also called *n*-dimensional euclidean space.

## 0.4.A Linear structure on $\mathbb{R}^n$

Given  $v_1, v_2 \in \mathbb{R}^n$ , we define their sum  $v_1 + v_2$  by adding the entries coordinate-wise. Given  $\lambda \in \mathbb{R}$  and  $v \in \mathbb{R}$ , we define  $\lambda v$  by multiplying each entry of v by  $\lambda$ . This endows  $\mathbb{R}^n$  with the structure of a vector space.

For each i = 1, ..., n, we define the *i*th standard basis vector, denoted  $e_i$ , to be the list whose ith entry is 1 and all other entries are 0.

$$e_1 = (1, 0, 0, \dots, 0, 0)$$
  
 $e_2 = (0, 1, 0, \dots, 0, 0)$   
 $\vdots$   
 $e_n = (0, 0, 0, \dots, 0, 1)$ 

Observe that

$$(\mathbf{h}_1,\ldots,\mathbf{h}_n)=\mathbf{h}_1\mathbf{e}_1+\mathbf{h}_2\mathbf{e}_2+\cdots+\mathbf{h}_n\mathbf{e}_n.$$

Thus the list  $e_1, \ldots, e_n$  is a basis for  $\mathbb{R}^n$ .

For any i = 1, ..., n, we let  $\pi_i : \mathbb{R}^n \to \mathbb{R}$  denote the *i*th *projection map*, given by

$$\pi_{i}(h_{1},\ldots,h_{n})=h_{i}.$$

This is a linear map.

## 0.4.B Norms on $\mathbb{R}^n$

Using the standard basis  $e_1, \ldots, e_n$  of  $\mathbb{R}^n$ , we can define the euclidean (or  $L^2$ ) and max (or  $L^{\infty}$ ) norms (cf. examples 0.3.9 and 0.3.10). Explicitly, if  $h = (h_1, \ldots, h_n) \in \mathbb{R}^n$ , we have the

following.

$$|\mathbf{h}|_2 = \sqrt{\mathbf{h}_1^2 + \dots + \mathbf{h}_n^2}$$
$$|\mathbf{h}|_{\infty} = \max\{|\mathbf{h}_1|, \dots, |\mathbf{h}_n|\}$$

For the most part, it doesn't matter which of these norms you use. We'll use |h| to denote either of them; in other words, you are free to interpret |h| as either  $|h|_2$  or  $|h|_{\infty}$ , whichever you like better. When we need to choose one over another, we'll explicitly specify this. Notice that, when n = 1, both of these norms are equal (both are just given by taking the absolute value of a real number).

## 0.5 Matrices

The following is a whirlwind review of some facts about matrices, mostly intended to establish notation. The reader is expected to remember definitions of matrix multiplication, determinants, and minors. Few proofs are included in this section; for more details, you are encouraged to reference a dedicated linear algebra text.

**Definition 0.5.1.** We write  $\mathcal{M}_{n \times m}$  for the set of all  $n \times m$  matrices (ie, the matrices with n rows and m columns). This is naturally a vector space, where addition and scalar multiplication are defined entrywise.

**Definition 0.5.2.** Let  $GL_n$  denote the subset of  $\mathcal{M}_{n \times n}$  consisting of invertible  $n \times n$  matrices.

Lemma 0.5.3. Let A be a matrix. Then all of the following numbers are equal.

- (1) The dimension of the span of the columns of A.
- (2) The dimension of the span of the rows of A.
- (3) The size of largest nonzero minor of A.

This integer is called the rank of A, and is denoted rank(A).

#### 0.5.A Matrix representations of linear maps

**Definition 0.5.4.** If V and W are both vector spaces, we write  $\mathcal{L}(V, W)$  for the set of all linear maps  $V \to W$ . This is naturally a vector space.

**Definition 0.5.5.** If V is a vector space, let GL(V) denote the set of invertible linear maps  $V \rightarrow V$ , regarded as a subset of  $\mathcal{L}(V, V)$ .

Throughout the rest of this section, we assume that U, V, and W are finite dimensional vector spaces.

**Definition 0.5.6.** Suppose B denotes a basis  $v_1, \ldots, v_n$  for V. Then any  $v \in V$  can be written uniquely as  $a_1v_1 + \cdots + a_nv_n$  for some scalars  $a_1, \ldots, a_n \in \mathbb{R}$ , and we define the *representation of* v *with respect to* B, denoted  $[v]_B$ , to be the column vector with n entries that records the scalars  $a_1, \ldots, a_n$ . In other words,

$$\left[\nu\right]_{B} = \begin{bmatrix} a_{1} \\ \vdots \\ a_{n} \end{bmatrix}.$$

**Lemma 0.5.7.** If B denotes a basis  $v_1, \ldots, v_n$  is a basis for V, then the function  $V \to \mathbb{R}^n$  given by  $v \mapsto [v]_B$  is an isomorphism, with inverse given by

$$\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \mapsto a_1 v_1 + \dots + a_n v_n.$$

**Definition 0.5.8.** Suppose  $l : V \to W$  is a linear map. Let B and C denote bases  $v_1, \ldots, v_m$  and  $w_1, \ldots, w_n$  for V and W, respectively. The *matrix representation of* l *with respect to* B *and* C, denoted  $[l]_{B,C}$ , is the  $n \times m$  matrix whose ith column of  $[l]_{B,C}$  is  $[l(v_i)]_C$ . In other words,

$$[\ell]_{B,C} = \begin{bmatrix} [\ell(\nu_1)]_C & \cdots & [\ell(\nu_m)]_C \end{bmatrix}.$$

**Lemma 0.5.9.** Suppose that V and W are m- and n-dimensional vector spaces, respectively, and that and B and C are bases for V and W, respectively. Then the function  $\ell \mapsto [\ell]_{B,C}$  is an isomorphism

$$\mathcal{L}(V,W) \longrightarrow \mathcal{M}_{n,m}.$$

**Lemma 0.5.10.** *If*  $l : V \to W$  *is a linear map between finite dimensional vector spaces and* B *and* C *are bases for* V *and* W*, respectively, then* 

$$[\ell]_{B,C}[\nu]_B = [\ell(\nu)]_C$$

for any  $v \in V$ .

**Lemma 0.5.11.** *If*  $\ell$  :  $U \to V$  *and*  $\ell' : V \to W$  *are linear maps, and* A, B *and* C *are bases for* U, V, *and* W, *respectively, then* 

$$[\ell' \circ \ell]_{A,C} = [\ell']_{B,C}[\ell]_{A,B}.$$

**Definition 0.5.12.** If  $\ell : V \to W$  is a linear map, then the *rank* of  $\ell$ , denoted rank( $\ell$ ) is defined to be the dimension of the range of  $\ell$ .

**Lemma 0.5.13.** *If*  $\ell$  :  $V \rightarrow W$  *is a linear map and* B *and* C *are bases for* V *and* W*, respectively, then* 

$$\operatorname{rank}(\ell) = \operatorname{rank}[\ell]_{B,C}.$$

**Definition 0.5.14.** If  $\ell : \mathbb{R}^m \to \mathbb{R}^n$  is a linear map, the *standard matrix representation* of  $\ell$ , denoted  $[\ell]$ , is the matrix of  $\ell$  with respect to the standard bases on  $\mathbb{R}^m$  and  $\mathbb{R}^n$ .

**Lemma 0.5.15.** If A is a  $n \times m$  matrix and  $\ell_A : \mathbb{R}^m \to \mathbb{R}^n$  is the linear map  $\ell_A(\nu) = A\nu$ , then

$$[\ell_A] = A.$$

#### 0.5.B Matrix norm

Observe that  $\mathcal{M}_{n \times m}$  is a finite dimensional vector space. If  $E_{j,i}$  denotes the matrix whose (j, i)-entry is 1 and all other entries are 0 for i = 1, ..., m and j = 1, ..., n, then the list of all of these mn matrices forms a basis for  $\mathcal{M}_{n \times m}$ . We can use this basis to define a  $L^2$  and max norm on  $\mathcal{M}_{n \times m}$ , as in examples 0.3.9 and 0.3.10. Explicitly, if A is a  $n \times m$  matrix whose (j, i)-entry (ie, the entry in row j and column i) is  $a_{j,i}$ , we have the following.

$$|A|_{2} = \sqrt{\sum_{j,i} a_{j,i}^{2}}$$
$$|A|_{\infty} = \max_{j,i} |a_{j,i}|$$

We know from exercise 0.3.11 that these two norms are equivalent. We'll write |A| to denote either of these norms; in other words, when you see |A|, you can interpret this to mean either  $|A|_2$  or  $|A|_{\infty}$ , whichever you like better. All of the following don't depend on which norm you're using.

**Lemma 0.5.16.** *The function* det :  $\mathfrak{M}_{n \times n} \to \mathbb{R}$  *is continuous.* 

*Proof.* Determinants can be computed using cofactor expansions, so det A is a polynomial in the entries of A.  $\Box$ 

**Corollary 0.5.17.** GL<sub>n</sub> *is an open subset of*  $M_{n \times n}$ *.* 

 $\begin{array}{l} \textit{Proof.} \ \text{Since det}: \mathfrak{M}_{n \times n} \to \mathbb{R} \text{ is continuous, and } \mathbb{R} \setminus \{0\} \text{ is an open subset of the codomain,} \\ \text{we see that } GL_n = det^{-1}(\mathbb{R} \setminus \{0\}) \text{ must be open in } \mathfrak{M}_{n \times n}. \end{array}$ 

**Lemma 0.5.18.** The function  $GL_n \to GL_n$  given by  $A \mapsto A^{-1}$  is a homeomorphism.

*Proof.* This function is its own inverse, so it is sufficient to show that  $A \mapsto A^{-1}$  is continuous. But

$$A^{-1} = \det(A)^{-1} \operatorname{adj}(A),$$

where adj(A) denotes the adjugate matrix of A. We know that det(A) is a polynomial in the entries of A. Each entry of adj(A) is a minor of A and is therefore also a polynomial in the entries of A. The lemma follows.

The following generalizes corollary 0.5.17.

**Lemma 0.5.19.** *Suppose*  $k \leq \min\{m, n\}$ *. Then* 

$$\mathsf{Z} = \{\mathsf{A} \in \mathcal{M}_{n \times m} : \operatorname{rank}(\mathsf{A}) < \mathsf{k}\}$$

*is a closed subset of*  $M_{n \times m}$ *.* 

*Proof.* Saying rank(A) < k is equivalent to insisting that all  $k \times k$  minors of A vanish. Each  $k \times k$  minor is a polynomial in the entries of A, so the zero set of any particular  $k \times k$  minor is a closed subset of  $\mathcal{M}_{n \times m}$ . We then take the intersection of the closed subsets corresponding to all possible  $k \times k$  minors, and recall the fact that intersections of closed subsets must be closed.

## 0.6 Operator norm

#### 0.6.A Basics

**Definition 0.6.1.** Suppose  $\ell : \mathbb{R}^m \to \mathbb{R}^n$  is a linear map. The *operator norm* of  $\ell$ , denoted  $\|\ell\|$ , is defined to be

$$\|\ell\| = \sup\{|\ell(\nu)| : |\nu| \leq 1\}.$$

#### 0 Preliminaries

If we're using the euclidean norms on  $\mathbb{R}^m$  and  $\mathbb{R}^n$ , then we get one operator norm, which we denote by  $\|\ell\|_2$  and call the *euclidean* (or L<sup>2</sup>) *operator norm* using the above definition. If we're using the max norms, then the above definition yields a different operator norm, which we denote by  $\|-\|_{\infty}$  and call the *max* (or L<sup> $\infty$ </sup>) *operator norm*. A lot of the basic properties of the operator norm work for either, so you can interpret the symbol  $\|-\|$  to refer to either of them, whichever you like better. If we really need to distinguish one from the other, we'll indicate this explicitly.

Another important remark is that the operator norm  $\|\ell\|$  is *not* the same as the norm of the standard matrix representation  $\|[\ell]\|$  as defined in section 0.5.B.

**Exercise 0.6.2.** Suppose  $\ell : \mathbb{R}^m \to \mathbb{R}^n$  is a linear map. Show that

$$\|\ell\| = \inf \{\lambda \in \mathbb{R} : |\ell(\nu)| \leq \lambda |\nu| \text{ for all } \nu \in \mathbb{R}^m \}.$$

In particular, this means that  $|\ell(v)| \leq ||\ell|| \cdot |v|$  for all  $v \in \mathbb{R}^m$ .

Here are some fundamental properties of the operator norm.

**Exercise 0.6.3.** Show that the operator norm is a norm on  $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$  in the sense of definition 0.3.1.

**Lemma 0.6.4.** Suppose  $\ell : \mathbb{R}^m \to \mathbb{R}^n$  is a nonzero linear map. Then there exists  $v_0 \in \mathbb{R}^m$  such that  $|v_0| = 1$  and  $|\ell(v_0)| = ||\ell||$ .

Proof. The "unit disk"

$$D = \{ x \in \mathbb{R}^m : |x| \leq 1 \}$$

is a nonempty compact subset of  $\mathbb{R}^m$ . The function  $\nu \mapsto |\ell(\nu)|$  is a continuous function  $D \to \mathbb{R}$ , so its image  $|\ell(D)|$  must also be a nonempty compact subset of  $\mathbb{R}$ . Thus there exists  $\nu_0 \in D$  such that

$$\ell(v_0)| = \sup |\ell(D)| = \|\ell\|.$$

Since  $\ell \neq 0$ , we know  $\|\ell\| \neq 0$  from exercise 0.6.3, which in turn means that  $v_0 \neq 0$ . To prove that  $|v_0| = 1$ , assume for a contradiction that  $|v_0| < 1$ , and let  $c = 1/|v_0| > 1$ . Then  $|cv_0| = |c||v_0| = 1$ , so  $cv_0 \in D$ . But then

$$|\ell(cv_0)| = |c\ell(v_0)| = c|\ell(v_0)| > |\ell(v_0)| = \sup |\ell(D)|,$$

which is absurd.

**Lemma 0.6.5.** Suppose  $\ell : \mathbb{R}^m \to \mathbb{R}^n$  is a linear map and  $|\ell(\nu)| = o(|\nu|)$  as  $\nu \to 0$ . Then  $\ell = 0$ .

*Proof.* The fact that  $|\ell(v)| = o(|v|)$  tells us that

$$\lim_{\nu \to 0} \frac{|\ell(\nu)|}{|\nu|} = 0.$$
 (0.6.6)

Suppose for a contradiction that  $\ell \neq 0$ . Using lemma 0.6.4, choose  $v_0 \in \mathbb{R}^m$  such that  $|v_0| = 1$  and  $|\ell(v_0)| = ||\ell||$ . For scalars c, observe that  $cv_0 \to 0$  as  $c \to 0$ , which means that equation (0.6.6) implies that

$$\lim_{c \to 0} \frac{|\ell(cv_0)|}{|cv_0|} = 0.$$

But observe that, for nonzero c, we have

$$\frac{|\ell(cv_0)|}{|cv_0|} = \frac{|c\ell(v_0)|}{|cv_0|} = \|\ell\|$$

so

$$0 = \lim_{c \to 0} \frac{|\ell(c\nu_0)|}{|c\nu_0|} = \lim_{c \to 0} \|\ell\| = \|\ell\|.$$

This contradicts exercise 0.6.3.

**Exercise 0.6.7.** If  $\ell : \mathbb{R}^n \to \mathbb{R}^n$  is an invertible linear map, show that

$$|\ell(v)| \ge \frac{|v|}{\|\ell^{-1}\|}$$

for all  $v \in \mathbb{R}^n$ . Conclude that  $\|\ell\| \ge \|\ell^{-1}\|^{-1}$ .

**Exercise 0.6.8** (Submultiplicativity). Suppose  $\ell : \mathbb{R}^m \to \mathbb{R}^n$  and  $\ell' : \mathbb{R}^n \to \mathbb{R}^p$  are both linear maps. Show that

$$\|\ell'\circ\ell\|\leqslant\|\ell'\|\cdot\|\ell\|.$$

Give an example to show that this inequality can be strict.

Here is one reason for preferring the max norms over the euclidean norms: there's an easy formula for the max operator norm in terms of the entries of a matrix representation.

**Proposition 0.6.9.** Let  $\ell : \mathbb{R}^m \to \mathbb{R}^n$  be a linear map and let  $A = [\ell]$  be its standard matrix

representation. Then  $\|\ell\|_{\infty}$  is the maximum absolute row sum of A. In other words,

$$\|\ell\|_{\infty} = \max_{j} \sum_{i=1}^{m} |a_{j,i}|$$

where  $a_{j,i}$  denotes the (j,i)-entry of A (ie, the entry in row j and column i).

Proof. Let M denote the maximum absolute row sum of A. In other words,

$$M = \max_{j} \sum_{i=1}^{m} |a_{j,i}|.$$

If  $h=(h_1,\ldots,h_m)\in \mathbb{R}^m$  and  $|h|_{\infty}\leqslant 1,$  then  $|h_i|\leqslant 1$  for all i and

$$|\ell(\mathbf{h})|_{\infty} = \max_{j} \left| \sum_{i=1}^{n} a_{j,i} \mathbf{h}_{i} \right| \leq \max_{j} \sum_{i=1}^{m} |a_{j,i}| = \mathbf{M}.$$
(0.6.10)

Taking the supremum over all h such that  $|h|_{\infty} \leq 1$  shows that  $||\ell||_{\infty} \leq M$ .

To prove equality, we will explicitly construct an  $h \in \mathbb{R}^m$  with  $|h|_{\infty} = 1$  such that  $|\ell(h)|_{\infty} = M$ . Let k be an integer (between 1 and n) which achieves the maximum absolute row sum of A. In other words, k is an integer such that

$$\sum_{i=1}^{m} |a_{k,i}| = M.$$

For each  $i = 1, \ldots, m$ , let

$$h_i = \begin{cases} 1 & \text{if } a_{k,i} \geqslant 0 \\ -1 & \text{if } a_{k,i} < 0 \end{cases}$$

and then consider the vector  $h = (h_1, ..., h_m) \in \mathbb{R}^m$ . Observe that  $|h|_{\infty} = 1$ , and notice that the kth entry of  $\ell(h) = Ah$  is

$$\sum_{i=1}^m a_{k,i}h_i = \sum_{i=1}^m |a_{k,i}| = M$$

by choice of h. Thus we have

$$M \leqslant |\ell(h)|_{\infty} \leqslant M$$

where the first inequality is from the definition of the max norm, and the second inequality

is from equation (0.6.10). Thus  $|\ell(h)|_{\infty} = M$ , completing the proof.

### 0.6.B Operator norm as a metric \*

Since the operator norm is a norm by exercise 0.6.3, we can regard  $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$  as a metric space by exercise 0.3.2. It turns out that many natural maps are continuous when we do this.

**Exercise 0.6.11** ("Evaluation is continuous"). Suppose  $v_0 \in \mathbb{R}^m$  is a fixed vector. Then the "evaluate at  $v_0$ " function  $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n) \to \mathbb{R}^n$  given by  $\ell \mapsto \ell(v_0)$  is continuous.

*Possible hint*. Since the "evaluate at  $v_0$ " function is linear, it is sufficient to check continuity at 0.

Exercise 0.6.12 ("Composition is continuous"). Show that the map

$$\mathcal{L}(\mathbb{R}^{n},\mathbb{R}^{p})\times\mathcal{L}(\mathbb{R}^{m},\mathbb{R}^{n})\longrightarrow\mathcal{L}(\mathbb{R}^{m},\mathbb{R}^{p})$$

given by  $(\ell', \ell) \mapsto \ell' \circ \ell$  is continuous.

**Exercise 0.6.13.** Suppose  $\ell : \mathbb{R}^m \to \mathbb{R}^n$  is linear and  $A = [\ell]$  is its standard matrix representation. Prove that

$$|A|_{\infty} \leq \|\ell\|_{\infty} \leq \mathfrak{m}|A|_{\infty}$$

*Possible hint*. Use proposition 0.6.9.

**Exercise 0.6.14.** Show that the isomorphism  $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n) \to \mathcal{M}_{n \times m}$  of lemma 0.5.9, given by  $\ell \mapsto [\ell]$ , is a homeomorphism.

Under the homeomorphism  $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^n) \to \mathcal{M}_{n \times n}$  of exercise 0.6.14, the set  $GL(\mathbb{R}^n)$  of invertible linear maps  $\mathbb{R}^n \to \mathbb{R}^n$  corresponds to the set  $GL_n$  of invertible  $n \times n$  matrices; since  $GL_n$  is an open subset of  $\mathcal{M}_{n \times n}$  by corollary 0.5.17, we conclude that  $GL(\mathbb{R}^n)$  is an open subset of  $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ . Similarly, the following is a consequence of combining lemma 0.5.18 and exercise 0.6.14.

**Exercise 0.6.15.** Show that the map  $GL(\mathbb{R}^n) \to GL(\mathbb{R}^n)$  given by  $\ell \mapsto \ell^{-1}$  is a homeomorphism.

# 0.7 Topological spaces

This is a bare-bones introduction to topological spaces, intended for readers who have seen metric spaces but not topological spaces before.

## 0.7.A Basics

**Definition 0.7.1.** Let X be a set. A *topology* on X is a set  $\tau_X$  of subsets of X, called *open* subsets, which satisfy the following axioms.

(T1)  $\emptyset$  and X are both open subsets.

(T2) The union of an arbitrary collection of open subsets is open.

(T3) The intersection of a finite collection of open subsets is open.

Also, a subset F such that  $X \setminus F$  is open is called a *closed* subset of X. A pair  $(X, \tau_X)$  consisting of a set X and a topology  $\tau$  on X is called a *topological space*. Often, we write simply "X" in place of the pair  $(X, \tau_X)$ .

#### Examples

Topological spaces come up throughout mathematics, and in general can be very very bizarre. We will not delve into the general theory of topological spaces, and won't see any of these bizarre examples. Instead, you are encouraged to content yourself with the following examples and constructions.

Here are some examples we've already seen.

**Example 0.7.2** (Metric spaces). If X is a metric space, the collection of subsets of X that are open with respect to the metric defines a topology on X. Thus, every metric space can be regarded as a topological space in a natural way. It's worth noticing that this process of regarding a metric space as a topological space "forgets information," since equivalent metrics on a set will define the same topology.

**Example 0.7.3.** Suppose V is a finite dimensional vector space. Then all norms on V are equivalent by theorem 0.3.12, so they all define the same topology on V by example 0.7.2 and exercise 0.3.8. This topology is called the *canonical topology* on V. Unless explicitly specified otherwise, we will always regard finite dimensional vector spaces as topological spaces using the canonical topology.

**Example 0.7.4.** Suppose X is any set. The *discrete topology* on X is the one where *all* subsets of X are declared to be open.

Here are a few ways of producing new topological spaces out of ones we already have.

**Example 0.7.5** (Subspace). If X is a topological space and S is a subset, then the *subspace topology* on S is

$$\tau_{S} := \{S \cap U : U \in \tau_{X}\}.$$

Further, S equipped with the subspace topology is called a *subspace* of X.

**Example 0.7.6** (Product spaces). Suppose X and Y are topological spaces. We define a topology  $\tau$  on the cartesian product X × Y by declaring a subset U to be open if it is a union of sets of the form V × W where U and V are open subsets of X and Y, respectively. This is called the *product topology* on X × Y.

**Example 0.7.7** (Quotient spaces). Suppose Y is any topological space, and let ~ be an equivalence relation on Y. For an element  $y \in Y$ , let [y] denote its equivalence class, and let  $X = Y / \sim$  denote the set of all equivalence classes. There is a natural function  $\pi : Y \to X$  given by sending a point  $y \in Y$  to its equivalence class [y]. We define a topology on X, called the *quotient topology*, by declaring a subset  $U \subseteq X$  to be open if and only if its preimage

$$\pi^{-1}(\mathbf{U}) = \{ \mathbf{y} \in \mathbf{Y} : [\mathbf{y}] \in \mathbf{U} \}$$

is open in Y.

And here is how subspaces and product spaces interact with metric spaces and finite dimensional vector spaces.

**Exercise 0.7.8.** Suppose X is a metric space and S is a subset. We can then regard X as a topological space as in example 0.7.2 and then we have a subspace topology  $\tau_1$  on S. On the other hand, we can restrict the metric on X to a metric on S to regard S itself as a metric space, and then forget the metric and just remember the topology  $\tau_2$  on S as in example 0.7.2. Show that  $\tau_1 = \tau_2$ .

**Exercise 0.7.9.** Suppose V is a finite dimensional vector space and U is a subspace. If we regard V as a topological space with the canonical topology of example 0.7.3, we can then give U the subspace topology  $\tau_1$ . On the other hand, we can also regard U as a vector space in its own right and give U the canonical topology  $\tau_2$ . Show that  $\tau_1 = \tau_2$ .

#### **Continuous functions**

**Definition 0.7.10** (Continuous functions). If X and Y are topological spaces, a function  $f : X \to Y$  is *continuous* if  $f^{-1}(U)$  is an open subset of X whenever U is an open subset of Y.

Continuous functions are stable under composition.

**Exercise 0.7.11.** If  $f : X \to Y$  and  $g : Y \to Z$  are continuous functions between topological spaces, then  $g \circ f$  is also continuous.

The constructions we discussed above come equipped with associated continuous maps.

- If X is a topological space and S is a subspace, the inclusion map  $i : S \to X$  is continuous.
- Suppose X and Y are topological spaces. Then the maps  $\pi_X : X \times Y \to X$  given by  $\pi_X(x, y) = x$  and  $\pi_Y : X \times Y \to Y$  given by  $\pi_Y(x, y) = y$  are both continuous.
- If Y is a topological space, ~ an equivalence relation on Y, and X = Y/ ~ the quotient space as in example 0.7.7 above, then the map  $\pi : Y \to X$  which carries a point  $y \in Y$  to its equivalence class is continuous.

The canonical topology on a finite dimensional vector space we discussed in example 0.7.3 above also has some nice continuity properties.

**Example 0.7.12.** If V is a finite dimensional vector space, then the addition map  $V \times V \rightarrow V$  and the scalar multiplication map  $\mathbb{R} \times V \rightarrow V$  are continuous functions (where V has the canonical topology and  $V \times V$  and  $\mathbb{R} \times V$  have the product topologies 0.7.6). This statement is precisely the same as exercises 0.3.3 and 0.3.4.

**Lemma 0.7.13.** Suppose  $\ell : V \to W$  is a linear map between two finite dimensional vector spaces. Then  $\ell$  is automatically continuous (with respect to the canonical topologies on V and W).

*Proof.* Choose a basis  $w_1, \ldots, w_r$  for  $\ell(V)$ , and extend it to a basis  $w_1, \ldots, w_n$  for W. For each  $w_1, \ldots, w_r$ , choose vectors  $v_1, \ldots, v_r$  such that  $\ell(v_i) = w_i$ . Linear independence of  $w_1, \ldots, w_r$  guarantees linear independence of  $v_1, \ldots, v_r$ . Choose a basis  $v_{r+1}, \ldots, v_m$  for the null space ker( $\ell$ ). Then  $v_1, \ldots, v_r, v_{r+1}, \ldots, v_m$  is a basis for V (exercise). Now consider the max norms with respect to the bases  $v_1, \ldots, v_m$  on V and  $w_1, \ldots, w_n$  on W, as in example 0.3.9. If  $v = a_1v_1 + \cdots + a_nv_m$ , we have

$$|\ell(v)|_{\infty} = |a_1w_1 + \dots + a_rw_r| = \max\{|a_1|, \dots, |a_r|\} \leq \max\{|a_1|, \dots, |a_n|\} = |v|_{\infty}$$

so exercise 0.3.5 tells us that  $\ell$  is continuous with respect to the max norm. Since the topologies determined by the max norms on V and W are precisely the canonical topologies (cf. example 0.7.3), we are done.

#### Open maps and homeomorphisms

**Definition 0.7.14** (Open map). Suppose X and Y are topological spaces and  $f : X \to Y$  is a function. Then f is *open* if f(U) is an open subset of Y whenever U is an open subset of X.

**Definition 0.7.15** (Homeomorphisms). Suppose X and Y are topological spaces and  $f : X \to Y$  is a function. Then f is a *homeomorphism* if it is bijective, continuous, and  $f^{-1} : Y \to X$  is continuous. More generally, f is a *homeomorphism onto its image* if is injective, continuous, and  $f^{-1} : f(X) \to X$  is continuous (where f(X) is regarded as a topological spaces with the subspace topology 0.7.5 that it inherits from Y).

Here is how these two notions are related to one another.

**Exercise 0.7.16.** Suppose  $f : X \to Y$  is injective and continuous.

- (a) Suppose f is open. Show that f is a homeomorphism onto its image.
- (b) Suppose  $f : X \to Y$  is a homeomorphism onto its image and that f(X) is an open subset of Y. Show that f is open.

**Exercise 0.7.17.** Let  $f : \mathbb{R} \to \mathbb{R}^2$  be the function f(x) = (x, 0). Show that f is *not* open, but that it *is* a homeomorphism onto its image.

**Example 0.7.18.** Suppose  $\ell : V \to W$  is an isomorphism of finite dimensional vector spaces. Then  $\ell$  is continuous by lemma 0.7.13. But  $\ell^{-1} : W \to V$  is also linear and hence continuous (again, by lemma 0.7.13), so  $\ell$  is a homeomorphism.

It's also possible to have continuous and open maps which are not homeomorphisms onto their images; such maps are necessarily not injective. Here is the key example.

**Exercise 0.7.19.** Let X and Y be topological spaces. Regard X × Y as a topological space with the product topology 0.7.6 and let  $\pi : X \times Y \to X$  be the projection map  $\pi(x, y) = x$ . Show that  $\pi$  is open.

#### 0.7.B Hausdorff spaces \*

**Definition 0.7.20** (Hausdorff). A topological space X is *Hausdorff* if, for every pair of distinct points  $x, y \in X$ , there exist disjoint open subsets U and V containing x and y, respectively.

**Example 0.7.21.** If X is a metric space, the corresponding topological space as in example 0.7.2 is automatically Hausdorff.

However, quotient spaces need not be Hausdorff.

**Exercise 0.7.22.** Let  $Y = \mathbb{R} \times \{\pm 1\}$  inside  $\mathbb{R}^2$ , and define an equivalence relation ~ by declaring that  $(a, 1) \sim (a, -1)$  for all  $a \neq 0$ . Show that  $X = Y/ \sim$  is not Hausdorff.

#### 0.7.C Compact and $\sigma\text{-compact spaces}$ $\star$

**Definition 0.7.23** (Open covers). An *open cover* of a topological space X is a collection  $\mathcal{U}$  of open subsets of X which cover X, in the sense that

$$\bigcup_{u\in \mathfrak{U}} u = x$$

A *subcover* of  $\mathcal{U}$  is a subset  $\mathcal{U}' \subseteq \mathcal{U}$  which is itself an open cover.

**Definition 0.7.24** (Compact). A topological space X is *compact* if every open subcover has a finite subcover.

**Definition 0.7.25** ( $\sigma$ -compact). A topological space X is  $\sigma$ -compact if it is a countable union of compact subspaces.

**Example 0.7.26.**  $\mathbb{R}^n$  is  $\sigma$ -compact, since

$$\mathbb{R}^n = \bigcup_{k=1}^{\infty} [-k,k]^n$$

and each hypercube  $[-k, k]^n$  is compact.

# 1 Single variable derivatives

The goal in this chapter is to formalize derivatives of functions whose input is a single real number and whose output is also a single real number. Probably many aspects of this theory will be familiar to you from your first exposure to single variable calculus. While there will likely be more formality here than is typical of introductory calculus courses, I strongly encourage you to use the intuition you developed during your calculus course as you go through this chapter.

# 1.1 Two definitions of the derivative

#### 1.1.A Slope of the tangent line

Let S be a subset of  $\mathbb{R}$  and  $a \in S$  an interior point. We would like to define the tangent line to the graph of a function  $f : S \to \mathbb{R}$  at a. Intuitively, the "tangent line" should be a line that "just barely touches" the graph of f at the point (a, f(a)). But it's not immediately clear how to formalize this intuitive definition.

One approach to formalization begins with noticing that this intuitive "tangent line" can be approximated by secant lines, and secant lines can be formalized without ambiguity. More precisely, for small values of h, the point a + h is close to a and still in S since a is an interior point of S. The slope of the secant line passing through the two points (a, f(a))and (a + h, f(a + h)) is computed by the quantity

$$\frac{f(a+h)-f(a)}{h},$$

since the "rise" is f(a + h) - f(a), and the "run" is (a + h) - a = h. This quantity is often called a *difference quotient*. See figure 1.1.1.

As h gets smaller and smaller, the secant line becomes a better and better approximation of our intuitive idea of the "tangent line" at a. So, we *define* the slope of the tangent line to be the limit of the slopes of the secant lines.

#### 1 Single variable derivatives



Figure 1.1.1: The graph of a function f is depicted in black. The secant line passing through the two points (a, f(a)) and (a + h, f(a + h)) is depicted in red. The "rise" between these points is f(a + h) - f(a), and the "run" is h. Thus the slope of the secant line is precisely the difference quotient (f(a + h) - f(a))/h.

**Definition 1.1.2.** A function  $f : S \to \mathbb{R}$  is *differentiable* at an interior point  $a \in S$  if

$$\lim_{h\to 0}\frac{f(a+h)-f(a)}{h}$$

exists. If this limit does exist, its value is called the *derivative of* f *at* a. We will usually denote this derivative as f'(a), but sometimes also as one of the following, where x is the variable that is being used to denote the argument to f.

$$\left. \frac{\mathrm{d}f}{\mathrm{d}x} \right|_{x=a} \quad \left. \frac{\mathrm{d}f}{\mathrm{d}x}(a) - \frac{\mathrm{d}}{\mathrm{d}x}f(x) \right|_{x=a}$$

**Exercise 1.1.3.** Use the definition of the derivative to determine whether or not each of the following functions f is differentiable at the given point a. If it is differentiable, calculate f'(a). Check your answers by making sure they agree with geometric intuition and/or rules you learned during your introductory calculus course.

- (a) f(x) = |x| at a = 0.
- (b) f(x) = x at a = 0.
- (c)  $f(x) = x^2$  at a = 0.
- (d)  $f(x) = x^2$  at a = 2.

(e) 
$$f(x) = \begin{cases} 1 & x \ge 0 \\ 0 & x < 0 \end{cases}$$
 at  $a = 0$ .

**Exercise 1.1.4** (Power rule for positive integer exponents). Prove that, if  $f(x) = x^n$  for a positive integer n, then  $f'(a) = na^{n-1}$ .

Possible hint. Recall the binomial theorem, which says that

$$(\mathfrak{a}+\mathfrak{h})^{\mathfrak{n}} = \sum_{k=0}^{\mathfrak{n}} \binom{\mathfrak{n}}{k} \mathfrak{a}^{k} \mathfrak{h}^{\mathfrak{n}-k}.$$

**Exercise 1.1.5** (Constant functions rule). Prove that, if  $f : S \to \mathbb{R}$  is *constant* (ie, there exists a real number c such that f(x) = c for all  $x \in S$ ), then f'(a) = 0 for all interior points  $a \in S$ .

**Exercise 1.1.6** (Differentiability implies continuity). Prove that if  $f : S \to \mathbb{R}$  is differentiable at an interior point  $a \in S$ , then f is also continuous at a.

**Exercise 1.1.7.** Suppose  $f : S \to \mathbb{R}$  is a function and  $a \in S$  is an interior point.

(a) If f is differentiable at a, must it be the case that

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a-h)}{2h}?$$

(b) If

$$\lim_{h\to 0}\frac{f(a+h)-f(a-h)}{2h}$$

exists, must it be the case that f is differentiable at a?

(c) What if we assume that

$$\lim_{h\to 0}\frac{f(a+h)-f(a-h)}{2h}$$

exists and that f is continuous at a?

Occasionally, it's useful to talk about differentiability from just one side.

**Definition 1.1.8** (One-sided differentiability). Suppose  $f : S \to \mathbb{R}$  is a function and  $a \in S$  is an interior point. Then f is *left differentiable at* a if

$$\lim_{h\to 0^-} \frac{f(a+h) - f(a)}{h}$$

exists, and then the value of this limit is called the *left derivative* of f at a. Similarly, f is *right differentiable at* a if

$$\lim_{h\to 0^+} \frac{f(a+h) - f(a)}{h}$$

exists, and then the value of this limit is called the *right derivative* of f at a.

**Exercise 1.1.9.** Suppose  $r \neq s$  are two distinct real numbers and let  $f : \mathbb{R} \to \mathbb{R}$  be the function

$$f(x) = \begin{cases} rx & \text{if } x \ge 0\\ sx & \text{if } x < 0 \end{cases}$$

Show that f is not differentiable at 0, but that it is both left differentiable and right differentiable at 0.

It's also useful to talk about differentiability not just at a single point, but everywhere on the domain.

**Definition 1.1.10.** A function  $f : S \to \mathbb{R}$  is *differentiable* if it is continuous, and it is differentiable at every interior point of S. Letting S° denote the interior of S, the function  $f' : S^{\circ} \to \mathbb{R}$  defined by  $a \mapsto f'(a)$  is called the *derivative* of f. Sometimes also use one of the following notations in place of f', where x is the variable that is being used to denote the argument to f.

$$\frac{df}{dx} \quad \frac{d}{dx}f(x)$$

*Caution.* Some authors define differentiability on a non-open subset slightly differently, so you should be cautious when you're reading other texts and see assertions about a function on a non-open subset of  $\mathbb{R}$  being "differentiable." Because of the potential for misunderstanding, we'll apply the above definition sparingly for non-open subsets. Specifically, we will only use this definition either when S is open, or when it is an interval.<sup>1</sup>

<sup>&</sup>lt;sup>1</sup>By *interval*, we mean an uncountable connected subset of  $\mathbb{R}$ . This means that intervals can be open, closed, or half-open; and they can be bounded or unbounded.

#### 1.1.B Best linear approximation

Let's now reinterpret the definition of the derivative in a way that will be better suited to generalization to the multivariable setting in chapter 2.

Suppose a function  $f : S \to \mathbb{R}$  is differentiable at an interior point  $a \in S$ . As we saw in the previous section, the tangent line to the graph at the point (a, f(a)) is a line with slope f'(a) passing through the point (a, f(a)). In other words, it is the graph of the function  $t : \mathbb{R} \to \mathbb{R}$  given by

$$\mathbf{t}(\mathbf{x}) = \mathbf{f}'(\mathbf{a})(\mathbf{x} - \mathbf{a}) + \mathbf{f}(\mathbf{a}).$$

Since t is "close" to f near a, we should have that

$$f(a+h) - f(a) \approx t(a+h) - t(a) = f'(a)h$$

for small values of h. See figure 1.1.11.



Figure 1.1.11: The graph of a function f is depicted in black, and its tangent line t at a is depicted in red. Then the vertical distance f(a+h) - f(a) is approximated by the vertical distance t(a+h)-t(a). The "rise" t(a+h)-t(a) can be computed as the slope times the "run." In other words, t(a+h) - t(a) = f'(a)h. As h gets smaller, this approximation of f(a+h) - f(a) gets better.

#### 1 Single variable derivatives

Notice that the function  $h \mapsto f'(a)h$  is a linear map  $\mathbb{R} \to \mathbb{R}$ , in the sense of linear algebra. The following result says that the function  $h \mapsto f'(a)h$  is the *only* linear map that is a good approximation to  $h \mapsto f(a + h) - f(a)$ . The statement uses little-oh notation; if you have not seen this before, take a look at definition 0.1.2.

**Proposition 1.1.12.** A function  $f : S \to \mathbb{R}$  is differentiable at an interior point  $a \in S$  if and only if there exists a linear map  $\ell : \mathbb{R} \to \mathbb{R}$  such that

$$|f(a+h) - f(a) - \ell(h)| = o(|h|) \text{ as } h \to 0.$$

*Moreover, if there exists such a linear map l, then l is uniquely determined: it must be given by* 

$$\ell(h) = f'(a)h$$

for all  $h \in \mathbb{R}$ .

