(1) The derivative as a linear transformation, (2) The inverse and implicit function theorems

Alan Chang

Mathcamp 2021 Week 4

Contents

1 Introduction

1.1 Course blurbs

1.1.1 The derivative as a linear transformation

Suppose we have a function $\mathbf{f} : \mathbb{R}^3 \to \mathbb{R}^2$. We can write $\mathbf{f}(x, y, z) = (g(x, y, z), h(x, y, z))$. In multivariable calculus, we learn about partial derivatives of $\mathbf{f}$. There are 6 different partial derivatives, which we can arrange into a matrix:

$$\begin{pmatrix} \partial_x g(x,y,z) & \partial_y g(x,y,z) & \partial_z g(x,y,z) \\ \partial_x h(x,y,z) & \partial_y h(x,y,z) & \partial_z h(x,y,z) \end{pmatrix} \tag{1.1}$$

If all the partial derivatives are continuous, then the matrix above is called the *total derivative* (or just *derivative*) of $\mathbf{f}$ and is denoted $\mathbf{f}'(x, y, z)$.

Arranging all the partial derivatives this way is not just for notational convenience. A 2×3 matrix represents a linear transformation $\mathbb{R}^3 \to \mathbb{R}^2$. In this class, we will see how to think of the derivatve of a $\mathbb{R}^m \to \mathbb{R}^n$ map as a linear transformation, and we will use this point of view to prove and interpret results such as the chain rule (in both single-variable and multivariable calculus).

Prerequisites: You should know the definition of the derivative from single-variable calculus. (You do need to know any multivariable calculus. Furthermore, this class does not overlap with Mark's multivariable class from Week 1.) You should be comfortable with matrix multiplication.

1.1.2 The inverse and implicit function theorems

If a function $f : \mathbb{R} \to \mathbb{R}$ satisfies $f'(x) \neq 0$ for all x, then the function f is invertible. In this class, we will look at a generalization of this to higher dimensions called the *inverse function theorem*: "If $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ is a function such that $\mathbf{f}'$ is continuous and $\det \mathbf{f}'(\mathbf{x}_0) \neq 0$, then $\mathbf{f}$ is locally a $\mathcal{C}^1$ homeomorphism near $\mathbf{x}_0$." (We will explain the precise meaning of this in class.)

It turns out the higher-dimensional situation is much harder than the one-dimensional situation. To understand the proof of the inverse function theorem, we will need tools such as the total derivative, linear algebra, and the Banach fixed-point theorem. We will also see a corollary of the inverse function theorem called the implicit function theorem, which allows us to describe solutions to system of equations as $\mathcal{C}^1$ submanifolds of Euclidean space.

Prerequisites: Week 2 "Introduction to analysis" and Week 4 "The derivative as a linear transformation" (or the equivalent to these classes)

1.2 References

Rudin, Principles of mathematical analysis, Chapter 9

(Almost any other analysis textbook that has a section on multivariable calculus should also cover this material.)

1.3 Guide for these notes

For the exercises, here is the difficulty scale:

-  : easy
-  : medium

-  : hard

The words "easy," "medium," and "hard" are not well-defined. Don't be afraid of difficult problems! It's by struggling with these exercises that you really learn.

Things labeled "Fun fact" are not needed for the class.

2 Day 1 (Derivative, Day 1)

2.1 Introduction

One goal of this class is to show how linear algebra is used to understand one of the most fundamental concepts in calculus: the derivative.

We will consider functions $\mathbf{f} : \mathbb{R}^m \to \mathbb{R}^n$. We will consider single-variable calculus and multivariable calculus. The "single-variable" in "single-variable calculus" refers to the case $m = 1$ (not to the case $n = 1$).

I assume everyone is already familiar with single-variable calculus. When we discuss single-variable calculus, the point is not to introduce new material, but to present material you already know in a new way. It may seem unnecessarily complicated to do this, but this new way of thinking is how you can generalize

If you have taken a non-proof-based multivariable calculus class, you have had this experience: The term "differentiable" was presented (maybe non-rigorously), but it was not emphasized, and then you quickly moved on to other topics.

2.2 Notation

Vectors will be written in boldface, e.g., "$\mathbf{x} \in \mathbb{R}^m$." When a function is written in boldface, it means its outputs are vectors, so we write "$f : \mathbb{R}^m \to \mathbb{R}$" and "$\mathbf{f} : \mathbb{R} \to \mathbb{R}^n$." (When handwriting in lecture, I use an arrow like in $\vec{x}$ instead of boldface.)