*Proof.* Let  $\ell$  be a linear map  $\ell(h) = ch$ , where c is some scalar. Observe that

$$\frac{|f(a+h) - f(a) - \ell(h)|}{|h|} = \left| \frac{f(a+h) - f(a) - ch}{h} \right|$$
$$= \left| \frac{f(a+h) - f(a)}{h} - c \right|.$$

This leads us to the following sequence of "if and only if" statements.

$$\begin{aligned} |f(a+h) - f(a) - \ell(h)| &= o(|h|) \iff \lim_{h \to 0} \frac{|f(a+h) - f(a) - \ell(h)|}{|h|} = 0 \\ \iff \lim_{h \to 0} \left| \frac{f(a+h) - f(a)}{h} - c \right| &= 0 \\ \iff \lim_{h \to 0} \left( \frac{f(a+h) - f(a)}{h} - c \right) = 0 \\ \iff \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} = c \end{aligned}$$

The result follows.

**Definition 1.1.13.** If  $f : S \to \mathbb{R}$  is differentiable at an interior point  $a \in S$ , we define  $df_a : \mathbb{R} \to \mathbb{R}$  to be the linear map  $df_a(h) = f'(a)h$ . The function  $df_a$  is sometimes called the *total derivative* or the *differential of f at a*.
Thus proposition 1.1.12 says that  $df_a$  is the only good linear approximation to the function

$$\mathbf{h} \mapsto \mathbf{f}(\mathbf{a} + \mathbf{h}) - \mathbf{f}(\mathbf{a}).$$

Being the *only good* linear approximation, it is also the *best* linear approximation, whence the name of this section. The following is an important version of this same idea that comes up frequently in proofs.

**Remark 1.1.14** ("Error in approximation" function). Suppose  $f : S \to \mathbb{R}$  is differentiable at an interior point  $a \in S$ . For sufficiently small values of h, we can define a "error in approximation" function r by

$$\mathbf{r}(\mathbf{h}) = \mathbf{f}(\mathbf{a} + \mathbf{h}) - \mathbf{f}(\mathbf{a}) - \mathbf{d}\mathbf{f}_{\mathbf{a}}(\mathbf{h}).$$

See figure 1.1.15. Rearranging the definition of r, we have

$$f(a+h) = f(a) + df_a(h) + r(h).$$

This equation just says that f(a + h) is the sum of the value of the tangent line (which is  $f(a) + df_a(h)$ ) and the error (which is r(h)). Notice that r(0) = 0. Moreover, f is differentiable hence continuous at a (cf. exercise 1.1.6), so r is also continuous at 0. Finally, r is "small." More precisely, we know from proposition 1.1.12 that |r(h)| = o(|h|) as  $h \to 0$ , which means that

$$\lim_{h\to 0}\frac{r(h)}{h}=0.$$

The following weak version of l'Hôpital's rule is an application of this idea.

**Exercise 1.1.16** (L'Hôpital's rule, weak version). Suppose  $f, g : S \to \mathbb{R}$  are differentiable at an interior point  $a \in S$ , that f(a) = g(a) = 0, and that  $g'(a) \neq 0$ . Show that

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{f'(a)}{g'(a)}.$$

*Possible hint*. If x = a + h, we have

$$f(x) = f(a+h) = f(a) + df_a(h) + r(h) = f'(a)h + r(h) = \left(f'(a) + \frac{r(h)}{h}\right)h,$$

where r is the "error in approximation" function for f, as defined in remark 1.1.14.

# 1 Single variable derivatives



Figure 1.1.15: The thick black curve is the graph of a function f and the thinner black line is its tangent line  $f(a) + df_a$  are depicted in black. If r is the "error in approximation" function, then r(h) is the the vertical distance indicated in red above.

# 1.2 Computing derivatives

In this section, we'll prove some of the basic rules of differentiation that one normally encounters in an introductory calculus course.

# 1.2.A Sum and scalar multiples rule

The following results can be proved using either the definition of the derivative 1.1.2 or the "best linear approximation" characterization of proposition 1.1.12. I strongly encourage you to try both methods for these exercises. When you use proposition 1.1.12, you may find it useful to have done exercise 0.1.6.

**Exercise 1.2.1** (Sum rule). Prove that, if  $f, g : S \to \mathbb{R}$  are both differentiable at an interior point  $a \in S$ , then f + g is also differentiable at a and

$$(f+g)'(a) = f'(a) + g'(a).$$

**Exercise 1.2.2** (Scalar multiples rule). Prove that, if c is a constant and  $f : S \to \mathbb{R}$  is differentiable at an interior point  $a \in S$ , then cf is also differentiable at a and

$$(cf)'(a) = cf'(a).$$

# 1.2.B Product and quotient rules

**Proposition 1.2.3** (Product rule). *Suppose*  $f, g : S \to \mathbb{R}$  *are both differentiable at*  $a \in S$ . *Then* fg *is also differentiable at* a *and* 

$$(fg)'(a) = g(a)f'(a) + f(a)g'(a).$$

*Proof.* The proof is a clever algebraic manipulation of difference quotients. The key trick is introducing a "cross term" of the form f(a)g(a + h), and then cancelling it out by also

introducing its negative. More precisely, for any sufficiently small real number h, we have

$$\frac{(fg)(a+h) - (fg)(a)}{h} = \frac{f(a+h)g(a+h) - f(a)g(a)}{h}$$
$$= \frac{f(a+h)g(a+h) - f(a)g(a+h) + f(a)g(a+h) - f(a)g(a)}{h}$$
$$= \frac{f(a+h)g(a+h) - f(a)g(a+h)}{h} + \frac{f(a)g(a+h) - f(a)g(a)}{h}$$
$$= g(a+h) \cdot \frac{f(a+h) - f(a)}{h} + f(a) \cdot \frac{g(a+h) - g(a)}{h},$$

where the cross term and its negative is indicated in blue. Taking the limit as  $h \to 0$  yields the result.

**Exercise 1.2.4.** Check your understanding of the above proof of the product rule.

- (a) Could you have instead introduced a cross term of the form f(a+h)g(a)? If so, rewrite the proof using this cross term. If not, explain why not.
- (b) At what point is exercise 1.1.6 tacitly used in the above proof?

**Exercise 1.2.5.** Reprove the scalar multiples rule (exercise 1.2.2) using the product rule.

**Exercise 1.2.6** (Quotient rule). Suppose  $f, g : S \to \mathbb{R}$  are both differentiable at an interior point  $a \in S$ , and that  $g(a) \neq 0$ . Prove that f/g is also differentiable at a and

$$(f/g)'(a) = \frac{g(a)f'(a) - f(a)g'(a)}{g(a)^2}.$$

Possible hint. Observe that

$$\frac{(f/g)(a+h)-(f/g)(a)}{h}=\frac{\frac{f(a+h)}{g(a+h)}-\frac{f(a)}{g(a)}}{h}.$$

Multiply both the numerator and denominator by g(a + h)g(a) to clear the denominators inside the numerator, and then introduce the cross term f(a)g(a). Be sure to notice at what point you use exercise 1.1.6.

**Exercise 1.2.7** (Power rule for integer exponents). Extend the result of exercise 1.1.4 to arbitrary integer exponents.

*Possible hint*. Use the quotient rule for negative integers. The exponent zero is a special but easy case (cf. exercise 1.1.5).

# 1.2.C Chain rule

**Proposition 1.2.8** (Chain rule). Suppose that S and T are both subsets of  $\mathbb{R}$ , that  $f : S \to T$  is differentiable at an interior point  $\mathfrak{a}$ , that  $f(\mathfrak{a})$  is an interior point of T, and that  $g : T \to \mathbb{R}$  is differentiable at  $f(\mathfrak{a})$ . Then the composite  $g \circ f : S \to \mathbb{R}$  is also differentiable at  $\mathfrak{a}$ , and

$$(\mathbf{g} \circ \mathbf{f})'(\mathbf{a}) = \mathbf{g}'(\mathbf{f}(\mathbf{a}))\mathbf{f}'(\mathbf{a}).$$

We will discuss two different (but similar) proofs of this important result. But first, here is an application. We will later prove a better version of this result (cf. theorem 1.3.21).

**Exercise 1.2.9** (Derivatives of inverses). Suppose  $f : S \to \mathbb{R}$  is injective, differentiable at an interior point  $a \in S$ , and that b = f(a) is an interior point of f(S). Show that, if the inverse  $f^{-1} : f(S) \to \mathbb{R}$  is differentiable at b, then

$$(f^{-1})'(b) = \frac{1}{f'(f^{-1}(b))}.$$

# First proof

The first proof uses the definition of the derivative definition 1.1.2, but there is a subtle technicality involved; so, let us discuss it informally first before diving into the formal proof.

The idea involved is again a clever rewriting of difference quotients. We introduce a cross term, which in this case is f(a + h) - f(a). But this time we introduce this cross term "multiplicatively" rather than "additively." More precisely, we have

$$\frac{(g \circ f)(a+h) - (g \circ f)(a)}{h} = \frac{g(f(a+h)) - g(f(a))}{h}$$
$$= \frac{g(f(a+h)) - g(f(a))}{f(a+h) - f(a)} \cdot \frac{f(a+h) - f(a)}{h}.$$

The difference quotient on the right is precisely f'(a). Thus, it would be sufficient to prove

$$\lim_{h \to 0} \frac{g(f(a+h)) - g(f(a))}{f(a+h) - f(a)} = g'(f(a)).$$
(1.2.10)

This formula looks an awful lot like the difference quotient that is used to define the derivative of g at f(a). In fact, since f is continuous at a by exercise 1.1.6, we have

#### 1 Single variable derivatives

 $f(a+h)\approx f(a)+h$  for small values of h, which means that

$$\frac{g(f(a+h)) - g(f(a))}{f(a+h) - f(a)} \approx \frac{g(f(a)+h) - g(f(a))}{f(a)+h - f(a)} = \frac{(g \circ f)(a+h) - (g \circ f)(a)}{h}$$

for small values of h. Taking the limit as  $h \rightarrow 0$  on the right-hand side yields precisely g'(f(a)), so equation (1.2.10) seems like a reasonable conjecture.

The problem is that equation (1.2.10) isn't true. Issues arise when f(a + h) = f(a) for arbitrarily small values of h, which leads to an infinite sequence of divisions by 0.

**Example 1.2.11.** Let

$$f(x) = \begin{cases} x^2 \sin(1/x) & \text{if } x \neq 0\\ 0 & \text{if } x = 0 \end{cases}$$

and let  $g(x) = x^2$ . Observe that g'(f(0)) = 0. You will show later (in exercise 1.2.25) that f is in fact differentiable at 0, so these functions do satisfy the hypothesis of the chain rule. However, the limit as  $h \to 0$  of

$$\frac{g(f(a+h)) - g(f(a))}{f(a+h) - f(a)} = \frac{(h^2 \sin(1/h))^2}{h^2 \sin(1/h)}$$

does not exist. For example, the sequence  $h_k = 1/k\pi$  converges to 0 as  $k \to \infty$ , but the expression above is undefined along this sequence.

We will get around the kind of issue that is described by the above example by carefully studying the expression

$$\frac{g(f(a+h)) - g(f(a))}{f(a+h) - f(a)}.$$
(1.2.12)

Observe that, when  $f(a + h) \neq f(a)$ , then this expression is precisely the difference quotient that calculates the slope of the secant line passing through (f(a), g(f(a))) and (f(a + h), g(f(a + h))). When f(a + h) = f(a), this secant line doesn't make sense; but what does make sense is the tangent line! So we'll remove the undefinedness of expression (1.2.12) by extending it so that it takes the value g'(f(a)) whenever f(a + h) = f(a).

*First proof of the chain rule.* Consider the function  $\epsilon$  defined for small values of h by

$$\epsilon(\mathbf{h}) = \mathbf{f}(\mathbf{a} + \mathbf{h}) - \mathbf{f}(\mathbf{a}).$$

See figure 1.2.13. You will verify in exercise 1.2.16 below that  $\epsilon$  is continuous at 0.



Figure 1.2.13: The graph of f is depicted in black. The value of the function t on a small nonzero input value h is the vertical distance depicted in red.

Next, define a function  $\sigma$  by the following.

$$\sigma(k) = \begin{cases} \frac{g(f(a)+k)-g(f(a))}{k} & \text{if } k \neq 0\\ g'(f(a)) & \text{if } k = 0 \end{cases}$$

In other words,  $\sigma$  is the function that, on some small nonzero input k, outputs the slope of the secant line connecting the two points (f(a), g(f(a)) and (f(a) + k, g(f(a) + k)). See figure 1.2.14. It follows from the definition of the derivative of g at f(a) that the function  $\sigma$  is continuous at 0.

You will verify in exercise 1.2.16 below that

$$\frac{(g \circ f)(a+h) - (g \circ f)(a)}{h} = \sigma(\varepsilon(h)) \cdot \frac{f(a+h) - f(a)}{h}$$
(1.2.15)

# 1 Single variable derivatives



Figure 1.2.14: The graph of g is depicted in black. The output of the function  $\sigma$  on a small nonzero input value k is the slope of the red secant line passing through the two points (f(a), g(f(a)) and (f(a) + k, g(f(a) + k)).

for all small values of h. Thus

$$\lim_{h \to 0} \frac{(g \circ f)(a+h) - (g \circ f)(a)}{h} = \lim_{h \to 0} \sigma(\epsilon(h)) \cdot \frac{f(a+h) - f(a)}{h}$$
$$\stackrel{(*)}{=} \sigma(\epsilon(0))f'(a)$$
$$= g'(f(a))f'(a)$$

where we have used the continuity of  $\epsilon$  and  $\sigma$  and the definition of the derivative of f at a for the step labeled (\*) above. This proves the chain rule.

Exercise 1.2.16. Verify the following facts used in the proof above.

- (a) Prove that  $\epsilon$  is continuous at 0.
- (b) Prove equation (1.2.15).

*Possible hint.* You might split up your proof of equation (1.2.15) into two cases: when  $f(a + h) \neq f(a)$ , and when f(a + h) = f(a).

### Second proof

Our second proof of the chain rule will use the "best linear approximation" interpretation of derivatives provided by proposition 1.1.12.

Second proof of the chain rule. By proposition 1.1.12, it is sufficient to prove that

$$g(f(a+h)) - g(f(a)) - dg_{f(a)}(df_a(h)) = o(|h|)$$
 as  $h \to 0$ .

Define the following "error in approximation" functions, as in remark 1.1.14.

$$r(h) = f(a+h) - f(a) - df_a(h)$$
  

$$s(k) = g(f(a)+k) - g(f(a)) - dg_{f(a)}(k)$$

Since

$$f(a+h) = f(a) + df_a(h) + r(h)$$

by definition of r, we have

$$g(f(a+h)) = g(f(a) + df_{a}(h) + r(h))$$
  
= g(f(a)) + dg\_{f(a)}(df\_{a}(h) + r(h)) + s(df\_{a}(h) + r(h))  
= g(f(a)) + dg\_{f(a)}(df\_{a}(h)) + dg\_{f(a)}(r(h)) + s(df\_{a}(h) + r(h))

where we use the definition of s on input  $k = df_a(h) + r(h)$  for the second step, and the fact that  $dg_{f(a)}$  is linear for the third. Thus

$$g(f(a+h)) - g(f(a)) - dg_{f(a)}(df_a(h)) = dg_{f(a)}(r(h)) + s(df_a(h) + r(h)).$$

We therefore want to prove that the right-hand side is o(|h|).

Since

$$dg_{f(a)}(r(h)) = g'(f(a)) \cdot r(h)$$

is just a scalar multiple of r, and r = o(|h|), we see that  $dg_{f(a)}(r) = o(|h|)$  also. So it is sufficient to prove that

$$s(df_a(h) + r(h)) = o(|h|).$$

We've now hit the technical part, so let's hit pause.

To see why this is technical, and for essentially the same reasons as the previous proof, notice that we're trying to prove that

$$\lim_{h\to 0}\frac{|s(df_{\mathfrak{a}}(h)+r(h)|)}{|h|}=0.$$

To prove this, it's tempting to introduce a multiplicative cross-term as follows.

$$\frac{|s(df_{\mathfrak{a}}(\mathfrak{h})+\mathfrak{r}(\mathfrak{h})|)}{|\mathfrak{h}|} = \frac{|s(df_{\mathfrak{a}}(\mathfrak{h})+\mathfrak{r}(\mathfrak{h})|)}{|df_{\mathfrak{a}}(\mathfrak{h})+\mathfrak{r}(\mathfrak{h})|} \cdot \frac{|df_{\mathfrak{a}}(\mathfrak{h})+\mathfrak{r}(\mathfrak{h})|}{|\mathfrak{h}|}$$

But notice that  $df_a(h) + r(h) = f(a + h) - f(a)$  by definition of r, so this cross term is problematic for the same reasons as before! If f(a + h) = f(a) for h arbitrary close to 0, we have an infinite sequence of divisions by zeroes that causes problems.

We'll deal with this problem in essentially the same way as before: namely, we extend the definition of

$$\frac{|s(df_a(h) + r(h)|)}{|df_a(h) + r(h)|}$$

so that it is also defined when f(a + h) = f(a).

Second proof of the chain rule, continued. Define  $\eta(k) = s(k)/k$  for nonzero k. If we set  $\eta(0) = 0$ , the fact that s(k) = o(|k|) as  $k \to 0$  implies that  $\eta$  is continuous at 0. Then  $s(k) = \eta(k)k$  for all k. Applying this with  $k = df_a(h) + r(h)$ , we see that

$$\frac{|\mathfrak{s}(df_{\mathfrak{a}}(\mathfrak{h})+\mathfrak{r}(\mathfrak{h}))|}{|\mathfrak{h}|} = \frac{|\eta(df_{\mathfrak{a}}(\mathfrak{h})+\mathfrak{r}(\mathfrak{h}))| \cdot |df_{\mathfrak{a}}(\mathfrak{h})+\mathfrak{r}(\mathfrak{h})|}{|\mathfrak{h}|}$$
$$\leqslant |\eta(df_{\mathfrak{a}}(\mathfrak{h})+\mathfrak{r}(\mathfrak{h}))| \cdot \left(|\mathfrak{f}'(\mathfrak{a})| + \frac{|\mathfrak{r}(\mathfrak{h})|}{|\mathfrak{h}|}\right)$$

Observe that r and therefore  $df_a + r$  is continuous at 0 (cf. remark 1.1.14). Moreover, we have just seen that  $\eta$  is continuous at  $(df_a + r)(0) = 0$ . Thus the composite  $\eta \circ (df_a + r)$  is also continuous at 0. Thus

$$\lim_{h\to 0} |\eta(df_{\mathfrak{a}}(h) + r(h))| \cdot \left( |f'(\mathfrak{a})| + \frac{|r(h)|}{|h|} \right) = 0,$$

where we have used the fact that |r(h)| = o(|h|) to see that the parenthetical expression has a finite limit (namely, |f'(a)|). So, by the squeeze theorem, we can therefore conclude that  $|s(df_a(h) + r(h))| = o(|h|)$ .

Notice that in the first proof, the trick to overcoming the technicality was defining the function  $\sigma$  at k = 0. In the second proof, the trick to overcoming the same technicality was defining  $\eta$  at k = 0. These two tricks are really the same trick, as shown by the following.

**Exercise 1.2.17.** Prove that the function  $\eta$  from the second proof and the function  $\sigma$  from the first proof are related by the equation

$$\sigma(k) = g'(f(a)) + \eta(k).$$

# 1.2.D Interior extremum theorem

We first recall the following definitions.

**Definition 1.2.18.** Suppose X is a set,  $f : X \to \mathbb{R}$  is a function, and  $a \in X$ .

- a is an *absolute maximum*, or just *maximum*, of f if  $f(x) \leq f(a)$  for all  $x \in X$ .
- a is an *absolute minimum*, or just *minimum*, of f if  $f(x) \ge f(a)$  for all  $x \in X$ .
- a is a *absolute extremum* or just *extremum*, if it is either an absolute maximum or an absolute minimum.

If X is a metric space<sup>2</sup>, we can also make the following definitions.

- a is a *local maximum* of f if there exists a neighborhood U of a in X such that  $f(x) \leq f(a)$  for all  $x \in U$ .
- a is a *local minimum* of f if there exists a neighborhood U of a in X such that  $f(x) \ge f(a)$  for all  $x \in U$ .
- a is a *local extremum* if it is either a local maximum or a local minimum.

The following is a key property of derivatives that you probably remember using frequently when you were first exposed to calculus.

**Exercise 1.2.19** (Interior extremum theorem). Suppose an interior point  $a \in S$  is a local extremum of a function  $f : S \to \mathbb{R}$ . If f is differentiable at a, prove that f'(a) = 0.

<sup>&</sup>lt;sup>2</sup>For the definition of local extremums, X could also be a topological space.

*Possible hint*. Consider the following "slope of secant" function (which we've seen before, during the first proof of the chain rule in section 1.2.C).

$$\sigma(h) = \begin{cases} \frac{f(a+h) - f(a)}{h} & \text{if } h \neq 0\\ f'(a) & \text{if } h = 0 \end{cases}$$

Since f is differentiable at x = a, we know that  $\sigma$  is continuous at h = 0. If a is a local minimum, what can you say about the values of  $\sigma$  for positive values of h? Negative values of h? Then consider the case when a is a local minimum.

But remember, the converse to the interior extremum theorem 1.2.19 is *not* true.

**Exercise 1.2.20.** Give an example of a function  $f : \mathbb{R} \to \mathbb{R}$  such that f'(0) = 0, but f does not have a local extremum at 0.

# **1.2.E Non-rational functions**

Using the rules we've developed so far, we can differentiate any rational function (ie, a quotient of a polynomial by a polynomial). In this section, we will simply state some facts about differentiation of other types of functions one frequently encounters. We will prove most of these a bit later; we are stating them now so that we can discuss a richer repertoire of examples as we go along. We've already seen the need for a richer repertoire of examples in example 1.2.11.

### **Power rule**

First off, here is the most general statement of the power rule.

**Proposition 1.2.21** (Power rule for real exponents). *If* r *is any real number and*  $f : (0, \infty) \to \mathbb{R}$  *is the function*  $f(x) = x^r$ *, then* f *is differentiable and*  $f'(x) = rx^{r-1}$ .

We will prove this for *rational* exponents in exercise 1.3.22. We won't prove it in full generality, because that would require a long tangential discussion of what  $x^r$  even means for irrational r. Actually, once we have a sufficiently rigorous definition of  $x^r$  for irrational r, it turns out that proposition 1.2.21 is actually a relatively straightforward *consequence* of exercise 1.3.22.

#### Exponential, sine, and cosine functions

Next up, we have the exponential, sine, and cosine functions. These are usually defined by their power series, as follows.

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^{n}}{n!}$$

$$\cos(x) = \sum_{n=0}^{\infty} \frac{(-1)^{n} x^{2n}}{(2n)!}$$

$$\sin(x) = \sum_{n=0}^{\infty} \frac{(-1)^{n} x^{2n+1}}{(2n+1)!}$$
(1.2.22)

We'll prove that power series can be differentiated term-by-term in theorem 1.3.28. This yields the following standard facts.

- (1) If  $f(x) = \exp(x)$ , then  $f'(x) = \exp(x)$ .
- (2) If f(x) = sin(x), then f'(x) = cos(x).
- (3) If  $f(x) = \cos(x)$ , then  $f'(x) = -\sin(x)$ .

**Remark 1.2.23.** Defining the exponential, sine, and cosine functions using power series is mathematically sound, but there are numerous facts about these functions that are not clear from the power series, and one needs to prove these facts separately. For example, it's useful knowing that, if  $e := \exp(1)$ , then  $\exp(x) = e^x$ . This fact is not immediately clear from the power series definition; it requires some proof, and we won't prove it here, but you should feel free to use it. It's obviously also useful to know that sines and cosines have something to do with trigonometry. Again, this is not clear from the power series definition; it requires some proof, but you should feel free to use it.

### **Examples**

Let's now discuss some examples using these functions.

**Exercise 1.2.24.** Let  $f, g : (0, \infty) \to \mathbb{R}$  be the functions given by

$$f(x) = \frac{\sin(x)}{x}$$
 and  $g(x) = \frac{\sin(x^2)}{x}$ .

Show that f and g are both differentiable, and that

$$\lim_{x\to\infty} f(x) = 0 = \lim_{x\to\infty} g(x).$$

Then show that

$$\lim_{x\to\infty} f'(x) = 0 \quad \text{but} \quad \lim_{x\to\infty} g'(x) \text{ does not exist.}$$

The following two examples both involve the expression sin(1/x). As we will see, functions involving this expression exhibit a wild variety of interesting pathological behavior.

**Exercise 1.2.25.** (a) Prove that the function

$$f(x) = \begin{cases} \sin(1/x) & \text{if } x \neq 0\\ 0 & \text{if } x = 0 \end{cases}$$

is discontinuous at 0.

*Possible hint*. Show the following.

$$\limsup_{x \to 0^{\pm}} \sin(1/x) = 1 \qquad \liminf_{x \to 0^{\pm}} \sin(1/x) = -1$$

(b) Prove that the function

$$f(x) = \begin{cases} x \sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

is continuous, but that it is not differentiable at 0.

(c) Prove that the function

$$f(x) = \begin{cases} x^2 \sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

is differentiable, but that the derivative f' is not continuous at 0.

.

The following example, again involving sin(1/x), illustrates a caveat to the interior extremum theorem exercise 1.2.19: even if f has a local extremum at a point, it need not be that the derivative "changes sign" at that point.

**Exercise 1.2.26.** Consider the function  $f : \mathbb{R} \to \mathbb{R}$  given by

$$f(x) = \begin{cases} x^4 (2 + \sin(1/x)) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

- (a) Show that f has an absolute minimum at 0.
- (b) Show that f is differentiable and calculate the derivative f'.
- (c) Show that, on every open interval of the form (a, 0) or (0, b), the derivative f' takes on both positive and negative values.

# 1.3 Differentiable functions

In this section, we'll discuss some properties of functions which are differentiable along an entire interval.

# 1.3.A Mean value theorem

The mean value theorem 1.3.3 is an extremely important foundational result. You may remember seeing and ignoring it during your first calculus class; at least, that's what I did, and I think that's okay. Its importance largely derives from the fact that it comes up incredibly frequently when proving things formally about derivatives. Before proceeding, I encourage you to look at figure 1.3.1 and ensure that you understand the geometric content of the statement of the mean value theorem.

Rolle's theorem 1.3.2 is technically a special case of the mean value theorem, but the general case can be derived from this special case.

**Exercise 1.3.2** (Rolle's theorem). Suppose a < b and let  $f : [a, b] \to \mathbb{R}$  be a differentiable function. If f(a) = f(b) = 0, prove that there exists  $c \in (a, b)$  such that f'(c) = 0.

*Possible hint.* Any function must fall into at least one of the following three cases: (i) f is constantly zero, (ii) f takes on positive values somewhere on [a, b], and (iii) f takes on negative values somewhere on [a, b]. Deal with the first case using exercise 1.1.5, and the latter two cases using the extreme value theorem<sup>3</sup> together with exercise 1.2.19.

<sup>&</sup>lt;sup>3</sup>The *extreme value theorem* says that if K is a compact metric space and  $f : K \to \mathbb{R}$  is continuous, then f achieves its global maximum and its global minimum somewhere on K (cf. [PM91, theorems 3.17 and 6.30]).

# 1 Single variable derivatives



Figure 1.3.1: The mean value theorem 1.3.3 states that, if we draw a secant line connecting two points (a, f(a)) and (b, f(b)) on the graph of a differentiable function f, then that secant line must be parallel to the tangent line of f at some point c in between a and b. The statement of the mean value theorem only asserts existence, not uniqueness; there could be multiple points c at which this happens.



Figure 1.3.4: On the left is depicted the graph of a function f and the secant line  $\ell$ . Subtracting  $\ell$  from the picture on the left yields the picture on the right.

**Exercise 1.3.3** (Mean value theorem). Suppose a < b and let  $f : [a, b] \to \mathbb{R}$  be a differentiable function. Prove that there exists  $c \in (a, b)$  such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

*Possible hint*. Let  $\ell$  be the secant line passing through (a, f(a)) and (b, f(b)). Then consider the function  $g = f - \ell$ . See figure 1.3.4. Then g(a) = g(b) = 0, so we can apply Rolle's theorem 1.3.2 to g.

We can use the mean value theorem to prove the following "if and only if upgrade" of exercise 1.1.5.

**Proposition 1.3.5.** *Suppose* I *is an interval and that*  $f : I \to \mathbb{R}$  *is differentiable. Then* f' = 0 *if and only if* f *is constant.* 

*Proof.* We showed in exercise 1.1.5 that f being constant implies f' = 0. Conversely, suppose we have a < b in I. We want to show that f(a) = f(b). Since f is differentiable on I, we can apply the mean value theorem on the smaller closed subinterval [a, b]. In other words, there exists a point  $c \in (a, b)$  such that

$$\frac{f(b) - f(a)}{b - a} = f'(c).$$

But f'(c) = 0 by assumption, so f(a) = f(b).

Here are some more consequences.

**Exercise 1.3.6.** Suppose I is an open interval and  $f : I \to \mathbb{R}$  is a differentiable function such that there exists a number M such that  $|f'(x)| \leq M$  for all  $x \in I$ . Show that  $|f(b) - f(a)| \leq M|b - a|$  for all  $a, b \in I$ .<sup>4</sup>

**Exercise 1.3.7** (Fixed points). Suppose  $f : \mathbb{R} \to \mathbb{R}$  is a function. A *fixed point* of f is a point  $\xi$ , such that  $f(\xi) = \xi$ .

- (a) Suppose f is differentiable and  $f'(x) \neq 1$  for all  $x \in \mathbb{R}$ . Prove that f has at most one fixed point.
- (b) Suppose f is differentiable and there exists a constant M < 1 such that  $|f'(x)| \leq M$  for all  $x \in \mathbb{R}$ .<sup>5</sup> Show that f has a unique fixed point.

*Possible hint.* Pick any  $x_0 \in \mathbb{R}$ , and then inductively define  $x_{n+1} = f(x_n)$  for all n. Prove inductively that  $|x_{n+1} - x_n| \leq M^n |x_1 - x_0|$  for all n. Deduce that the sequence  $x_0, x_1, x_2, \ldots$  is Cauchy, and let  $\xi = \lim x_n$ . Then show that  $\xi$  is a fixed point of f.

(c) Suppose  $f(x) = x + 1/(1 + \exp(x))$ . Show that f has no fixed points, and that f is differentiable with 0 < f'(x) < 1 for all  $x \in \mathbb{R}$ . Explain why this does not contradict part (b).

# 1.3.B Monotonicity

We begin by recalling the following definition.

**Definition 1.3.8.** Suppose  $f : S \to \mathbb{R}$  is a function and  $x, y \in S$ .

- f is *increasing* if x < y implies  $f(x) \leq f(y)$ .
- f is *decreasing* if x < y implies  $f(x) \ge f(y)$ .
- f is *strictly increasing* if x < y implies f(x) < f(y).
- f is *strictly decreasing* if x < y implies f(x) > f(y).

Further, f is *monotone* if it is either increasing or decreasing, and *strictly monotone* if it is either strictly increasing or strictly decreasing.

<sup>&</sup>lt;sup>4</sup>Using the vocabulary of definition 0.3.13, this result can be rephrased by saying that, if f' is bounded, then  $|f(b) - f(a)| \leq ||f'||_{sup} |b - a|$ .

<sup>&</sup>lt;sup>5</sup>Using the vocabulary of definition 0.3.13, this hypothesis can be rephrased by saying that  $\|f'\|_{sup} < 1$ .

Here are some important properties of strictly monotone continuous functions.

**Exercise 1.3.9.** Suppose I is an interval and  $f : I \to \mathbb{R}$  is continuous. Prove that f is injective if and only if f is strictly monotone.

**Exercise 1.3.10.** Prove that, if I is an open interval and  $f : I \to \mathbb{R}$  is continuous and strictly monotone, then f(I) is also an open interval. Conclude that f is an open map (ie, that the image of any open subset of the domain is an open subset of  $\mathbb{R}$ ).

*Possible hint*. Suppose f is strictly increasing and I = (a, b). Define  $c = \lim_{x \to a^+} f(x)$  and  $d = \lim_{x \to b^-} f(x)$  and use the intermediate value theorem to prove that f(I) = (c, d). Then deal with the case when f is strictly decreasing.

The proofs of the following standard relationships between derivatives and monotonicity are similar to the proof of proposition 1.3.5.

**Exercise 1.3.11** (Derivatives and monotonicity). Suppose I is an interval and that  $f : I \to \mathbb{R}$  is differentiable. Prove the following.

(a)  $f' \ge 0$  if and only if f is increasing.

(b)  $f' \leq 0$  if and only if f is decreasing.

**Exercise 1.3.12** (Derivatives and strict monotonicity). Suppose I is an interval and that  $f : I \to \mathbb{R}$  is differentiable. Prove the following.

(a) If f' > 0, then f is strictly increasing.

(b) If f' < 0, then f is strictly decreasing.

It's worth noticing that exercise 1.3.12 cannot be upgraded to if and only if statements.

**Exercise 1.3.13.** Give an example of a strictly increasing differentiable function  $f : \mathbb{R} \to \mathbb{R}$  for which there exists some  $a \in \mathbb{R}$  such that f'(a) = 0.

And again, there are pathologies involving monotonicity than can be exhibited by functions involving sin(1/x). Here is an example of a function whose derivative is positive at a single point, but it is not monotone near that point.

**Exercise 1.3.14.** Let  $f : \mathbb{R} \to \mathbb{R}$  denote the function

$$f(x) = \begin{cases} x + 2x^2 \sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

- (a) Show that f is differentiable and calculate f'. In particular, show that f'(0) = 1.
- (b) Show that f is not monotone on any open interval containing 0.

# 1.3.C Darboux's theorem $\star$

Derivatives are often "close" to being continuous, but in a weird way. They cannot have "simple" discontinuities, which means that when they are discontinuous at all, they are discontinuous in wild ways. Darboux's theorem is one way in which this principle manifests itself. Before getting to the precise statement, we make the following definition.

**Definition 1.3.15.** Suppose I is an interval and  $f : I \to \mathbb{R}$  is a function. Then f *has the intermediate value property* if, for every a < b in I and every  $y \in \mathbb{R}$  strictly between f(a) and f(b), there exists  $x \in (a, b)$  such that f(x) = y.<sup>6</sup>

Using this vocabulary, the intermediate value theorem asserts that every continuous function has the intermediate value property [PM91, theorem 3.3]. Simple examples of discontinuous functions don't have the intermediate value property (cf. exercise 1.3.16 below). But, it turns out that some fairly bizarre discontinuous functions do have the intermediate value property (cf. exercise 1.3.17 below).

**Exercise 1.3.16.** Let I be an interval and  $f : I \to \mathbb{R}$  a function.

(a) Recall that f has a *removable discontinuity* at a point  $a \in I$  if

$$\lim_{x \to a} f(x)$$

exists, but the value of this limit does not equal f(a). Show that, if f has a removable discontinuity, then f does *not* have the intermediate value property.

(b) Recall that f has a *jump discontinuity* at a point  $a \in I$  if the two one-sided limits

$$\lim_{x\to a^-} f(x) \text{ and } \lim_{x\to a^+} f(x)$$

both exist, but their values are not equal to each other. Show that, if f has a jump discontinuity, then f does *not* have the intermediate value property.

<sup>&</sup>lt;sup>6</sup>Functions which have the intermediate value property are sometimes also called *Darboux functions*.

**Exercise 1.3.17.** Consider the function  $f : \mathbb{R} \to \mathbb{R}$  defined by

$$f(x) = \begin{cases} \sin(1/x) & \text{if } x \neq 0\\ 0 & \text{if } x = 0. \end{cases}$$

We know from exercise 1.2.25 that this function is discontinuous at 0. Show that it has the intermediate value property.

*Possible hint*. First show that for any interval I containing 0, we have f(I) = [-1, 1]. Then explain how this implies the intermediate value property.

Now we know that the derivative f' of a differentiable function f need not be continuous (cf. exercise 1.2.25). Darboux's theorem asserts that, even though f' might not be continuous, it must still have the intermediate value property! This means, for example, that derivatives cannot have removable or jump discontinuities (because, as we saw in exercise 1.3.16, functions with these kinds of discontinuities do not have the intermediate value property).

**Theorem 1.3.18** (Darboux). *Suppose* I *is an open interval and*  $f : I \to \mathbb{R}$  *is differentiable. Then* f' *has the intermediate value property.* 

*Proof.* Suppose a < b are two elements of I and m is a real number strictly between f'(a) and f'(b). We want to show that there exists  $c \in (a, b)$  such that f'(c) = m. This statement somewhat resembles the statement of the mean value theorem 1.3.3, and indeed, our proof will also somewhat resemble the proof we gave of the mean value theorem.

Consider the differentiable function  $g : I \to \mathbb{R}$  defined by g(x) = f(x) - mx. See figure 1.3.19. Observe that f'(c) = m if and only if g'(c) = 0. Thus we would like to show that there exists  $c \in (a, b)$  such that g'(c) = 0.

There are two cases two consider: either (i) f'(a) < m < f'(b) or (ii) f'(a) > m > f'(b). Suppose we are in case (i). By the extreme value theorem, g attains its minimum on the compact interval [a, b] at some point  $c \in [a, b]$ . We claim that c must be an interior point of [a, b], and then interior extremum theorem 1.2.19 will imply that g'(c) = 0.

If c = a were the minimum of g on [a, b], then we would have

$$\frac{g(a+h)-g(a)}{h} > 0$$

for all sufficiently small positive values of h. This would mean that  $g'(a) \ge 0$ , but we know that g'(a) = f'(a) - m < 0, which is a contradiction. Similarly, if c = b were the minimum



Figure 1.3.19: On the left is depicted the graph of a function f, the two tangent lines at a and b, and the graph of the line  $\ell(x) = mx$  for some m in between f'(a) and f'(b). Subtracting  $\ell$  from this picture yields the picture on the right.

of g on [a, b], then we would have

$$\frac{g(b+h)-g(b)}{h} < 0$$

for all sufficiently small negative values of h. This would imply that  $g'(b) \le 0$ , but we know that g'(b) = f'(b) - m > 0.

In case (ii), we instead let c be the *maximum* of g on [a, b], and argue similarly to show that c must be an interior point.

Here is an example application.

**Exercise 1.3.20.** Suppose I is an open interval and  $f : I \to \mathbb{R}$  is a differentiable function such that  $f'(x) \neq 0$  for all  $x \in I$ . Show that f is strictly monotone.

# 1.3.D Inverse function theorem

**Theorem 1.3.21.** Suppose that I is an open interval, that  $f : I \to \mathbb{R}$  is differentiable, and that  $f'(x) \neq 0$  for all  $x \in I$ . Then f is strictly monotone, the inverse function  $f^{-1} : f(I) \to \mathbb{R}$  is differentiable, and

$$(f^{-1})'(y) = \frac{1}{f'(f^{-1}(y))}$$

for all  $y \in f(I)$ .

*Proof.* Exercise 1.3.20 tells us that f is strictly monotone, so exercise 1.3.10 tells us that f is an open map and that f(I) is an open interval. In particular, f being open means that  $f^{-1}$  is continuous. To see that it is differentiable, fix  $y \in f(I)$  and let  $x = f^{-1}(y)$ . For each k, define

$$h(k) = f^{-1}(y+k) - f^{-1}(y).$$

Rearranging, this is equivalent to all of the following.

$$f^{-1}(y + k) = f^{-1}(y) + h(k)$$
  

$$f^{-1}(y + k) = x + h(k)$$
  

$$y + k = f(x + h(k))$$
  

$$k = f(x + h(k)) - y$$
  

$$k = f(x + h(k)) - f(x)$$

Thus

$$\frac{f^{-1}(y+k) - f^{-1}(y)}{k} = \frac{h(k)}{k} = \frac{h(k)}{f(x+h(k)) - f(x)} = \frac{1}{\left(\frac{f(x+h(k)) - f(x)}{h(k)}\right)}$$

Since  $f^{-1}$  is continuous at y, we have  $\lim_{k\to 0} h(k) = 0$ . Thus the denominator on the far right above tends to  $f'(x) = f'(f^{-1}(y))$  as  $k \to 0$ . Since the limit of the denominator is nonzero by assumption, the result follows from the quotient rule for limits.