Unless otherwise stated, a "vector" is a column vector. Also, because writing actual column vectors in the middle of a sentence looks bad, when we write $(1, 3, 5)$, we mean the vector $\begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix}$.

A $n \times m$ matrix is a matrix with n rows and m columns. The set of $n \times m$ matrices is denoted $\mathbb{R}^{n \times m}$.

Example 2.1. $f(x) = x^2$. The name of the function is "f", so it is correct to say "f is a function." Often people will say "$f(x)$ is a function" but this is actually an abuse of notation. "$f(x)$" denotes the value of the function when the input is x.

Example 2.2. $T(x) = \begin{pmatrix} x & x^2 & x^3 \\ x^4 & x^5 & x^6 \end{pmatrix}$. Then

1. for each x, $T(x)$ is a linear transformation $\mathbb{R}^3 \to \mathbb{R}^2$.
2. for each x, $T(x) : \mathbb{R}^3 \to \mathbb{R}^2$.
3. $T : \mathbb{R} \to \mathbb{R}^{2\times 3}$.

However, if you interchange $T(x)$ and T in any of the statements above, they become false.

2.3 Linear and affine functions

We define linear functions.

Definition 2.3 ("Abstract" definition)**.** Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a function. We say T is *linear* if it satisfies $T(c\mathbf{x} + d\mathbf{y}) = cT(\mathbf{x}) + dT(\mathbf{y})$ for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$ and $c, d \in \mathbb{R}$.

Definition 2.4 ("Concrete" definition)**.** Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a function. We say T is *linear* if there exists a matrix $A \in \mathbb{R}^{n\times m}$ such that $T(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^m$.

A fundamental result from linear algebra (which we will not prove) is that the two definitions are equivalent.

Example 2.5. Here are examples of linear functions in special cases.

1. $f : \mathbb{R} \to \mathbb{R}$ is a linear function if and only if it is of the form $f(x) = ax$, where $a \in \mathbb{R}$.
2. $\mathbf{f} : \mathbb{R} \to \mathbb{R}^n$ is a linear function if and only if it is of the form $\mathbf{f}(x) = \mathbf{a}x$, where $\mathbf{a} \in \mathbb{R}^n$. (The expression $\mathbf{a}x$ denotes scalar multiplication, where x is the scalar and $\mathbf{a}$ is the vector.)
3. $f : \mathbb{R}^m \to \mathbb{R}$ is an linear function if and only if it is of the form $f(\mathbf{x}) = \mathbf{a} \cdot \mathbf{x}$, where $\mathbf{a} \in \mathbb{R}^m$. (The expression $\mathbf{a} \cdot \mathbf{x}$ denotes the dot product.)

Next we define affine functions.

Definition 2.6. Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a function. We say T is *affine* if there exists a $n \times m$ matrix A and a vector $b \in \mathbb{R}^n$ such that $T(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$ for all $\mathbf{x} \in \mathbb{R}^m$.

Remark 2.7. Warning: What we call "affine" is what most high school math classes call "linear."

The main theme of this class is:

$$\text{"Functions behave like their affine approximations."} \tag{2.1}$$

The statement is in quotes because it is not a rigorous mathematical statement but we will see several instances of this.

2.4 1-dimension

Definition 2.8. We give three definitions of the derivative of a function $\mathbb{R} \to \mathbb{R}$.

1. Let $f : \mathbb{R} \to \mathbb{R}$ and $x \in \mathbb{R}$. We say f is *differentiable* at x if

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \tag{2.2}$$

   exists. The number (2.2) is called the *derivative* of f at x and it is denoted $f'(x)$.

2. Let $f : \mathbb{R} \to \mathbb{R}$ and $x \in \mathbb{R}$. We say f is *differentiable* at x if there exists a number $a \in \mathbb{R}$ such that

$$\lim_{h \to 0} \frac{|f(x+h) - f(x) - ah|}{|h|} = 0 \tag{2.3}$$

   The number a is called the *derivative* of f at x and it is denoted $f'(x)$.