**Exercise 1.3.22** (Power rule for fractional exponents). Extend the result of exercise 1.2.7 to arbitrary rational exponents (ie, to exponents that can be written in the form m/n where m and n are both integers).

**Exercise 1.3.23.** If  $f(x) = \ln(x)$ , prove that f'(x) = 1/x. You can use the facts that  $\ln : (0, \infty) \to \mathbb{R}$  is the inverse of the exponential function, and that the exponential function is its own derivative.

# 1.3.E Uniform limits \*

Derivatives interact somewhat strangely with limits. Here are some examples to illustrate that the limit of the derivatives is *not* necessarily the derivative of the limit.

**Exercise 1.3.24.** For any positive integer n, let  $f_n : \mathbb{R} \to \mathbb{R}$  be the function

$$f_n(x) = \frac{\sin(nx)}{n}.$$

(a) Show that  $f_n$  is differentiable.

(b) Show that the  $f_n$  converge uniformly to 0.

(c) Show that the derivatives  $f'_n$  do *not* converge pointwise.

**Exercise 1.3.25.** For any positive integer n, let  $f_n : \mathbb{R} \to \mathbb{R}$  be the function

$$f_n(x) = \sqrt{x^2 + \frac{1}{n}}.$$

(a) Show that  $f_n$  is differentiable.

(b) Show that the  $f_n$  converge uniformly to the absolute value function f(x) = |x|.

(c) Show that the derivatives  $f'_n$  converge pointwise but *not* uniformly.

In fact, to get well-behaved interacted between limits and derivatives, what we really need is uniform convergence *of the derivatives*.

**Theorem 1.3.26.** Suppose U is an open subset and  $f_n : U \to \mathbb{R}$  is a differentiable function for each n = 0, 1, ... such that  $\lim f_n = f$  pointwise. If the derivatives  $f'_n$  converge uniformly on compact subsets of U, then f is differentiable and  $\lim f'_n = f'$ .

*Proof.* Set  $g = \lim f'_n$ . Fix  $a \in U$  and consider the following "slope of secant" functions (as in the first proof of the chain rule in section 1.2.C).

$$\sigma_{n}(h) = \begin{cases} \frac{f_{n}(a+h) - f_{n}(a)}{h} & \text{if } h \neq 0\\ f'_{n}(a) & \text{if } h = 0 \end{cases}$$
$$\sigma(h) = \begin{cases} \frac{f(a+h) - f(a)}{h} & \text{if } h \neq 0\\ g(a) & \text{if } h = 0 \end{cases}$$

Continuity of  $\sigma_n$  and  $\sigma$  away from h = 0 is clear. At h = 0, we know that  $\sigma_n$  is continuous because  $f_n$  is differentiable at a. Note moreover that  $\sigma_n$  converges pointwise to  $\sigma$ .

In fact,  $\sigma_n$  converges *uniformly* to  $\sigma$  on a compact interval I containing 0. To see this, observe that for any m, n, and  $h \in I \setminus \{0\}$ , we have

$$\sigma_{\mathfrak{m}}(\mathfrak{h}) - \sigma_{\mathfrak{n}}(\mathfrak{h}) = \frac{(f_{\mathfrak{m}}(\mathfrak{a} + \mathfrak{h}) - f_{\mathfrak{n}}(\mathfrak{a} + \mathfrak{h})) - (f_{\mathfrak{m}}(\mathfrak{a}) - f_{\mathfrak{n}}(\mathfrak{a}))}{\mathfrak{h}} = f'_{\mathfrak{m}}(\xi) - f'_{\mathfrak{n}}(\xi),$$

where for the last step we have applied the mean value theorem to the function  $f_m - f_n$  on the closed interval between a and a + h to find  $\xi$  strictly between a and a + h. Moreover, we have  $\sigma_m(0) - \sigma_n(0) = f'_m(a) - f'_n(a)$ . In other words, for every  $h \in I$ , there exists  $\xi \in a + I := \{a + x : x \in I\}$  such that  $\sigma_m(h) - \sigma_n(h) = f'_m(\xi) - f'_n(\xi)$ . Thus

$$\|\sigma_m - \sigma_n\|_{sup,I} \leqslant \|f'_m - f'_n\|_{sup,a+I}.$$

Since the derivatives converge uniformly on the compact subset a + I, they are uniformly Cauchy. It follows from the above inequality that the  $\sigma_n$  are also uniformly Cauchy, and therefore uniformly convergent. But  $\sigma$  is the pointwise limit of the  $\sigma_n$ , so it must be the uniform limit.

Since the  $\sigma_n$  are continuous on I, and uniform limits of continuous functions are also continuous, we conclude that  $\sigma$  is also continuous. This proves that f is differentiable at a and  $f'(a) = g(a) = \lim f'_n(a)$ .

**Remark 1.3.27.** The hypotheses of this theorem can be made weaker by only assuming that the functions  $f_n$  converge *just at a single point* rather than pointwise on all of U, but one still needs to assume uniform convergence of the derivatives on compact subsets [Rud76, theorem 7.17]. The idea of the proof of this generalization is similar to the proof above, but the flow of the argument gets obscured by more technical details.

This theorem lets us prove that we can differentiate power series term-by-term.

**Theorem 1.3.28.** Suppose the power series  $\sum_{n=0}^{\infty} a_n x^n$  has a nonzero radius of convergence R. Then the function  $f: (-R, R) \to \mathbb{R}$  given by

$$f(x) = \sum_{n=0}^{\infty} \alpha_n x^n$$

is differentiable and

$$f'(x) = \sum_{n=0}^{\infty} n a_n x^{n-1}.$$

Moreover, the radius of convergence of the power series  $\sum_{n=0}^{\infty} na_n x^{n-1}$  is also R.

Proof. Let

$$f_n(x) = a_0 + a_1 x + \dots + a_n x^n$$

be the sequence of the partial sums of the power series. Then clearly  $\lim f_n = f$  pointwise. The conclusion of theorem 1.3.28 would therefore follow by applying theorem 1.3.26 to the sequence  $f_n$ , provided we first show that  $f'_n$  converges uniformly on compact subsets of (-R, R). To prove this, it is sufficient to show that the radius of convergence of the power series

$$\sum_{n=0} na_n x^{n-1}$$

is also R. Let R' denote the radius of convergence of this power series; we want to show that R' = R. Observe that

$$\frac{1}{\mathsf{R}'} = \limsup_{n \to \infty} |\mathfrak{n}\mathfrak{a}_n|^{1/(n-1)} = \limsup_{n \to \infty} (|\mathfrak{n}\mathfrak{a}_n|^{1/n})^{n/(n-1)}.$$

We know that  $\lim n^{1/n} = 1$  and that  $\lim n/(n-1) = 1$ . We also know that

$$\frac{1}{R} = \limsup_{n \to \infty} |a_n|^{1/n}.$$

Putting all of this together, we conclude that R = R'.

**Exercise 1.3.29.** Verify that, if the exponential, sine, and cosine functions are defined by their power series as in equation (1.2.22), then their derivatives are calculated by the formulas we stated in section 1.2.E.

# **1.4** C<sup>k</sup> hierarchy

### 1.4.A Continuous differentiability

We saw in section 1.3.C that the derivative of a differentiable function is always "close" to being continuous, though in a slightly bizarre way. Insisting that the derivative actually be continuous is often useful for ruling out pathological behavior.

**Definition 1.4.1.** Suppose U is an open subset of  $\mathbb{R}$  and  $f : U \to \mathbb{R}$  is differentiable. If the derivative  $f' : U \to \mathbb{R}$  is also continuous, then f is said to be *continuously differentiable*.

For example, we saw in exercise 1.3.14 that  $f(x) = x + 2x^2 \sin(1/x)$  defined a differentiable function  $f : \mathbb{R} \to \mathbb{R}$  such that  $f'(0) \neq 0$ , but which is not monotone on any interval containing that point. This is rather bizarre behavior, and it cannot happen with continuously differentiable functions.

**Exercise 1.4.2.** Suppose  $f : U \to \mathbb{R}$  is continuously differentiable, and  $a \in S$  is a point such that  $f'(a) \neq 0$ . Show that there exists an open interval containing a on which f is strictly monotone.

To appreciate the following fact, it's worth noting the following: there exist *non-constant* differentiable functions  $f : U \to \mathbb{R}$  such that f'(x) = 0 for all  $x \in U \cap \mathbb{Q}$ ! These extremely bizarre functions were first found by Pompeiu [Pom07]. This pathological behavior is also ruled out by insisting that the derivative be continuous.

**Exercise 1.4.3.** Suppose  $f : U \to \mathbb{R}$  is continuously differentiable. Show that, if f'(x) = 0 for all  $x \in U \cap \mathbb{Q}$ , then f is constant.

Unfortunately, insisting that the derivative of a function be continuous does not rule out all possible bizarre behavior. For example, in exercise 1.2.26, we saw that  $f(x) = x^4 (2 + \sin(1/x))$  defined a differentiable function f with an absolute minimum at 0, but the derivative doesn't just simply "change sign" at 0. In fact, it turns out that function is also continuously differentiable (cf. exercise 1.4.4), so merely insisting on continuous differentiability does not quite manage to rule out that kind of pathological behavior. However, as it turns out, the derivative of f' (called the "second derivative" of f, and denoted f"), turns out *not* to be continuous.

**Exercise 1.4.4.** Let  $f : \mathbb{R} \to \mathbb{R}$  be the function from exercise 1.2.26. Show that f' is differentiable (so, in particular, f is continuously differentiable), but that f'' is not continuous.

This leads us to thinking about iterated derivatives.

# **1.4.B** C<sup>k</sup> functions

**Definition 1.4.5.** Let U be an open subset of  $\mathbb{R}$  and  $f : U \to \mathbb{R}$  a function. We then define the following.

- f is said to be C<sup>0</sup> if it is continuous.
- f is said to be C<sup>1</sup> if it is continuously differentiable.

# 1 Single variable derivatives

- f is said to be C<sup>2</sup> if it is differentiable and f' is C<sup>1</sup>. In this case, the derivative (f')' of f' is called the *second derivative* of f, and is denoted either f" or f<sup>(2)</sup>.
- f is said to be  $C^3$  if it is differentiable and f' is  $C^2$ . In this case, the second derivative (f')'' of f' is called the *third derivative* of f and is denoted  $f^{(3)}$ .
- f is said to be C<sup>4</sup> if it is differentiable and f' is C<sup>3</sup>. In this case, the third derivative  $(f')^{(3)}$  of f' is called the *fourth derivative* of f and is denoted  $f^{(4)}$ .

Inductively, for any positive integer k, f is said to be  $C^k$  if it is differentiable and f' is  $C^{k-1}$ . In this case, the (k-1)st derivative  $(f')^{(k-1)}$  of f' is called the *kth derivative* of f and is denoted  $f^{(k)}$ .

First up, let's prove that all reasonable ways of combining functions in a particular differentiability class stays within that differentiability class.

**Exercise 1.4.6.** Prove that the set of all  $C^k$  functions  $U \to \mathbb{R}$  is a vector space.

*Possible hint*. Induct on k.

**Exercise 1.4.7.** (a) Prove that the product of any two  $C^k$  functions is also  $C^k$ .

(b) If f, g : U  $\rightarrow \mathbb{R}$  are C<sup>k</sup> functions, show that

$$(fg)^{(k)} = \sum_{i=0}^{k} \binom{k}{i} f^{(i)} g^{(k-i)}.$$

**Exercise 1.4.8.** Show that the composite of two C<sup>k</sup> functions is also C<sup>k</sup>.

*Possible hint*. Induct on k. The base case k = 0 is clear. For the inductive step, use the chain rule 1.2.8 and exercise 1.4.7(a).

**Exercise 1.4.9.** Suppose  $f: U \to \mathbb{R}$  is  $C^k$  and  $f(x) \neq 0$  for all  $x \in U$ . Show that 1/f is also  $C^k$ .

*Possible hint.* You might first show that the function g(x) = 1/x is  $C^k$  for all k. Then  $1/f = g \circ f$ , so you can apply exercise 1.4.8.

**Proposition 1.4.10.** Suppose I is an open interval,  $f : I \to \mathbb{R}$  is injective,  $C^k$  for some  $k \ge 1$ , and that  $f'(x) \neq 0$  for all  $x \in I$ . Then  $f^{-1} : f(I) \to U$  is also  $C^k$ .

*Proof.* Since f' is nonzero on I, we know from theorem 1.3.21 that  $f^{-1}$  is differentiable and that

$$(f^{-1})'(a) = \frac{1}{f'(f^{-1}(a))}.$$

In other words,  $(f^{-1})'$  is the composite of three functions:  $f^{-1}$ , f', and the function  $x \mapsto 1/x$ .

We will inductively prove that  $f^{-1}$  is  $C^r$  for all  $0 \le r \le k$ . Certainly  $f^{-1}$  is  $C^0$  (ie, continuous), since it is even differentiable, as we saw above. Inductively, suppose  $f^{-1}$  is  $C^r$  for some  $0 \le r < k$ . We know that f is  $C^k$  and therefore  $C^r$ , and also that  $x \mapsto 1/x$  is  $C^r$  (cf. exercise 1.4.9), so  $(f^{-1})'$  is the composite of three  $C^r$  functions. Thus exercise 1.4.8 tells us that  $(f^{-1})'$  is  $C^r$ , which shows that  $f^{-1}$  is  $C^{r+1}$ . This completes the induction.

Observe that exercise 1.1.6 tells us that we have an infinite chain of implications

$$\cdots \Longrightarrow C^3 \Longrightarrow C^2 \Longrightarrow C^1 \Longrightarrow C^0$$

The following shows that all of these implications are strict; in other words, for every k, there exists a  $C^k$  function which is not  $C^{k+1}$ .

Exercise 1.4.11. For any non-negative integer k, let

$$f_k(x) = x^k |x|.$$

See figure 1.4.12. Prove that  $f_k$  is  $C^k$ , but the (k + 1)st derivative  $f^{(k+1)}$  does not exist. *Possible hint*. First prove that  $f'_{k+1} = (k+2)f_k$ .



Figure 1.4.12: Graphs of the functions  $f_k(x) = x^k |x|$  for k = 0, 1, 2, 3.

**Exercise 1.4.13.** For any positive integer k, let  $f_k : \mathbb{R} \to \mathbb{R}$  be the function given by

$$f_k(x) = \begin{cases} x^{k+1} \sin(1/x) & \text{if } x \neq 0\\ 0 & \text{if } x = 0. \end{cases}$$

Show that the kth derivative  $f^{(k)}$  exists, but it is not continuous.

# 1.4.C Taylor's theorem **\***

Taylor's theorem gives us a way of approximating a  $C^k$  function by a polynomial of at most degree k. When k = 1, the theorem says nothing more than proposition 1.1.12.

**Theorem 1.4.14** (Taylor). Suppose I is an open interval,  $f : I \to \mathbb{R}$  is  $C^k$  for some non-negative integer k, and  $a \in I$ . Then there exists a unique polynomial  $p_k$  at most k, called the degree k Taylor polynomial of f at a, such that  $|f(a+h) - p_k(h)| = o(|h|^k)$  as  $h \to 0$ . Moreover, we have

$$p_{k}(h) = f(a) + f'(a)h + \frac{f''(a)}{2!}h^{2} + \frac{f^{(3)}(a)}{3!}h^{3} + \dots + \frac{f^{(k)}(a)}{k!}h^{k}.$$
 (1.4.15)

*Proof.* Let  $p_k(h)$  be defined as equation (1.4.15) and set  $r(h) = f(a + h) - p_k(h)$ . Observe that r is  $C^k$  by exercise 1.4.6. By direct calculation, we have  $r^{(i)}(0) = 0$  for all i = 0, 1, ..., k. If k = 0, then the fact that |r(h)| = o(1) follows from the continuity (ie,  $C^0$ -ness) of r and the fact that r(0) = 0; so we can assume for the rest of the proof that  $k \ge 1$ .

Suppose h > 0. By iterating the mean value theorem, we have

$$\begin{aligned} r(h) &= r(h) - r(0) \\ &= r'(h_1)h = (r'(h_1) - r'(0))h \\ &= r''(h_2)h_1h = (r''(h_2) - r''(0))h_1h \\ &\vdots \\ &= r^{(k-1)}(h_{k-1})h_{k-2}h_{k-3}\cdots h_2h_1h \end{aligned}$$

where  $0 < h_{k-1} < \cdots < h_2 < h_1 < h.$  This means that

$$\begin{split} \frac{|\mathbf{r}(\mathbf{h})|}{|\mathbf{h}|^{k}} &= \left| \frac{\mathbf{r}^{(k-1)}(\mathbf{h}_{k-1})\mathbf{h}_{k-2}\cdots\mathbf{h}_{2}\mathbf{h}_{1}\mathbf{h}}{\mathbf{h}^{k}} \right| \\ &\leqslant \left| \frac{\mathbf{r}^{(k-1)}(\mathbf{h}_{k-1})\mathbf{h}^{k-1}}{\mathbf{h}^{k}} \right| \\ &= \left| \frac{\mathbf{r}^{(k-1)}(\mathbf{h}_{k-1})}{\mathbf{h}} \right| \\ &\leqslant \left| \frac{\mathbf{r}^{(k-1)}(\mathbf{h}_{k-1})}{\mathbf{h}_{k-1}} \right| \\ &= \left| \frac{\mathbf{r}^{(k-1)}(\mathbf{h}_{k-1}) - \mathbf{r}^{(k-1)}(\mathbf{0})}{\mathbf{h}_{k-1}} \right|. \end{split}$$

Now as  $h \to 0^+$ , we also have that  $h_{k-1} \to 0$ , so this final expression tends to  $r^{(k)}(0) = 0$ . By the squeeze theorem, we conclude that

$$\lim_{h\to 0^+}\frac{|\mathbf{r}(h)|}{|\mathbf{h}|^k}=\mathbf{0}.$$

One argues similarly with h < 0 to prove that

$$\lim_{h\to 0^-}\frac{|\mathbf{r}(h)|}{|\mathbf{h}|^k}=\mathbf{0}.$$

In this case, we have  $h < h_1 < h_2 < \cdots < h_{k-1} < 0$ . This proves that  $|r(h)| = o(|h|^k)$ .

To prove that  $p_k$  is unique, suppose q is any polynomial of degree at most k such that  $|f(a+h)-q(h)|=o(|h|^k).$  Then

$$|p_k(h) - q(h)| \leq |f(a+h) - p_k(h)| + |f(a+h) - q(h)|$$

so exercises 0.1.6 and 0.1.7 imply that  $|p_k(h) - q(h)| = o(|h|^k)$ . But  $p_k - q$  is a polynomial of degree at most k, and a nonzero polynomial of degree at most k cannot be  $o(|h|^k)$  as  $h \to 0$ . So we must have  $p_k = q$ .

**Remark 1.4.16.** It's worth remarking that Taylor's theorem does not use the fact that  $f^{(k)}$  is continuous; it merely requires that  $f^{(k)}$  exists, ie, that f is k *times differentiable*. This is evident in the proof above.

We can now return to our discussion of the pathology exhibited by the function

 $f(x) = x^4 (2 + \sin(1/x))$  from exercises 1.2.26 and 1.4.4, and prove that such pathologies cannot occur under some hypotheses. We prepare for this discussion with the following lemma.

**Lemma 1.4.17.** Suppose I is an open interval,  $f : I \to \mathbb{R}$  is  $C^k$  for some positive integer k, that

$$f(a) = f'(a) = f^{(2)}(a) = \cdots = f^{(k-1)}(a) = 0,$$

and that  $f^{(k)}(a) > 0$ . Then there exists an  $\varepsilon > 0$  such that f is positive on  $(a, a + \varepsilon)$ , and on  $(a - \varepsilon, a)$ , f is positive if k is even and negative if k is odd. Moreover, if  $f^{(k)}(a) < 0$ , then all instances of "positive" in the conclusion get replaced by "negative" and vice versa.

*Proof.* Let us suppose that  $f^{(k)}(a) > 0$ . By Taylor's theorem 1.4.14, we see that

$$f(a+h) = \frac{f^{(k)}(a)}{k!}h^k + r(h)$$

where  $|r(h)| = o(|h|^k)$  as  $h \to 0$ . Dividing through by  $h^k$ , we see that

$$\frac{f(a+h)}{h^k} = \frac{f^{(k)}(a)}{k!} + \frac{r(h)}{h^k}.$$

Since  $f^{(k)}(a) > 0$ , we also have  $f^{(k)}(a)/k! > 0$ . Since  $\lim_{h\to 0} |r(h)/h^k| = 0$ , there exists  $\varepsilon > 0$  such that  $|r(h)/h^k| < \varepsilon$  for all  $|h| < \varepsilon$ . Then, if  $0 < |h| < \varepsilon$ , we have

$$\frac{f(a+h)}{h^k} = \frac{f^{(k)}(a)}{k!} + \frac{r(h)}{h^k} > 0.$$

Now note that if k is even, then  $h^k$  is always positive, so the lemma follows. On the other hand, if k is odd, then  $h^k$  is positive for positive h and negative for negative h, and again the lemma follows. The proof when  $f^{(k)}(a) < 0$  is similar.

**Exercise 1.4.18.** Suppose f is  $C^k$  for some  $k \ge 1$ , that

$$f'(a) = f^{(2)}(a) = \cdots = f^{(k-1)}(a) = 0$$

and  $f^{(k)}(a) \neq 0$ . Then there exists an  $\epsilon > 0$  such that f is strictly monotone on  $(a - \epsilon, a)$  and on  $(a, a + \epsilon)$ .

*Possible hint*. Apply lemma 1.4.17 to f'. Then use exercise 1.3.12.

**Example 1.4.19.** Let  $f : \mathbb{R} \to \mathbb{R}$  be the function of exercises 1.2.26 and 1.4.4. Then f is twice differentiable and f'(0) = f''(0) = 0, but it is not thrice differentiable at 0. In other words, it doesn't eventually have a nonzero derivative at 0, so we're not able to apply exercise 1.4.18.

The difference  $f(a + h) - p_k(h)$  is often called a "remainder," since it's what's left over after we approximate f by its Taylor polynomial. It turns out that if f is  $C^{k+1}$ , we can use the mean value theorem to express this remainder in terms of the (k + 1)st derivative.

**Theorem 1.4.20** (Taylor's theorem with remainder, Lagrange form). Suppose I is an open interval,  $f : I \to \mathbb{R}$  is  $C^{k+1}$  for some non-negative integer k and  $a \in U$ . Then for any h such that  $a + h \in I$ , there exists  $\xi$  between a and a + h such that

$$f(a+h) - p_k(h) = \frac{f^{(k+1)}(\xi)}{(k+1)!}h^{k+1},$$

where  $p_k$  is the degree k Taylor polynomial of f at a.

*Proof.* When k = 0, we have  $p_0(h) = f(a)$  and the statement follows immediately from the mean value theorem 1.3.3. So we can assume that  $k \ge 1$ . Let  $r(h) = f(a + h) - p_k(h)$ . Fix h > 0 and consider the function

$$g(t) = r(t) - \frac{r(h)}{h^{k+1}}t^{k+1}.$$

Plugging in t = h, we see that g(h) = 0. As we noted in the proof of Taylor's theorem 2.5.32, we have  $r^{(i)}(0) = 0$  for i = 0, 1, ..., k, from which it follows that  $g^{(i)}(0) = 0$  for i = 0, 1, ..., k. Moreover, we have

$$g^{(k+1)}(t) = f^{(k+1)}(t) - \frac{(k+1)!r(h)}{h^{k+1}},$$

because  $p_k$  is a polynomial of degree at most k, which means that its (k + 1)st derivative vanishes.

Since g(0) = g(h) = 0, Rolle's theorem 1.3.2 tells us that there exists  $h_1$  between 0 and h such that  $g'(h_1) = 0$ . Since  $g'(0) = g'(h_1) = 0$ , Rolle's theorem again gives us  $h_2$  between 0 and  $h_2$  such that  $g^{(2)}(h_2) = 0$ . Inductively, we find a sequence  $0 < h_{k+1} < h_k < h_{k-1} < \cdots < h_1 < h$  such that  $g^{(i)}(h_i) = 0$  for all  $i = 0, \dots, k, k+1$ . Letting  $\xi = h_{k+1}$ , we find that

$$0 = g^{(k+1)}(\xi) = f^{(k+1)}(\xi) - \frac{(k+1)!r(h)}{h^{k+1}},$$

and rearranging this equation yields precisely

$$f(a+h) - p_k(h) = r(h) = \frac{f^{(k+1)}(\xi)}{(k+1)!}h^{k+1}.$$

The proof when h < 0 is analogous.

# 1.4.D Smooth functions

Insisting that a function be  $C^k$  for larger and larger values of k rules out more and more pathological behavior. This suggests the following.

**Definition 1.4.21.** A function  $f : U \to \mathbb{R}$  is said to be  $C^{\infty}$ , or *infinitely differentiable*, or *smooth*, if it is  $C^k$  for all k.

It's a little annoying that there are three different words that all mean the same thing, but that's just how it is; all three are commonly used in the mathematical literature, so it's best to get used to all of them. We'll usually use the word "smooth."

It follows immediately from exercises 1.4.6 to 1.4.9 and proposition 1.4.10 that sums, scalar multiples, products, reciprocals, composites, and inverses of smooth functions are also smooth.

The class of smooth functions is very well-behaved. On the one hand, as we have already seen, smoothness rules out lots of pathological behavior. For example, functions like exercise 1.3.14 are not smooth, and in fact, if the derivative of a smooth function f is nonzero at a single point, then f must be strictly monotone in a neighborhood of that point (cf. exercise 1.4.2). We've also seen that functions like the one from exercise 1.2.26 and example 1.4.19 are not smooth, and that if f is smooth and a is a local extremum and eventually f has a nonzero derivative at a, then in fact the derivative of f must "change sign" at a (cf. exercise 1.4.18).

On the other hand, the class of smooth functions is not overly restrictive. There is a vast array of smooth functions, and we can tailor smooth functions to almost arbitrary specifications.

#### Diffeomorphisms

Our first example of this principle is the following smooth bijection  $(-1, 1) \rightarrow \mathbb{R}$  whose inverse is also smooth. A smooth bijection with a smooth inverse is also called a "diffeo-

morphism." Roughly, the fact that there exists diffeomorphism  $(-1,1) \rightarrow \mathbb{R}$  "smoothly stretch out" (-1,1) to all of  $\mathbb{R}$ , and also "smoothly shrink" all of  $\mathbb{R}$  down to (-1,1).

**Example 1.4.22.** Consider the function  $f : (-1, 1) \to \mathbb{R}$  defined by

$$f(\mathbf{x}) = \frac{\mathbf{x}}{1 - \mathbf{x}^2}.$$

See figure 1.4.23. Since this is a quotient of two smooth functions and the denominator is nonzero on (-1, 1), this function is also smooth. We can calculate that

$$f'(x) = \frac{1 + x^2}{(1 - x^2)^2}$$

and we can see from this formula that the derivative is always strictly positive, so f is strictly increasing (cf. exercise 1.3.12). In particular, it is injective. Moreover, we have

$$\lim_{x \to -1^+} f(x) = -\infty \quad \text{and} \ \lim_{x \to 1^-} f(x) = +\infty$$

so the intermediate value theorem guarantees that f is bijective. Thus the inverse function  $f^{-1} : \mathbb{R} \to (-1, 1)$  exists. Finally, since f' is always nonzero, it follows from proposition 1.4.10 that  $f^{-1}$  must be smooth. It is possible to verify that

$$f^{-1}(x) = \frac{-1 + \sqrt{4x^2 + 1}}{2x}$$

by check that composing this formula with f yields the identity, but, thanks to proposition 1.4.10, we don't actually need this formula for  $f^{-1}$  at all to know that  $f^{-1}$  is smooth.

Of course, there's nothing special about the open interval (-1, 1). In fact, we can construct a diffeomorphism between any pair of open intervals!

**Exercise 1.4.24.** (a) For real numbers a < b, construct a diffeomorphism  $f : (a, b) \rightarrow \mathbb{R}$ .

- (b) Show that the exponential function is a diffeomorphism  $\mathbb{R} \to (0, \infty)$ .
- (c) Construct a diffeomorphism  $\mathbb{R} \to (\mathfrak{a}, \infty)$  for any real number  $\mathfrak{a}$ .
- (d) Construct a diffeomorphism  $\mathbb{R} \to (-\infty, \mathfrak{a})$  for any real number  $\mathfrak{a}$ .
- (e) If I and I' are both open intervals, show that there exists a diffeomorphism  $f: I \to I'$ .

#### 1 Single variable derivatives



Figure 1.4.23: The graph of a diffeomorphism  $(-1, 1) \rightarrow \mathbb{R}$ .

# Infinitely flat functions

We next discuss an example of a smooth function which gets "infinitely flat" at 0, but is not constantly equal to 0. But first, a preliminary remark.

**Remark 1.4.25.** Here is a fact that you might recognize:

$$\lim_{u \to \infty} \frac{u^m}{e^u} = 0 \tag{1.4.26}$$

for any non-negative integer m. We'll use this fact, and also the related fact that

$$\lim_{x\to 0^+}\frac{e^{-1/x}}{x^m}=0$$

for any non-negative integer m, which follows immediately from equation (1.4.26) by making the substitution u = 1/x.

You probably remember proving equation (1.4.26) using l'Hôpital's rule in your calculus class. We haven't proved l'Hôpital's rule (except for the weak version in exercise 1.1.16, which you can check is not enough to prove equation (1.4.26)), but you can find proofs of l'Hôpital's rule in many places (eg, [PM91, theorem 4.15], [Rud76, theorem 5.13], or even Wikipedia). The proof is a clever application of the mean value theorem 1.3.3.

An alternative proof of equation (1.4.26) uses the power series definition of the exponential
function. The idea is to notice that

$$\frac{e^{\mathfrak{u}}}{\mathfrak{u}^{\mathfrak{m}}} = \frac{1}{\mathfrak{u}^{\mathfrak{m}}} + \dots + \frac{1}{(\mathfrak{m}-1)!\mathfrak{u}} + \sum_{k=0}^{\infty} \frac{\mathfrak{u}^{k}}{(\mathfrak{m}+k)!}$$

As  $u \to \infty$ , the first few summands all tend to 0, and the series at the end tends to  $\infty$ . Equation (1.4.26) follows from this; details are omitted, but you should be able to work them out yourself if you're interested.

**Exercise 1.4.27.** Consider the function  $f : \mathbb{R} \to \mathbb{R}$  defined as follows.

$$f(x) = \begin{cases} e^{-1/x} & \text{if } x > 0\\ 0 & \text{if } x \leqslant 0 \end{cases}$$

See figure 1.4.28.



Figure 1.4.28: The graph of the function f from exercise 1.4.27 that is "infinitely flat" at 0.

- (a) Use remark 1.4.25 to show that f'(0) = 0.
- (b) Check that

$$f'(x) = \begin{cases} \frac{e^{-1/x}}{x^2} & \text{if } x > 0\\ 0 & \text{if } x \leqslant 0. \end{cases}$$

You've already computed f'(0). You can compute formulas for the derivative f'(x) for x > 0 and for x < 0 using the usual rules for differentiation.

# 1 Single variable derivatives

(c) Use remark 1.4.25 to show that f''(0) = 0, and then check that

$$f''(x) = \begin{cases} p_2(x) \cdot \frac{e^{-1/x}}{x^4} & \text{if } x > 0\\ 0 & \text{if } x \leqslant 0 \end{cases}$$

where  $p_2$  is the function  $p_2(x) = 1 - 2x$ .

(d) Inductively, prove that

$$f^{(k)}(x) = \begin{cases} p_k(x) \cdot \frac{e^{-1/x}}{x^{2k}} & \text{if } x > 0\\ 0 & \text{if } x \leqslant 0 \end{cases}$$

where  $p_k$  is a polynomial function of degree k - 1 such that  $p_k(0) = 1$ .

Thus f is an example of a smooth function that is "infinitely flat" at x = 0, in the sense that  $f^{(k)}(x) = 0$  for all k, but is not constantly equal to 0.

#### **Bump functions**

Next up, we have "bump functions." We begin with the following definition.

**Definition 1.4.29** (Support of a real-valued function). Suppose X is a metric space<sup>7</sup> and  $f : X \to \mathbb{R}$  is a function. The *support* of f, denoted supp(f), is the closure of the set of points where f is nonzero. In other words,

$$\operatorname{supp}(f) = \overline{\{x \in X : f(x) \neq 0\}}.$$

A "bump function" is a smooth function with compact support.

**Example 1.4.30.** Consider the function  $\psi : \mathbb{R} \to \mathbb{R}$  given by the following.

$$\psi(\mathbf{x}) = \begin{cases} e^{-1/(1-\mathbf{x}^2)} & \text{if } \mathbf{x} \in (-1,1) \\ 0 & \text{otherwise.} \end{cases}$$

See figure 1.4.31. It is clear from the definition of  $\psi$  that the support of  $\psi$  is the closed interval [-1, 1]. Moreover, observe that, if f is the "infinitely flat" function from exercise 1.4.27 and

<sup>&</sup>lt;sup>7</sup>For the definition of support, X could more generally be a topological space.

 $p : \mathbb{R} \to \mathbb{R}$  is the polynomial  $p(x) = 1 - x^2$ , then  $\psi = f \circ p$ . Thus  $\psi$  is smooth.



Figure 1.4.31: The graph of the function  $\psi$  from example 1.4.30.

**Exercise 1.4.32.** For any a < b in  $\mathbb{R}$ , construct a bump function whose support is [a, b].

**Exercise 1.4.33.** Let F be a closed subset of  $\mathbb{R}$ . Construct a smooth function which is zero on F and nonzero outside of F.

*Possible hint*. Recall that the complement of F is a countable disjoint union of open intervals (cf. [PM91, theorem 6.17]).

# **Bridge functions**

Next up, we'll discuss an example of a "bridge" function. It will be constant on  $(-\infty, -1]$ , and also constant on  $[1, \infty)$ , but the values on these two intervals is different; and on (-1, 1), the function "smoothly bridges the gap" between the two values.

In discussing this example, we will invoke the fundamental theorem of calculus, even though we have not proved it (or even defined integrals, for that matter). If you're unhappy with using things you haven't proved, you can find rigorous discussions of integration in many places (eg, [PM91, chapter 5], [Rud76, chapter 6], etc). Alternatively, you might also be interested in [GO03, chapter 3, example 12], which gives an example of a "bridging function" that involves no integrals (but does involve a double exponential).

**Example 1.4.34.** Let  $\psi$  be the bump function from example 1.4.30, and then consider

$$\eta(x) = \int_{-1}^{x} \psi(t) \, dt.$$

# 1 Single variable derivatives

Then

$$\eta'(x) = \frac{d}{dx} \int_{-1}^{x} \psi(t) \, dt = \psi(x)$$

by the fundamental theorem of calculus. Thus  $\eta$  is smooth, since  $\eta' = \psi$  is smooth. Note moreover that  $\eta$  is increasing, since  $\eta' = \psi \ge 0$ . Finally, it is clear from the geometric interpretation of integrals as "area under the curve" that  $\eta$  is constantly equal to 0 for all  $x \le -1$  and that it is constantly equal to  $\eta(1)$  for all  $x \ge 1$ .

**Exercise 1.4.35.** For a < b in  $\mathbb{R}$ , let I = [a, b] and let U be an open subset containing I. Prove that there exists a bump function  $f : \mathbb{R} \to \mathbb{R}$  such that f(x) = 1 for all  $x \in K$  and f(x) = 0 for all  $x \notin U$ .

*Possible hint*. The characteristic function

$$\chi_I(x) = \begin{cases} 1 & \text{if } x \in I \\ 0 & \text{if } x \notin I \end{cases}$$

is close to having all of the properties we want, but it's not smooth; it's clearly not even continuous. "Bridge" the discontinuities.

In this chapter, we will study derivatives of functions  $S \to \mathbb{R}^n$  where S is a subset of  $\mathbb{R}^m$ and m and n are arbitrary positive integers. As it turns out, the most important case is when n = 1. In other words, the "multi" in the name of this chapter (as opposed to the "single" in the name of the previous chapter) is really referring to the fact that we might have multiple *inputs* (ie, m > 1), rather than to the fact that we might have multiple outputs (ie, n > 1). The single variable case is actually quite important for the multivariable case; we'll often use results from chapter 1 to prove their multivariable counterparts.

We'll begin by analyzing a particular example to build up some geometric intuition in section 2.1, before proceeding with the abstract discussion of multivariable derivatives.

# 2.1 Introductory example

Recall that we started off our discussion of single variable derivatives starting with a very geometric idea of tangent lines to graphs. For large values of m (and n), graphs become hard to visualize and it is not so clear what "tangent" should mean. But there is at least one multivariable situation where, by exerting some strain on the three-dimensional visualization sectors of our brain, we can geometrically formalize what "tangent" might mean. This is the m = 2, n = 1 situation. The graph of such a function is a surface in  $\mathbb{R}^3$ , and its "tangent" at a point should be a plane.

By analyzing a specific function  $f : \mathbb{R}^2 \to \mathbb{R}$ , we can get some valuable insights into multivariable derivatives; the analysis will presage many of the concepts we will discuss later in the chapter. Any function would do, but let's focus on the function  $f : \mathbb{R}^2 \to \mathbb{R}$  defined by

$$f(x,y) = x^2 + y^2$$

for our analysis.

#### Describing the graph of f

First off, let's try to get a solid understanding of the graph of f. The graph  $\Gamma$  is a subset of  $\mathbb{R}^2 \times \mathbb{R} = \mathbb{R}^3$  defined by

$$\Gamma = \{(x, y, z) : z = f(x, y)\} = \{(x, y, z) : z = x^2 + y^2\}.$$

It's a little hard to visualize in three dimensions immediately, so we will start by looking at two-dimensional slices of  $\Gamma$ , until we've seen enough slices that we have a sense of the three-dimensional geometry.

If we consider the "vertical" slice obtained by setting x = 0, we find a parabola  $z = y^2$ . Similarly, if we consider the slice obtained by setting x = 1, we obtain the parabola  $z = 1+y^2$ . Fixing x = 5, we obtain the parabola  $z = 25 + y^2$ . In fact, we can see that no matter what value of x we fix, the resulting slice is always a parabola, just translated up from the origin by differing amounts; see figure 2.1.1.



Figure 2.1.1: Here are three vertical slices of the graph of the function  $f(x, y) = x^2 + y^2$ . From left to right, they are the x = 0, x = 1, and x = 2 slices, respectively.

Slicing "vertically in the other direction" gives similar results. If we fix y = 0, we end up with the parabola  $z = x^2$ , and if we fix y = 2, we end up with the parabola  $z = x^2 + 4$ .

It's also worth considering the "level sets," ie, the "horizontal" slices of  $\Gamma$  obtained by fixing various values of z. When z = 0, there is just the single point x = y = 0, ie, the origin. Fixing z = 1, the slice is given by  $1 = x^2 + y^2$ , which is a circle of radius 1. Fixing z = 17, the slice is  $17 = x^2 + y^2$ , which is circle of radius  $\sqrt{17}$ . In fact, all of the level sets are circles, and these circles form a sort of "topographic map" style picture of the graph of f. See figure 2.1.2.

Stitching all of these two-dimensional slices together in our minds, we can see that the



Figure 2.1.2: Here is a "topographic map" style picture of the function  $f(x, y) = x^2 + y^2$ . Each circle is a level set of f. The innermost circle is the set of points such that f(x, y) = 1 (this is a circle of radius 1). The next circle from the inside (drawn in gray) is the set of points such that f(x, y) = 2 (this is a circle of radius  $\sqrt{2}$ ). In general, the kth circle from the inside is the set of points such that f(x, y) = k. It is a circle of radius  $\sqrt{k}$ , and the circles where  $\sqrt{k}$  is an integer are drawn in black rather than gray. The fact that the circles get "closer together" as k increases is an indication that the function grows more and more rapidly as we get further from the origin.

graph of f is a "big parabolic bowl" inside  $\mathbb{R}^3$ . See figure 2.1.3.