3. Let $f : \mathbb{R} \to \mathbb{R}$ and $x \in \mathbb{R}$. We say f is *differentiable* at x if there exists a number $a \in \mathbb{R}$ and a function $\mathsf{E} : \mathbb{R} \to \mathbb{R}$ such that

$$f(x+h) = f(x) + ah + \mathsf{E}(h) \text{ and } \lim_{h \to 0} \frac{|\mathsf{E}(h)|}{|h|} = 0. \tag{2.4}$$

   The number a is called the *derivative* of f at x and it is denoted $f'(x)$.

It is not hard to see that these three definitions Definition 2.8(1), Definition 2.8(2), and Definition 2.8(3) are equivalent. You are asked to do this Exercise 3.1. (Actually, we also discussed this in class.)

Let us discuss these three definitions. The first definition (Definition 2.8(1)) is what people usually see in a single-variable calculus class. It does not generalize to higher dimensions, so we will have very little use of this definition. The other two definitions (Definition 2.8(2) and Definition 2.8(3)) do generalize to higher dimensions.

Remark 2.9. We will work with the third definition (Definition 2.8(3)) a lot, so it is helpful to discuss it in detail.

This definition has the following interpretation: We think x as fixed and h as varying. We think of $h \mapsto f(x) + ah$ (as a function of h) as an "affine approximation" of $h \mapsto f(x+h)$. We can think of E as the "error" of this approximation.

The color coding in (2.4) is for the constant term, the linear term, and the error term. This will be used throughout these notes.

Let us look more at the condition

$$\lim_{h \to 0} \frac{|\mathsf{E}(h)|}{|h|} = 0. \tag{2.5}$$

This condition implies $\lim_{h\to 0} |\mathsf{E}(h)| = 0$, but the converse is not true. For example the function $\mathsf{E}(h) = h$ satisfies $\lim_{h\to 0} |\mathsf{E}(h)| = 0$ but not $\lim_{h\to 0} \frac{|\mathsf{E}(h)|}{|h|} = 0$.

The condition (2.5) means the error "becomes small very quickly as $h \to 0$" and it captures the geometric idea that "$f(x+h)$ and $f(x) + f'(x)h$ (viewed as functions of h) are tangent at $h = 0$."

Example 2.10. Consider the function $f(x) = (x+1)^2$. Let us consider an affine approximation at $x = 0$. We have

$$f(0+h) = f(h) = (h+1)^2 = 1 + 2h + h^2. \tag{2.6}$$

Note that the error term above satisfies $\lim_{h\to 0} \frac{|h^2|}{|h|} = \lim_{h\to 0} |h| = 0$. Thus, $1 + 2h$ is a good affine approximation, and we do indeed have $f'(0) = 2$.

If we had tried to write

$$f(h) = (h+1)^2 = 1 + h + (h + h^2). \tag{2.7}$$

instead, then the error term has the property $\lim_{h\to 0} \frac{|h+h^2|}{|h|} = \lim_{h\to 0} |1+h| = 1 \neq 0$. Thus, $1 + h$ is a bad affine approximation, and we have $f'(0) \neq 1$.

2.5 The sum rule

Theorem 2.11. *If $f, g : \mathbb{R} \to \mathbb{R}$ are differentiable at x, then so is $f + g$, and furthermore $(f+g)'(x) = f'(x) + g'(x)$.*

Proof. We may assume without loss of generality that $x = 0$. Suppose $f(0) = a, f'(0) = b, g(0) = c, g'(0) = d$. Then

$$f(h) = a + bh + \mathsf{E}_f(h) \tag{2.8}$$

$$g(h) = c + dh + \mathsf{E}_g(h) \tag{2.9}$$

Our goal is to determine $(f+g)'(0)$. By adding together the two equations above,

$$(f+g)(h) = (a+c) + (b+d)h + \mathsf{E}_f(h) + \mathsf{E}_g(h) \tag{2.10}$$

The "error" in this affine approximation satisfies

$$\lim_{h\to 0} \frac{|\mathsf{E}_f(h) + \mathsf{E}_g(h)|}{|h|} = 0 \tag{2.11}$$

which proves that $(f+g)'(0) = b + d = f'(0) + g'(0)$. □

2.6 The product rule

Theorem 2.12. *If $f, g : \mathbb{R} \to \mathbb{R}$ are differentiable at x, then so is fg, and furthermore $(fg)'(x) = f(x)g'(x) + f'(x)g(x)$.*

Proof. We may assume without loss of generality that $x = 0$. Suppose $f(0) = a, f'(0) = b, g(0) = c, g'(0) = d$. Then

$$f(h) = a + bh + \mathsf{E}_f(h) \tag{2.12}$$

$$g(h) = c + dh + \mathsf{E}_g(h) \tag{2.13}$$

Multiplying these two together

$$f(h)g(h) = [a + bh + \mathsf{E}_f(h)]\,[c + dh + \mathsf{E}_g(h)] \tag{2.14}$$

$$= ac + (ad + bc)h + bdh^2 + (c + dh)\mathsf{E}_f(h) + (a + bh)\mathsf{E}_g(h) + \mathsf{E}_f(h)\mathsf{E}_g(h) \tag{2.15}$$

It remains to show that the error term satisfies

$$\lim_{h\to 0} \frac{|bdh^2 + (c + dh)\mathsf{E}_f(h) + (a + bh)\mathsf{E}_g(h) + \mathsf{E}_f(h)\mathsf{E}_g(h)|}{|h|} = 0 \tag{2.16}$$

You are asked to do this in Exercise 3.2 □

2.7 Chain rule

Theorem 2.13 (Single-variable chain rule)**.** *Fix f, g, x. Suppose f is differentiable at x and g is differentiable at $f(x)$. Then $(g \circ f)'(x) = g'(f(x))f'(x)$.*

Proof, part 1. By definition of differentiability, we have

$$f(x + h) = f(x) + f'(x)h + \mathsf{E}_f(h) \tag{2.17}$$

$$g(f(x) + k) = g(f(x)) + g'(f(x))k + \mathsf{E}_g(k) \tag{2.18}$$

(Recall that x is fixed, and h and k vary.) For the rest of the proof, we let k be the following function of h:

$$k = f'(x)h + \mathsf{E}_f(h). \tag{2.19}$$

This gives

$$(g \circ f)(x + h) = g(f(x) + f'(x)h + \mathsf{E}_f(h)) \tag{2.20}$$

$$= g(f(x) + k) \tag{2.21}$$

$$= g(f(x)) + g'(f(x))k + \mathsf{E}_g(k) \tag{2.22}$$

$$= g(f(x)) + g'(f(x))(f'(x)h + \mathsf{E}_f(h)) + \mathsf{E}_g(k) \tag{2.23}$$

$$= g(f(x)) + g'(f(x))f'(x)h + g'(f(x))\mathsf{E}_f(h) + \mathsf{E}_g(k) \tag{2.24}$$

Now we show the error terms satisfies

$$\lim_{h \to 0} \frac{|g'(f(x))\mathsf{E}_f(h) + \mathsf{E}_g(k)|}{|h|} = 0 \tag{2.25}$$

You are asked to do this in Exercise 3.3. We will also do this tomorrow ☐

3 Day 1 exercises

Recommended: Exercise 3.2, Exercise 3.5

Exercise 3.1. ( ) Show that the three definitions of the derivative in Definition 2.8 are equivalent.

Exercise 3.2. ( ) Finish the proof of the product rule by proving (2.16). (Hint: Using the triangle inequality, break the expression up into several parts.)

Exercise 3.3. ( ) Finish the proof of the chain rule by proving (2.25). We will also do this tomorrow.

Exercise 3.4. ( ) Using Definition 2.8(3), show that for the function $f(x) = \frac{1}{x}$, we have $f'(x) = -\frac{1}{x^2}$.

Exercise 3.5. ( ) This exercise will be useful for tomorrow. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be

$$f(x,y) = \begin{cases} \dfrac{xy}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases} \tag{3.1}$$

(See a plot of the function here: https://www.wolframalpha.com/input/?i=plot+z%3Dxy%2F%28x%5E2%2By%5E2%29)

1. Let $\mathbf{e}_1 = (1,0), \mathbf{e}_2 = (0,1), \mathbf{0} = (0,0)$. Compute

$$\lim_{h \to 0} \frac{f(h\mathbf{e}_1) - f(\mathbf{0})}{h} \tag{3.2}$$

and

$$\lim_{h \to 0} \frac{f(h\mathbf{e}_2) - f(\mathbf{0})}{h} \tag{3.3}$$

2. Let $\mathbf{v} = (v_1, v_2) \in \mathbb{R}^2$ be any vector. Compute

$$\lim_{h \to 0} \frac{f(h\mathbf{v}) - f(\mathbf{0})}{h} \tag{3.4}$$

4 Day 2 (Derivative, Day 2)

4.1 A quick note on bounding the error terms

We have already seen a lot of "error terms" yesterday. You may get the impression that "analysis is all about bounding error terms" or "analysis has a lot of ugly error terms." But I would suggest that you wait until you have more experience with analysis before making these conclusions. It is true that in analysis, it is common to have error terms that you need to bound. But for many of these, there are standard techniques for dealing with them, so for someone experienced with these arguments, it is very clear how to bound the error terms that we have been dealing with.