# Tangent plane at a = (2, 1)

Now that we understand what the graph of f looks like, let's try to figure out what the "tangent plane" T at a point a = (2, 1) will look like. Of course, it's a plane inside  $\mathbb{R}^3$  that passes through the point (2, 1, f(2, 1)) = (2, 1, 5). To specify which plane it actually is, we start by describing some lines that lie on this plane.

Consider the y = 1 slice of  $\Gamma$ , which is a vertical slice containing the point a. We know that the graph of  $\Gamma$  along this slice is the parabola  $z = x^2 + 1$ . So, slicing the tangent plane T along y = 1 should yield the tangent line to this function of one variable, which we know



Figure 2.1.3: The graph of the function  $f(x, y) = x^2 + y^2$ . "Vertical" slices (ie, slices along planes that are parallel to either the *xz*-plane or the *yz*-plane) are parabolas. "Horizontal slices" (ie, slices along planes parallel to the *xy*-plane) are circles.

how to compute from chapter 1. The slope of this tangent line is

$$\left. \frac{\mathrm{d}}{\mathrm{d}x} f(x,1) \right|_{x=2} = 4.$$

This quantity is called a *partial derivative of* f *at* a, and will be denoted  $(\partial f/\partial x)(a)$ . Thus the tangent line is the line parametrized by

$$h \mapsto (2, 1, 5) + h \cdot (1, 0, 4).$$

Similarly, we can consider the x = 2 slice, which is another vertical slice containing a. We know that the graph  $\Gamma$  along this slice is the parabola  $z = 4 + y^2$ . The slope of the corresponding tangent line is

$$\frac{\partial f}{\partial y}(a) = \left. \frac{df(2,y)}{dy} \right|_{y=1} = 2.$$

Thus this tangent line is the line parametrized by

$$k \mapsto (2, 1, 5) + k \cdot (0, 1, 2).$$

See figure 2.1.4.



Figure 2.1.4: The graph of the function  $f(x, y) = x^2 + y^2$ , together with the point (a, f(a)) = (2, 1, 5) and two of its tangent lines: one parallel to the *xz*-plane, and the other parallel to the *yz*-plane.

These two tangent lines uniquely determine the entire tangent plane T (ie, T is the plane containing these two lines). One way of describing T that immediately falls out of the parametrizations we've found of the tangent lines is the plane parametrized by

$$\begin{bmatrix} h \\ k \end{bmatrix} \mapsto \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix} + \begin{bmatrix} h \\ k \\ 4h+2k \end{bmatrix}$$

The important part of this expression is the bottom entry on the far right, the 4h + 2k. The function  $\mathbb{R}^2 \to \mathbb{R}$  given by  $(h, k) \mapsto 4h + 2k$  is what we will call the *total derivative* or the

*differential* of f and denote by  $df_a$ .

Notice that  $df_a$  is a linear map  $\mathbb{R}^2 \to \mathbb{R}$ . Moreover, its graph is a plane passing through the origin that is parallel to T. In other words, if we take the graph of  $df_a$  and translate it over to the point (2, 1, 5), we obtain exactly the tangent plane T. Speaking more loosely, the derivative  $df_a : \mathbb{R}^2 \to \mathbb{R}$  records all of the "slopey information" about the tangent plane T.

Notice moreover that, if  $e_1, e_2$  are the standard basis vectors of  $\mathbb{R}^2$  (cf. section 0.4), then

$$df_{a}(e_{1}) = 4 = \frac{\partial f}{\partial x}(a)$$
$$df_{a}(e_{2}) = 2 = \frac{\partial f}{\partial u}(a)$$

so the standard matrix representation  $[df_a]$  (cf. definition 0.5.14) of the linear map  $df_a$  is

$$[df_{\mathfrak{a}}] = \begin{bmatrix} 4 & 2 \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x}(\mathfrak{a}) & \frac{\partial f}{\partial y}(\mathfrak{a}) \end{bmatrix}.$$
(2.1.5)

# $df_{\alpha}$ as an approximation

Consider the function  $\varepsilon:\mathbb{R}^2\to\mathbb{R}$  defined by

$$(\mathbf{h},\mathbf{k})\mapsto |\mathbf{f}(\mathbf{a}+(\mathbf{h},\mathbf{k}))-\mathbf{f}(\mathbf{a})-\mathbf{df}_{\mathbf{a}}(\mathbf{h},\mathbf{k})|.$$

Then

$$\varepsilon(\mathbf{h},\mathbf{k}) = |(2+\mathbf{h})^2 + (1+\mathbf{k})^2 - 5 - 4\mathbf{h} - 2\mathbf{k}| = \mathbf{h}^2 + \mathbf{k}^2.$$

Notice that  $\varepsilon$  gets small rapidly as  $(h, k) \rightarrow 0$ . More precisely, the claim is that

$$\lim_{(h,k)\to 0}\frac{\epsilon(h,k)}{|(h,k)|}=0,$$

ie, that  $\epsilon(h, k) = o(|(h, k)|)$  as  $(h, k) \rightarrow 0$ .

In fact, for the claim, it doesn't matter whether |(h, k)| denotes the euclidean norm  $|(h, k)|_2 = \sqrt{h^2 + k^2}$  or the max norm  $|(h, k)|_{\infty} = \max\{|h|, |k|\}$  (cf. section 0.4). If it's the euclidean norm, we have

$$\lim_{(h,k)\to 0} \frac{\varepsilon(h,k)}{|(h,k)|_2} = \lim_{(h,k)\to 0} = \frac{h^2 + k^2}{\sqrt{h^2 + k^2}} = \lim_{(h,k)\to 0} \sqrt{h^2 + k^2} = 0.$$

If instead it's the max norm, observe that

$$\frac{\varepsilon(h,k)}{|(h,k)|_{\infty}} = \frac{h^2 + k^2}{|(h,k)|_{\infty}} \leq \frac{2(|(h,k)|_{\infty})^2}{|(h,k)|_{\infty}} = 2|(h,k)|_{\infty}$$

so taking the limit as  $(h, k) \rightarrow 0$  and applying the squeeze theorem yields the same result.

Since  $\epsilon$  gets small rapidly, we can say that the function  $(h, k) \mapsto f(a) + df_a(h, k)$  is a good approximation of the function  $(h, k) \mapsto f(a + (h, k))$  for small vectors (h, k). Said differently, letting x = a + (h, k), the function  $x \mapsto f(a) + df_a(x - a)$  is a good approximation of f near a.

### Overview

In what follows, we will turn the above example on its head and *define* the differential  $df_a$  to be a linear function which yields a good approximation of f near a point a, and then we will prove in theorem 2.3.34 that partial derivatives (ie, derivatives along various "slices") can be used to compute the standard matrix representation of the the differential.

This might seem a bit "backwards" given our analysis of the example above, but there is a good reason for doing this; it turns out that there are some bizarre functions where partial derivatives make sense, but tangent planes do not (cf. exercises 2.3.28, 2.3.30 and 2.3.32).

We'll be using a lot of linear algebra in this chapter. As mentioned in section 0.4 and stated , the notation |-| will denote either the euclidean or the max norm on  $\mathbb{R}^n$ , and you can choose to interpret it to be whichever of the two norms you like better (in the calculation we did above, the euclidean norm was a little easier; but, in general, I find the max norm to be far more convenient). When it makes a difference which norm on  $\mathbb{R}^n$  we have in mind, we'll specify this explicitly. We'll also need some facts about the operator norm as we go along; I encourage you to at least skim through section 0.6.A before proceeding.

# 2.2 Definition of the derivative

Throughout, S will denote a subset of  $\mathbb{R}^m$ .

**Definition 2.2.1** (Differentiability at a point). A function  $f : S \to \mathbb{R}^n$  is *differentiable at* an interior point  $a \in S$  if there exists a linear map  $\ell : \mathbb{R}^m \to \mathbb{R}^n$  such that

$$|f(a+h) - f(a) - \ell(h)| = o(|h|)$$
 as  $h \to 0$ .

It turns out that there exists at most one linear map that has the above property.

**Lemma 2.2.2.** Suppose  $f: S \to \mathbb{R}^m$  is differentiable at an interior point  $a \in S$ . If  $\ell$  and  $\ell'$  are both linear maps  $\mathbb{R}^m \to \mathbb{R}^n$  such that

$$|f(a+h) - f(a) - \ell(h)| = o(|h|)$$
 and  $|f(a+h) - f(a) - \ell'(h)| = o(|h|)$  as  $h \to 0$ ,

then  $\ell = \ell'$ .

*Proof.* Let  $\phi = \ell' - \ell$ . Then  $\phi$  is also a linear map  $\mathbb{R}^m \to \mathbb{R}^n$ . Moreover, observe that

$$\begin{split} |\phi(h)| &= |\ell'(h) - \ell(h)| \\ &= |(f(a+h) - f(a) - \ell(h)) - (f(a+h) - f(a) - \ell'(h)| \\ &\leq |f(a+h) - f(a) - \ell(h)| + |f(a+h) - f(a) - \ell'(h)|. \end{split}$$

Since both  $|f(a + h) - f(a) - \ell(h)|$  and  $|f(a + h) - f(a) - \ell'(h)|$  are o(|h|), exercises 0.1.6 and 0.1.7 imply that  $|\phi(h)| = o(|h|)$  also. Then lemma 0.6.5 implies that  $\phi = 0$ .

**Exercise 2.2.3.** Look at the exercises from chapter 0 that are invoked in the proof above (namely, exercises 0.1.6, 0.1.7 and 0.6.3) and do any of them that you haven't already done.

Lemma 2.2.2 tells us that the following definition makes sense.

**Definition 2.2.4.** Suppose  $f : S \to \mathbb{R}^n$  is differentiable at an interior point  $a \in S$ . Then the *total derivative* or the *differential of* f *at* a, denoted  $df_a$ , is the unique linear map  $\mathbb{R}^m \to \mathbb{R}^n$  satisfying

$$|f(a+h) - f(a) - df_a(h)| = o(|h|)$$
 as  $h \to 0$ .

*Pedantic remark.* Since we're using |-| to refer indiscriminately to both the euclidean and max norms, it might be worth pointing out that the above definition is independent of which norm you have in mind. More precisely, for any linear function  $\ell : \mathbb{R}^m \to \mathbb{R}^n$ , we have  $|f(a + h) - f(a) - \ell(h)|_2 = o(|h|_2)$  if and only if  $|f(a + h) - f(a) - \ell(h)|_{\infty} = o(|h|_{\infty})$ . You might try proving this; the key is exercise 0.3.11. Thus, if df<sub>a</sub> exists, it is a "good" approximation for f(a + h) - f(a), independently of whether we're measuring distances using the euclidean norm or the max norm.

We've now defined the derivative, but it's fairly difficult to use this definition in practice. There are only a few examples that can be computed directly from this definition. Here are some that I think are instructive. **Exercise 2.2.5.** Suppose  $\ell : \mathbb{R}^m \to \mathbb{R}^n$  is a linear map. Prove that  $d\ell_a = \ell$  for all  $a \in \mathbb{R}^m$ .

**Exercise 2.2.6.** Let  $\mu : \mathbb{R}^2 \to \mathbb{R}$  be the multiplication map  $\mu(x, y) = xy$ . Prove that

$$d\mu_{(a,b)}(h,k) = bh + ak$$

for all  $(a, b) \in \mathbb{R}^2$  and  $(h, k) \in \mathbb{R}^2$ .

**Exercise 2.2.7.** Suppose  $v, w \in \mathbb{R}^n$  and  $f : \mathbb{R} \to \mathbb{R}^n$  is given by f(t) = v + tw. Calculate  $df_a$  for any  $a \in \mathbb{R}$ . What is  $df_a(1)$ ?

We will develop a bit more theory in order to compute more effectively. Meanwhile, we can prove the following directly from the definition.

**Exercise 2.2.8.** If  $f : S \to \mathbb{R}^n$  is differentiable at an interior point  $a \in S$ , show that f must be continuous at a.

Of course, the converse to exercise 2.2.8 is false. We've seen single variable examples; here is a multivariable example.

**Exercise 2.2.9.** Consider the function  $f : \mathbb{R}^2 \to \mathbb{R}$  given by

$$f(x,y) = \sqrt{|xy|}.$$

See figure 2.2.10. Show that f is continuous at the origin, but that it is not differentiable at the origin.

*Possible hint*. If f were differentiable at the origin, then there would exist  $a, b \in \mathbb{R}$  such that

$$\lim_{(\mathbf{h},\mathbf{k})\to \mathbf{0}}\frac{|\mathbf{f}(\mathbf{h},\mathbf{k})-\mathbf{a}\mathbf{h}-\mathbf{b}\mathbf{k}|}{|(\mathbf{h},\mathbf{k})|}=\mathbf{0}.$$

Think about what happens with this limit when you approach the origin along the line k = 0, along the line h = 0, and along the line h = k.



Figure 2.2.10: The graph of the function  $f(x, y) = \sqrt{|xy|}$ . The graph has four "leaves," which look a bit like the "leaves" of the Sydney Opera House, or of the Lotus Temple in New Delhi.

# 2.3 Computing derivatives

# 2.3.A Sum and scalar multiples rule

**Exercise 2.3.1** (Sum rule). Prove that, if  $f, g : S \to \mathbb{R}^n$  are both differentiable at an interior point  $a \in S$ , then f + g is also differentiable at a and

$$d(f+g)_{\mathfrak{a}} = df_{\mathfrak{a}} + dg_{\mathfrak{a}}.$$

**Exercise 2.3.2** (Scalar multiples rule). Prove that, if c is a constant and  $f : S \to \mathbb{R}^n$  is differentiable at an interior point  $a \in S$ , then cf is also differentiable at a and

$$d(cf)_{a} = c \cdot df_{a}$$

# 2.3.B Chain rule

The proof of the following is essentially the same as the second proof of the single variable chain rule given in section 1.2.C. Since this proof is basically a repeat, some details are omitted; you are asked to fill in the details in exercise 2.3.5 below.

**Theorem 2.3.3** (Chain rule). Suppose that S and T are subsets of  $\mathbb{R}^m$  and  $\mathbb{R}^n$ , respectively, that  $f: S \to \mathbb{R}^n$  is differentiable at an interior point a, that  $f(S) \subseteq T$  and f(a) is an interior point of T, and that  $g: T \to \mathbb{R}^p$  is differentiable at f(a). Then the composite  $g \circ f: S \to \mathbb{R}^p$  is also differentiable at a, and

$$\mathrm{d}(\mathfrak{g}\circ f)_{\mathfrak{a}}=\mathrm{d}\mathfrak{g}_{f(\mathfrak{a})}\circ\mathrm{d}f_{\mathfrak{a}}.$$

*Proof.* Define the "error in approximation" functions

$$r(h) = f(a+h) - f(a) - df_a(h)$$
  

$$s(k) = g(f(a)+k) - g(f(a)) - dg_{f(a)}(k)$$

and then observe that

$$g(f(a+h)) - g(f(a)) - dg_{f(a)}(df_a(h)) = dg_{f(a)}(r(h)) + s(df_a(h) + r(h)).$$
(2.3.4)

This means that

$$|g(f(a+h)) - g(f(a)) - dg_{f(a)}(df_a(h))| \leq |dg_{f(a)}(r(h))| + |s(df_a(h) + r(h))|.$$

Since  $dg_{f(a)}$  is linear and |r(h)| = o(|h|), exercise 2.3.5 guarantees that  $|dg_{f(a)}(r(h))| = o(|h|)$ also. Thus, by exercise 0.1.6, it is sufficient to prove that  $|s(df_a(h) + r(h))| = o(|h|)$ . To do this, define  $\eta(k) = |s(k)|/|k|$  and then notice that

$$\begin{aligned} \frac{|s(df_{\mathfrak{a}}(\mathfrak{h}) + r(\mathfrak{h}))|}{|\mathfrak{h}|} &= \eta(df_{\mathfrak{a}}(\mathfrak{h}) + r(\mathfrak{h})) \cdot \frac{|df_{\mathfrak{a}}(\mathfrak{h}) + r(\mathfrak{h})|}{|\mathfrak{h}|} \\ &\leq \eta(df_{\mathfrak{a}}(\mathfrak{h}) + r(\mathfrak{h})) \left( \|df_{\mathfrak{a}}\| + \frac{|r(\mathfrak{h})|}{|\mathfrak{h}|} \right) \end{aligned}$$

where we have used exercise 0.6.2 for the inequality. We now take the limit as  $h \to 0$ . Using the facts that  $\eta \circ (df_a + r)$  is continuous at 0 and that |r(h)| = o(|h|), we obtain the result.

**Exercise 2.3.5.** (a) Prove equation (2.3.4).

- (b) Do exercise 0.6.2 if you haven't already.
- (c) Suppose  $r : \mathbb{R}^m \to \mathbb{R}^n$  is a function such that |r(h)| = o(|h|) as  $h \to 0$ . If  $\ell : \mathbb{R}^n \to \mathbb{R}^p$  is linear, prove that  $|\ell(r(h))| = o(|h|)$  also. *Possible hint.* Even though the operator norm does not appear anywhere in the statement

of this exercise, you might find the concept useful (especially exercise 0.6.2).

(d) Prove that  $\eta \circ (df_a + r)$  is continuous at 0.

# 2.3.C Differentiability by components

Recall that we asserted at the beginning of this chapter that the "multi" in the name of this chapter refers to the number of inputs (rather than the number of outputs). This section is where we show that understanding the n = 1 case is "enough" to understand general n.

**Definition 2.3.6** (Component functions). For any function  $f : S \to \mathbb{R}^n$ , we define the jth *component function*  $f_j : S \to \mathbb{R}$  to be the composite  $\pi_j \circ f$ , where  $\pi_j$  is the jth projection map (cf. section 0.4).

**Proposition 2.3.7.** A function  $f : S \to \mathbb{R}^n$  is differentiable at an interior point  $a \in S$  if and only if the component function  $f_j : S \to \mathbb{R}$  is differentiable at a for all j = 1, ..., n. Moreover,

$$df_{a}(h) = \begin{bmatrix} df_{1,a}(h) \\ \vdots \\ df_{n,a}(h) \end{bmatrix}.$$
 (2.3.8)

*Proof.* If f is differentiable at a, then the composite  $f_j = \pi_j \circ f$  is also differentiable at a by the chain rule 2.3.3. Moreover, since  $\pi_j$  is linear, we know that  $d\pi_{j,f(\alpha)} = \pi_j$  by exercise 2.2.5. Thus, by the chain rule,

$$df_{j,a}(h) = (d\pi_{j,f(a)} \circ df_a)(h) = \pi_j(df_a(h))$$

for all j, which is precisely equation (2.3.8). For the converse, we turn equation (2.3.8) on its head. Suppose  $f_j$  is differentiable at a for all j, and let  $\ell : \mathbb{R}^m \to \mathbb{R}^n$  be the linear map

$$\ell(h) = \begin{bmatrix} df_{1,a}(h) \\ \vdots \\ df_{n,a}(h) \end{bmatrix}.$$

We now want to show that  $|f(a + h) - f(a) - \ell(h)| = o(|h|)$  as  $h \to 0$ . Observe that the component functions of  $h \mapsto f(a + h) - f(a) - \ell(h)$  are given by

$$\nu \mapsto \pi_j \left( f(a+h) - f(a) - \ell(h) \right) = f_j(a+h) - f_j(a) - df_{j,a}(h)$$

since  $\pi_j$  is linear, and we know that  $|f_j(a+h) - f_j(a) - df_{j,a}(h)| = o(|h|)$ . By exercise 2.3.9 below applied with the function  $r(h) = f(a+h) - f(a) - \ell(h)$ , we conclude that  $|f(a+h) - f(a) - \ell(h)| = o(|h|)$ .

**Exercise 2.3.9.** Show that if S is a neighborhood of 0 in  $\mathbb{R}^m$  and  $r : S \to \mathbb{R}^n$  is a function such that  $|r_j(h)| = o(|h|)$  as  $h \to 0$  for all j = 1, ..., n, then |r(h)| = o(|h|) also.

*Possible hint*. This is easy if you're using the max norm. If you're using the euclidean norm, you may find it useful to do exercise 0.3.11 first.

**Exercise 2.3.10.** Suppose S is a subset of  $\mathbb{R}$  and  $f : S \to \mathbb{R}^n$  is differentiable at an interior point a. Show that the standard matrix representation  $[df_a]$  of the linear map  $df_a : \mathbb{R} \to \mathbb{R}^n n$  is given by

$$[df_{\mathfrak{a}}] = \begin{bmatrix} f_{1}'(\mathfrak{a}) \\ \vdots \\ f_{\mathfrak{n}}'(\mathfrak{a}) \end{bmatrix},$$

where  $f'_j(a)$  is the derivative of the single variable function  $f_j: S \to \mathbb{R}$  in the sense of chapter 1.

# 2.3.D Product and quotient rules

We showed in the previous section that it's enough to understand the n = 1 case, ie, the case of real-valued (rather than  $\mathbb{R}^n$ -valued) functions. So in this section and the next, we'll focus our attention on real-valued functions.

Note that we can combine real-valued functions in more ways than the ones we've already discussed; specifically, we can multiply and divide them. As in chapter 1, we have product and quotient rules to deal with this. As it turns out, we can actually derive these using the multivariable chain rule 2.3.3!

**Proposition 2.3.11** (Product rule). *Suppose*  $f, g : S \to \mathbb{R}$  *are both differentiable at an interior* 

point  $a \in S$ . Then fg is also differentiable at a and

$$d(fg)_a = g(a)df_a + f(a)dg_a.$$

*Proof.* Let  $f \times g$  denote the function  $S \to \mathbb{R}^2$  defined by

$$(f \times g)(x) = (f(x), g(x)).$$

Observe that the first and second component functions of  $f \times g$  are f and g, respectively; thus it follows from proposition 2.3.7 that

$$d(f \times g)_{a}(h) = (df_{a}(h), dg_{a}(h)).$$

Let  $\mu$  :  $\mathbb{R}^2 \to \mathbb{R}$  denote the multiplication map  $\mu(x, y) = xy$  that we considered in exercise 2.2.6. Then  $fg = \mu \circ (f \times g)$ , so by the chain rule 2.3.3, we have that

$$d(fg)_{a}(h) = d\mu_{(f(a),g(a))}(d(f \times g)_{a}(h))$$
$$= d\mu_{f(a),g(a)}(df_{a}(h), dg_{a}(h))$$
$$= g(a)df_{a}(h) + f(a)dg_{a}(h),$$

where we have used the calculation from exercise 2.2.6 for the final step.

**Exercise 2.3.12.** Suppose  $f : S \to \mathbb{R}$  is differentiable at an interior point  $a \in S$  and that  $f(a) \neq 0$ . Show that 1/f is also differentiable at a, and

$$\mathrm{d}(1/\mathrm{f})_{\mathfrak{a}} = -\frac{\mathrm{d}\mathrm{f}_{\mathfrak{a}}}{\mathrm{f}(\mathfrak{a})^2}.$$

*Possible hint.* Consider the function  $\iota : \mathbb{R} \setminus \{0\} \to \mathbb{R}$  given by  $\iota(x) = 1/x$ . This is a single variable function, so we know  $d\iota_x$  from chapter 1. Now note that  $1/f = \iota \circ f$ .

**Exercise 2.3.13** (Quotient rule). Suppose  $f, g : S \to \mathbb{R}$  is differentiable at an interior point  $a \in S$  and that  $g(a) \neq 0$ . Show that f/g is also differentiable at a, and that

$$d(f/g)_{\mathfrak{a}} = \frac{g(\mathfrak{a})df_{\mathfrak{a}} - f(\mathfrak{a})dg_{\mathfrak{a}}}{g(\mathfrak{a})^2}$$

*Possible hint*. Combine exercise 2.3.12 and proposition 2.3.11.

# 2.3.E Directional and partial derivatives

We now formalize the "slicing" we did in section 2.1.

**Definition 2.3.14** (Directional derivatives). Suppose  $f : S \to \mathbb{R}$  is a function,  $a \in S$  is an interior point, and  $h \in \mathbb{R}^m$  is a vector. The *directional derivative of* f *at* a *with respect to* h, denoted  $\partial_h f(a)$ , is defined to be the limit

$$\lim_{t\to 0}\frac{f(a+th)-f(a)}{t}.$$

If  $h = e_i$  is the ith standard basis vector (cf. section 0.4), then this is called the *i*th partial *derivative of* f at a and denoted  $\partial_i f(a)$  (instead of  $\partial_{e_i} f(a)$ ). If we're using  $x_i$  to denote the *i*th component of the input of f, then  $\partial_i f(a)$  is sometimes also denoted by one of the following.

$$f_{x_i}(a) = \frac{\partial f}{\partial x_i}\Big|_{x=a} = \frac{\partial}{\partial x_i}f(x)\Big|_{x=a} = \frac{\partial f}{\partial x_i}(a)$$

You may remember from multivariable calculus that partial derivatives behave a great deal like single variable derivatives. The reason for this is the following important remark, which reinterprets a directional derivative as a single variable derivative.

**Remark 2.3.15.** For  $a \in S$  an interior point and  $h \in \mathbb{R}^m$ , consider the function  $\xi : \mathbb{R} \to \mathbb{R}^m$  given by  $\xi(t) = a + th$ . This is the "line through a in the direction of the vector h." Note that  $0 = \xi^{-1}(a)$  is an interior point of  $\xi^{-1}(S)$  since a is a interior point of S. If  $f : S \to \mathbb{R}$  is a function, then  $f \circ \xi$  is a single variable function  $\xi^{-1}(S) \to \mathbb{R}$ , and we have

$$\partial_{\mathbf{h}} \mathbf{f}(\mathbf{a}) = (\mathbf{f} \circ \boldsymbol{\xi})'(\mathbf{0}),$$

just by writing out the definition of both sides. A bit more geometrically,  $f \circ \xi$  can be thought of as the restriction of f to the line passing through a in the direction of h. This is a single variable function, and the derivative of this single variable function at 0 is the directional derivative of f at a with respect to h.

**Example 2.3.16.** Consider the function  $f : \mathbb{R}^2 \to \mathbb{R}$  given by  $f(x, y) = \sin(xy^2)$ . Fix  $(a, b) \in \mathbb{R}^2$  and define  $\xi : \mathbb{R} \to \mathbb{R}^2$  by  $\xi(t) = (a + t, b)$ . In other words,  $\xi$  is the line through (a, b) in the direction of the vector (1, 0), as in remark 2.3.15. Then

$$(f \circ \xi)(t) = f(a+t,b) = \sin((a+t)b^2).$$

Then

$$\begin{split} \frac{\partial f}{\partial x}(a,b) &= (f \circ \xi)'(0) \\ &= \left. \frac{d}{dt} \left( \sin((a+t)b^2) \right) \right|_{t=0} \\ &= b^2 \cos((a+t)b^2) \big|_{t=0} \\ &= b^2 \cos(ab^2). \end{split}$$

**Exercise 2.3.17.** With notation as in example 2.3.16, compute  $(\partial f/\partial y)(a, b)$ .

**Exercise 2.3.18.** Suppose  $f : S \to \mathbb{R}$  is a function and  $a \in S$  is an interior point. Prove the following.

- (a)  $\partial_0 f(\alpha) = 0$ .
- (b) If  $\partial_h f(a)$  exists for some  $h \in \mathbb{R}^m$ , then  $\partial_{\lambda h} f(a)$  also exists for all  $\lambda \in \mathbb{R}$  and

$$\partial_{\lambda h} f(a) = \lambda \partial_h f(a).$$

We can use remark 2.3.15 to prove some directional derivative analogs of results we proved in chapter 1.

**Exercise 2.3.19** (Product rule for directional derivatives). Suppose  $f, g : S \to \mathbb{R}$  are functions,  $a \in S$  is an interior point, and  $h \in \mathbb{R}^m$  is a vector such that  $\partial_h f(a)$  and  $\partial_h g(a)$  both exist. Prove that  $\partial_h(fg)(a)$  also exists and that

$$\partial_{\mathbf{h}}(\mathbf{f}g)(\mathbf{a}) = g(\mathbf{a})\partial_{\mathbf{h}}(\mathbf{f})(\mathbf{a}) + f(\mathbf{a})\partial_{\mathbf{h}}g(\mathbf{a}).$$

*Possible hint*. Use remark 2.3.15 and the single variable product rule 1.2.3.

**Exercise 2.3.20** (Quotient rule for directional derivatives). Suppose  $f, g : S \to \mathbb{R}$  are functions,  $a \in S$  is an interior point such that  $g(a) \neq 0$  and  $\partial_h f(a)$  and  $\partial_h g(a)$  both exist for some  $h \in \mathbb{R}^m$ . Show that  $\partial_h(f/g)(a)$  also exists and

$$\partial_{h}(f/g)(a) = \frac{g(a)\partial_{h}f(a) - f(a)\partial_{h}g(a)}{g(a)^{2}}.$$

*Possible hint*. Use remark 2.3.15 and the single variable quotient rule 1.2.6.

**Exercise 2.3.21** (Interior extremum theorem for directional derivatives). Suppose  $f : S \to \mathbb{R}$  is a function and an interior point  $a \in S$  is a local extremum of f. Show that, if  $\partial_h f(a)$  exists for some  $h \in \mathbb{R}^m$ , then  $\partial_h f(a) = 0$ .

*Possible hint*. Use remark 2.3.15 and the single variable interior extremum theorem 1.2.19.

Just as the converse to the single variable interior extremum theorem 1.2.19 is false (cf. exercise 1.2.20), so too is the converse to the interior extremum theorem for directional derivatives 2.3.21, except that there are now even more ways for the converse to fail. Here are two examples to illustrate this.

**Example 2.3.22.** Consider the function  $f : \mathbb{R}^2 \to \mathbb{R}$  given by

$$f(x,y) = x^2 - y^2.$$

Let  $\xi(t) = 0 + te_1 = (t, 0)$ . Then  $f \circ \xi$  has a minimum at 0, so  $(f \circ \xi)'(0) = (\partial f)/(\partial x)(0) = 0$ . Similarly, if  $\xi(t) = 0 + te_2 = (0, t)$ , then  $f \circ \xi$  has a maximum at 0, so  $(f \circ \xi)'(0) = (\partial f)/(\partial y)(0) = 0$  also. But clearly these two facts together mean that 0 cannot be a local extremum of f. See figure 2.3.23. This kind of a point is sometimes called a "saddle."

**Exercise 2.3.24.** Consider the function  $f : \mathbb{R}^2 \to \mathbb{R}$  given by

$$f(x, y) = (y - x^2)(y - 3x^2).$$

See figure 2.3.25 for the graph of f.

(a) Show that the origin is *not* a local extremum of f.

*Possible hint*. Consider the values of f on points of the form (0, b) and  $(a, 2a^2)$  near the origin. You might consider identifying where these points are on the graph depicted in figure 2.3.25 to get a more visual sense of what's going on.

(b) Show that f has a local minimum when restricted to an arbitrary line through the origin. Conclude that  $\partial_h f(a) = 0$  for all  $h \in \mathbb{R}^2$ .

Now for the main result of this section: partial derivatives help us compute derivatives. Note that we will generalize this result slightly in theorem 2.3.34 below.

**Proposition 2.3.26.** *If*  $f : S \to \mathbb{R}$  *is differentiable at an interior point*  $a \in S$  *and*  $h \in \mathbb{R}^m$ *, then* 

$$df_a(h) = \partial_h f(a).$$



Figure 2.3.23: The graph of the function  $f(x, y) = x^2 - y^2$ . Vertical slices along planes parallel to the xz-plane are downwards facing parabolas, while those along planes parallel to the yz-plane are upwards facing parabolas.

Thus, the standard matrix representation  $[df_a]$  of the total derivative  $df_a : \mathbb{R}^m \to \mathbb{R}$  is a  $1 \times m$  matrix whose ith entry is the ith partial derivative  $\partial_i f(a)$ .

$$[df_{a}] = \begin{bmatrix} \partial_{1}f(a) & \partial_{2}f(a) & \cdots & \partial_{m}f(a) \end{bmatrix}$$

*Proof.* If h = 0, we know that  $\partial_0 f(a) = 0$  by exercise 2.3.18 and that  $df_a(0) = 0$  since  $df_a$  is linear, so we are done. Thus we can assume that h is nonzero. Let r be the "error in approximation" function

$$\mathbf{r}(\mathbf{k}) = \mathbf{f}(\mathbf{a} + \mathbf{k}) - \mathbf{f}(\mathbf{a}) - \mathbf{df}_{\mathbf{a}}(\mathbf{k}),$$

so that  $|\mathbf{r}(k)| = o(|k|)$  as  $k \to 0$ . Letting k = th, we see that

$$\mathbf{r}(\mathsf{th}) = \mathbf{f}(\mathsf{a} + \mathsf{th}) - \mathbf{f}(\mathsf{a}) - \mathsf{df}_{\mathsf{a}}(\mathsf{th}) = \mathbf{f}(\mathsf{a} + \mathsf{th}) - \mathbf{f}(\mathsf{a}) - \mathsf{tdf}_{\mathsf{a}}(\mathsf{h})$$



Figure 2.3.25: The graph of the function  $f(x, y) = (y - x^2)(y - 3x^2)$ .

by linearity of  $df_a$ , so

$$\frac{r(th)}{t} = \frac{f(a+th) - f(a) - tdf_a(h)}{t} = \frac{f(a+th) - f(a)}{t} - df_a(h).$$

Then

$$\lim_{t\to 0} \left| \frac{f(a+th) - f(a)}{t} - df_a(h) \right| = \lim_{t\to 0} \frac{|r(th)|}{|t|} = \lim_{t\to 0} \frac{|r(th)|}{|th|} \cdot |h| = 0$$

where we used the fact that |r(k)| = o(|k|) as  $k \to 0$  for the last step. This tells us that  $\partial_h f(a)$  exists and equals  $df_a(h)$ . In particular, applying this with  $h = e_i$ , we see that  $df_a(e_i) = \partial_i f(a)$ , which yields the standard matrix representation.

**Remark 2.3.27.** If  $f: S \to \mathbb{R}$  is differentiable at an interior point  $a \in S$ , the column vector

$$[df_{a}]^{\mathsf{T}} = \begin{bmatrix} \vartheta_{1}f(a)\\ \vartheta_{2}f(a)\\ \vdots\\ \vartheta_{\mathfrak{m}}f(a) \end{bmatrix} \in \mathbb{R}^{\mathfrak{m}}$$

is often called the *gradient* of f at a, and is denoted  $\nabla f(a)$ . If  $h \in \mathbb{R}^m$  is any vector, then

$$[df_{a}(h)] = [df_{a}]h = [df_{a}]^{\mathsf{T}} \cdot h = \nabla f(a) \cdot h,$$

where the  $\cdot$  denotes the dot product.

Now for the important caveat: unfortunately, the converse to proposition 2.3.26 is *not* true: the existence of all partial derivatives is not sufficient to guarantee differentiability. Even the existence of all *directional* derivatives is not sufficient to guarantee differentiability. In fact, we cannot guarantee differentiability even if all directional derivatives exist *and* the function  $h \mapsto \partial_h f(a)$  is linear.

Here are three examples to drive home these points. They're useful examples to study closely; you'll gain valuable intuition about multivariable derivatives by doing so.

**Exercise 2.3.28.** Consider the function  $f : \mathbb{R}^2 \to \mathbb{R}$  defined as follows.

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

Show that the partial derivatives of f with respect to x and y both exist at the origin, but that f is not even continuous at the origin (so it cannot be differentiable, by exercise 2.2.8). *Possible hint*. Slice along the line y = rx for various r. See figure 2.3.29.

**Exercise 2.3.30.** Consider the function  $f : \mathbb{R}^2 \to \mathbb{R}$  given by

$$f(x,y) = \begin{cases} \frac{x^3}{x^2 + y^2} & \text{if } (x,y) \neq 0\\ 0 & \text{if } (x,y) = 0. \end{cases}$$

See figure 2.3.31. Show that f is continuous and that  $\partial_h f(0)$  exists for all nonzero  $h \in \mathbb{R}^2$ , but that f is not differentiable at the origin.

*Possible hint.* Calculate a formula for  $\partial_{(h,k)}f(0)$  for all  $(h,k) \in \mathbb{R}^2$ , and show that the function  $\mathbb{R}^2 \to \mathbb{R}$  given by  $(h,k) \mapsto \partial_{(h,k)}f(0)$  is not linear. Explain why proposition 2.3.26 then implies that f is *not* differentiable at 0.



Figure 2.3.29: Two views of the graph of the function f from exercise 2.3.28. The one on the left is obtained from the one on the right by a  $90^{\circ}$  rotation counterclockwise around the *z*-axis.

**Exercise 2.3.32.** Consider the function  $f : \mathbb{R}^2 \to \mathbb{R}$  given by

$$f(x,y) = \begin{cases} |x| & \text{if } y = x^2 \\ 0 & \text{otherwise} \end{cases}$$

- (a) Doodle a picture of f by hand.
- (b) Show that f is continuous at the origin.
- (c) Show that all directional derivatives of f at the origin exist and are equal to 0.

*Possible hint*. First show it for the partial derivatives, using remark 2.3.15. Then fix  $r \neq 0$  and show that  $\partial_{(1,r)} f(0) = 0$ , again using remark 2.3.15. Then use exercise 2.3.18 to conclude.

(d) Show that f is not differentiable at the origin.

*Possible hint.* Find a differentiable function  $\gamma : \mathbb{R} \to \mathbb{R}^2$  such that  $\gamma(0) = 0$  and  $f \circ \gamma$  is not differentiable at the origin. Use the chain rule 2.3.3 to conclude that f cannot be differentiable at the origin.

Thankfully, this phenomenon is fairly pathological. Theorem 2.5.1 below proves that the partial derivatives existing *and being continuous* is sufficient to guarantee differentiability.



Figure 2.3.31: The graph of the function f from exercise 2.3.30.

# 2.3.F Jacobian matrix

**Definition 2.3.33** (Jacobian matrix). Suppose  $f : S \to \mathbb{R}^n$  is differentiable at an interior point  $a \in S$ . The *Jacobian matrix* of f at a, denoted f'(a), is the  $n \times m$  matrix whose (j, i)-entry is  $\partial_i f_j(a)$ . In other words,

$$f'(a) = \begin{bmatrix} \partial_1 f_1(a) & \partial_2 f_1(a) & \cdots & \partial_m f_1(a) \\ \partial_1 f_2(a) & \partial_2 f_2(a) & \cdots & \partial_m f_2(a) \\ \vdots & \vdots & \ddots & \vdots \\ \partial_1 f_n(a) & \partial_2 f_n(a) & \cdots & \partial_m f_n(a) \end{bmatrix}$$

Note that propositions 2.3.7 and 2.3.26 tell us that all of these partial derivatives exist.

**Theorem 2.3.34.** If  $f : S \to \mathbb{R}^n$  is differentiable at an interior point  $a \in S$ , then

$$df_{\mathfrak{a}}(\mathfrak{h}) = \begin{bmatrix} \partial_{\mathfrak{h}} f_{1}(\mathfrak{a}) \\ \cdots \\ \partial_{\mathfrak{h}} f_{\mathfrak{n}}(\mathfrak{a}) \end{bmatrix}.$$

In particular,  $[df_{\alpha}] = f'(\alpha)$ .

Exercise 2.3.35. Prove theorem 2.3.34.

*Possible hint*. Use propositions 2.3.7 and 2.3.26.

Using theorem 2.3.34, many of the rules for differentiation we've seen can be rewritten in forms that you might be more familiar with from single variable calculus. For example, since composition of linear maps corresponds to multiplication of the corresponding matrix representations, the chain rule 2.3.3 states that

$$(g \circ f)'(a) = [d(g \circ f)_a] = [dg_{f(a)} \circ df_a] = [dg_{f(a)}][df_a] = g'(f(a))f'(a)$$

Though the equality  $(g \circ f)'(a) = g'(f(a))f'(a)$  looks superficially identical to the equality from the single variable chain rule 1.2.8, it's worth noticing that the terms in this multivariable version are *matrices*. In particular, this means that that order of the multiplication g'(f(a))f'(a) cannot be interchanged (this was a non-issue in the single variable case).

Similarly, the matrix version of the multivariable product rule 2.3.11 states that

$$(fg)'(a) = g(a)f'(a) + f(a)g'(a),$$

which again looks superficially identical to the single variable product rule 1.2.3, but again, it is worth noting that g(a) and f(a) are scalars, whereas f'(a) and g'(a) are matrices.

# 2.3.G Rank of a differentiable map

**Definition 2.3.36** (Rank of a differentiable map). Suppose  $f : S \to \mathbb{R}^n$  is differentiable at an interior point  $a \in S$ . The rank of the linear map  $df_a : \mathbb{R}^m \to \mathbb{R}^n$  (or, equivalently, of the matrix f'(a)) is also called the *rank of* f *at* a, denoted rank<sub>a</sub>(f). Note that we always have

$$\operatorname{rank}_{\mathfrak{a}}(f) \leq \min\{\mathfrak{m},\mathfrak{n}\}.$$

We then make the following definitions in various special cases.

- We say that f is *immersive at* a if  $rank_a(f) = m$  (equivalently, if  $df_a$  is injective).
- We say that f is *submersive at* a if  $rank_a(f) = n$  (equivalently, if  $df_a$  is surjective).
- We say that f is *étale at* a if rank<sub>a</sub>(f) = m = n (equivalently, if df<sub>a</sub> is an isomorphism).
- We say that a is a *regular point* of f if rank<sub>a</sub>(f) = min{m, n}. Otherwise, we say that a is a *critical point* of f.

We also say that a point  $b \in \mathbb{R}^n$  is a *regular value* of f if every  $a \in f^{-1}(b)$  is a regular point of f, and that it is a *critical value* otherwise.

There are lots of relationships between these notions, depending on how the dimension of the domain, m, compares with the dimension of the codomain, n. You should convince yourself of the following facts.

- If m > n...
  - f cannot be immersive at any point.
  - a is a regular point of f if and only if f is submersive at a.
- If m < n...
  - f cannot be submersive at any point.
  - a is a regular point of f if and only if f is immersive at a.
- If m = n...
  - a is a regular point of f if and only if f is immersive at a, if and only if f is submersive at a, if and only if f is étale at a.

*Caution.* Some mathematicians use a different definition of "regular point." They define a point to be regular if  $\operatorname{rank}_{\alpha}(f) = n$ , so that, by definition, it's *always* true (for all m, n regardless of how these two integers compare) that f is a submersion at a if and only if a is a regular point of f. If  $m \ge n$ , there's no difference between these alternative definitions and our definitions. But, if m < n, there is a difference: these other mathematicians would say that all points are critical in this case, but that isn't true for us with the definitions we've made above. So, if you see the words "regular point" and "critical point" in another text and it's possible that m < n, it's worth looking back to make sure you know what the author means!



Figure 2.3.38: The function  $f(r, \theta) = (r \cos \theta, r \sin \theta)$  of example 2.3.37 transforms the pictures on the right to the pictures on the left.

Unimportant remark. While we're on the topic of terminology, it may be worth mentioning that the word "étale" is not commonly used in real analysis or in differential geometry; in fact, as far as I'm aware, there is no word in these fields that means that the derivative is invertible. But the word "étale" is used in algebraic geometry to mean that the derivative is invertible, so I decided to import this word from algebraic geometry. That said, if you do go on to study algebraic geometry at some point, it's worth noting the algebraic geometry words corresponding to "immersive" and "submersive" are "unramified" and "smooth," respectively. This usage of "smooth" in algebraic geometry is unrelated to the real analysis and differential geometry usage of smooth that we'll encounter later on.

**Example 2.3.37.** Let  $f : \mathbb{R}^2 \to \mathbb{R}^2$  be the map  $f(r, \theta) = (r \cos \theta, r \sin \theta)$ . See figure 2.3.38.

We compute partial derivatives.

$$\frac{\partial f_1}{\partial r} = \cos \theta \quad \frac{\partial f_2}{\partial r} = \sin \theta \quad \frac{\partial f_1}{\partial \theta} = -r \sin \theta \quad \frac{\partial f_2}{\partial \theta} = r \cos \theta.$$

Thus

$$f'(r, \theta) = egin{bmatrix} \cos \theta & -r \sin \theta \ \sin \theta & r \cos \theta \end{bmatrix}.$$

Since this is a square matrix,  $df_{(r,\theta)}$  has non-maximal rank if and only if its determinant vanishes. But

$$\det f'(r,\theta) = r\cos^2\theta + r\sin^2\theta = r,$$

so the critical locus (ie, the set of all critical points) of f is the vertical line r = 0, and f is étale away from this line.

**Exercise 2.3.39** (Parametrization of a cusp). Consider the function  $f : \mathbb{R} \to \mathbb{R}^2$  defined by

$$f(\mathbf{x}) = (\mathbf{x}^2, \mathbf{x}^3).$$

See figure 2.3.40 for a picture of the image of f. The "pointy" part at the origin is sometimes called a cusp. Calculate f'(a) for all  $a \in \mathbb{R}$ . Show that 0 is a critical point of f, and that f is immersive away from 0.

**Exercise 2.3.41** (Parametrization of a node). Consider the function  $f:\mathbb{R}\to\mathbb{R}^2$  defined by

$$f(x) = (x^2 - 1, x^3 - x).$$

See figure 2.3.42 for a picture of the image of f. The self-intersecting part at the origin is called a node. Calculate f'(a) for all  $a \in \mathbb{R}$ , and show that f is immersive everywhere.

**Exercise 2.3.43.** Show that an interior point  $a \in S$  is a critical point of  $f : S \to \mathbb{R}$  if and only if  $\partial_i f(a) = 0$  for all i.

**Exercise 2.3.44.** Let  $f : \mathbb{R}^3 \to \mathbb{R}$  be the function

$$f(x,y,z) = x^2 + y^2 + z^2$$

Calculate f'(a) for all  $a \in \mathbb{R}^3$ , and find all critical points of f. Also, describe the level sets of f. In other words, for fixed  $r \in \mathbb{R}$ , describe the set of all  $a \in \mathbb{R}^3$  such that f(a) = r.



Figure 2.3.40: The image of the function  $f(x) = (x^2, x^3)$ . The origin is f(0).



Figure 2.3.42: The image of the function  $f(x) = (x^2 - 1, x^3 - x)$ . We have f(0) = (-1, 0) and  $f(\pm 1) = (0, 0)$ .

**Exercise 2.3.45.** Find the critical points of the function  $f : \mathbb{R}^3 \to \mathbb{R}^2$  given by

$$f(x, y, z) = (xy, z).$$

**Exercise 2.3.46.** Find the critical points of the function  $f : \mathbb{R}^3 \to \mathbb{R}^2$  given by

$$f(x, y, z) = (x + y^2, y + z^2).$$

# 2.4 Differentiable functions

Throughout this section, U denotes an open subset of  $\mathbb{R}^m$ . In this section, we'll discuss functions that are differentiable at every point in U.

**Definition 2.4.1.** A function  $f : U \to \mathbb{R}^n$  is *differentiable* if it is differentiable at every  $a \in U$ . The *total derivative* of f, denoted df, is the function  $U \to \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$  given by  $a \mapsto df_a$ . Sometimes the function df is also called the *differential* of f.

**Definition 2.4.2.** If  $f: U \to \mathbb{R}$  and  $h \in \mathbb{R}^m$  is a vector such that  $\partial_h f(a)$  exists for all  $a \in U$ , then the *directional derivative* of f with respect to h is the function  $\partial_h f$  given by  $a \mapsto \partial_h f(a)$ . When  $h = e_i$  is one of the standard basis vectors, this function is called the *ith partial derivative* of f and is denoted  $\partial_i f$ . If we're using the variable  $x_i$  to denote the *i*th component of the input of f, then  $\partial_i f$  is sometimes also denoted by one of the following.

$$f_{x_i} = \frac{\partial f}{\partial x_i}$$

# 2.4.A Vanishing derivatives

First off, here are some definitions that are rather useful when studying functions that are differentiable everywhere on an open set.

**Definition 2.4.3.** If  $a, b \in \mathbb{R}^n$ , the *straight line path* from a to b is the function

$$\gamma(t) = (1-t)a + tb.$$

Observe that  $\gamma(0) = a$  and  $\gamma(1) = b$ .

**Exercise 2.4.4.** Show that the straight line path from a to b is differentiable.

**Definition 2.4.5.** A subset S of  $\mathbb{R}^n$  is *convex* if the straight line path between any two points in S is contained in S. In other words, S is convex if, whenever  $a, b \in S$  and  $\gamma$  is the straight line path from a to b, then  $\gamma(t) \in S$  for all  $t \in [0, 1]$ .

Roughly speaking, the role of intervals in single variable calculus is played by convex sets in multivariable calculus. For example, here is a multivariable generalization of proposition 1.3.5. We will use the single variable version to prove this multivariable version.

**Proposition 2.4.6.** *Suppose* U *is convex and*  $f : U \to \mathbb{R}^n$  *is a differentiable function. Then* df = 0 *if and only if* f *is constant.* 

*Proof.* Taking w = 0 in exercise 2.2.7 shows that, if f is constant, then df = 0. For the converse, we can assume without loss of generality that n = 1 (cf. exercise 2.4.7). We want to show that, for any  $a, b \in U$ , we have f(a) = f(b). Let  $\gamma$  be the straight line path from a to b. Convexity of U tells us that  $\gamma(t) \in U$  for all  $t \in [0, 1]$ . Also,  $\gamma$  is differentiable by exercise 2.4.4. Thus  $f \circ \gamma$  is a differentiable function  $[0, 1] \rightarrow \mathbb{R}$  and

$$(f \circ \gamma)'(t) = f'(\gamma(t))\gamma'(t) = 0$$

where we used the (matrix version of the) chain rule 2.3.3 at the second step, and the fact that df = 0 for the final step. Thus it follows from proposition 1.3.5 that  $f \circ \gamma$  is constant, which means that

$$f(a) = f(\gamma(0)) = f(\gamma(1)) = f(b).$$

**Exercise 2.4.7.** Explain why proving that df = 0 implies that f is constant for n = 1 implies the same fact for general n.

In fact, it's not necessary that the domain be convex in order for the conclusion of proposition 2.4.6 to be true. Before getting to the general statement, here is a useful intuition-building exercise.

**Exercise 2.4.8.** Let U denote  $\mathbb{R}^2$  minus the non-negative x-axis. In other words,

$$\mathbf{U} = \mathbb{R}^2 \setminus \{ (\mathbf{x}, \mathbf{0}) \in \mathbb{R}^2 : \mathbf{x} \ge \mathbf{0} \}.$$

(a) Show that U is not convex.

(b) Suppose  $f: U \to \mathbb{R}$  is a differentiable function such that df = 0. Show that f is constant.

Possible hint. Any pair of points in U can be joined by two straight line paths...

**Exercise 2.4.9.** Let U be a connected<sup>1</sup> open subset of  $\mathbb{R}^m$ .

(a) Show that, for any  $a, b \in U$ , there exists a finite sequence of points

$$a = a_0, a_1, \ldots, a_k = b$$

such that the image of the straight line path between  $a_i$  and  $a_{i+1}$  is entirely contained in U for all i.

*Possible hint*. For  $a \in U$ , consider the set S of all points  $a' \in U$  such that there exists a finite sequence of points  $a = a_0, a_1, ..., a_k = a'$  such that the image of the straight line path connecting  $a_i$  to  $a_{i+1}$  is entirely contained in U for all i. Show that S is open, and also that the complement of S is open. Since U is connected and S is nonempty, it must be that the complement of S is empty.

(b) If  $f: U \to \mathbb{R}^n$  is differentiable, show that df = 0 if and only if f is constant.

# 2.4.B Bounded derivatives

Here is a multivariable version of exercise 1.3.6. Again, we will actually use the single variable version to prove the multivariable version.

**Proposition 2.4.10.** Suppose U is convex and  $f : U \to \mathbb{R}^n$  is a differentiable function. Suppose further that there exists a constant M such that  $\|df_x\| \leq M$  for all  $x \in U$ . Then

$$|f(b) - f(a)| \le M|b - a|$$

for all  $a, b \in U$ .

*Proof.* We'll prove this only when |-| denotes the max norm. The proof for the euclidean norm is a bit trickier: see [Rud76, theorem 9.19]. Suppose first that n = 1. If a = b, there is nothing to do, so we can assume that  $a \neq b$ . Let

$$h = \frac{b-a}{|b-a|},$$

<sup>&</sup>lt;sup>1</sup>Recall that a metric space X is *connected* if it cannot be written as the union of two disjoint open subsets.

ie, h is the vector pointing from a towards b normalized to unit length. Consider the function  $\xi(t) = a + th$  as in remark 2.3.15. Then

$$(\mathbf{f} \circ \boldsymbol{\xi})'(\mathbf{t}) = \partial_{\mathbf{h}} \mathbf{f}(\mathbf{a} + \mathbf{t}\mathbf{h}) = \mathbf{d} \mathbf{f}_{\mathbf{a} + \mathbf{t}\mathbf{h}}(\mathbf{h})$$

by theorem 2.3.34, which means that

$$|(\mathbf{f} \circ \boldsymbol{\xi})'(\mathbf{t})| = |\mathbf{d}\mathbf{f}_{a+\mathbf{t}\mathbf{h}}(\mathbf{h})| \leq ||\mathbf{d}\mathbf{f}_{a+\mathbf{t}\mathbf{h}}|||\mathbf{h}| \leq M|\mathbf{h}| = M,$$

where we used the fact that h has unit length for the final step. Notice that  $\xi(0) = a$  and  $\xi(|b - a|) = b$ , so by applying exercise 1.3.6 to  $f \circ \xi$ , we see that

$$|f(b) - f(a)| \leq M ||b - a| - 0| = M |b - a|.$$

This proves the statement for n = 1, and using this we can now prove the statement for general n. Note that  $df_{j,a} = \pi_j \circ df_a$  and  $||\pi_j|| = 1$  so

$$\|df_{j,a}\| \leq \|\pi_j\| \|df_a\| \leq M$$

by exercise 0.6.8. Applying proposition 2.4.10 to the component functions  $f_i$ , so

$$|f_j(b) - f_j(a)| \leq M|b - a|$$

for all  $a, b \in B$  and all j = 1, ..., n. Taking the maximum over all j proves the result.  $\Box$ 

- **Exercise 2.4.11.** (a) The statement of the proposition 2.4.10 requires that the domain of f is convex. Where in the proof is this used?
- (b) Prove that  $\|\pi_i\| = 1$ .
- (c) If you haven't already done it, do exercise 0.6.8.

# 2.4.C Inverse function theorem

The inverse function theorem is an extremely important foundational result. It comes up frequently in real analysis and differential geometry, and has inspired significant developments in algebraic geometry and number theory. **Definition 2.4.12.** A function  $f : U \to \mathbb{R}^n$  is *étale* if it is differentiable and étale at every  $a \in U$ .

**Theorem 2.4.13** (Inverse function theorem). Suppose  $f : U \to \mathbb{R}^n$  is étale. Then f is open. Moreover, for every  $a \in U$ , there exists an open neighborhood V of a such that  $f|_V$  is injective, the inverse function  $(f|_V)^{-1}$  is differentiable, and

$$d((f|_V)^{-1})_y = (df_{(f|_V)^{-1}(y)})^{-1}$$

for all  $y \in f(V)$ .

The proof will require an extremely lengthy discussion. For the time being, let us content ourselves with thinking about the statement of the theorem.

First off, the statement invites comparison with the single variable version we established in theorem 1.3.21. If I is an open interval in  $\mathbb{R}$ , then  $f : I \to \mathbb{R}$  is étale if and only if it is differentiable and f'(x) never vanishes, so the hypotheses of the theorem above and the single variable version match up exactly. The multivariable version above asserts that f must be open and "locally injective." On the other hand, the single variable version asserts that f is strictly monotone; and we've seen that strict monotonicity implies that f is open and injective (cf. exercises 1.3.9 and 1.3.10). The assertion on the differentiability of the inverse and the formula for the derivative of the inverse match up exactly.

Thus, the single variable version is a stronger statement (when it applies) because it guarantees full injectivity everywhere on the domain, whereas the multivariable version above only guarantees "local injectivity" near every point of the domain. This is the only difference between the two statements, and here is an example to show that we cannot expect full injectivity in the multivariable setting.

**Exercise 2.4.14.** Consider the function  $f : \mathbb{R}^2 \to \mathbb{R}^2$  given by

$$f(x, y) = (e^x \cos y, e^x \sin y).$$

- (a) Compute df and verify that f is étale.
- (b) Show that f is *not* injective. In fact, for any point (a, b) in the range, show that there exist infinitely many points in the domain which map to (a, b).
- (c) Find an open neighborhood V of the origin such that  $f|_V$  is injective.
**Exercise 2.4.15.** Recall the function  $f(x) = x + 2x^2 \sin(1/x)$  that we saw in exercise 1.3.14. As we've seen, f'(0) = 1 but f is not monotone (hence not injective, by exercise 1.3.9) on any open neighborhood of 0. Explain why this does not violate the inverse function theorem. In other words, show that f is not étale on any open neighborhood of 0.

## 2.4.D Proof of the inverse function theorem $\star$

This section is devoted to the proof of the inverse function theorem 2.4.13. It's a long and arduous proof.

#### Differentiability of the local inverse

As it turns out, once a local inverse has been constructed, proving that it is differentiable is not terribly difficult. The proof of this part is similar in spirit to the single variable version (cf. the proof of theorem 1.3.21).

**Proposition 2.4.16.** Suppose  $f : U \to \mathbb{R}^n$  is differentiable, open, and injective. Then  $f^{-1}$  is differentiable and

$$d(f^{-1})_y = (df_{f^{-1}(y)})^{-1}$$

for all  $y \in f(U)$ .

*Proof.* Fix  $x \in U$  and y = f(x). Let

$$s(k) = f^{-1}(y+k) - f^{-1}(k) - df_x^{-1}(k).$$

We want to show that |s(k)| = o(|k|) as  $k \to 0$ . In fact, since df<sub>x</sub> is an invertible linear map, it is sufficient to show that  $|df_x(s(k))| = o(|k|)$  (cf. exercise 2.3.5 part (c)).

Define

$$h(k) = f^{-1}(y+k) - f^{-1}(y).$$

Since f is injective, we see that h(k) = 0 if and only if k = 0. Moreover,  $f^{-1}$  is continuous since f is open, which implies that h is also continuous. Now by rewriting s in terms of h,

#### 2 Multivariable derivatives

we have the following.

$$\begin{split} s(k) &= h(k) - df_x^{-1}(k) \\ df_x(s(k)) &= df_x(h(k)) - k \\ &= df_x(h(k)) - y + k + y \\ &= df_x(h(k)) - f(x + h(k)) + f(x). \end{split}$$

In other words, if we consider the "error in approximation" function

$$\mathbf{r}(\mathbf{h}) = \mathbf{f}(\mathbf{x} + \mathbf{h}) - \mathbf{f}(\mathbf{x}) - \mathbf{d}\mathbf{f}_{\mathbf{x}}(\mathbf{h}),$$

then

$$df_{x}(s(k)) = -r(h(k))$$

for all k. This means that, for all  $k \neq 0$ , we have

$$\frac{|df_{x}(s(k))|}{|k|} = \frac{|r(h(k))|}{|k|} = \frac{|r(h(k))|}{|h(k)|} \cdot \frac{|h(k)|}{|k|}$$

where we have used the fact that  $h(k) \neq 0$  since  $k \neq 0$ . Now, since |r(h)| = o(|h|) as  $h \to 0$  and since the function h is continuous at k = 0, we have that

$$\lim_{k\to 0}\frac{|\mathbf{r}(\mathbf{h}(k))|}{|\mathbf{h}(k)|}=0.$$

Thus it is sufficient to show that there exists a constant M such that  $|h(k)| \leq M|k|$  for all sufficiently small values of k.

Since |r(h)| = o(|h|) as  $h \to 0$ , there exists  $\delta > 0$  such that  $|r(h)| \le |h|$  for all  $|h| < \delta$ . Then we have

$$\begin{split} |h| &\ge |r(h)| = |f(x+h) - f(x) - df_x(h)| \\ &\ge |df_x(h)| - |f(x+h) - f(a)| \\ &\ge \frac{|h|}{\|df_x^{-1}\|} - |f(x+h) - f(x)| \end{split}$$

where we have used exercise 0.6.7 for the final step. After rearranging, this becomes

$$|h| \leq \left(\frac{1}{\|df_x^{-1}\|} - 1\right)^{-1} |f(x+h) - f(x)|.$$

110

Set

$$M=\left(\frac{1}{\|df_x^{-1}\|}-1\right)^{-1}$$

Now if h = h(k), we have f(x + h(k)) - f(x) = k, so the last inequality displayed above reads

$$|h(k)| \leqslant M|f(x+h(k)) - f(x)| = M|k|.$$

#### Reductions

For the remainder of this proof,  $f : U \to \mathbb{R}^n$  will denote an étale map. In particular, this means that m = n, ie, that U is an open subset of  $\mathbb{R}^n$ . We'll closely follow Terry Tao's exposition of Saint Raymond's proof [Sai02].

Let's begin by making a couple of helpful definitions.

**Definition 2.4.17.** For any point  $a \in U$ , we make the following definitions.

- We say that f is *locally injective at* a if there exists an open neighborhood V of a such that f|<sub>V</sub> is injective.
- We say that f is *locally surjective at* a if f(a) is an interior point of f(V) for every neighborhood V of a.

**Exercise 2.4.18.** Show that f is open if and only if f is locally surjective at every  $a \in U$ .

In other words, to prove the inverse function theorem, we want to show that f is locally injective and locally surjective at every point  $a \in U$ . We can replace f with the function  $x \mapsto f(x + a)$  in order to assume that a = 0. Then, we can replace f with the function  $x \mapsto f(x) - f(0)$  in order to assume that f(0) = 0. Finally, since  $df_0$  is invertible, we can replace f with  $df_0^{-1} \circ f$  in order to assume that  $df_0 = id$ .

**Exercise 2.4.19.** Check that you understand the "we can assume"s in the previous paragraph. More precisely, suppose  $a \in U$ ,  $\ell : \mathbb{R}^n \to \mathbb{R}^n$  is a linear map,  $\alpha_0 \in \mathbb{R}^n$ , and  $\alpha : \mathbb{R}^n \to \mathbb{R}^n$  is the function

$$\alpha(\mathbf{x}) = \ell(\mathbf{x}) + \alpha_0.$$

Prove the following facts.

(a) Show that  $\alpha$  is bijective, and calculate d $\alpha$  and d $\alpha^{-1}$ .

(b) Let  $\tilde{f} = \alpha \circ f$ .

- Show that f is étale if and only if f is étale.
- Show that f is locally injective at a if and only if  $\tilde{f}$  is locally injective at a.
- Show that f is locally surjective at a if and only if  $\tilde{f}$  is locally surjective at a.

(c) Let  $\tilde{f} : \alpha^{-1}(U) \to \mathbb{R}^n$  be the composite  $\tilde{f} = f \circ \alpha$ .

- Show that f is étale if and only if f is étale.
- Show that f is locally injective at a if and only if  $\tilde{f}$  is locally injective at  $\alpha^{-1}(a)$ .
- Show that f is locally surjective at a if and only if  $\tilde{f}$  is locally surjective at  $\alpha^{-1}(a)$ .
- (d) Figure out what  $\ell$  and  $\alpha_0$  are for each of the "we can assume"s of the previous paragraph.

In other words, we want to show that, if f is an étale map such that f(0) = 0 and  $df_0 = id$ , then f is locally injective and locally surjective at 0. Local surjectivity is easier; local injectivity is very, very difficult.

## Local surjectivity

Since f(0) = 0 and  $df_0 = id$ , we have |f(h) - h| = o(|h|) as  $h \to 0$  by the definition of the derivative, so there exists  $\delta > 0$  such that

$$|f(h) - h| \le |h|/2$$
 (2.4.20)

for all  $|h| < \delta$ . Replacing f with the function  $x \mapsto f(x/\delta)$ , we can assume without loss of generality that  $\delta = 1$  (cf. exercise 2.4.19). We now claim that for all 0 < r < 1, we have

$$B(0, r/3) \subseteq f(B(0, r)).$$
 (2.4.21)

Suppose  $b \in B(0, r/3)$ . Let  $\nu : \mathbb{R}^m \to \mathbb{R}$  be the differentiable function given by  $\nu(x) = x_1^2 + \cdots x_n^2$  and then consider the function  $\sigma : \overline{B}(0, r) \to \mathbb{R}$  given by

$$\sigma(\mathbf{x}) = \mathbf{v}(\mathbf{f}(\mathbf{x}) - \mathbf{b}).$$

Since  $\overline{B}(0, r)$  is compact, this function must attain its minimum at a point  $a \in \overline{B}(0, r)$ . We will show that  $a \in B(0, r)$  and that f(a) = b, thus proving equation (2.4.21).

First of all, we must have  $a \in B(0, r)$ . Suppose for a contradiction that |a| = r. Then by equation (2.4.20) we would have

$$r/2 \ge |f(a) - a| \ge |a| - |f(a)| = r - |f(a)| > 2r/3,$$

where we've used the reverse triangle inequality for the second step. This is evidently contradiction, since r/2 < 2r/3. In other words, we have proved that a is an interior point of the domain of  $\sigma$ .

Now we prove that f(a) = b. The chain rule 2.3.3 tells us that

$$d\sigma_a = d\nu_{f(a)-b} \circ df_a.$$

Since the interior point a is the absolute minimum of  $\sigma$ , the interior extremum theorem 2.3.21 and theorem 2.3.34 tells us that  $d\sigma_a = 0$ , which means that  $df_a(\mathbb{R}^n) \subseteq \ker(d\nu_{f(a)-b})$ . Since  $df_a$  is invertible, this means that  $d\nu_{f(a)-b} = 0$ . But we have

$$\nu'(x) = \begin{bmatrix} 2x_1 & 2x_2 & \cdots & 2x_n \end{bmatrix}$$

so  $dv_{f(a)-b} = 0$  if and only if f(a) - b = 0. We have now proved equation (2.4.21).

Notice that equation (2.4.21) implies that f is locally surjective at 0. Indeed, any open neighborhood V of 0 contains an open ball of the form B(0, r) for 0 < r < 1, and then

$$f(\mathbf{V}) \supseteq f(B(0, \mathbf{r})) \supseteq B(0, \mathbf{r}/3),$$

so f(0) = 0 is an interior point of f(V).

#### Local injectivity under assumptions

Proving local injectivity in general is very challenging, but is a bit easier if we add some hypotheses to the statement of the theorem. In this section, we look at some of these arguments under additional assumptions.

Assuming n = 1. As we've already noted, we've already proved not just local injectivity but full injectivity when n = 1 in theorem 1.3.21. In any case, it's worth isolating the argument of injectivity and inspecting it closely.

Suppose we have  $a_1 < a_2$  such that  $f(a_1) = f(a_2)$ . We'll prove that this leads to a

contradiction using two different but very similar arguments. The first argument is very quick. The mean value theorem guarantees that there exists  $\xi$  between  $a_1$  and  $a_2$  such that  $f'(\xi) = 0$ , which contradicts our assumption that f is étale.

The second argument avoids using the mean value theorem basically by proving the mean value theorem for this particular situation. The idea is to set  $b := f(a_1) = f(a_2)$  and consider the function

$$\sigma(\mathbf{x}) = (\mathbf{f}(\mathbf{x}) - \mathbf{b})^2$$

Since  $\sigma$  is a continuous function on the compact interval  $[a_1, a_2]$ , the extreme value theorem guarantees that it attains its maximum at some point  $\xi \in [a_1, a_2]$ . Observe that  $\sigma(a_1) = \sigma(a_2) = 0$  and  $\sigma \ge 0$ , so if  $\sigma(\xi) = 0$ , then  $\sigma = 0$  constantly, so f is constantly equal to b, which means the derivative vanishes on  $(a_1, a_2)$  by proposition 2.4.6, contradicting our assumption that f is étale. Thus  $\sigma(\xi)$  must be an interior point of the interval  $(a_1, a_2)$  and  $\sigma(\xi) \ne 0$ , ie,  $f(\xi) \ne b$ . Then the interior extremum theorem 1.2.19 plus the chain rule tells us that

$$0 = \sigma'(\xi) = 2(f(\xi) - b)f'(\xi).$$

Since  $f(\xi) \neq b$ , this says that  $f'(\xi) = 0$ , which again contradicts our assumption that f is étale.

This second argument is useful because the proof in the general setting involves maximizing a multivariable function that's very similar to the single variable function  $\sigma$  above. But before we get to the general proof, here is another proof under additional hypotheses.

Assuming continuous differentiability. We haven't formally defined it yet, but f is said to be "continuously differentiable" if all of the partial derivatives of the component functions  $\partial_i f_j$  are continuous (cf. definition 2.5.4). The proof that f is locally injective at 0 is a little easier if we assume that f is continuously differentiable, but still a little tricky. The proof here is essentially the one from [Spi65, chapter 2].

Since  $df_0 = id$ , we know that  $\partial_i f_j(0) = \delta_{i,j}$ . Since  $\partial_i f_j$  is continuous at 0, there exists  $\delta > 0$  such that

$$|\partial_{\mathfrak{i}}f_{\mathfrak{j}}(0) - \partial_{\mathfrak{i}}f_{\mathfrak{j}}(x)| = |\delta_{\mathfrak{i},\mathfrak{j}} - \partial_{\mathfrak{i}}f_{\mathfrak{j}}(x)| \leqslant \frac{1}{2n}.$$

Let  $V = B(0, \delta)$ . We will show that  $f|_V$  is injective.

Let g(x) = x - f(x). It turns out that it is sufficient to prove that  $||dg_x|| \le 1/2$  for all

 $x \in V$ . Indeed, if we can prove this, then proposition 2.4.10 would tell us that

$$|g(x_1) - g(x_2)| \leq \frac{1}{2}|x_1 - x_2|$$

for all  $x_1, x_2 \in V$ . But

$$|g(x_1) - g(x_2)| = |(x_1 - f(x_1)) - (x_2 - f(x_2))| \ge |x_1 - x_2| - |f(x_1) - f(x_2)|$$

by the reverse triangle inequality. This leads to the following.

$$\begin{split} |\mathbf{x}_1 - \mathbf{x}_2| - |\mathbf{f}(\mathbf{x}_1) - \mathbf{f}(\mathbf{x}_2)| &\leq \frac{1}{2} |\mathbf{x}_1 - \mathbf{x}_2| \\ \frac{1}{2} |\mathbf{x}_1 - \mathbf{x}_2| &\leq |\mathbf{f}(\mathbf{x}_1) - \mathbf{f}(\mathbf{x}_2)| \\ |\mathbf{x}_1 - \mathbf{x}_2| &\leq 2 |\mathbf{f}(\mathbf{x}_1) - \mathbf{f}(\mathbf{x}_2)| \end{split}$$

This shows that  $f(x_1) = f(x_2)$  if and only if  $x_1 = x_2$ , or, in other words, that  $f|_V$  is injective.

We'll prove that  $\|df_x\| \leq 1/2$  only for the max norm (and, correspondingly, the max operator norm). Observe that  $\partial_i g_j(x) = \delta_{i,j} - \partial_i f_j(x)$ , so we have

$$|g'(\mathbf{x})|_{\infty} = \max_{i,j} |\partial_i g_j(\mathbf{x})| \leqslant \frac{1}{2n}$$

for all  $x \in V$ . Using exercise 0.6.13, we see that  $\|dg_x\|_{\infty} \leq 1/2$  for all  $x \in V$ .

*Unimportant remark.* It is also true that  $||dg_x||_2 \leq 1/2$ . To prove this, we would use the fact that  $||\ell||_2 \leq \sqrt{mn}|A|_{\infty}$  when A is the standard matrix representation of a linear map  $\ell : \mathbb{R}^m \to \mathbb{R}^n$  (cf. [GV13, section 2.3.2]), instead of exercise 0.6.13. In any case, we didn't prove proposition 2.4.10 for the eucldiean norm situation, so perhaps its better to stick with the max norm here anyway.

#### Local injectivity

Incomplete. I'll write this up eventually. The continuously differentiable case we proved above is really sufficient almost all of the time (and is all we'll use as we go forward), but if you're really curious about the general case, look at Terry Tao's blog post.

## **2.5** C<sup>k</sup> hierarchy

Throughout this section, U denotes an open subset of  $\mathbb{R}^m$ .

## 2.5.A Continuous differentiability

#### **Continuity of partials**

The following theorem is the promised sort-of-converse to proposition 2.3.26. Recall from exercise 2.3.28 above that  $\partial_i f$  existing for all i is *not* sufficient to guarantee differentiability. But, insisting that  $\partial_i f$  exist *and be continuous* turns out to be enough to guarantee differentiability of f. The proof is harder than you might expect.

**Theorem 2.5.1.** Suppose  $f : U \to \mathbb{R}$  is a function such that the *i*th partial derivative  $\partial_i f : U \to \mathbb{R}$  exists for all i = 1, ..., m. If  $a \in U$  is a point such that  $\partial_i f$  is continuous at a for all *i*, then *f* is differentiable at *a*.

*Proof.* If f is to be differentiable at a, we know from theorem 2.3.34 what  $df_a$  needs to be in terms of the partial derivatives; so let's turn this on its head by writing down the linear map defined by the partial derivatives, and then proving that that linear map is in fact  $df_a$ .

More precisely, we mean the following. Let  $\ell$  be the linear map  $\mathbb{R}^m \to \mathbb{R}$  such that

$$[\ell] = \begin{bmatrix} \partial_1 f(\alpha) & \cdots & \partial_m f(\alpha) \end{bmatrix}.$$

In other words,  $\ell : \mathbb{R}^m \to \mathbb{R}$  is the linear map given by

$$\ell(h_1,\ldots,h_m) = h_1 \partial_1 f(a) + \cdots + h_m \partial_m f(a).$$

We will show that

$$|f(a+h) - f(a) - \ell(h)| = o(|h|)$$
 as  $h \to 0$ ,

which implies that f is differentiable at a and that  $df_a = \ell$ . In other words, we will show that, for any  $\epsilon > 0$ , there exists a  $\delta > 0$  such that, for all  $h \in \mathbb{R}^m$  with  $|h| < \delta$ , we have

$$|f(a+h) - f(a) - \ell(h)| \leq |h|\epsilon.$$
(2.5.2)

Let  $\delta_0 > 0$  be small enough that  $B(a, \delta_0) \subseteq U$ . We will soon make it smaller to get inequality (2.5.2) to hold, but for the moment it is sufficient that  $B(a, \delta_0) \subseteq U$ . Suppose



Figure 2.5.3: A picture of a,  $a_1$  and  $a_2 = a + h$ , as defined in the proof of theorem 2.5.1. Defining  $\xi_1$  and  $\xi_2$  using the mean value theorem as in the proof, the point  $a + \xi_1 e_1$  is somewhere along the horizontal line joining a and  $a + h_1$ , and the point  $a + h_1 + \xi_2 e_2$  is somewhere on the vertical line joining  $a + h_1$  and a + h.

 $h=(h_1,\ldots,h_m)\in \mathbb{R}^m \text{ with } |h|<\delta_0. \text{ Then consider the following vectors in } \mathbb{R}^m.$ 

$$a_{0} = a$$

$$a_{1} = a + (h_{1}, 0, 0, ..., 0, 0)$$

$$a_{2} = a + (h_{1}, h_{2}, 0, ..., 0, 0)$$

$$\vdots$$

$$a_{m-1} = a + (h_{1}, h_{2}, h_{3}, ..., h_{m-1}, 0)$$

$$a_{m} = (h_{1}, h_{2}, h_{3}, ..., h_{m-1}, h_{m}) = a + h$$

The vectors  $a_0, \ldots, a_m$  define a sort of "zigzag path" from a to a + h. See figure 2.5.3. Clearly  $a_i \in B(a, \delta_0)$  for all  $i = 1, \ldots, m$ .

Notice that

$$f(a+h)-f(h)=\sum_{i=1}^m \left(f(a_i)-f(a_{i-1})\right).$$

117

Let us fix i temporarily. Notice that  $a_i = a_{i-1} + h_i e_i$ , so

$$f(a_i) - f(a_{i-1}) = f(a_{i-1} + h_i e_i) - f(a_{i-1}).$$

Suppose  $h_i \neq 0$ . We can then apply the mean value theorem 1.3.3 to the single variable function

$$t \mapsto f(a_{i-1} + te_i)$$

on the closed interval between 0 and  $h_i$ . This function is well-defined and differentiable on this interval: its derivative at a value t is  $\partial_i f(a_{i-1} + te_i)$ , by definition of partial derivatives. The mean value theorem 1.3.3 says that there exists  $\xi_i$  between 0 and  $h_i$  such that

$$\partial_{i}f(a_{i-1} + \xi_{i}e_{i}) = \frac{f(a_{i-1} + h_{i}e_{i}) - f(a_{i-1})}{h_{i}} = \frac{f(a_{i}) - f(a_{i-1})}{h_{i}},$$

or, in other words,

$$f(a_i) - f(a_{i-1}) = \xi_i \partial_i f(a_{i-1} + \xi_i e_i)$$

If  $h_i = 0$ , we set  $\xi_i = 0$  as well. See figure 2.5.3 again. Unfixing i, we find that

$$f(a+h)-f(a) = \sum_{i=1}^{m} (f(a_i)-f(a_{i-1})) = \sum_{i=1}^{m} \xi_i \partial_i f(a_{i-1}+\xi_i e_i).$$

Thus

$$|f(a+h) - f(a) - \ell(h)| \leq \sum_{i=1}^{m} |\xi_i| \cdot |\partial_i(a_{i-1} + \xi_i e_i) - \partial_i f(a)|.$$

Since  $\partial_i f$  is continuous at a, there exists  $\delta_i > 0$  such that  $|\partial_i(a+k) - \partial_i(a)| < \varepsilon/m$  for all  $|k| < \delta_i$ . Let  $\delta = \min\{\delta_0, \delta_1, \dots, \delta_m\}$ . If  $|h| < \delta$ , then

$$|\partial_{i}(a_{i-1} + \xi_{i}e_{i}) - \partial_{i}f(a)| < \varepsilon/m$$

which means that

$$|f(a+h)-f(a)-\ell(h)|\leqslant \sum_{i=1}^m \frac{|\xi_i|\varepsilon}{m}=\frac{\varepsilon}{m}\sum_{i=1}^m |\xi_i|\leqslant |h|\varepsilon,$$

where we use the fact that  $|\xi_i| \leq |h_i|$  for all i for the final step.

*Unimportant remark.* If you're using the euclidean norm, you'll have to apply exercise 0.3.11 for the final step of the proof above. This is another example of why the max norm is often easier to deal with.

#### **Continuous differentiability**

**Definition 2.5.4** (Continuously differentiable functions). We say that f is *continuously differentiable* if  $\partial_i f_j : U \to \mathbb{R}$  is continuous for all i = 1, ..., m and j = 1, ..., n. Theorem 2.5.1 guarantees that such a function is, in particular, differentiable.

As we saw in the single variable case, continuously differentiable functions have lots of nice properties. For example, recall the calculation we did in example 2.3.37 showing that the critical locus was a line inside  $\mathbb{R}^2$ , which is a closed subset. This fact is a special case of the following.

**Exercise 2.5.5.** Suppose  $f : U \to \mathbb{R}^n$  is continuously differentiable. Show that the "critical locus" of f (ie, the set of all critical points of f) is a closed subset of U.

Here is a multivariable version of the differentiable-but-not-continuously-differentiable function from exercise 1.2.25.

**Exercise 2.5.6.** Let  $f : \mathbb{R}^2 \to \mathbb{R}$  denote the following function.

$$f(x,y) = \begin{cases} x^2 \sin(1/x) + y^2 \sin(1/y) & \text{if } x, y \neq 0\\ x^2 \sin(1/x) & \text{if } x \neq 0 \text{ and } y = 0\\ y^2 \sin(1/y) & \text{if } x = 0 \text{ and } y \neq 0\\ 0 & \text{if } x = y = 0 \end{cases}$$

Show that f is differentiable, but not continuously differentiable.