See also this blog post by Terry Tao. https://terrytao.wordpress.com/career-advice/theres-more-to-mathematics-than-rigour-and-proofs/

4.2 Finishing the proof of the chain rule

Proof of Theorem 2.13, part 2. We need to show (2.25). First,

$$\frac{|g'(f(x))\mathsf{E}_f(h)|}{|h|} = |g'(f(x))| \cdot \frac{|\mathsf{E}_f(h)|}{|h|} \tag{4.1}$$

For the second term, we first note that

$$k = f'(x)h + \mathsf{E}_f(h) = f'(x)h + \frac{\mathsf{E}_f(h)}{h}h \qquad \text{so} \qquad \lim_{h\to 0} k = 0 \tag{4.2}$$

If $k = 0$, then $\frac{|\mathsf{E}_g(k)|}{|h|} = 0$. If $k \neq 0$, then

$$\frac{|\mathsf{E}_g(k)|}{|h|} = \frac{|\mathsf{E}_g(k)|}{|k|} \cdot \frac{|f'(x)h + \mathsf{E}_f(h)|}{|h|} \leq \frac{|\mathsf{E}_g(k)|}{|k|} \left(|f'(x)| + \frac{|\mathsf{E}_f(h)|}{|h|} \right) \tag{4.3}$$

Thus

$$\lim_{h\to 0} \frac{|\mathsf{E}_g(k)|}{|h|} = 0 \tag{4.4}$$

□

4.3 The case $\mathbb{R} \to \mathbb{R}^n$

(We did not cover this section in class.)

Definition 4.1. We give three definitions of the derivative of a function $\mathbb{R} \to \mathbb{R}^n$.

1. Let $\mathbf{f} : \mathbb{R} \to \mathbb{R}^n$ and $x \in \mathbb{R}$. We say $\mathbf{f}$ is *differentiable* at x if

$$\lim_{h \to 0} \frac{\mathbf{f}(x+h) - \mathbf{f}(x)}{h} \tag{4.5}$$

exists. The vector (4.5) is called the *derivative* of $\mathbf{f}$ at x and it is denoted $\mathbf{f}'(x)$.

2. Let $\mathbf{f} : \mathbb{R} \to \mathbb{R}^n$ and $x \in \mathbb{R}$. We say $\mathbf{f}$ is *differentiable* at x if there exists a vector $a \in \mathbb{R}^n$ such that

$$\lim_{h \to 0} \frac{|\mathbf{f}(x+h) - \mathbf{f}(x) - \mathbf{a}h|}{|h|} = 0 \tag{4.6}$$

The vector $\mathbf{a}$ is called the *derivative* of $\mathbf{f}$ at x and it is denoted $\mathbf{f}'(x)$.

3. Let $\mathbf{f} : \mathbb{R} \to \mathbb{R}^n$ and $x \in \mathbb{R}$. We say $\mathbf{f}$ is *differentiable* at x if there exists a vector $a \in \mathbb{R}^n$ and a function $\mathsf{E} : \mathbb{R} \to \mathbb{R}^n$ such that

$$\mathbf{f}(x+h) = \mathbf{f}(x) + \mathbf{a}h + \mathsf{E}(h) \text{ and } \lim_{h \to 0} \frac{|\mathsf{E}(h)|}{|h|} = 0. \tag{4.7}$$

The vector $\mathbf{a}$ is called the *derivative* of $\mathbf{f}$ at x and it is denoted $\mathbf{f}'(x)$.

Theorem 4.2. *Let $\mathbf{f} : \mathbb{R} \to \mathbb{R}^n$ and $x \in \mathbb{R}$. Then $\mathbf{f}$ is differentiable at x if and only if f_i is differentiable at x for each i, and furthermore,*

$$\mathbf{f}'(x) = (f_1'(x), \ldots, f_n'(x)). \tag{4.8}$$

4.4 The case $\mathbb{R}^m \to \mathbb{R}$

Now we begin multivariable calculus. Let $\mathbf{e}_1, \ldots, \mathbf{e}_m$ be the standard basis of $\mathbb{R}^m$. (So $\mathbf{e}_1 = (1, 0, 0, \ldots, 0)$, $\mathbf{e}_2 = (0, 1, 0, \ldots, 0)$, etc.)