**Exercise 2.5.7.** You might recall proving at some point in a previous class that  $\mathbb{R}^2$  and  $\mathbb{R}$  have the same cardinality: in other words, that there exists a bijection  $f : \mathbb{R}^2 \to \mathbb{R}$ . Show that any such bijection cannot be continuously differentiable.

*Possible hint*. Suppose for a contradiction that  $f : \mathbb{R}^2 \to \mathbb{R}$  is a continuously differentiable bijection. Let U be an open subset of  $\mathbb{R}^2$  such that  $\partial_1 f(x) \neq 0$  for all  $x \in U$ , and then consider the function  $g : U \to \mathbb{R}^2$  given by  $g(x_1, x_2) = (f(x_1, x_2), x_2)$ .

**Proposition 2.5.8.** Suppose  $f : U \to \mathbb{R}^n$  is continuously differentiable, open, and injective. Then  $f^{-1}$  is also continuously differentiable.

*Proof.* We know from proposition 2.4.16 that  $f^{-1}$  is differentiable and that

$$d(f^{-1})_y = (df_{f^{-1}(y)})^{-1}$$

for all  $y \in U$ . Taking standard matrix representations, we get

$$[d(f^{-1})_{y}] = [(df_{f^{-1}(y)})^{-1}] = [df_{f^{-1}(y)}]^{-1}.$$

By theorem 2.3.34, the (j, i)-entry on the left hand side is  $\partial_i (f^{-1})_j (y)$ , so this equality says that  $\partial_i (f^{-1})_j : f(U) \to \mathbb{R}$  is equal to the following composite.

$$f(U) \xrightarrow{f^{-1}} U \xrightarrow{x \mapsto [df_x]} GL_n \xrightarrow{A \mapsto A^{-1}} GL_n \xrightarrow{(j,i)-entry} \mathbb{R}$$

Here  $GL_n$  denotes the set of invertible  $n \times n$  matrices as in section 0.5.B. Each of the functions being composed here is continuous (cf. exercise 2.2.8 and lemma 0.5.18), so we conclude that  $\partial_i (f^{-1})_j$  is continuous.

#### Continuous differentiability and continuity of the derivative **\***

For this section, we will need to make substantial use of the continuity properties of the operator norm. But it's also not terribly important. The main result of this section, proposition 2.5.11, is aesthetically rather satisfying, and can occasionally streamline proofs.

**Lemma 2.5.9.** Suppose  $f: U \to \mathbb{R}^n$  is differentiable. Then  $df: U \to \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$  is continuous if and only if  $df_j: U \to \mathcal{L}(\mathbb{R}^m, \mathbb{R})$  is continuous for all j = 1, ..., n.

*Proof.* We will use proposition 2.3.7, which tells us that f is differentiable if and only if  $f_j$  is differentiable for all j = 1, ..., n, and moreover that  $df_{j,a} = \pi_j \circ df_a$  for all  $a \in U$ . In other words, the following diagram commutes.



You will verify in exercise 2.5.10 below that the map "postcompose with  $\pi_j$ " map  $\pi_j \circ -$ :  $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n) \to \mathcal{L}(\mathbb{R}^m, \mathbb{R})$  is continuous. Thus, if df is continuous, then df<sub>j</sub> is also continuous.

Conversely, suppose that  $df_j$  is continuous for all j. Let us show that df is continuous at any  $a \in U$ . Suppose  $\epsilon > 0$ . Since  $df_j$  is continuous at a, there exists  $\delta_j > 0$  such that

$$\|df_{j,x}-df_{j,\alpha}\|<\frac{\varepsilon}{\sqrt{n}}$$

for all  $x \in U$  such that  $|x - a| < \delta_j$ . Let  $\delta := \min\{\delta_1, \dots, \delta_n\}$ . Then  $\|df_x - df_a\| < \varepsilon$  for all  $x \in U$  such that  $|x - a| < \delta$ , as you will verify in exercise 2.5.10 below.

- **Exercise 2.5.10.** (a) Check that the "postcompose with  $\pi_j$ " map  $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n) \to \mathcal{L}(\mathbb{R}^m, \mathbb{R})$  (ie, the one given by  $\phi \mapsto \pi_j \circ \phi$ ) is continuous. Alternatively, prove the more general statement given in exercise 0.6.12 and then derive this statement.
- (b) With notation as in the latter part of the proof above, verify that  $\|df_a df_b\| < \epsilon$ .

**Proposition 2.5.11.** Suppose  $f: U \to \mathbb{R}^n$  is differentiable. Then f is continuously differentiable if and only if  $df: U \to \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$  is continuous.

*Proof.* Thanks to lemma 2.5.9, we can assume without loss of generality that n = 1. If f is differentiable, then proposition 2.3.26 tells us that  $\partial_i f = df(e_i)$ . In other words, the following diagram commutes.

$$U \xrightarrow{df} \mathcal{L}(\mathbb{R}^m, \mathbb{R})$$

$$\downarrow^{"evaluate at e_i"}$$

$$\partial_i f \xrightarrow{} \mathbb{R}$$

The "evaluate at  $e_i$ " map is continuous by exercise 0.6.11. So, if df is continuous, then  $\partial_i f$  must also be continuous, as it is the composite of two continuous functions. The converse is left as an exercise; see exercise 2.5.12 below.

- **Exercise 2.5.12.** (a) Elaborate on the first sentence of the above proof (in other words, explain why proving theorem 2.5.1 for n = 1 is sufficient to prove it for all n).
- (b) If you haven't already done it, do exercise 0.6.11.
- (c) Complete the proof above by showing that df :  $U \to \mathcal{L}(\mathbb{R}^m, \mathbb{R})$  is continuous.

**Remark 2.5.13.** This reformulation of continuous differentiability is often convenient in proofs. For example, in the proof of proposition 2.5.8, we calculated that

$$d(f^{-1})_y = (df_{f^{-1}(y)})^{-1}.$$

This says that the  $d(f^{-1}) = inversion \circ df \circ f^{-1}$ .

$$\begin{array}{c} f(U) \xrightarrow{d(f^{-1})} GL(\mathbb{R}^n) \\ f^{-1} \downarrow & \uparrow \text{inversion} \\ U \xrightarrow{df} GL(\mathbb{R}^n) \end{array}$$

We know that  $f^{-1}$  is continuous; df is continuous by proposition 2.5.11, and inversion is continuous by exercise 0.6.15. Thus  $d(f^{-1})$  is continuous, so proposition 2.5.11 implies that  $f^{-1}$  is continuously differentiable.

## **2.5.B** C<sup>k</sup> functions

**Definition 2.5.14** ( $C^k$  functions). We say that a function  $f : U \to \mathbb{R}^n$  is  $C^0$  if it is continuous. Then, inductively, for any positive integer k, we say that a function  $f : U \to \mathbb{R}^n$  is  $C^k$  if the partial derivatives  $\partial_i f_i$  all exist and are  $C^{k-1}$  functions  $U \to \mathbb{R}$ .

By definition,  $f : U \to \mathbb{R}^n$  is  $C^1$  if and only if it is continuously differentiable. It is clear from the definitions that a function  $f : U \to \mathbb{R}^n$  is  $C^k$  if and only if its component functions  $f_j : U \to \mathbb{R}$  are  $C^k$ . Here are some results showing that reasonable ways of combining  $C^k$ functions still yield  $C^k$  functions.

**Exercise 2.5.15.** Show that the set of all  $C^k$  functions  $U \to \mathbb{R}^n$  form a vector space.

**Exercise 2.5.16.** Suppose U is an open subset  $\mathbb{R}^m$ , that V is an open subset of  $\mathbb{R}^n$ , and that  $f: U \to \mathbb{R}^n$  and  $g: V \to \mathbb{R}^p$  are both  $C^k$ , and that  $f(U) \subseteq V$ . Show that  $g \circ f$  is also  $C^k$ .

**Exercise 2.5.17.** Suppose f, g : U  $\rightarrow \mathbb{R}$  are both C<sup>k</sup>. Show that fg is also C<sup>k</sup>.

**Exercise 2.5.18.** Suppose  $f : U \to \mathbb{R}$  is  $C^k$  and  $f(x) \neq 0$  for all  $x \in U$ . Show that 1/f is also  $C^k$ .

*Possible hint*. Note that 1/f is f composed with the single variable inversion function.

**Exercise 2.5.19.** Suppose  $f : U \to \mathbb{R}^n$  is étale, injective, and  $C^k$  for some  $k \ge 1$ . Show that  $f^{-1}$  is also  $C^k$ .

*Possible hint.* This is a multivariable version of proposition 1.4.10, and can be proved by using a similar induction. Some ingredients you might think about combining are the inverse function theorem 2.4.13 and proposition 2.5.8 and lemma 0.5.18 (and their proofs).

#### 2.5.C Equality of mixed partials \*

**Definition 2.5.20** (Mixed partials). Suppose  $f : U \to \mathbb{R}$  is  $\mathbb{C}^2$ . The partial derivative  $\partial_i f : U \to \mathbb{R}$  for any i = 1, ..., m is a continuously differentiable function, so its jth partial derivative  $\partial_j(\partial_i f)$  is a continuous function  $U \to \mathbb{R}$  for any j = 1, ..., m. We denote the function  $\partial_j(\partial_i f)$  more succinctly as  $\partial_{j,i} f$ , or as

$$\frac{\partial^2 f}{\partial x_i x_i}$$

**Theorem 2.5.21** (Equality of mixed partials). Suppose  $f : U \to \mathbb{R}$  is  $C^2$ . Then, for all i, j = 1, ..., m, we have  $\partial_{j,i} f = \partial_{i,j} f$ .

Proof. Observe that

$$\begin{split} \partial_{j,i} f(a) &= \lim_{t_j \to 0} \frac{\partial_i f(a + t_j e_j) - \partial_i f(a)}{t_j} \\ &= \lim_{t_j \to 0} \frac{\lim_{t_i \to 0} \frac{f(a + h_i e_i + t_j e_j) - f(a + t_j e_j)}{t_i} - \lim_{t_i \to 0} \frac{f(a + t_i e_i) - f(a)}{t_i}}{t_j} \\ &= \lim_{t_j \to 0} \lim_{t_i \to 0} \frac{f(a + t_i e_i + t_j e_j) - f(a + t_i e_i) - f(a + t_j e_j) + f(a)}{t_i t_j}. \end{split}$$

Notice that, if we unwind the definition of  $\partial_{i,j} f(a)$  in the same way, we end up taking a double limit of the same expression, but the order of the limits is interchanged. In other words, if we define

$$\sigma(t_i,t_j) = \frac{f(a+t_ie_i+t_je_j) - f(a+t_ie_i) - f(a+t_je_j) + f(a)}{t_it_j},$$

then

$$\partial_{j,i} f(a) = \lim_{t_j \to 0} \lim_{t_i \to 0} \sigma(t_i, t_j) \quad \text{whereas} \quad \partial_{i,j} f(a) = \lim_{t_i \to 0} \lim_{t_j \to 0} \sigma(t_i, t_j).$$

So, to prove the theorem, it is sufficient to prove that

$$\partial_{j,i}f(a) = \lim_{h \to 0} \sigma(t,t).$$

Fix a nonzero real number t and consider the function

$$\alpha(s) = f(a + se_i + te_j) - f(a + se_i).$$

Notice that

$$\sigma(t,t) = \frac{\alpha(t) - \alpha(0)}{t^2}$$

and that  $\alpha$  is differentiable with

$$\alpha'(s) = \partial_i f(a + se_i + te_j) - \partial_i f(a + se_i).$$

Applying the mean value theorem 1.3.3 to  $\alpha$  on the closed interval between 0 and h, there exists  $\xi_i(t)$ , strictly between 0 and t, such that

$$\sigma(t,t) = \frac{\alpha(t) - \alpha(0)}{t^2} = \frac{\alpha'(\xi_i(t))}{t} = \frac{\partial_i f(a + \xi_i(t)e_i + te_j) - \partial_i f(a + \xi_i(t)e_i)}{t}.$$

Now consider the function

$$\beta(s) = \partial_i f(a + \xi_i(t)e_i + se_j).$$

Then  $\beta$  is differentiable with

$$\beta'(s) = \partial_{j,i}f(a + \xi_i(t)e_i + se_j),$$

so, applying the mean value theorem 1.3.3 to  $\beta$  on the closed interval between 0 and h, we find that there exists  $\xi_i(t)$ , strictly between 0 and t, such that

$$\sigma(t,t) = \frac{\beta(t) - \beta(0)}{t} = \beta'(\xi_j(t)) = \partial_{j,i}f(a + \xi_i(t)e_i + \xi_j(t)e_j).$$

Now notice that, since both  $\xi_i(t)$  and  $\xi_j(t)$  are between 0 and h, we have  $\xi_i(t), \xi_j(t) \to 0$ 

as  $h \to 0$ . Since  $\partial_{i,i} f$  is continuous at a, we have

$$\lim_{t\to 0} \sigma(t,t) = \lim_{t\to 0} \partial_{j,i} f(a + \xi_i(t)e_i + \xi_j(t)e_j) = \partial_{j,i} f(a).$$

**Remark 2.5.22.** If  $f : U \to \mathbb{R}$  is  $C^2$ , the *hessian* of f at a, denoted f''(a), is the matrix of second partial derivatives.

$$f''(a) = \begin{bmatrix} \partial_{1,1}f(a) & \partial_{1,2}f(a) & \cdots & \partial_{1,m}f(a) \\ \vdots & \vdots & \ddots & \vdots \\ \partial_{m,1}f(a) & \partial_{m,2}f(a) & \cdots & \partial_{m,m}f(a) \end{bmatrix}$$

Theorem 2.5.21 is equivalent to the assertion that f''(a) is a symmetric matrix.

There are many versions of this theorem, all with slightly different hypotheses. For instance, there is the version due to Peano [Rud76, theorem 9.41]. In any case, it's worth remarking that some continuity condition on the mixed partials is crucial; it is not enough that the mixed partials exist.

**Exercise 2.5.23.** Consider the following function  $f : \mathbb{R}^2 \to \mathbb{R}$ .

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

(a) Calculate all of the following.

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = \frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 f}{\partial x \partial y}$$

(b) Show that the mixed partials are discontinuous at the origin.

(c) Show that the mixed partials disagree at the origin.

#### 2.5.D Taylor's theorem \*

The main difficulty in the multivariable Taylor's theorem is notation; we've already done most of the hard work of the proof with the single variable versions of theorems 1.4.14 and 1.4.20. Let us first discuss some standard notational conventions that make the statement of Taylor's theorem more readable.

#### Multi-index notation

**Definition 2.5.24** (Multi-index notation). A *multi-index* is an m-tuple  $\alpha = (\alpha_1, ..., \alpha_m)$  where  $\alpha_i$  is a non-negative integer. For a multi-index  $\alpha$  and any non-negative integer  $\ell$ , we define the following.

$$\begin{aligned} |\alpha| &= \alpha_1 + \alpha_2 + \dots + \alpha_n \\ \alpha! &= \alpha_1! \alpha_2! \cdots \alpha_m! \\ \binom{\ell}{\alpha} &= \frac{\ell!}{\alpha!} = \frac{\ell!}{\alpha_1! \cdots \alpha_m!} \end{aligned}$$

If  $\alpha$  and  $\beta$  are both multi-indices, we define the sum  $\alpha + \beta$  componentwise; then we have  $|\alpha + \beta| = |\alpha| + |\beta|$ . We let 0 denote the multi-index (0, 0, ..., 0), and for each i = 1, ..., m, we let  $1_i$  denote the multi-index that is 1 in position i and zero everywhere else.

If  $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_m) \in \mathbb{R}^m$ , we define

$$x^{\alpha} = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_m^{\alpha_m}$$

In particular, we have  $x^0 = 1$  and  $x^{1_i} = x_i$ . Also, if  $f: U \to \mathbb{R}$  is  $C^{|\alpha|}$ , we define

$$\partial^{\alpha} f = \partial_1^{\alpha_1} \partial_2^{\alpha_2} \cdots \partial_m^{\alpha_m} f,$$

where we use the convention that  $\partial^0 f = f$ . In particular, we have  $\partial^{1_i} f = \partial_i f$ .

We now establish some rules of doing algebra with multi-indices.

**Exercise 2.5.25.** Suppose  $\alpha$ ,  $\beta$  are multi-indices and  $x = (x_1, \ldots, x_m) \in \mathbb{R}^m$ . Prove that

$$x^{\alpha}x^{\beta} = x^{\alpha+\beta}.$$

**Exercise 2.5.26** (Equality of mixed partials, multi-index version). Suppose  $\alpha$ ,  $\beta$  are multi-indices and  $f: U \to \mathbb{R}$  is  $C^{|\alpha+\beta|}$ . Prove that

$$\partial^{\alpha}\partial^{\beta}f = \partial^{\alpha+\beta}f.$$

*Possible hint*. First consider the case when  $\alpha = 1_i$  for some i = 1, ..., m.

**Theorem 2.5.27** (Multinomial theorem). *For any*  $x = (x_1, ..., x_m) \in \mathbb{R}^m$ *, we have* 

$$(\mathbf{x}_1 + \dots + \mathbf{x}_m)^{\ell} = \sum_{|\alpha| = \ell} {\ell \choose \alpha} \mathbf{x}^{\alpha}.$$

The proof of the multinomial theorem is omitted because we won't actually have much use for it. If you're interested, try it yourself; if you get stuck, you can find a proof on Wikipedia.

There is a convenient combinatorial interpretation of  $\binom{\ell}{\alpha}$  when  $\ell = |\alpha|$  which generalizes the interpretation of binomial coefficients. For this reason, the  $\binom{\ell}{\alpha}$  are often called "multinomial coefficients."

**Proposition 2.5.28.** Suppose  $\alpha$  is a multi-index and  $\ell = |\alpha|$ . Then the multinomial coefficient  $\binom{\ell}{\alpha}$  is the number of ways of placing  $\ell$  objects into m bins, with  $\alpha_1$  objects in the first bin,  $\alpha_2$  objects in the second bin, and so on. In particular,  $\binom{\ell}{\alpha}$  is a positive integer.

*Proof.* The number of ways of choosing  $\alpha_1$  objects out of the  $\ell$  to place in the first bin is

$$\binom{\ell}{\alpha_1}.$$

Now we have  $l - \alpha_1$  objects remaining, and the number of ways of choosing  $\alpha_2$  of them to be placed into the second bin is

$$\binom{\ell-\alpha_1}{\alpha_2}$$

Continuing in this way, the total number of ways of placing the l objects into the first m bins is the product of all of these binomial coefficients

$$\binom{\ell}{\alpha_1}\binom{\ell-\alpha_1}{\alpha_2}\cdots\binom{\ell-\alpha_1-\cdots-\alpha_{m-1}}{\alpha_m}.$$

But if we expand out each of these binomial coefficients, this product is

$$\frac{\ell!}{\alpha_1!(\ell-\alpha_1)!}\frac{(\ell-\alpha_1)!}{\alpha_2!(\ell-\alpha_1-\alpha_2)!}\cdots\frac{(\ell-\alpha_1-\cdots-\alpha_{m-1})!}{\alpha_m!(\ell-\alpha_1-\cdots-\alpha_m)!} = \binom{\ell}{\alpha}$$

because the numerator in each of the fractions starting from the second cancels with one of the terms in the denominator of the previous fraction, and because  $|\alpha| = \ell$  means that  $(\ell - \alpha_1 - \cdots - \alpha_m)! = 1$ .

**Corollary 2.5.29.** *Suppose*  $\alpha'$  *is a multi-index and*  $|\alpha'| = \ell + 1$ *. Then* 

$$\binom{\ell+1}{\alpha'} = \sum_{\alpha'=\alpha+1_i} \binom{\ell}{\alpha},$$

where the sum is over all  $\alpha$  such that  $\alpha' = \alpha + 1_i$  for some i.

*Proof.* Suppose we want to place  $\ell + 1$  objects into m bins with  $\alpha'_i$  objects in the ith bin for all i. On the one hand, proposition 2.5.28 tells us that the left hand side is the number of ways of doing this. On the other hand, if  $\alpha'_i \ge 1$ , then there is a unique multi-index  $\alpha$  such that  $\alpha' = \alpha + 1_i$ . We can place the first object in bin i, then the number of ways of distributing the remaining  $\ell$  objects is  $\binom{\ell}{\alpha}$ . Summing over all such i yields precisely the sum on the right hand side.

Multi-index notation also gives us compact notation for multivariable polynomials.

**Definition 2.5.30.** A *polynomial* p in m variables is a function  $\mathbb{R}^m \to \mathbb{R}$  of the form

$$p(x) = \sum_{\alpha} c_{\alpha} x^{\alpha}$$

where  $x = (x_1, ..., x_m) \in \mathbb{R}^m$ , the sum is over all multi-indices, and for every multi-index  $\alpha$ , we have  $c_{\alpha} \in \mathbb{R}$  and  $c_{\alpha} = 0$  for all but finitely many  $\alpha$  (ie, the sum displayed above is finite). The *degree* of p is the maximum  $|\alpha|$  such that  $c_{\alpha} \neq 0$ . If p = 0, we define its degree to be  $-\infty$ .

**Exercise 2.5.31.** Suppose p is a polynomial of degree at most k in m variables. If  $|p(h)| = o(|h|^k)$  as  $h \to 0$ , show that p = 0.

#### Statement of Taylor's theorem

We're now ready to state the multivariable version of Taylor's theorem.

**Theorem 2.5.32** (Taylor). Suppose B is an open ball,  $f : B \to \mathbb{R}$  is  $C^k$  for some non-negative integer k, and  $a \in U$ . Then there exists a unique polynomial  $p_k$  of degree at most k in m variables, called the degree k Taylor polynomial of f at a, such that  $|f(a+h) - p_k(h)| = o(|h|^k)$  as  $h \to 0$ . Moreover, we have

$$p_{k}(h) = \sum_{|\alpha| \leq k} \frac{\partial^{\alpha} f(a)}{\alpha!} h^{\alpha}.$$
(2.5.33)

Finally, if f is  $C^{k+1}$ , then for any h such that  $a + h \in B$ , there exists  $\xi$  on the line segment between a and a + h such that

$$f(a+h)-p_{k}(h)=\sum_{|\alpha|=k+1}\frac{\partial^{\alpha}f(\xi)}{\alpha!}h^{k+1}.$$

While the notation is compact and aesthetically pleasing, there's a lot of information packed into very few symbols. I exhort you to work through the following example before proceeding. It may start feeling tedious at some point, but power through it.

**Exercise 2.5.34.** Let  $f : \mathbb{R}^2 \to \mathbb{R}$  be the function  $f(x, y) = \sin(xy)$ . Calculate the degree 3 Taylor polynomial of f at the origin.

#### Proof of Taylor's theorem

*Proof of Taylor's theorem* 2.5.32. We begin with a calculation. Fix  $h = (h_1, ..., h_m) \in \mathbb{R}^m$  such that  $a + h \in B$  and consider the function  $\gamma(t) = a + th$ . Note that  $\gamma'(t) = h$ , and all higher derivatives of  $\gamma$  vanish. Then  $g = f \circ \gamma$  is a single variable  $C^k$  function, and we claim that

$$g^{(\ell)}(t) = \sum_{|\alpha|=\ell} \binom{\ell}{\alpha} (\partial^{\alpha} f \circ \gamma)(t) h^{\alpha}$$
(2.5.35)

for all  $t \in [0, 1]$  and  $\ell = 0, 1, ..., k$ . When  $\ell = 0$ , equation (2.5.35) is just the definition of g. We'll actually prove equation (2.5.35) in general by induction on  $\ell$ , and  $\ell = 0$  suffices for the base case for this induction; that said, because the notation gets a bit dense, it is instructive to look at a couple of small values of  $\ell$  separately to help us understand the general inductive step.

The case when  $\ell = 1$  follows quickly from the chain rule. Indeed, we have  $\gamma'(t) = h$ , so

$$g'(t)=f'(\gamma(t))\gamma'(t)=\sum_{i=1}^m \vartheta_i f(\gamma(t))h_i,$$

using theorem 2.3.34 for the second step. This is precisely equivalent to equation (2.5.35) for  $\ell = 1$ .

Now consider the case when  $\ell = 2$ . For any i = 1, ..., m, we can calculate the derivative

of  $\partial_i f \circ \gamma$  using the chain rule. We have

$$\begin{aligned} (\partial_i f \circ \gamma)'(t) &= (\partial_i f)'(\gamma(t))\gamma'(t) \\ &= \sum_{j=1}^m \partial_{j,i} f(\gamma(t)) h_j \end{aligned}$$

which means that

$$g''(t) = \sum_{i=1}^{n} \sum_{j=1}^{m} \vartheta_{j,i} f(\gamma(t)) h_{j} h_{i}$$
$$= \sum_{j < i} 2\vartheta_{j,i} f(\gamma(t)) h_{j} h_{i} + \sum_{i} \vartheta_{i,i} f(\gamma(t)) h_{i}^{2}$$

where we have used the equality of mixed partials 2.5.21 for the second step. This too is equivalent to equation (2.5.35), because a multi-index  $\alpha$  such that  $|\alpha| = 2$  either is 2 in one entry and 0 everywhere else (in which case  $\alpha! = 2$  so  $\binom{2}{\alpha} = 1$ ), or else is 1 in two entries and 0 everywhere else (in which case  $\alpha! = 1$  so  $\binom{2}{\alpha} = 2$ ).

Now for the general inductive step. Suppose we know equation (2.5.35) for some non-negative integer  $\ell$ . The chain rule then gives us the second step in the following.

$$g^{(\ell+1)}(t) = \frac{d}{dt} \sum_{|\alpha|=\ell} {\ell \choose \alpha} (\partial^{\alpha} \circ \gamma)(t) h^{\alpha}$$
$$= \sum_{|\alpha|=\ell} {\ell \choose \alpha} ((\partial^{\alpha} f)'(\gamma(t))h) h^{\alpha}$$
$$= \sum_{|\alpha|=\ell} \sum_{i=1}^{m} {\ell \choose \alpha} \partial^{\alpha+1} f(\gamma(t)) h^{\alpha+1} f^{\alpha}$$
$$= \sum_{|\alpha|=\ell} \sum_{i=1}^{m} {\ell \choose \alpha} (\partial^{\alpha+1} f \circ \gamma)(t) h^{\alpha+1} f^{\alpha}$$
$$= \sum_{|\alpha'|=\ell+1} {\ell+1 \choose \alpha'} (\partial^{\alpha'} f \circ \gamma)(t) h^{\alpha'}$$

We have used theorem 2.3.34 together with equality of mixed partials (in the multi-index form of exercise 2.5.26) for the third equality, and corollary 2.5.29 for the final equality. This completes the induction and proves equation (2.5.35).

We'll prove the "with remainder" form of theorem 2.5.32 first, and then use it to prove the "without remainder" form, as in [Kö93, section 2.4]. Suppose f is  $C^{k+1}$ . Then the single variable function g defined above is also  $C^{k+1}$ . The single variable version of Taylor's theorem with remainder 1.4.14 tells us that

$$g(1) = \sum_{\ell=0}^{k} \frac{g^{(\ell)}(0)}{\ell!} + \frac{g^{(k+1)}(s)}{(k+1)!}$$

for some  $s \in (0, 1)$ . But then, applying equation (2.5.35), this equation says precisely that

$$f(a+h) = p_k(h) + \sum_{|\alpha|=k+1} \frac{\partial^{\alpha} f(\xi)}{\alpha!} h^{\alpha}$$

for  $\xi = a + sh$ , where  $p_k$  is the polynomial defined in equation (2.5.33).

Now let us prove the existence part of the "without remainder" form. In other words, we want to prove that, if f is  $C^k$  and  $p_k$  is defined as in equation (2.5.33), then  $|f(a+h)-p_k(h)| = o(|h|^k)$  as  $h \to 0$ . If k = 0, this follows immediately from the definition of continuity of f at a, so we can assume that  $k \ge 1$ . For any nonzero  $h \in \mathbb{R}^m$ , we can apply the "with remainder" form of Taylor's theorem that we've already proved to see that

$$f(a+h) = p_{k-1}(h) + \sum_{|\alpha|=k} \frac{\partial^{\alpha} f(\xi)}{\alpha!} h^{\alpha} = p_k(h) + \sum_{|\alpha|=k} \frac{\partial^{\alpha} f(\xi) - \partial^{\alpha} f(a)}{\alpha!} h^{\alpha}$$

where  $\xi$  is some point on the line segment between a and a + h. In other words, we have

$$f(a+h) - p_k(h) = \sum_{|\alpha|=k} \frac{\partial^{\alpha} f(\xi) - \partial^{\alpha} f(a)}{\alpha!} h^{\alpha}$$

and we want to prove that this is "small." Notice that

$$\frac{|f(a+h)-p_k(h)|}{|h|^k}\leqslant \sum_{|\alpha|=k}\frac{|\partial^{\alpha}f(\xi)-\partial^{\alpha}f(a)|\cdot|h^{\alpha}|}{\alpha!|h|^k}\leqslant \sum_{|\alpha|=k}\frac{|\partial^{\alpha}f(\xi)-\partial^{\alpha}f(a)|}{\alpha!}.$$

Here the first inequality is just the triangle inequality, and the second inequality is because

$$|\mathbf{h}^{\alpha}| = |\mathbf{h}_{1}^{\alpha_{1}} \cdots \mathbf{h}_{m}^{\alpha_{m}}| \leq |\mathbf{h}|^{k}.$$

131

#### 2 Multivariable derivatives

Choose  $\epsilon > 0$ . Since f is  $C^k$ , we know that  $\partial^{\alpha} f$ , and therefore also  $\partial^{\alpha} / \alpha!$ , is continuous for all multi-indices  $\alpha$  such that  $|\alpha| = k$ . There are only finitely many such multi-indices, so there exists a  $\delta > 0$  such that

$$\sum_{|\alpha|=k} \frac{|\partial^{\alpha} f(x) - \partial^{\alpha} f(\alpha)|}{\alpha!} < \epsilon$$

for all  $x \in B(a, \delta)$ . Now if  $|h| < \delta$ , then  $\xi \in B(a, \delta)$ , so we have

$$\frac{|f(a+h)-p_{k}(h)|}{|h|^{k}} \leqslant \sum_{|\alpha|=k} \frac{|\partial^{\alpha}f(\xi)-\partial^{\alpha}f(a)|}{\alpha!} < \varepsilon$$

This proves that  $|f(a + h) - p_k(h)| = o(|h|^k)$  as  $h \to 0$ . Finally, the uniqueness assertion of theorem 2.5.32 follows from exercise 2.5.31, just as in the proof of the single variable version 1.4.14.

#### 2.5.E Smooth functions

**Definition 2.5.36** (Smoothness). We say that a function  $f : U \to \mathbb{R}^n$  is  $C^{\infty}$ , or *infinitely differentiable*, or *smooth*, if it is  $C^k$  for all k.

In section 1.4.D, we formulated the principle that smooth single variable functions can be tailored to almost arbitrary specifications. The same principle remains true of smooth multivariable functions, and we can often bootstrap up from the single variable case. Here are some examples.

#### **Bump functions**

Recall the definition of support in definition 1.4.29. Just like in the single variable setting, a "bump function" is a smooth function with compact support. In fact, we can use single variable bump functions to construct multivariable bump functions.

**Exercise 2.5.37.** Let  $\psi$  :  $\mathbb{R} \to \mathbb{R}$  denote the single variable bump function from example 1.4.30. Then define  $\Psi$  :  $\mathbb{R}^m \to \mathbb{R}$  by

$$\Psi(\mathbf{x}_1,\ldots,\mathbf{x}_m)=\psi(\mathbf{x}_1)\psi(\mathbf{x}_2)\cdots\psi(\mathbf{x}_m).$$

See figure 2.5.38. Show that  $\Psi$  is a bump function supported on the hypercube  $[-1, 1]^m$ .



Figure 2.5.38: The graph of the function  $\Psi(x, y) = \psi(x)\psi(y)$ , where  $\psi$  is the single variable bump function from example 1.4.30.

**Exercise 2.5.39.** Let  $\psi : \mathbb{R} \to \mathbb{R}$  denote the single variable bump function from example 1.4.30. Then define  $\Psi : \mathbb{R}^m \to \mathbb{R}$  by

$$\Psi(x_1,\ldots,x_m)=\psi(x_1^2+\cdots+x_m^2).$$

Show that  $\Psi$  is a bump function supported on the closed unit ball  $\overline{B}(0, 1)$ .

#### Paths

**Definition 2.5.40** (Paths). Given  $a, b \in \mathbb{R}^m$ , a *path* starting at a and ending at b is a continuous function  $\gamma : [0, 1] \to \mathbb{R}^m$  such that  $\gamma(0) = a$  and  $\gamma(1) = b$ .

- We say that  $\gamma$  is *smooth* if it is the restriction to [0, 1] of a smooth function  $(-\epsilon, 1+\epsilon) \rightarrow \mathbb{R}^m$  for some  $\epsilon > 0$ . This means that we can talk about the iterated derivatives  $\gamma^{(k)}(t)$  not just for  $t \in (0, 1)$ , but also for t = 0, 1.
- We say that  $\gamma$  is *inside* or *contained in* an open subset U of  $\mathbb{R}^m$  if  $\gamma(t) \in U$  for all

 $t \in [0, 1].$ 

We met straight line paths between any two points of  $\mathbb{R}^m$  in definition 2.4.3, and then in definition 2.4.5, we defined a subset  $S \subseteq \mathbb{R}^m$  to be convex by requiring that all straight line paths between pairs of points of S are inside S.

**Exercise 2.5.41.** Check that the straight line path between any two points in  $\mathbb{R}^m$  is smooth.

**Exercise 2.5.42** (Slowing down smooth paths). Suppose  $\gamma$  is a smooth path in  $\mathbb{R}^m$ . Show that there exists a smooth path  $\gamma_{\text{slow-start}}$  with the same start and end as  $\gamma$  and the same image as  $\gamma$ , such that

$$\gamma_{\text{slow-start}}^{(k)}(0) = 0$$

for all  $k \ge 1$ . Then show that there also exists a smooth path  $\gamma_{\text{slow-stop}}$  with the same start and end as  $\gamma$  and the same image as  $\gamma$ , such that

$$\gamma_{\text{slow-stop}}^{(k)}(1) = 0$$

for all  $k \ge 1$ .

*Possible hint.* To slow down the start of the path, modify the "infinitely flat" function of exercise 1.4.27 to find a function  $f : \mathbb{R} \to \mathbb{R}$  such that  $f^{(k)}(0) = 0$  for all  $k \ge 0$  and f(1) = 1, and then consider  $\gamma_{\text{slow-start}} = \gamma \circ f$ .

**Definition 2.5.43** (Concatentation of paths). If  $\gamma_1$  and  $\gamma_2$  are paths in  $\mathbb{R}^m$  and  $\gamma_2$  starts where  $\gamma_1$  ends, we define their *concatenation*  $\gamma_1 \star \gamma_2 : [0, 1] \to \mathbb{R}^m$  as follows.

$$(\gamma_1 \star \gamma_2)(t) = \begin{cases} \gamma_1(2t) & \text{if } t \leqslant 1/2\\ \gamma_2(2t-1) & \text{if } t > 1/2 \end{cases}$$

In other words,  $\gamma_1 \star \gamma_2$  "travels along  $\gamma_1$  at twice the speed," and then "travels along  $\gamma_2$  at twice the speed."

Observe that if  $\gamma_1$  and  $\gamma_2$  are both inside U, then  $\gamma_1 \star \gamma_2$  is also inside U. However, even if  $\gamma_1 \star \gamma_2$  are both smooth, their concatenation need not be. To concatenate smoothly, we can use the following.

**Exercise 2.5.44** (Smooth concatenation). Suppose  $\gamma_1$  and  $\gamma_2$  are smooth paths in  $\mathbb{R}^m$  and  $\gamma_2$  starts where  $\gamma_1$  ends. Using the notation of exercise 2.5.42 and definition 2.5.43, show

that

$$\gamma = \gamma_{1, slow-stop} \star \gamma_{2, slow-start}$$

is smooth.

**Exercise 2.5.45.** Suppose U is a connected open subset of  $\mathbb{R}^m$ . Show that, for any pair of points  $a, b \in U$ , there exists a smooth path inside U starting at a and ending at b.

*Possible hint*. Use exercise 2.5.44 and the first part of exercise 2.4.9.

# 3 Manifolds

Roughly speaking, a "manifold" is a locally flat geometric object without pre-ordained local coordinate systems. Before we formalize what we mean by this, it's useful to philosphize a bit. First of all, let's think about what "locally flat" means. Consider, for example, the surface of the earth. On the one hand, we know that it's a sphere (approximately, at least). On the other hand, we also know from personal experience that the curvature of the earth is irrelevant in situationsn where everything of concern is within a few miles of us. The earth looks flat. It is in this sense that the surface of the earth is "locally flat."

When we have such a locally flat geometric object, it's often useful to introduce "local coordinate systems." For example, we talk about things being "in front of" us or being "behind" us in our everyday lives, despite the fact that, if you go far enough forward, you'll eventually end up behind where you started! In other words, "in front of" and "behind" make no sense if we're thinking globally (ie, if we're thinking at the level of the entirety of the earth). They only make sense "locally," but despite that, they're incredibly useful notions in everyday life. Similar considerations apply with "left" and "right."

It's also useful to allow local coordinate systems to change depending on the situation we're in. Returning to the example of "in front of" and "behind," notice that these notions don't refer to absolute directions: if you turn 90° in some direction, what was "in front of" you before you turned isn't "in front of" you anymore. After your 90° turn, there's now a new useful notion of "in front of." In other words, changing the situation you're in has made a different local coordinate system useful.

Let's now return to the sentence we started with, that "manifolds are locally flat geometric objects without pre-ordained local coordinate systems." We'll see below how to formalize "locally flat" mathematically. The way we'll avoid having pre-ordained local coordinate systems is by remembering *all possible choices* of local coordinate systems!

It's also worth remarking that this chapter marks a somewhat significant turning point in our meditation of derivatives and tangents. We began chapter 1 with the observation that we could formalize our intuitive idea of "tangent line" using derivatives. In the next chapter, we'll see that, in the abstract setting of manifolds, we can formalize the idea of "tangent vector," and then use it to define an even further generalization of the derivative.

## 3.1 Definition of a manifold

At this point, we will need to use the language of topological spaces. If you've seen metric spaces but not topological spaces, you should skim through section 0.7.A (in particular, at least definitions 0.7.1, 0.7.10, 0.7.14 and 0.7.15 and example 0.7.2).

#### 3.1.A Charts

Throughout this section, let X be a topological space. Recall that we want for X to be "locally flat." Formally, this means that we can cover X with open subsets, each of which is "flat" in the sense that it is homeomorphic to an open subset of  $\mathbb{R}^n$ . Notice that, if we remember not just the fact that each open subset U in that cover is homeomorphic to  $\mathbb{R}^n$ , but the actual homeomorphism, then we can transfer the usual coordinate system on  $\mathbb{R}^n$  through the homeomorphism onto U. In other words, a homeomorphism between U and an open subset of  $\mathbb{R}^n$  tells us both that U is "flat," and gives us a coordinate system on U. This is precisely what's accomplished by the following definition.

**Definition 3.1.1** (Chart). A *chart* on X is a pair (U, x) consisting of an open subset  $U \subseteq X$  and a continuous, open, and injective function  $x : U \to \mathbb{R}^n$  for some n.

- The integer n is called the *dimension* of the chart, and write dim(U, x) to denote it.
- The function x is called the *coordinate function* of the chart.
- If  $p \in X$ , we say that the chart (U, x) *contains* p if  $p \in U$ .
- For all i = 1,..., n, we let x<sub>i</sub> := π<sub>i</sub> ∘x. For any p ∈ U, the real numbers x<sub>1</sub>(p),..., x<sub>n</sub>(p) are called the *coordinates of* p *with respect to* (U, x).

**Remark 3.1.2.** If (U, x) is a chart on X and  $p \in U$  is a point, we can "recenter the chart at p" by replacing x with the function x' = x - x(p). In other words, if x' = x - x(p), then (U, x') is also a chart, and we have x'(p) = 0.