Definition 4.3. Let $f : \mathbb{R}^m \to \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^m$. For $j \in \{1, \ldots, m\}$, the jth partial derivative of f at $\mathbf{x}$ is

$$\partial_j f(\mathbf{x}) = \lim_{t \to 0} \frac{f(\mathbf{x} + t\mathbf{e}_j) - f(\mathbf{x})}{t}. \tag{4.9}$$

In other words, it is the derivative of f, when viewed as a function of only x_j (and all other variables kept fixed).

Example 4.4. If $f(x_1, x_2) = x_1 x_2^2 + x_2$, then $\partial_1 f(x_1, x_2) = x_2^2$, and $\partial_2 f(x_1, x_2) = 2x_1 x_2 + 1$.

If you have studied multivariable calculus, then you have definitely seen the definition above. It might be tempting to define "f is differentiable at $\mathbf{x}$" to mean "all the partial derivatives $\partial_j f$ exists at $\mathbf{x}$," but this is not how things are done. We will see why later. The actual definition of differentiability makes no mention of partial derivatives.

Definition 4.5. We give two definitions of the derivative of a function $\mathbb{R}^m \to \mathbb{R}$.

1. Let $f : \mathbb{R}^m \to \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^m$. We say f is *differentiable* at $\mathbf{x}$ if there exists a vector $\mathbf{a} \in \mathbb{R}^m$ such that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{|f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) - \mathbf{a}\cdot\mathbf{h}|}{|\mathbf{h}|} = 0 \tag{4.10}$$

   The vector $\mathbf{a}$ is called the *gradient* of f at $\mathbf{x}$ and it is denoted $\nabla f(\mathbf{x})$.

2. Let $f : \mathbb{R}^m \to \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^m$. We say f is *differentiable* at $\mathbf{x}$ if there exists a vector $\mathbf{a} \in \mathbb{R}^m$ and a function $\mathsf{E} : \mathbb{R}^m \to \mathbb{R}$ such that

$$f(\mathbf{x}+\mathbf{h}) = f(\mathbf{x}) + \mathbf{a}\cdot\mathbf{h} + \mathsf{E}(\mathbf{h}) \text{ and } \lim_{\mathbf{h}\to\mathbf{0}} \frac{|\mathsf{E}(\mathbf{h})|}{|\mathbf{h}|} = 0. \tag{4.11}$$

   The vector $\mathbf{a}$ is called the *gradient* of f at $\mathbf{x}$ and it is denoted $\nabla f(\mathbf{x})$.

Remark 4.6. Recall $\lim_{\mathbf{h}\to\mathbf{0}} \frac{|\mathsf{E}(\mathbf{h})|}{|\mathbf{h}|} = 0$ means:

$$\forall \varepsilon > 0, \exists \delta > 0, \forall \mathbf{h} \in \mathbb{R}^m, \text{ if } |\mathbf{h}| < \delta \text{ then } \frac{|\mathsf{E}(\mathbf{h})|}{|\mathbf{h}|} < \varepsilon \tag{4.12}$$

We do not need this ε, δ definition in this class.

4.5 The relationship between ∇f and $\partial_j f$

Definition 4.5 does not tell you how to actually calculate ∇f. The following theorem does that.

Theorem 4.7. *If $f : \mathbb{R}^m \to \mathbb{R}$ is differentiable at $\mathbf{x}$, then $\nabla f(\mathbf{x}) = (\partial_1 f(\mathbf{x}), \dots, \partial_m f(\mathbf{x}))$.*

Proof. Fix f and $\mathbf{x}$. Let $\mathbf{a} = (a_1, \dots, a_m) = \nabla f(\mathbf{x})$. Recall that $\mathbf{e}_1, \dots, \mathbf{e}_m$ is the standard basis of $\mathbb{R}^m$. This implies $\mathbf{a}\mathbf{e}_j = a_j$

From the definition of differentiability, we know

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{|f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) - \mathbf{a}\cdot\mathbf{h}|}{|\mathbf{h}|} = 0 \tag{4.13}$$

which implies (by letting $\mathbf{h} = t\mathbf{e}_j$)

$$\lim_{t\to 0} \frac{|f(\mathbf{x}+t\mathbf{e}_j) - f(\mathbf{x}) - a_j t|}{|t|} = 0 \tag{4.14}$$

which is precisely the statement that $\partial_j f(\mathbf{x}) = a_j$. □

Warning: Theorem 4.7 does *not* say: "If the partial derivatives $\partial_1 f, \dots, \partial_m f$ all exist at $\mathbf{x}$, then f is differentiable at $\mathbf{x}$." It also does not say: "If the directional derivatives $\partial_{\mathbf{v}} f$ all exist at $\mathbf{x}$, then f is differentiable at $\mathbf{x}$." That statement is **FALSE**. See Exercise 5.3 and Exercise 5.4. If you plot those examples, you will see that there is no good tangent plane approximation to those functions at the origin.

Here is a fact which we will not prove.

Theorem 4.8. *Let $f : \mathbb{R}^m \to \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^m$. If the partial derivatives $\partial_1 f, \dots, \partial_m f$ all exist and **are continuous at** $\mathbf{x}$, then f is differentiable at $\mathbf{x}$.*

4.6 The general case $\mathbb{R}^m \to \mathbb{R}^n$

Definition 4.9. We give two definitions of the derivative of a function $\mathbb{R}^m \to \mathbb{R}^n$.

1. Let $\mathbf{f} : \mathbb{R}^m \to \mathbb{R}^n$ and $\mathbf{x} \in \mathbb{R}^m$. We say $\mathbf{f}$ is *differentiable* at $\mathbf{x}$ if there exists a matrix $A \in \mathbb{R}^{n \times m}$ such that

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{|\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x}) - A\mathbf{h}|}{|\mathbf{h}|} = 0 \tag{4.15}$$

   The matrix A is called the *derivative* (or *total derivative*) of $\mathbf{f}$ at $\mathbf{x}$ and it is denoted $\mathbf{f}'(\mathbf{x})$.

2. Let $\mathbf{f} : \mathbb{R}^m \to \mathbb{R}^n$ and $\mathbf{x} \in \mathbb{R}^m$. We say $\mathbf{f}$ is *differentiable* at $\mathbf{x}$ if there exists a matrix $A \in \mathbb{R}^{n \times m}$ and a function $\mathsf{E} : \mathbb{R}^m \to \mathbb{R}^n$ such that

$$\mathbf{f}(\mathbf{x}+\mathbf{h}) = \mathbf{f}(\mathbf{x}) + A\mathbf{h} + \mathsf{E}(\mathbf{h}) \text{ and } \lim_{\mathbf{h} \to \mathbf{0}} \frac{|\mathsf{E}(\mathbf{h})|}{|\mathbf{h}|} = 0. \tag{4.16}$$

   The matrix A is called the *derivative* (or *total derivative*) of $\mathbf{f}$ at $\mathbf{x}$ and it is denoted $\mathbf{f}'(\mathbf{x})$.

We have defined the derivative of $\mathbf{f}$. This definition does not say anything about what the entries are.

Theorem 4.10. *If $\mathbf{f} : \mathbb{R}^m \to \mathbb{R}^n$ is differentiable at $\mathbf{x}$, then entry in row i column j of the matrix $\mathbf{f}'(\mathbf{x})$ is $\partial_j f_i(\mathbf{x})$.*

Proof. The proof is similar to the proof of Theorem 4.7. □

The following is the analogue of Theorem 4.8 (also stated without proof).

Theorem 4.11. *Let $\mathbf{f} : \mathbb{R}^m \to \mathbb{R}^n$. Suppose for every i and j, the partial derivative $\partial_i f_j$ exists and is a continuous function. Then $\mathbf{f}$ is differentiable.*

Example 4.12. If $\mathbf{f} : \mathbb{R}^m \to \mathbb{R}$, then $f'(\mathbf{x})$ is a row vector. It is equal to the transpose of $\nabla f(\mathbf{x})$, which is a column vector.