**Exercise 3.1.3.** Let  $X = \mathbb{R}^2$  and let U denote  $\mathbb{R}^2$  minus the non-negative real axis. Let polar :  $U \to \mathbb{R}^2$  be the function which sends a point p to its polar representation  $(r(p), \theta(p))$ ,

where r(p) = |p| and  $\theta(p)$  is the angle in radians strictly between 0 and  $2\pi$  formed between P and the non-negative real axis. Show that (U, polar) is a chart. Calculate the coordinates of  $(1, 1) \in \mathbb{R}^2$  with respect to (U, polar).

#### 3.1.B Compatibility of charts

Charts formalize the idea of local coordinate systems; the following formalizes the idea of *changing* local coordinate systems.

**Definition 3.1.4** (Transition function). Suppose (U, x) and (U', x') are charts on X. The *transition function* from (U, x) to (U', x') is the function  $x' \circ x^{-1}$ .

$$\mathbf{x}(\mathbf{U} \cap \mathbf{U}') \xrightarrow{\mathbf{x}^{-1}} \mathbf{U} \cap \mathbf{U}' \xrightarrow{\mathbf{x}'} \mathbf{x}'(\mathbf{U} \cap \mathbf{U}')$$

**Definition 3.1.5** (Compatibility of charts). The two charts (U, x) and (U', x') on X are *smoothly compatible*, or just *compatible*, if both of the transition maps between (U, x) and (U', x') are smooth.

**Exercise 3.1.6.** Let  $id : \mathbb{R}^2 \to \mathbb{R}^2$  denote the identity map and polar :  $U \to \mathbb{R}^2$  the polar representation map from exercise 3.1.3. Show that ( $\mathbb{R}^2$ , id) and (U, polar) are compatible charts.

**Exercise 3.1.7.** Suppose (U, x) is a chart and  $p \in X$  is a point. Consider the recentered chart (U, x') where x' = x - x(p), as in remark 3.1.2. Show that (U, x) and (U, x') are compatible.

**Exercise 3.1.8.** Show that the relation "is compatible with" is *not* an equivalence relation on the set of all charts.

*Possible hint*. The function  $x \mapsto x^3$  is a smooth homeomorphism  $\mathbb{R} \to \mathbb{R}$ , but its inverse is not differentiable.

**Exercise 3.1.9.** Suppose (U, x) and (U', x') are compatible charts and that the intersection  $U \cap U'$  is nonempty. Show that  $\dim(U, x) = \dim(U', x')$ .

*Possible hint.* Pick a point  $p \in U \cap V$ , and consider the derivatives of  $x' \circ x^{-1}$  at x(p) and of  $x \circ x'^{-1}$  at x'(p).

## 3.1.C Atlases and manifolds

**Definition 3.1.10** (Atlas). An *atlas* on X is a collection A of compatible charts such that covers X. In other words, a collection A of charts is an atlas if every pair of charts is compatible, and

$$\bigcup_{(\mathbf{U},\mathbf{x})\in\mathcal{A}}\mathbf{U}=\mathbf{X}$$

If  $\mathcal{A}'$  is also an atlas and  $\mathcal{A}' \subseteq \mathcal{A}$ , we say that  $\mathcal{A}'$  is a *sub-atlas* of  $\mathcal{A}$ . An atlas is *maximal* if there is no strictly larger atlas containing it.

**Exercise 3.1.11** (Existence and uniqueness of maximal atlases). Prove that every atlas A is contained in a unique maximal atlas.

*Possible hint.* Let A' be the set of all charts that are compatible with all of the charts in A. Prove that A' is an atlas (caution: keep exercise 3.1.8 in mind). Then prove that it is maximal, and that it is the only maximal atlas containing A.

This leads us to the following definition.

**Definition 3.1.12** (Manifold). A *smooth manifold*, or just a *manifold*, is a pair  $(X, A_X)$  where X is a topological space and  $A_X$  is a maximal atlas. Often, we just write "X" in place of the pair  $(X, A_X)$ . Sometimes the maximal atlas  $A_X$  is called the *manifold structure* or the *smooth structure* on X.

Exercise 3.1.11 tells us that any not-necessarily-maximal atlas on a topological space X can be extended uniquely to a maximal atlas, thus giving X the structure of a manifold. When constructing examples, we will usually give an explicit non-maximal atlas and then use exercise 3.1.11 to extend it to a maximal atlas; it is difficult to describe maximal atlases explicitly. But it is often convenient for theorem statements and abstract proofs to assume that the atlas that the manifold comes equipped with is already maximal, which is why the definition is made the way it is.

*Unimportant remark.* For any  $k = 0, 1, 2, ..., \infty$ , we could define charts to be  $C^k$ -compatible by requiring that the transition functions between them are  $C^k$  (rather than smooth, as in definition 3.1.5). This would lead to a definition of  $C^k$ -manifolds. In these notes, we'll just stick to the  $k = \infty$  case.

#### 3.1.D Dimension

**Definition 3.1.13** (Dimension). Suppose X is a manifold and  $p \in X$  is a point. By exercise 3.1.9, there is a unique integer n such that  $\dim(U, x) = n$  for every chart  $(U, x) \in A_X$  containing p. This integer is called the *dimension of X at* p, and is denoted  $\dim_p(X)$ .

If there exists a single integer n such that  $\dim_p(X) = n$  for all  $p \in X$ , we say that X is *equidimensional* and that the *dimension of* X is n.

## 3.2 Examples

#### 3.2.A First examples

**Example 3.2.1.** The single chart ( $\mathbb{R}^n$ , id), where id :  $\mathbb{R}^n \to \mathbb{R}^n$  is the identity map, defines an atlas on  $\mathbb{R}^n$ . We can extend this to a maximal atlas, called the *euclidean atlas* on  $\mathbb{R}^n$ . Unless explicitly specified otherwise, we always regard  $\mathbb{R}^n$  as a manifold by equipping it with the euclidean atlas. By exercise 3.1.6, the polar representation chart (U, polar) is an element of the euclidean atlas.

**Exercise 3.2.2.** Let  $f : \mathbb{R} \to \mathbb{R}$  be the function  $f(x) = x^3$ . Note that f is smooth and strictly increasing (hence also injective and open, by exercises 1.3.9 and 1.3.10). Show that ( $\mathbb{R}$ , f) is *not* a member of the euclidean atlas on  $\mathbb{R}$ .

**Exercise 3.2.3.** Suppose U is an open subset of  $\mathbb{R}^n$  and  $f : U \to \mathbb{R}^n$  is a function. Show that the pair (U, f) is a chart in the euclidean atlas on  $\mathbb{R}^n$  if and only if f is smooth, injective, and étale.

**Exercise 3.2.4** (The circle  $S^1$ ). Let  $S^1$  denote the set of points in  $\mathbb{R}^2$  along the unit circle. In other words,

$$S^1 = \{(a, b) \in \mathbb{R}^2 : a^2 + b^2 = 1\}.$$

- (a) Do exercise 0.7.8. This shows that both ways that we might think to regard S<sup>1</sup> as a topological space are actually the same.
- (b) Let U = S<sup>1</sup> \{(0,1)}. For p ∈ S<sup>1</sup>, let x(p) denote the x-coordinate of the point where the line y = −1 intersects the line passing through (0,1) and p. See figure 3.2.5. If p = (a, b), find a formula for x(p) in terms of a and b.
- (c) Show that  $x : U \to \mathbb{R}$  is a homeomorphism. Thus (U, x) is a chart on S<sup>1</sup>.



Figure 3.2.5: Given a point  $p \in U = S^1 \setminus \{(0,1)\}$ , we draw a line connecting p to (0,1), indicated in red. The point where that line intersects the line y = -1 (in black) is sometimes called the *projection of* p *from* (0,1) *onto*  $\mathbb{R} \times \{-1\}$ . The x-coordinate of this point is x(p).

*Possible hint*. Use geometry to write down a formula for the inverse function to x.

- (d) Let  $U' = S^1 \setminus \{(0, -1)\}$ . For  $p \in U'$ , let x'(p) denote the x-coordinate of the point where the line y = 1 intersects the line passing through (0, -1) and p. Show that  $x' : U' \to \mathbb{R}$  is also a homeomorphism.
- (e) Show that (U, x) and (U', x') are compatible charts. Thus  $\mathcal{A} = \{(U, x), (U', x')\}$  is an atlas on S<sup>1</sup>. Extending  $\mathcal{A}$  to a maximal atlas  $\mathcal{A}_{S^1}$ , we have defined a manifold structure on S<sup>1</sup>.
- (f) Let  $V = \{(a, b) \in S^1 : a > 0\}$  be the "right half of the circle." If  $y : V \to \mathbb{R}$  is the function y(a, b) = b, show that (V, y) is a chart and that it is compatible with the charts in the atlas  $\mathcal{A}$  above. Thus  $(V, y) \in \mathcal{A}_{S^1}$ .
- (g) Can you think of any other charts in the maximal atlas  $A_{S^1}$ , besides the three discussed above?

**Exercise 3.2.6** (The n-sphere  $S^n$ ). Let  $S^n$  denote the set of points in  $\mathbb{R}^{n+1}$  of norm 1. Generalize exercise 3.2.4 and show that "projection from the north pole" and "projection from the south pole" define an atlas on  $S^n$ .

**Example 3.2.7** (Matrices). There is an isomorphism of vector spaces  $x : \mathcal{M}_{n \times m} \to \mathbb{R}^{mn}$  where, if  $A \in \mathcal{M}_{n \times m}$  has entries  $a_{i,j}$ , then

$$x(A) = \begin{bmatrix} a_{1,1} \\ a_{1,2} \\ \vdots \\ a_{1,m} \\ a_{2,1} \\ \vdots \\ a_{n,m} \end{bmatrix}.$$

If we endow  $\mathcal{M}_{n \times m}$  with the canonical topology, then x is automatically a homeomorphism by example 0.7.18. In other words, the single chart  $(\mathcal{M}_{n \times m}, x)$  is an atlas on  $\mathcal{M}_{n \times m}$ . Extending it to a maximal atlas using exercise 3.1.11, we see that we can regard  $\mathcal{M}_{n,m}$ as a manifold. It is mn-dimensional. The same process allows us to regard *any* finite dimensional vector space (not just  $\mathcal{M}_{n \times m}$ ) as a manifold.

#### 3 Manifolds

**Exercise 3.2.8** (Graphs of continuous functions). Suppose U is an open subset of  $\mathbb{R}^m$  and  $f: U \to \mathbb{R}^n$  is a continuous function.

(a) There are two ways we might think to regard  $U \times \mathbb{R}^n$  as a topological space. One way is to use the product topology 0.7.6. The second way is to regard  $U \times \mathbb{R}^n$  as a subset of  $\mathbb{R}^m \times \mathbb{R}^n = \mathbb{R}^{m+n}$  and then give  $U \times \mathbb{R}^n$  the subspace topology 0.7.5. Show that these two topologies are the same.

(b) Let

$$\Gamma = \{ (x, y) \in U \times \mathbb{R}^n : f(x) = y \}$$

be the graph of f, regarded as a subspace of  $U \times \mathbb{R}^n$ . Show that the function  $\pi : \Gamma \to \mathbb{R}^m$  given by  $\pi(x, y) = x$  is continuous, injective, and open.

Thus the single chart  $(\Gamma, \pi)$  defines an atlas on  $\Gamma$ . Extending this to a maximal atlas, we can regard  $\Gamma$  as a manifold.

**Exercise 3.2.9** (Open submanifolds). Let X be a manifold and  $U \subseteq X$  an open subset. Show that

$$\mathcal{A}_{U} = \{ (V, x) \in \mathcal{A}_{X} : V \subseteq U \}$$

is a maximal atlas on U. We call U, equipped with this maximal atlas, an *open submanifold* of X.

**Exercise 3.2.10** (Product manifold). Let X and Y be manifolds. Let  $\mathcal{A}$  be the set of charts of the form  $(U \times V, x \times y)$ , where  $(U, x : U \to \mathbb{R}^m) \in \mathcal{A}_X$ ,  $(V, y : V \to \mathbb{R}^n) \in \mathcal{A}_Y$ , and  $x \times y$  denotes the natural map  $U \times V \to \mathbb{R}^m \times \mathbb{R}^n = \mathbb{R}^{m+n}$ . Show that  $\mathcal{A}$  is an atlas on  $X \times Y$ . We call  $X \times Y$ , equipped with the corresponding maximal atlas, the *product manifold*.

#### 3.2.B Möbius strip

Let B denote the box  $[0, 1] \times (0, 1)$  inside  $\mathbb{R}^2$ , and let ~ be the equivalence relation on B where  $(a, b) \sim (a', b')$  if we either have a = a' and b = b', or else we have a = 0, a' = 1, and b' = 1 - b. See figure 3.2.11. We will denote the equivalence class of a point  $(a, b) \in B$  as [a, b].

**Exercise 3.2.12.** Check that ~ is an equivalence relation, and that every equivalence class contains either one or two points.


Figure 3.2.11: Above is a picture of the box  $B = [0, 1] \times (0, 1)$  and the equivalence relation  $\sim$  described above. Geometrically, this equivalence relation identifies the vertical black line on the left of the box with the vertical black line on the right of the box, but "with a twist." For example, the point (0, 1/4) on the left is identified with the point (1, 3/4) on the right. If we pick up the box and actually glue together the edges as indicated by the equivalence relation, we end up with the Möbius strip as pictured below.



## 3 Manifolds

The *Möbius strip* M is the set of equivalence classes in B. In other words, we take the box  $[0, 1] \times (0, 1)$  and "glue" its left and right edges, with a twist. If you've ever made a Möbius strip out of paper, you can probably see that we've formalized exactly that same process. See figure 3.2.11, again.

We regard M as a topological space using the quotient topology, as in example 0.7.7. To give X the structure of a manifold, we will construct an atlas consisting of two charts. But, before proceeding with the formalism, let's describe the intuition. Imagine making a cut in the Möbius strip corresponding to a vertical cut in the box B (ie, a cut perpendicular to the central red circle in figure 3.2.11). Once we make such a cut, we can untwist the Möbius strip and we're left with an open rectangle, which we can regard as an open subset of  $\mathbb{R}^2$ . In other words, we've defined a chart on M. The points along the cut are missing from this chart; but if we construct a second chart by making a second separate cut, we will have constructed two charts that cover M.

Let's formalize this now. The easiest place to cut is the line along which we glued the Möbius strip in the first place. In other words, we let

$$\mathsf{U} = \{[\mathfrak{a}, \mathfrak{b}] \in \mathsf{M} : \mathfrak{a} \in (0, 1)\}$$

and let  $x : U \to \mathbb{R}^2$  be the function x([a, b]) = (a, b).

**Exercise 3.2.13.** Show that U is an open subset of M, and that the function  $x : U \to \mathbb{R}^2$  is well-defined, continuous, injective, and open.

The second chart is a bit harder to describe mathematically, but the intuitive idea is the same. We'll cut the Möbius strip along the line corresponding the vertical line a = 1/2 in B. Formally, let

$$V = \{[a, b] \in M : a \neq 1/2\}$$

and let  $y:V\to \mathbb{R}^2$  be the function defined by

$$y([a,b]) = \begin{cases} (a,b) & \text{if } a < 1/2 \\ (a-1,1-b) & \text{if } a > 1/2 \end{cases}$$

**Exercise 3.2.14.** Show that V is also an open subset of M, and that  $y : V \to \mathbb{R}^2$  is well-defined, continuous, injective, and open.

**Exercise 3.2.15.** Show that (U, x) and (V, y) are compatible.

Thus  $\mathcal{A} = \{(\mathbf{U}, \mathbf{x}), (\mathbf{V}, \mathbf{y})\}$  is an atlas on the Möbius strip M. Letting  $\mathcal{A}_{M}$  denote the corresponding maximal atlas, we obtain the structure of a manifold on M.

**Exercise 3.2.16.** Fix a constant  $c \in [0, 1)$ , and let  $U_c$  be the subset of M where we've cut out the vertical line a = c. Show that  $U_c$  is an open subset of M. Write down a formula for a well-defined function  $x_c : U_c \to \mathbb{R}^2$  which is continuous, injective, and open; the image of  $x_c$  should be an open box in  $\mathbb{R}^2$ , and you should recover the function x above when c = 0, and the function y when c = 1/2. Then show that  $(U_c, x_c) \in \mathcal{A}_M$  for all  $c \in [0, 1)$ .

### 3.2.C More quotient examples

Here is another example of a manifold you probably recognize (perhaps by the name "donut" rather than "torus") which can be constructed as a quotient space.

**Exercise 3.2.17** (Torus). Let B denote the box  $[0, 1] \times [0, 1]$  inside  $\mathbb{R}^2$  and let ~ denote the equivalence relation generated by  $(a, 0) \sim (a, 1)$  and  $(0, b) \sim (1, b)$  for all  $a, b \in [0, 1]$ . The *torus* T is defined to be the corresponding quotient space. Describe an atlas on T.

But sometimes quotient spaces can be very difficult to visualize directly; the only way to have geometric intuition in these cases is to remember the space we started with before quotienting, and the equivalence relation we defined on it.

**Exercise 3.2.18** (Projective line). Let  $S^1$  denote the circle as in exercise 3.2.4. Define an equivalence relation ~ on  $S^1$  which identifies antipodal points; in other words, every point on  $S^1$  is declared to be equivalent to the point that is diametrically opposite it. The *projective line*  $\mathbb{P}^1$  is defined to be the corresponding quotient space. Describe an atlas on  $\mathbb{P}^1$ .

**Exercise 3.2.19** (Projective plane). Let  $S^2$  denote the 2-sphere as in exercise 3.2.6. Define an equivalence relation ~ on  $S^2$  which identifies antipodal points; in other words, every point on  $S^2$  is declared to be equivalent to the point that is diametrically opposite it. The *projective plane*  $\mathbb{P}^2$  is defined to be the corresponding quotient space. Describe an atlas on  $\mathbb{P}^2$ .

#### 3.2.D Submanifolds

**Definition 3.2.20.** Suppose S is a subset of a manifold X. Then S is a *submanifold* of X if there exist charts  $(U, x) \in A_X$  covering S such that

$$S \cap U = \{p \in U : x_1(p) = x_2(p) = \cdots = x_k(p) = 0\}$$

## 3 Manifolds

for some non-negative integer  $k \leq \dim(U, x)$ . Such a chart (U, x) is said to be *adapted* to *S*, and the integer k is called the *codimension of S in* (U, x).

We'll justify this terminology in a moment by showing that submanifolds are actually manifolds in their own right (cf. proposition 3.2.23). But first, let's discuss some examples.

**Example 3.2.21.** Consider the set  $S = \{(0, y) \in \mathbb{R}^2 : -1 < x < 1\}$  inside  $\mathbb{R}^2$ . In other words, S is the open interval (-1, 1) along the y-axis inside  $\mathbb{R}^2$ . Then S is a submanifold. Indeed, let U be the open box  $(-1, 1) \times (-1, 1)$  inside  $\mathbb{R}^2$ , and  $x : U \to \mathbb{R}^2$  the inclusion map. Then (U, x) is a chart on  $\mathbb{R}^2$ , and

$$S \cap U = \{ p \in S : x_1(p) = 0 \}$$

so (U, x) is adapted to S and S has codimension 1 in (U, x).

**Exercise 3.2.22.** Show that each of the following is a submanifold of  $\mathbb{R}^2$ .

- (a) The y-axis.
- (b) The open interval (-1, 1) along the x-axis.
- (c) The line y = x.

We now make good on our promise of proving that submanifolds are in fact manifolds. The proof is a little long because there are a lot of details to check, but there are no real tricks or surprises.

**Proposition 3.2.23.** Let S be a submanifold of a manifold X. Suppose  $(U, x) \in A_X$  is adapted to S and S has codimension k in (U, x). Let  $x|_S : S \cap U \to \mathbb{R}^{n-k}$  be the function defined by

$$\mathbf{x}|_{\mathbf{S}}(\mathbf{p}) = (\mathbf{x}_{k+1}(\mathbf{p}), \dots, \mathbf{x}_{n}(\mathbf{p})).$$

Regarding S as a topological space using the subspace topology, the pair  $(S \cap U, x|_S)$  is a chart on S, and the set of all charts of this form is an atlas on S. Thus S becomes a manifold if we equip it with the corresponding maximal atlas  $A_S$ . Finally, if  $(U, x) \in A_X$  is a chart adapted to S such that S has codimension k in S, then

$$\dim_p(S) = \dim_p(X) - k$$

for all  $p \in S \cap U$ .

*Proof.* Observe that, by definition,  $x|_S$  is the composite

$$S \cap U \longrightarrow U \xrightarrow{x} \mathbb{R}^n \xrightarrow{\pi} \mathbb{R}^{n-k}$$

where the first map is the inclusion, and the final map  $\pi : \mathbb{R}^n \to \mathbb{R}^{n-k}$  projects onto the final n-k coordinates (ie,  $\pi(x_1, \ldots, x_n) = (x_{k+1}, \ldots, x_n)$ ). Each of these maps is continuous, so the composite  $x|_S$  is continuous as well. The proof that  $x|_S$  is injective is left as an exercise (cf. exercise 3.2.24).

Let us show that  $x|_S$  is open. By definition of the subspace topology 0.7.5, every open subset of  $S \cap U$  is of the form  $S \cap V$  where V is an open subset of U. Suppose  $p \in S \cap V$ . We want to show that  $x|_S(p)$  is an interior point of  $x|_S(S \cap V)$ . Since V is open in U and  $x : U \to \mathbb{R}^n$  is an open map, we know that x(p) is an interior point of x(V), ie, there exists  $\varepsilon > 0$  such that every point of  $\mathbb{R}^n$  within  $\varepsilon$  of x(p) is inside x(V). We claim that the  $\varepsilon$  ball around  $x|_S(p)$  is contained in  $x|_S(S \cap V)$ . Suppose we have  $a = (a_{k+1}, \ldots, a_n) \in \mathbb{R}^{n-k}$ such that  $|a - x|_S(p)| < \varepsilon$ . We will show that  $a \in x|_S(S \cap V)$ .

Notice that, since (U, x) is adapted to S, the vector  $x(p) \in \mathbb{R}^n$  is the same as  $x|_S(p) \in \mathbb{R}^{n-k}$  except that it is prepended with k zeroes. Let  $\iota : \mathbb{R}^{n-k} \to \mathbb{R}^n$  be the map that preprends vectors in  $\mathbb{R}^{n-k}$  with a string of k zeroes. In other words,

$$\iota(b_{k+1},\ldots,b_n) = (\overbrace{0,\ldots,0}^{k \text{ times}}, b_{k+1},\ldots,b_n).$$

Thus  $x(p) = \iota(x|_S(p))$ . Now by the definition of the euclidean norm, we see that  $|\iota(a) - x(p)| = |a - x|_S(p)|$ , which means that  $|\iota(a) - x(p)| < \varepsilon$ . This means that  $\iota(a) \in x(V)$  due to our choice of  $\varepsilon$ . Since x is injective, there exists a unique  $q \in V$  such that  $x(q) = \iota(a)$ . This means that the first k coordinates of q are zero, so  $q \in S \cap U$  since (U, x) is adapted to S. Moreover

$$\mathbf{x}|_{\mathbf{S}}(\mathbf{q}) = \pi(\mathbf{x}(\mathbf{q})) = \pi(\mathbf{\iota}(\mathbf{a})) = \mathbf{a}$$

which shows that  $a \in x|_S(S \cap V)$ .

Next up, suppose (U, x) and (U', x') are both charts in  $\mathcal{A}_X$  that are adapted to S. We want to show that  $(S \cap U, x|_S)$  and  $(S \cap U', x'|_S)$  are compatible. If  $S \cap U \cap U'$  is empty, there is nothing to do, so we assume that  $S \cap U \cap U'$  is nonempty. Let k and k' be the codimensions of S in the two charts (U, x) and (U', x'), respectively. Once we prove compatibility, it follows from exercise 3.1.9 that we must have k = k'.

## 3 Manifolds

The proof of compatibility is encapsulated by the following diagram.

$$\begin{array}{c} x|_{S}(S \cap U \cap U') \xrightarrow{(x|_{S})^{-1}} S \cap U \cap U' \xrightarrow{x'|_{S}} x'|_{S}(S \cap U \cap U') \\ \downarrow & \downarrow & \uparrow \\ x(U \cap U') \xrightarrow{x^{-1}} U \cap U' \xrightarrow{x'} x'(U \cap U') \end{array}$$

Here  $\iota$  denotes the "prepend with k zeroes" map, and  $\pi'$  is the "project onto the final k' coordinates" map. You should stare at the above diagram until you've convinced yourself that it "commutes," ie, that the result of following any two paths of arrows that start and end at the same place ends up being the same. (The fact that the square on the left of the diagram commutes is because (U, x) is adapted to S, and the fact that square on the right commutes is by definition of the function  $x'|_{S}$ .)

In other words, the diagram tells us that the transition function  $x'|_{S} \circ (x|_{S})^{-1}$  is equal to  $\pi' \circ (x' \circ x^{-1}) \circ \iota$ . Since  $(\mathbf{U}, \mathbf{x})$  and  $(\mathbf{U}', \mathbf{x}')$  are compatible, we know that  $x' \circ x^{-1}$  is smooth. Thus the transition function  $x'|_{S} \circ (x|_{S})^{-1}$  is smooth.

Since S is a submanifold, we know that it can be covered by charts in X that are adapted to S. Thus we've just shown that the set

$$\mathcal{A} = \{ (S \cap U, x|_S) : (U, x) \in \mathcal{A}_X \text{ is adapted to } S \}$$

is an atlas on S. The assertion about  $\dim_{p}(S)$  follows immediately.

**Exercise 3.2.24.** Show that  $x|_S$  is injective.

We've used the word "submanifold" once before, in exercise 3.2.9 where we defined "open submanifolds." Let us now show that this earlier usage of the word "submanifold" was justified and does not produce any conflicts with our new definitions.

Exercise 3.2.25. Suppose X is a manifold and U is an open subset.

- (a) Show that U is a submanifold (in the sense of definition 3.2.20).
- (b) Show that the maximal atlas on U constructed by proposition 3.2.23 is equal to the atlas described in exercise 3.2.9.

*Possible hint*. Show that a chart  $(V, x) \in A_X$  is adapted to U if and only if  $V \subseteq U$ , and that U has codimension 0 in any such chart.

Here is another property you would hope would be true.

**Exercise 3.2.26** ("Submanifolds of submanifolds are submanifolds"). Let X be a manifold, S a submanifold of X, and T a submanifold of S. Show that T is a submanifold of X.

## 3.2.E Submanifolds of euclidean space

We now have two important results which let us produce lots of examples of submanifolds of euclidean space out of smooth functions (in the sense of definition 2.5.36).

First up, we have the following important theorem which goes by many names; we will call it the "regular level set theorem," but some other names include the "preimage theorem" or the "submersion theorem." We will later generalize this theorem.

**Theorem 3.2.27** (Regular level set theorem). Suppose U is an open subset of  $\mathbb{R}^m$  and  $f: U \to \mathbb{R}^n$  is a smooth function where  $m \ge n$ . If  $q \in \mathbb{R}^n$  is a regular value of f, then  $S = f^{-1}(q)$  is a submanifold of U of dimension m - n.

*Proof.* By replacing f with the function  $x \mapsto f(x) - q$ , we can assume without loss of generality that q = 0. Fix  $p \in S$ . We want to find a chart adapted to S containing p. Since 0 is a regular value and  $m \ge n$ , we know that f is submersive at p. This means that the  $n \times m$  matrix f'(p) has a non-vanishing  $n \times n$  minor, ie, there exist n distinct integers  $i_1, \ldots, i_n$  between 1 and m such that

$$\det \begin{bmatrix} \partial_{i_1} f_1(p) & \cdots & \partial_{i_n} f_1(p) \\ \vdots & \ddots & \vdots \\ \partial_{i_1} f_n(p) & \cdots & \partial_{i_n} f_n(p) \end{bmatrix} \neq 0.$$
(3.2.28)

By permuting the coordinates, we can assume without loss of generality that  $i_1 = 1, i_2 = 2, ..., i_n = n$  (cf. exercise 3.2.29). Consider the function  $y : U \to \mathbb{R}^m$  such that, if  $x = (x_1, ..., x_m) \in U$ , then

$$y(x) = (f_1(x), \dots, f_n(x), x_{n+1}, \dots, x_m).$$

## 3 Manifolds

Then

$$y'(x) = \begin{bmatrix} \partial_1 f_1(x) & \cdots & \partial_n f_1(x) & \partial_{n+1} f_1(x) & \cdots & \partial_m f_1(x) \\ \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ \partial_1 f_n(x) & \cdots & \partial_n f_n(x) & \partial_{n+1} f_n(x) & \cdots & \partial_m f_n(x) \\ 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1 \end{bmatrix}$$

In other words, the  $(m - n) \times m$  submatrix on the bottom left of g'(x) is all zeroes, and the  $(m - n) \times (m - n)$  submatrix on the bottom right is the identity matrix. This means that det y'(x) is equal, up to sign, to the determinant of the  $n \times n$  in the top left, which we know from equation (3.2.28) is nonzero when x = p. Thus y is étale at p.

Since f is smooth, so is y, so there exists an open neighborhood V of p such that  $y|_V$  is étale. By the inverse function theorem 2.4.13, we can replace V with a smaller open neighborhood of p in order to further assume that  $y|_V$  is injective. Then exercise 3.2.3 tells us that (V, y) is a chart in the euclidean atlas containing p. Moreover,

$$S \cap V = \{p' \in V : f(p') = 0\} = \{p' \in V : y_1(p') = \dots = y_n(p') = 0\}$$

because  $y_j = f_j$  for all j = 1, ..., n. This shows that (V, y) is adapted to S, and that S has codimension n at p in  $\mathbb{R}^m$ .

**Exercise 3.2.29.** Suppose  $\sigma$  is a permutation of the set  $\{1, \ldots, m\}$ . Let  $f_{\sigma} : \mathbb{R}^m \to \mathbb{R}^m$  be the function which "permutes the coordinates of elements of  $\mathbb{R}^m$  according to  $\sigma$ ," ie,

$$f_{\sigma}(x_1,\ldots,x_m) = (x_{\sigma(1)},\ldots,x_{\sigma(m)}).$$

Show that a subset S is a submanifold of  $\mathbb{R}^m$  if and only if  $f_{\sigma}(S)$  is a submanifold of  $\mathbb{R}^m$ .

**Exercise 3.2.30.** Recall from exercises 3.2.4 and 3.2.6 that the n-sphere S<sup>n</sup> is defined by

$$S^n = \{x \in \mathbb{R}^{n+1} : |x| = 1\}.$$

- (a) Use the regular level set theorem 3.2.27 to show that  $S^1$  is a submanifold of  $\mathbb{R}^2$ .
- (b) Show that the maximal atlas on S<sup>1</sup> produced by proposition 3.2.23 coincides with the maximal atlas from exercise 3.2.4.

(c) Generalize parts (a) and (b) to  $S^n$  for arbitrary n.

**Proposition 3.2.31.** Suppose U is an open subset of  $\mathbb{R}^m$  and  $f: U \to \mathbb{R}^n$  is a smooth function. Then the graph of f is a submanifold of  $\mathbb{R}^{m+n}$ .

*Proof.* We derive this from the regular level set theorem 3.2.27. Consider the function  $g: U \times \mathbb{R}^n \to \mathbb{R}^n$  given by g(x, y) = f(x) - y. Then the graph of f is equal to  $g^{-1}(0)$ , so the result will follow from theorem 3.2.27 provided we can show that 0 is a regular value of g.

In fact, g is submersive at all points, so all of its values are regular! To see this, we calculate dg. Note that if  $x = (x_1, ..., x_m) \in U$  and  $y = (y_1, ..., y_n) \in \mathbb{R}^n$ , then

 $g(x,y) = g(x_1,\ldots,x_m,y_1,\ldots,y_n) = (f_1(x_1,\ldots,x_m) - y_1,\ldots,f_n(x_1,\ldots,x_m) - y_n)$ 

by definition of g, which means that

$$\frac{\partial g_k}{\partial x_i}(x,y) = \frac{\partial f_k}{\partial x_j}(x) = \partial_j f_k(x) \quad \text{and} \quad \frac{\partial g_k}{\partial y_j}(x,y) = -\delta_{j,k} = \begin{cases} -1 & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases}$$

where i = 1, ..., m and j, k = 1, ..., n. In other words, g'(x, y) is the following  $n \times (m + n)$  matrix.

$$g'(x,y) = \begin{vmatrix} \partial_1 f_1(x) & \partial_2 f_1(x) & \cdots & \partial_m f_1(x) & -1 & 0 & \cdots & 0 \\ \partial_1 f_2(x) & \partial_2 f_2(x) & \cdots & \partial_m f_2(x) & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \partial_1 f_n(x) & \partial_2 f_n(x) & \cdots & \partial_m f_n(x) & 0 & 0 & \cdots & -1 \end{vmatrix} = \begin{bmatrix} f'(x) & -I \end{bmatrix}$$

where I denotes the  $n \times n$  identity matrix. Since I has rank n, so does g'(x, y), proving that g is submersive everywhere.

An important cautionary remark is that sometimes, a subspace that is a manifold can fail to be a submanifold. For example, you might notice that we showed in exercise 3.2.8 that the graph of a *continuous* function between two euclidean spaces is a subspace of euclidean space and is also a manifold; and then we showed in proposition 3.2.31 that the graph of a *smooth* function is a submanifold of euclidean space. And indeed, the graphs of continuous-but-not-smooth functions provide examples of subspaces that are manifolds but not submanifolds.

**Exercise 3.2.32.** Let  $f : \mathbb{R} \to \mathbb{R}$  be the function f(x) = |x|. Show that the graph of f is not a submanifold of  $\mathbb{R}^2$ .

## 3.3 Smooth functions

We will now define what it means for a function between two manifolds to be smooth. Recall that in chapter 2 we saw that the case when the codomain is  $\mathbb{R}$  was the most important. The same will be true here. You may find that the number of definitions of "smooth" gets slightly out of hand; if you do, just remember that, whenever multiple definitions of "smooth" make sense, they'll all agree.

### 3.3.A Smooth real-valued functions

## Definition of smoothness for real-valued functions

**Definition 3.3.1** (Smooth real-valued function). Let X be a manifold. A function  $f : X \to \mathbb{R}$  is said to be *smooth* if, for any chart  $(U, x) \in \mathcal{A}_X$ , the function  $f \circ x^{-1}$  is smooth.

$$\mathbf{x}(\mathbf{U}) \xrightarrow{\mathbf{x}^{-1}} \mathbf{U} \xrightarrow{\mathbf{f}} \mathbb{R}$$

There are two disconcerting things about this definition. First, there is the issue that checking smoothness of a particular function seems next to impossible: the maximal atlas  $A_X$  will have *many, many* charts, some that we may not even know about, and the definition makes it seem like we need to check smoothness on all of them! Thankfully, we don't need to check on all possible charts. As soon as we check smoothness on any atlas, even if it's not maximal, we're done.

**Lemma 3.3.2.** Let X be a manifold and  $f : X \to \mathbb{R}$  a function. Suppose that there exists an atlas  $\mathcal{A} \subseteq \mathcal{A}_X$  such that  $f \circ x^{-1}$  is smooth for any  $(\mathbf{U}, \mathbf{x}) \in \mathcal{A}$ . Then f is smooth.

*Proof.* Suppose  $(V, y) \in A_X$ . For any  $(U, x) \in A$ , we have maps as follows.



Since (V, y) and (U, x) are both elements of the maximal atlas  $A_X$ , we know that the transition function  $x \circ y^{-1}$  is smooth. Since  $(U, x) \in A$ , we know by assumption that  $f \circ x^{-1}$  is smooth. Thus

$$f \circ y^{-1} = (f \circ x^{-1}) \circ (x \circ y^{-1})$$

is a smooth function  $y(U \cap V) \to \mathbb{R}$ .

Now note that, since A is an atlas, its charts cover all of X, so

$$\bigcup_{(\mathbf{U},\mathbf{x})\in\mathcal{A}} \mathbf{y}(\mathbf{U}\cap \mathbf{V}) = \mathbf{y}(\mathbf{V}).$$

Since  $f \circ y^{-1}$  is smooth when restricted to each the open subset  $y(U \cap V)$  for all  $(U, x) \in A$ , we conclude that  $f \circ y^{-1}$  must be smooth on all of y(V).

The second disconcerting thing is the following. Suppose  $U \subseteq \mathbb{R}^m$  is an open subset and  $f: U \to \mathbb{R}$  is a function. On the one hand, we defined what it means for f to be smooth back in definition 2.5.36. On the other hand, we can regard U as an open submanifold of  $\mathbb{R}^n$  (cf. exercise 3.2.9) and then definition 3.3.1 gives us another definition of what it means for f to be smooth. A priori, it could be that these definitions conflict. Thankfully, they do not.

**Exercise 3.3.3.** Suppose  $U \subseteq \mathbb{R}^m$  is an open subset and  $f : U \to \mathbb{R}$  is a function. Show that f is smooth in the sense of definition 2.5.36 if and only if it is smooth in the sense of definition 3.3.1.

*Possible hint*. You might find lemma 3.3.2 useful.

From here on out, we'll use the fact that these two definitions of smoothness coincide without comment.

### Combining smooth real-valued functions

**Definition 3.3.4.** If X is a manifold, we write O(X) for the set of smooth functions  $X \to \mathbb{R}$ .

Here is the standard set of stability properties we would want.

**Exercise 3.3.5.** (a) Show that O(X) is a vector space.

(b) If f,  $g \in O(X)$ , show that  $fg \in O(X)$  also.

(c) If  $f \in O(X)$  and  $f(x) \neq 0$  for all x, show that  $1/f \in O(X)$  also.

*Possible hint.* You might decide to do part (d) first, and then use it to prove (c). Or not, you could also just prove (c) directly.

(d) Suppose  $f \in O(X)$ , U is an open subset of  $\mathbb{R}$ ,  $g : U \to \mathbb{R}$  is smooth, and  $f(X) \subseteq U$ . Show that  $g \circ f \in O(X)$ .

*Unimportant remark.* if you've taken abstract algebra, you might be interested to notice that the first two stability properties in exercise 3.3.5 show that O(X) is an  $\mathbb{R}$ -algebra (ie, that it is simultaneously a commutative ring and a vector space, and that these two structures are "compatible" with each other).

### Smooth approximations of characteristic functions of compact subsets \*

Here is a generalization of exercise 1.4.35 which proves, roughly speaking, that we can find a smooth approximation of the characteristic function of a compact subset of a manifold. You may want to look at definition 0.7.24 for the definition of compactness for topological spaces.

**Proposition 3.3.6.** Suppose K is a compact subset of a manifold X and U is an open neighborhood of K. Then there exists a smooth function  $f \in O(X)$  such that  $0 \le f(x) \le 1$  for all  $x \in \mathbb{R}$ , f(x) = 1 for all  $x \in K$ , and f(x) = 0 for all  $x \notin U$ .

*Proof.* Suppose (V, x) is a chart containing some  $p \in K$ . Recentering, we can assume that x(p) = 0 (cf. remark 3.1.2). By shrinking V, we can assume that  $\overline{V} \subseteq U$  (cf. exercise 3.3.7). Then x(V) is an open neighborhood of  $0 = x(p) \in \mathbb{R}^n$ , so there exists  $\varepsilon > 0$  such that  $B(0, \varepsilon) \subseteq x(V)$ . Let  $\psi : \mathbb{R}^n \to \mathbb{R}$  be a bump function supported on  $\overline{B}(0, \varepsilon)$  (cf. exercise 2.5.39). Then  $\psi \circ x$  is a smooth function on V (cf. exercise 3.3.7) which is supported entirely inside V (namely, on  $x^{-1}(\overline{B}(0, \varepsilon))$ ), and is zero everywhere else on V. Define  $f_p : X \to \mathbb{R}$  by

$$f_{p}(x) = \begin{cases} \psi \circ x & \text{if } x \in V \\ 0 & \text{if } x \notin V. \end{cases}$$

Then  $f_p$  is smooth (cf. exercise 3.3.7), and it is strictly positive on an open neighborhood  $V_p$  of p contained inside U.

We can construct such a function  $f_p$  for every  $p \in K$ . Since K is compact, there exist finitely many  $p_1, \ldots, p_n \in K$  such that  $K \subseteq V_{p_1} \cup \cdots \cup V_{p_n}$ . Thus  $f_{p_1} + \cdots + f_{p_n}$  is strictly positive on K, and zero outside U. Since K is compact, there exists  $\delta > 0$  such that  $f_{p_1} + \cdots + f_{p_n} \ge \delta$ 

on K (cf. exercise 3.3.7). Let g be a bridging function such that g(x) = 1 for  $x \ge \delta$  and g(x) = 0 for  $x \le 0$  (cf. example 1.4.34), and then set  $f = g \circ (f_{p_1} + \dots + f_{p_n})$ .

Exercise 3.3.7. Fill in the missing details in the above proof.

- (a) Prove that we can "shrink V in order to assume that  $\bar{V}\subseteq U."$
- (b) Modify the bump function of exercise 2.5.39 so that it has support  $\overline{B}(0, \epsilon)$ .
- (c) Explain why  $\psi \circ x$  is a smooth function  $V \to \mathbb{R}$ .
- (d) Explain why  $f_p$  is a smooth function  $X \to \mathbb{R}$ .
- (e) If  $h : K \to \mathbb{R}$  is a continuous function such that h(x) > 0 for all  $x \in K$ , show that there exists  $\delta > 0$  such that  $h(x) \ge \delta$  for all x.

*Possible hint*. Let  $\delta = \inf\{h(x) : x \in K\}$  and recall the extreme value theorem.

## 3.3.B Smooth functions between manifolds

Here is a definition that may look a bit terrifying at first.

**Definition 3.3.8** (Smoothness). Suppose X and Y are manifolds and  $f : X \to Y$  is a function. Then f is *smooth* if, for every chart  $(V, y) \in A_Y$  and every chart  $(U, x) \in A_X$  such that  $f(U) \subseteq V$ , the function  $y \circ f \circ x^{-1} : x(U) \to y(V)$  is smooth.



Once again, we don't need to check all possible charts. The statement is a bit complicated, but it's exactly the statement you might hope for.

**Exercise 3.3.9.** Suppose X and Y are manifolds and  $f : X \to Y$  is a function. Suppose further that there exists an atlas  $\mathcal{A} \subseteq \mathcal{A}_Y$  such that, for every  $(V, y) \in \mathcal{A}$ , there exists an atlas

## 3 Manifolds

 $\mathcal{A}' \subseteq \mathcal{A}_{f^{-1}(V)}$ , such that for every  $(U, x) \in \mathcal{A}'$ , the function  $y \circ f \circ x^{-1} : x(U) \to y(V)$  is smooth. Show that f is smooth.

We should also note that we haven't redefined the word "smooth" in any conflicting ways. More precisely, we have the following two exercises.

**Exercise 3.3.10.** If  $U \subseteq \mathbb{R}^m$  is an open subset and  $f : U \to \mathbb{R}^n$  is a function, show that f is smooth in the sense of definition 2.5.36 if and only if it is smooth in the sense of definition 3.3.8.

**Exercise 3.3.11.** Show that a function  $f : X \to \mathbb{R}$  is smooth in the sense of definition 3.3.1 if and only if it is smooth in the sense of definition 3.3.8.

From here on out, we'll use the fact that all of these definitions of smoothness coincide without comment.

Here is an important example of a smooth function.

**Exercise 3.3.12.** Let  $f : \mathbb{R} \to S^1$  be the function  $f(t) = (\cos(t), \sin(t))$ . Show that f is smooth.

### **Combining smooth maps**

It doesn't make sense to add or multiply maps that take values in a manifold. The only thing that makes sense is composition, and indeed, we can compose.

**Exercise 3.3.13.** Suppose X, Y, Z are all manifolds and  $f : X \to Y$  and  $g : Y \to Z$  are smooth. Show that  $g \circ f$  is also smooth.

## Reduction to smooth real-valued functions

Finally, we show that the case when the codomain is  $\mathbb{R}$  is the most important case. We do this in two steps.

**Proposition 3.3.14.** Suppose X is a manifold and  $f : X \to \mathbb{R}^n$  is a function. Then f is smooth if and only if the component functions  $f_i = \pi_i \circ f$  are smooth for all i = 1, ..., n.

*Proof.* The single chart ( $\mathbb{R}^n$ , id) is a sub-atlas of the euclidean atlas on  $\mathbb{R}^n$ , so exercise 3.3.9 says that f is smooth if and only if, for every  $(U, x) \in \mathcal{A}_X$ , the composite  $f \circ x^{-1} : x(U) \to \mathbb{R}^n$  is smooth. This is a function from an open subset of euclidean space into euclidean space,

so we know from the definition of smoothness in chapter 2 that  $f \circ x^{-1}$  is smooth if and only if its component functions  $\pi_i \circ (f \circ x^{-1})$  are smooth for all i. But

$$\pi_{\mathfrak{i}} \circ (\mathfrak{f} \circ \mathfrak{x}^{-1}) = (\pi_{\mathfrak{i}} \circ \mathfrak{f}) \circ \mathfrak{x}^{-1} = \mathfrak{f}_{\mathfrak{i}} \circ \mathfrak{x}^{-1},$$

and  $f_i \circ x^{-1}$  is smooth for all  $(U, x) \in A_X$  if and only if  $f_i$  is smooth, again by exercise 3.3.9.  $\Box$ 

**Proposition 3.3.15.** *Suppose*  $f : X \to Y$  *is a function between two manifolds. Then the following are equivalent.* 

- (a) f is smooth.
- (b) For every open subset  $V \subseteq Y$  and every smooth function  $g : V \to \mathbb{R}$ , the composite  $g \circ f|_{f^{-1}(V)}$  is a smooth function  $f^{-1}(V) \to \mathbb{R}$ .

*Proof.* One direction follows immediately from exercise 3.3.13. Conversely, suppose f has the property described in (b). Choose a chart  $(V, y) \in A_Y$  of dimension n. By assumption,  $y_i \circ f|_{f^{-1}(V)}$  is a smooth function  $f^{-1}(V) \to \mathbb{R}$  for all i = 1, ..., n. By proposition 3.3.14, this means that  $y \circ f|_{f^{-1}(V)}$  is smooth. Since this is true for all charts  $(V, y) \in A_Y$ , we conclude that f is smooth.

*Unimportant remark.* Suppose  $f : X \to Y$  is a smooth function between manifolds. For any  $g \in O(Y)$ , we have  $g \circ f \in O(X)$ . Thus  $g \mapsto g \circ f$  defines a function  $f^{\sharp} : O(Y) \to O(X)$ , which is actually a homomorphism of  $\mathbb{R}$ -algebras.

## 3.3.C Paths

Incomplete. Gist: we can do basically all we did in section 2.5.E on manifolds, too.

### 3.3.D Diffeomorphisms

Diffeomorphisms are how we formalize the idea that two manifolds are "basically the same."

**Definition 3.3.16** (Diffeomorphism). Suppose X and Y are manifolds. A function  $f : X \to Y$  is a *diffeomorphism* if it is smooth, bijective, and the inverse function  $f^{-1} : Y \to X$  is also smooth. If there exists a diffeomorphism  $f : X \to Y$ , then X and Y are *diffeomorphic*.

## 3 Manifolds

**Exercise 3.3.17.** Regard  $\mathbb{R}$  as a manifold in two different ways: once with the euclidean atlas, and next with the maximal atlas containing ( $\mathbb{R}$ ,  $x \mapsto x^3$ ). Show that these two manifolds are diffeomorphic.

**Exercise 3.3.18.** If X is a manifold and  $(U, x) \in A_X$ , prove that  $x : U \to x(U)$  is a diffeomorphism.

**Definition 3.3.19** (Local diffeomorphism). Suppose X and Y are manifolds. A function  $f : X \to Y$  is a *local diffeomorphism* if, for every  $p \in X$ , there exists an open neighborhood U of p such that f(U) is an open subset of Y and  $f|_U$  is a diffeomorphism  $U \to f(U)$ .

**Exercise 3.3.20.** Consider the smooth function  $f : \mathbb{R} \to S^1$  given by  $f(t) = (\cos(t), \sin(t))$  from exercise 3.3.12. Show that f is a local diffeomorphism.

Exercise 3.3.21. Show that local diffeomorphisms are smooth and open.

# **4 Tangent Spaces**

# 4.1 Tangent space

Throughout this section, we let X denote a manifold and  $p \in X$  a point. We also let  $n := \dim_p(X)$ .

## 4.1.A Curves based at a point

**Definition 4.1.1.** A *curve in* X *based at* p is a smooth function  $\gamma : (-\epsilon, \epsilon) \to X$  for some  $\epsilon > 0$  such that  $\gamma(0) = p$ .

Suppose that  $(U, x) \in A_X$  is a chart containing p and that  $\gamma : (-\epsilon, \epsilon) \to \mathbb{R}$  is a curve based at p. Since U is an open subset of X and  $\gamma$  is continuous, the preimage  $\gamma^{-1}(U)$  is an open subset of  $(-\epsilon, \epsilon)$  containing 0, so there exists  $0 < \epsilon' \leq \epsilon$  such that the image of the smaller interval  $(-\epsilon', \epsilon)$  is entirely contained inside U. We let  $x \circ \gamma$  denote the composite function

$$(-\epsilon',\epsilon') \longrightarrow U \longrightarrow \mathbb{R}^n.$$

We can then take its derivative at 0 as in chapter 2. This derivative  $d(x \circ \gamma)_0$  is a linear map  $\mathbb{R} \to \mathbb{R}^n$ , so its matrix representation

$$(\mathbf{x} \circ \boldsymbol{\gamma})'(\mathbf{0})$$

is a column vector in  $\mathbb{R}^n$ .

**Definition 4.1.2.** Two curves  $\gamma_1, \gamma_2$  in X based at p are *equivalent* if  $d(x \circ \gamma_1)_0 = d(x \circ \gamma_2)_0$  for every chart  $(U, x) \in \mathcal{A}_X$  containing p.

As usual, the first point of order is to check that we don't need to check all possible charts.

**Exercise 4.1.3.** Show that, if  $\gamma_1, \gamma_2$  are two curves in X based at p and there exists a chart  $(U, x) \in A_X$  containing p and

$$\mathbf{d}(\mathbf{x}\circ\boldsymbol{\gamma}_1)_0=\mathbf{d}(\mathbf{x}\circ\boldsymbol{\gamma}_2)_0,$$

then  $\gamma_1$  is equivalent to  $\gamma_2$ .

**Definition 4.1.4.** We define the *tangent space of* X *at* p, denoted  $T_pX$ , to be the set of equivalence classes of curves in X based at p. Elements of  $T_pX$  are called *tangent vectors* at p. If  $\gamma$  is a curve based at p, we will write  $v_{\gamma}$  for the corresponding tangent vector in  $T_pX$ .

### 4.1.B Vector space structure

We will now put a vector space structure on the tangent space  $T_pX$ . There are a lot of details involved in the process, so we begin with a high-level overview. If we choose a chart  $(U, x) \in A_X$  containing p, there is a bijection  $\sigma : T_pX \to \mathbb{R}^n$  (cf. lemma 4.1.5). This means that we can define vector space operations on  $T_pX$  by "pulling back along  $\sigma$ ." More precisely, this mean the following: given  $v \in T_pX$  and  $\lambda \in \mathbb{R}$ , we define

$$\lambda v := \sigma^{-1}(\lambda \sigma(v)),$$

and given  $v_1, v_2 \in T_p X$ , we define

$$v_1 + v_2 := \sigma^{-1}(\sigma(v_1) + \sigma(v_2)).$$

We can then show that these operations define the structure of a vector space on  $T_p V$  by showing that all of the vector space axioms are satisfied (cf. lemma 4.1.7). The final step is to show that these operations do not depend on the choice of chart (U, x) (cf. exercise 4.1.11). Thus there is a canonical, coordinate-free, vector space structure on  $T_p X$ .

**Lemma 4.1.5.** Suppose  $(U, x) \in A_X$  is a chart containing p. There is a well-defined bijection  $\sigma : T_pX \to \mathbb{R}^n$  given by  $v_{\gamma} \mapsto [d(x \circ \gamma)_0] = d(x \circ \gamma)_0(1)$ .

*Proof.* The fact that  $\sigma$  is well-defined follows immediately from definition 4.1.2. For a vector  $w \in \mathbb{R}^n$ , define  $\gamma_w : (-\varepsilon, \varepsilon) \to X$  by

$$\gamma_{w}(t) = x^{-1}(x(p) + tw)$$

where  $\epsilon$  is chosen to be small enough that  $x(p) + tw \in x(U)$  for all  $t \in (-\epsilon, \epsilon)$  (cf. exercise 4.1.6). Then  $\gamma_w(0) = x^{-1}(x(p)) = p$ , so  $\gamma_w$  is a curve in X based at p. Notice that

$$(\mathbf{x} \circ \gamma_{w})(\mathbf{t}) = \mathbf{x}(\mathbf{x}^{-1}(\mathbf{x}(\mathbf{p}) + \mathbf{t}w)) = \mathbf{x}(\mathbf{p}) + \mathbf{t}w$$

so  $[d(x \circ \gamma_w)_0] = w$  (cf. exercise 4.1.6). It follows from this that  $w \mapsto v_{\gamma_w}$  is inverse to  $\sigma$ , proving that  $\sigma$  is bijective (cf. exercise 4.1.6).

**Exercise 4.1.6.** (a) Explain why there must exist  $\varepsilon > 0$  such that  $x(p) + tw \in x(U)$  for all  $t \in (-\varepsilon, \varepsilon)$ .

- (b) Explain why  $[d(x \circ \gamma)_0] = w$ .
- (c) Explain how it follows from (b) that  $w \mapsto v_{\gamma_w}$  is inverse to  $\sigma$ .

**Lemma 4.1.7.** Suppose  $(U, x) \in A_X$  is a chart containing p and  $\sigma : T_pX \to \mathbb{R}^n$  is the bijection of lemma 4.1.5. Then  $T_pX$ , equipped with the addition and scalar multiplication operations obtained by pulling back along  $\sigma$ , is a vector space. The zero element is the equivalence class of the constant curve at p.

*Proof.* This follows immediately from the fact that  $\mathbb{R}^n$  is a vector space and that  $\sigma$  is bijective, but writing out all of the details is fairly tedious. To explain the idea, let's show that vector addition is commutative. For  $v_1, v_2 \in T_p X$ , we have

$$v_1 + v_2 = \sigma^{-1}(\sigma(v_1) + \sigma(v_2))$$
  
=  $\sigma^{-1}(\sigma(v_2) + \sigma(v_1))$   
=  $v_2 + v_1$ 

where we used the definition of addition in  $T_pX$  on the first and last steps, and the fact that addition in  $\mathbb{R}^n$  is commutative for the middle step.

Let  $\gamma$  denote the constant curve  $\gamma(t) = p$  for all t. Then  $x \circ p$  is also a constant function which always takes the value x(p), so its derivative is 0. Thus, for any  $v \in T_pX$ , we have

$$v + v_{\gamma} = \sigma^{-1}(\sigma(v) + \sigma(v_{\gamma}))$$
$$= \sigma^{-1}(\sigma(v) + 0)$$
$$= \sigma^{-1}(\sigma(v))$$
$$= v,$$

proving that  $v_{\gamma}$  is the zero element of  $T_p X$ .

The rest of the proof is left to you (cf. exercise 4.1.8)

**Exercise 4.1.8.** Check all of the other axioms that need to be checked to ensure that  $T_pX$  is a vector space.<sup>1</sup>

**Exercise 4.1.9.** Suppose  $(U, x) \in A_X$  is a chart containing p and  $\sigma : T_pX \to \mathbb{R}^n$  is the bijection of lemma 4.1.5. If  $T_pX$  is equipped with the vector space operations obtained by pulling back along  $\sigma$ , then  $\sigma : T_pX \to \mathbb{R}^n$  is a linear map. Therefore, it is an isomorphism of vector spaces, and dim  $T_pX = n$ .

**Lemma 4.1.10.** Suppose  $(U, x), (U', x') \in A_X$  are both charts containing p and that  $\sigma, \sigma' : T_p X \to \mathbb{R}^n$  are the bijections from lemma 4.1.5 corresponding to (U, x), (U', x'), respectively. Then  $\sigma' = d(x' \circ x)_{x(p)} \circ \sigma$ , where  $x' \circ x$  denotes the transition map (cf. definition 3.1.4).



*Proof.* Let  $\gamma$  be a curve in X based at p. Then

$$\mathbf{x}' \circ \mathbf{\gamma} = \mathbf{x}' \circ (\mathbf{x}^{-1} \circ \mathbf{x}) \circ \mathbf{\gamma} = (\mathbf{x}' \circ \mathbf{x}) \circ (\mathbf{x} \circ \mathbf{\gamma}),$$

so the chain rule 2.3.3 says that

$$d(x' \circ \gamma)_{0} = d(x' \circ x)_{(x \circ \gamma)(0)} \circ d(x \circ \gamma)_{0} = d(x' \circ x)_{x(p)} \circ d(x \circ \gamma)_{0}.$$

Evaluating at 1, we find that

$$\sigma'(\nu_{\gamma}) = d(x' \circ \gamma)_{0}(1) = d(x' \circ x)_{x(p)}(d(x \circ \gamma)_{0}(1)) = d(x' \circ x)_{x(p)}(\sigma(\nu_{\gamma})).$$

Since this is true for all  $\gamma$ , we conclude that  $\sigma' = d(x' \circ x)_{x(p)} \circ \sigma$ .

<sup>&</sup>lt;sup>1</sup>During a class I took with him, George Bergman once said something along the following lines: "If you're sure you can write the details of a proof, you probably don't need to. But if you're not sure you can formalize the details, you need to do it." I encourage you to apply this principle for exercise 4.1.8.

**Exercise 4.1.11.** Suppose  $(U, x), (U', x') \in A_X$  are both charts containing p and that  $\sigma, \sigma' : T_p X \to \mathbb{R}^n$  are the bijections from lemma 4.1.5 corresponding to (U, x), (U', x'), respectively. Show that the vector space operations on  $T_p X$  defined by pulling back along  $\sigma$  are the same as those defined by pulling back along  $\sigma'$ .

*Possible hint*. Use lemma 4.1.10. The key is that  $d(x' \circ x^{-1})_{x(p)}$  is an invertible linear map.

This completes our construction of a vector space structure on  $T_pX$ . Notice that  $T_pX$  is an example of a finite dimensional vector space that does not come equipped with any canonical basis! More precisely, this means the following. If we choose a chart  $(U, x) \in A_X$ containing p, the bijection  $\sigma : T_pX \to \mathbb{R}^n$  of lemma 4.1.5 is an isomorphism of vector spaces (cf. exercise 4.1.9), so  $\sigma^{-1}(e_1), \ldots, \sigma^{-1}(e_n)$  is a basis of  $T_pX$ . But then, if we choose a different chart  $(U', x') \in A_X$  containing p and  $\sigma'$  is the corresponding isomorphism  $T_pX \to \mathbb{R}^n$ , then  $\sigma'^{-1}(e_1), \ldots, \sigma'^{-1}(e_n)$  will in general be a different basis of  $T_pX$ , even though the vector space structures defined by pulling back along  $\sigma$  and  $\sigma'$  are the same! So, for a general manifold X and point p, since there's no "canonical" choice of a chart containing p, there is also no "canonical" choice of a basis on  $T_pX$ . I strongly encourage you to work through the following exercise to make sense of this.

**Exercise 4.1.12.** Consider the unit circle S<sup>1</sup>, regarded manifold using the charts (U, x) and (U', x') defined in exercise 3.2.4. Any point  $p \in S^1$  can be written as  $(\cos \theta, \sin \theta)$  for some  $\theta \in \mathbb{R}$ .

(a) Observe that  $\gamma : \mathbb{R} \to S^1$  given by

$$\gamma(t) = (\cos(\theta + t), \sin(\theta + t))$$

is a curve in S<sup>1</sup> based at p. Explain why  $v_{\gamma}$  is a basis for  $T_pS^1$ .

- (b) Suppose  $p \in U$  and let  $\sigma : T_pS^1 \to \mathbb{R}$  be the isomorphism of lemma 4.1.5 corresponding to the chart (U, x). Find  $\lambda \in \mathbb{R}$  such that  $\sigma(\lambda v_{\gamma}) = 1$ .
- (c) Suppose  $p \in U'$  and let  $\sigma' : T_pS^1 \to \mathbb{R}$  be the isomorphism of lemma 4.1.5 corresponding to the chart (U', x'). Find  $\lambda' \in \mathbb{R}$  such that  $\sigma'(\lambda' v_{\gamma}) = 1$ .

So, if  $p \in U \cap U'$ , we see that there are three possible bases we might choose on  $T_pS^1$  ( $v_\gamma$ , or  $\lambda v_\gamma$ , or  $\lambda' v_\gamma$ ), none of these is any more "canonical" than another. This is what is meant when we say that  $T_pS^1$  has no canonical choice of basis.

#### 4 Tangent Spaces

**Remark 4.1.13.** Suppose U is an open subset of X containing p regarded as an open submanifold (cf. exercise 3.2.9). A curve  $\gamma$  in U based at p is also automatically a curve in X, which means there is a natural "inclusion" map  $T_pU \rightarrow T_pX$  given by  $\nu_{\gamma} \mapsto \nu_{\gamma}$ . But we can "shrink" curves in X to curves in U without changing their equivalence class, so  $T_pU \rightarrow T_pX$  is actually bijective. Moreover, if we choose a chart in  $A_U$  containing p, we can use this chart to define the vector space structure on both  $T_pU$  and  $T_pX$ , so  $T_pU \rightarrow T_pX$  is actually an isomorphism of vector spaces. Since this isomorphism is tautological, we sometimes write  $T_pU = T_pX$ .

**Remark 4.1.14.** There is one situation where we do have a canonical choice of a chart:  $\mathbb{R}^n$ . Namely, the identity map id :  $\mathbb{R}^n \to \mathbb{R}^n$  defines a chart. If  $\gamma$  is a curve in  $\mathbb{R}^n$  based at a point p, then  $d(id \circ \gamma)_0 = d\gamma_0$ , so the isomorphism  $\sigma : T_p \mathbb{R}^n \to \mathbb{R}^n$  is given by  $\sigma(\nu_{\gamma}) = [d\gamma_0]$ . We will call this the "canonical isomorphism"  $T_p \mathbb{R}^n \to \mathbb{R}^n$ , which we sometimes also write as  $T_p \mathbb{R}^n = \mathbb{R}^n$ . But this is fairly abusive and it's worth remembering that  $T_p \mathbb{R}^n$  is not literally equal to  $\mathbb{R}^n$ . The former is equivalence classes of curves, the latter is column vectors, and the relationship between the two is given by taking the derivative at 0.

## 4.1.C Derivations \*

There is another perspective on tangent vectors which is often useful, though it is decidedly less geometric at first glance. It begins with the following definition.

**Definition 4.1.15.** Let X be a manifold and  $p \in X$  a point. A *derivation* on X based at p is a linear function  $\partial : O(X) \to \mathbb{R}$  such that

$$\partial(\mathbf{f}\mathbf{g}) = \partial(\mathbf{f})\mathbf{g}(\mathbf{p}) + \mathbf{f}(\mathbf{p})\partial(\mathbf{g}).$$

We let  $\text{Der}_{p}(X)$  denote the set of all derivations on X based at p.

The following is not very hard to prove.

**Exercise 4.1.16.** Show that  $Der_{p}(X)$  is a vector space.

It turns out that  $T_pX$  and  $Der_p(X)$  are isomorphic, as we will soon prove.

**Remark 4.1.17.** Sometimes, people define the tangent space to be  $\text{Der}_p(X)$ . It's not very geometric, but it at least makes the vector space operations on the tangent space fairly obvious (cf. exercise 4.1.16); compare this with our geometric definition of the tangent

space in terms of curves, where we had to do a lot of work to regard the tangent space as a vector space (cf. section 4.1.B). That said, even if one defines the tangent space to be  $\text{Der}_{p}(X)$ , it's not clear how to produce examples of derivations, or to prove that  $\text{Der}_{p}(X)$  is finite dimensional. One way or another, one needs to show that derivations are somehow linked to curves. This is what we do next.

**Definition 4.1.18.** Suppose  $\gamma$  is a curve in X based at p. If  $f : X \to \mathbb{R}$  is a smooth function, then  $f \circ \gamma$  is a single variable function  $(-\epsilon, \epsilon) \to \mathbb{R}$ . We define  $\partial_{\gamma} : \mathcal{O}(X) \to \mathbb{R}$  by

$$\partial_{\gamma}(\mathbf{f}) = (\mathbf{f} \circ \gamma)'(\mathbf{0}).$$

**Lemma 4.1.19.** If  $\gamma$  is a curve in X based at p, the function  $\partial_{\gamma}$  is a derivation.

*Proof.* Suppose  $f, g \in O(X)$  and  $\lambda \in \mathbb{R}$ . To see that  $\partial_{\gamma}$  is linear, we use the sum and scalar multiple rules from exercises 1.2.1 and 1.2.2.

$$\begin{aligned} \partial_{\gamma}(f + \lambda g) &= ((f + \lambda g) \circ \gamma)'(0) \\ &= ((f \circ \gamma) + \lambda(g \circ \gamma))'(0) \\ &= (f \circ \gamma)'(0) + \lambda(g \circ \gamma)'(0) \\ &= \partial_{\gamma}(f) + \lambda \partial_{\gamma}(g) \end{aligned}$$

To see that it is a derivation, we use the single variable product rule 1.2.3.

$$\begin{aligned} \partial_{\gamma}(fg) &= (fg \circ \gamma)'(0) \\ &= ((f \circ \gamma)(g \circ \gamma))'(0) \\ &= (f \circ \gamma)'(0)(g \circ \gamma)(0) + (f \circ \gamma)(0)(g \circ \gamma)'(0) \\ &= \partial_{\gamma}(f)g(p) + f(p)\partial_{\gamma}(g) \end{aligned}$$

This completes the proof.

**Lemma 4.1.20.** If  $\gamma_1$  and  $\gamma_2$  are equivalent curves in X based at p, then  $\vartheta_{\gamma_1} = \vartheta_{\gamma_2}$ .

*Proof.* Choose a chart  $(U, x) \in A_X$  containing p and let  $\gamma$  be a curve in X based at p. Then

$$\begin{aligned} \partial_{\gamma}(f) &= (f \circ \gamma)'(0) \\ &= d(f \circ \gamma)_{0}(1) \\ &= d(f \circ x^{-1} \circ x \circ \gamma)_{0}(1) \\ &= d(f \circ x^{-1})_{x(p)} \left( d(x \circ \gamma)_{0}(1) \right) \end{aligned}$$

where we used the chain rule 2.3.3 at the last step. Since  $\partial_{\gamma}(f)$  only depends on  $d(x \circ \gamma)_0$ , we see that equivalent curves must give rise to the same derivation.

)

**Theorem 4.1.21.** *The function*  $T_p X \to Der_p(X)$  *given by*  $v_{\gamma} \mapsto \partial_{\gamma}$  *is an isomorphism.* 

**Definition 4.1.22.** If  $v \in T_p X$ , we let  $\partial_v$  denote the corresponding element in  $\text{Der}_p(X)$ .

## 4.2 Pushforward

### 4.2.A Definition of the pushforward

Let  $f : X \to Y$  be a smooth map of manifolds and  $p \in X$  a point. Given a curve  $\gamma$  in X based at p, observe that  $f \circ \gamma$  is a curve in Y based at f(p).

**Lemma 4.2.1.** Let  $f : X \to Y$  be a smooth map of manifolds and  $p \in X$  a point. If  $\gamma_1$  and  $\gamma_2$  are equivalent curves in X based at p, then  $f \circ \gamma_1$  and  $f \circ \gamma_2$  are equivalent curves in Y based at f(p).

This allows us to make the following definition. We will shortly see that this is the "ultimate" generalization of derivatives.

**Definition 4.2.2.** Let  $f : X \to Y$  be a smooth map of manifolds and  $p \in X$  a point. If  $\gamma$  is a curve in X based at p, we define its *pushforward*  $f_{*,p}(\nu_{\gamma}) \in T_{f(p)}Y$  to be the equivalence class of the curve  $f \circ \gamma$ . This defines the *pushforward map*  $f_{*,p} : T_pX \to T_{f(p)}Y$ . If p is clear from context, we also write just  $f_*$  instead of  $f_{*,p}$ .

**Proposition 4.2.3.** Let  $f : X \to Y$  be a smooth map of manifolds and  $p \in X$  a point. The pushforward map  $f_* : T_pX \to T_{f(p)}Y$  is linear.

The following result states that the pushforward generalizes the derivative at a point.

**Proposition 4.2.4.** *Let* U *is an open subset of*  $\mathbb{R}^m$ ,  $p \in U$  *is a point, and*  $f : U \to \mathbb{R}^n$  *is a smooth map. Then, using the identifications of remarks* **4.1.13** *and* **4.1.14***, we have*  $f_* = df_p$ .



*Proof.* Let  $\sigma$  denote the isomorphisms  $T_p\mathbb{R}^m \to \mathbb{R}^m$  and  $T_p\mathbb{R}^n \to \mathbb{R}^n$  of remark 4.1.14. Suppose  $\gamma$  is a curve in U based at p. Then

$$(\mathrm{df}_{\mathrm{p}}\circ\sigma)(\nu_{\gamma})=\mathrm{df}_{\mathrm{p}}(\mathrm{d}\gamma_{0}(1))=(\mathrm{df}_{\mathrm{p}}\circ\mathrm{d}\gamma_{0})(1)=\mathrm{d}(\mathrm{f}\circ\gamma)_{0}(1),$$

where we used the chain rule 2.3.3 for the final step. On the other hand, we also have

$$(\sigma \circ f_*)(\nu_{\gamma}) = \sigma(\nu_{f \circ \gamma}) = d(f \circ \gamma)_0.$$

**Remark 4.2.5.** Sometimes, the map  $f_* : T_pX \to T_pY$  is denoted  $df_p$ , similar to the notation we were using in chapter 2. This creates no conflicts, thanks to proposition 4.2.4.

In light of proposition 4.2.4, the following is a generalization of the chain rule.

**Proposition 4.2.6.** Suppose  $f : X \to Y$  and  $g : Y \to Z$  are smooth maps of manifolds, and  $p \in X$  is a point. Then  $g_* \circ f_* = (g \circ f)_*$ .



*Proof.* If  $\gamma$  is a curve in X based at p, then

$$g_*(f_*(v_{\gamma})) = g_*(v_{f \circ \gamma}) = v_{g \circ f \circ \gamma} = (g \circ f)_*(v_{\gamma}).$$

This proof looks absurdly easy, and that's because it is. You might be wondering why the proof of this generalization of the chain rule looks so easy, while the proof of the chain rule in chapter 2 was significantly more work. It's because the pushforward is so abstractly defined that it's basically useless and uncomputable until we know that it actually

## 4 Tangent Spaces

generalizes the derivative from chapter 2 and therefore can be computed using techniques from chapter 2. This was the content of proposition 4.2.4, and notice that we used the chain rule 2.3.3 to prove proposition 4.2.4.

## 4.2.B Rank of a smooth map

In light of proposition 4.2.4, the following generalizes definition 2.3.36.

**Definition 4.2.7.** Suppose  $f : X \to Y$  is a smooth map between manifolds and  $p \in X$  is a point. The rank of the linear map  $f_* : T_pX \to T_{f(p)}Y$  is also called the *rank of* f *at* p, denoted rank<sub>p</sub>(f). Note that rank<sub>p</sub>(f)  $\leq \min\{\dim_p(X), \dim_{f(p)}(Y)\}$ .

- If  $\operatorname{rank}_p(f) = \dim_p(X)$ , then we say that f is *immsersive at* p. This is equivalent to requiring that  $f_*$  is an injective map  $T_pX \to T_pY$ .
- If  $\operatorname{rank}_{p}(f) = \dim_{f(p)}(Y)$ , then we say that f is *submersive at* p. This is equivalent to requiring that  $f_{*}$  is a surjective map  $T_{p}X \to T_{p}Y$ .
- If  $\operatorname{rank}_p(f) = \dim_p(X) = \dim_{f(p)}(Y)$ , then we say that f is *étale at* p. This is equivalent to requiring that  $f_*$  is an isomorphism  $T_pX \to T_pY$ .
- If  $\operatorname{rank}_{p}(f) = \min\{\dim_{p}(X), \dim_{f(p)}(Y)\}\)$ , then we say that p is a *regular point* of f.
- If  $\operatorname{rank}_{p}(f) < \min\{\dim_{p}(X), \dim_{f(p)}(Y)\}\)$ , then we say that p is a *critical point* of f.

**Definition 4.2.8.** Suppose  $f : X \to Y$  is a smooth map between manifolds.

- If f is immersive at every  $p \in X$ , then f is an *immersion*.
- If f is submersive at every  $p \in X$ , then f is a *submersion*.
- if f is étale at every  $p \in X$ , then f is *étale*.

**Exercise 4.2.9.** Find the critical points of the map  $f: S^2 \to \mathbb{R}^2$  given by f(x, y, z) = (xy, z).

# Bibliography

- [Kö93] Konrad Königsberger. Analysis 2. Springer-Lehrbuch. Grundwissen Mathematik. Springer-Verlag, 1993. ISBN: 3-540-54723-1.
- [Axl15] Sheldon Axler. *Linear algebra done right*. Third. Undergraduate Texts in Mathematics. Springer, 2015. DOI: 10.1007/978-3-319-11080-6.
- [GMM13] Pedro M. Gadea, Jaime Muñoz Masqué, and Ihor V. Mykytyuk. Analysis and algebra on differentiable manifolds. Second. Problem Books in Mathematics. Springer, 2013. DOI: 10.1007/978-94-007-5952-7.
- [GO03] Bernard R. Gelbaum and John M. H. Olmsted. *Counterexamples in analysis*. Corrected reprint of the second (1965) edition. Dover Publications, Inc., 2003.
  ISBN: 0-486-42875-3.
- [GV13] Gene H. Golub and Charles F. Van Loan. *Matrix computations*. Fourth. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, 2013. ISBN: 978-1-4214-0859-0.
- [Hef19] Jim Hefferon. Linear algebra. Third. 2019. URL: http://joshua.smcvt.edu/ linearalgebra/.
- [PM91] Murray H. Protter and Charles B. Morrey Jr. A first course in real analysis. Second. Undergraduate Texts in Mathematics. Springer-Verlag, 1991. DOI: 10.1007/978-1-4419-8744-0.
- [Pom07] Dimitrie D. Pompeiu. "Sur les fonctions dérivées". In: *Mathematische Annalen* 63.3 (1907), pp. 326–332. ISSN: 0025-5831. DOI: 10.1007/BF01449201.
- [Pug15] Charles C. Pugh. *Real mathematical analysis*. Second. Undergraduate Texts in Mathematics. Springer, 2015. DOI: 10.1007/978-3-319-17771-7.
- [Rud76] Walter Rudin. Principles of mathematical analysis. Third. International Series in Pure and Applied Mathematics. McGraw-Hill Book C, 1976.

# Bibliography

[Sai02]	Jean Saint Raymond. "Local inversion for differentiable functions and the
	Darboux property". In: Mathematika 49.1-2 (2002), 141–158 (2004). ISSN: 0025-
	5793. doi: 10.1112/S0025579300016132.

- [Spi05] Michael Spivak. A Comprehensive Introduction to Differential Geometry. Third.Vol. I. Publish or Perish, Inc, 2005. ISBN: 0-914098-83-7.
- [Spi65] Michael Spivak. *Calculus on manifolds*. W. A. Benjamin, Inc., 1965.

# Index

adjugate matrix, 19 atlas, 134 maximal atlas, 134 "bridge" function, 72, 73 bump function, 71, 128 canonical topology, 23, 25, 137 Cauchy-Schwarz inequality, 13 chain rule, 39, 84 chart, 133  $C^{\infty}$ , *see* smooth circle, 135, 157 C<sup>k</sup>, 61, 118 cofactor expansion, 19 compact, 27 compatibile, 133 component function, 85 connected, 103 constant function, 31, 50, 101 continuous, 24 continuously differentiable, 59, 115 coordinates, 133 critical point, 96, 99, 162 critical locus, 60, 99 curve, 153

## cusp, <mark>99</mark>

Darboux function, see intermediate value property Darboux's theorem, 54 derivation, 158 derivative, 32, 101 derivative at a point, 30, 82 diffeomorphism, 68, 151 local diffeomorphism, 151 difference quotient, 29 differentiable, 32, 101 differentiable at a point, 30, 81 differential, see derivative dimension of a manifold, 135 directional derivative, 88 discontinuity jump discontinuity, 53 removable discontinuity, 53 equivalence of curves, 153 of norms, 12, 14 "error in approximation" function, 35, 43,

#### 84

## Index

étale, 96, 162 euclidean space, 15 extreme value theorem, *see* extremum absolute extremum, 45 absolute maximum, 45 absolute minimum, 45, 49 extreme value theorem, 49, 54 interior extremum theorem, 45, 54, 90 local extremum, 45, 46 local maximum, 45 local minimum, 45

fundamental theorem of calculus, 72

GL<sub>n</sub>, 16, 19 GL(V), 17, 22 gradient, 93

Hausdorff, 26 hessian, 121 homeomorphism, 19, 22, 26, 136, 137 homeomorphism onto its image, 26

immersion, 162 immersive, 96, 162 infinitely differentiable, *see* smooth "infinitely flat" function, 71 "infinitely flat" functions, 69 interior extremum theorem, *see* extremum intermediate value property, 53, 54 intermediate value theorem, 53, 68 interval, 32 inverse function, 39, 62

jacobian matrix, 94

L<sup>2</sup> norm, 13, 16, 18 L'Hôpital's rule, 35, 69 linear approximation, 33  $L^{\infty}$  norm, see max norm local diffeomorphism, see diffeomorphism, local diffeomorphism  $\mathcal{L}(V, W), 17$ manifold, 134 max norm, 13, 16, 18 maximal atlas, see atlas, maximal atlas mean value theorem, 49, 50, 114, 120 Rolle's theorem, 49 minor, 19 mixed partials, 119 equality of mixed partials, 119  $M_{n \times m}$ , 16, 137 Möbius strip, 138 monotone, 51–53 decreasing, 51, 52 increasing, 51, 52 strictly decreasing, 51, 52 strictly increasing, 51, 52, 68 strictly monotone, 51, 52, 55 multi-index, 122 multinomial, 123 n-sphere, 137 node, 99 norm, 11 open open map, 26, 52, 133

open subset, 23 open cover, 27

Index

subcover, 27 operator norm, 20, 85, 116  $\mathbb{P}^1$ , see projective line  $\mathbb{P}^2$ , see projective plane partial derivative, 88 power rule, 31, 38, 46, 56 power series, 47 product product metric, 11, 12, 22 product topology, 24 product rule, 37, 87, 89 projection map, 15, 85 projective line, 141 projective plane, 141 pushforward, 160 quotient rule, 38, 88, 90 quotient space, 24 rank of a differentiable function, 96 of a linear map, 18 of a matrix, 16 of a smooth map, 162 regular point, 96, 162 remainder, 66 representation of a linear map, 17 of a vector, 17 standard matrix representation, 18, 22 Rolle's theorem, see mean value theorem

S<sup>1</sup>, *see* circle saddle, 90

scalar multiples rule, 37, 84 σ-compact, 27 smooth, 67, 128, 146, 148 smooth manifold, see manifold S<sup>n</sup>, *see* n-sphere standard basis, 15 standard matrix representation, see representation, standard matrix representation subcover, see open cover, subcover submersion, 162 submersive, 96, 162 submultiplicativity, 21 subspace, 23 subspace topology, 23 sum rule, 37, 84 support, 71 tangent space, 154 topological space, 23, 133 topology, 23 torus, 141 total derivative, see derivative transition function, 133