Example 4.13. Let $\mathbf{g} : \mathbb{R}^3 \to \mathbb{R}^2$. We can write

$$\mathbf{g}(\mathbf{x}) = \mathbf{g}(x_1, x_2, x_3) = \begin{pmatrix} g_1(x_1, x_2, x_3) \\ g_2(x_1, x_2, x_3) \end{pmatrix} \tag{4.17}$$

Then

$$\mathbf{g}'(\mathbf{x}) = \begin{pmatrix} \partial_1 g_1(x_1, x_2, x_3) & \partial_2 g_1(x_1, x_2, x_3) & \partial_3 g_1(x_1, x_2, x_3) \\ \partial_1 g_2(x_1, x_2, x_3) & \partial_2 g_2(x_1, x_2, x_3) & \partial_3 g_2(x_1, x_2, x_3) \end{pmatrix} \tag{4.18}$$

4.7 Multivariable chain rule

Theorem 4.14 (Multivariable chain rule)**.** *Fix $\mathbf{f} : \mathbb{R}^\ell \to \mathbb{R}^m$, $\mathbf{g} : \mathbb{R}^m \to \mathbb{R}^n$ and $\mathbf{x} \in \mathbb{R}^\ell$. Suppose $\mathbf{f}$ is differentiable at $\mathbf{x}$ and $\mathbf{g}$ is differentiable at $\mathbf{f}(\mathbf{x})$. Then $(\mathbf{g} \circ \mathbf{f})'(x) = \mathbf{g}'(\mathbf{f}(\mathbf{x}))\mathbf{f}'(\mathbf{x})$.*

Remark 4.15. Note that $\mathbf{g}'(\mathbf{f}(\mathbf{x}))\mathbf{f}'(\mathbf{x})$ denotes matrix multiplication and that $\mathbf{g}'(\mathbf{f}(\mathbf{x})) \in \mathbb{R}^{n\times m}$ and $\mathbf{f}'(\mathbf{x}) \in \mathbb{R}^{m\times\ell}$, so the matrix multiplication is possible.

Proof sketch. Everything in the proof of the single-variable chain rule (Theorem 2.13) works until line (2.24). (Typographically, we only need to change everything to vectors.) For the part afterwards, a few parts need to be changed to use something called the "operator norm." Exercise 5.5 defines the operator norm and asks you to work out the proof carefully. □

Example 4.16. Let $g : \mathbb{R}^2 \to \mathbb{R}$ and $\mathbf{f} : \mathbb{R} \to \mathbb{R}^2$ be a function. We have $(g \circ \mathbf{f})(t) = g(f_1(t), f_2(t))$. We have

$$g'(x_1, x_2) = \begin{pmatrix} \partial_1 g(x_1, x_2) & \partial_2 g(x_1, x_2) \end{pmatrix} \qquad \mathbf{f}'(t) = \begin{pmatrix} f_1'(t) \\ f_2'(t) \end{pmatrix} \tag{4.19}$$

By the multivariable chain rule Theorem 4.14, we have

$$(g \circ \mathbf{f})'(t) = \begin{pmatrix} \partial_1 g(\mathbf{f}(t)) & \partial_2 g(\mathbf{f}(t)) \end{pmatrix} \begin{pmatrix} f_1'(t) \\ f_2'(t) \end{pmatrix} = \partial_1 g(\mathbf{f}(t)) f_1'(t) + \partial_2 g(\mathbf{f}(t)) f_2'(t). \tag{4.20}$$

This is a form of the chain rule that is more often taught in multivariable calculus classes. This is more commonly written

$$\frac{dg}{dt} = \frac{\partial g}{\partial x}\frac{dx}{dt} + \frac{\partial g}{\partial y}\frac{dy}{dt} \tag{4.21}$$

where the left-hand side is understood to mean $\frac{d}{dt}[g(x(t), y(t))]$. See Exercise 5.6 for another example of the chain rule.

4.8 Some applications

4.8.1 Inverse function theorem

If $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ is an affine function $f(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$, then $\mathbf{f}$ is invertible if and only if A is an invertible matrix.

If $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ is any $\mathcal{C}^1$ function, then the invertibility of $\mathbf{f}'$ at a particular point $\mathbf{x}_0$ tells us something about the invertible of $\mathbf{f}$ near $\mathbf{x}_0$.

We will see more about this in the follow-up class, starting tomorrow.

4.8.2 Change of variables

If $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ is an affine function $f(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$ and $S \subset \mathbb{R}^n$, then

$$(\text{volume of } f(S)) = |\det A|(\text{volume of } S). \tag{4.22}$$

This motivates the change of variables formula. If $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$, and $\mathbf{y} = \mathbf{f}(\mathbf{x})$, then

$$dy_1 \cdots dy_n = \det \mathbf{f}'(\mathbf{x})\, dx_1 \cdots dx_n \tag{4.23}$$

or put another way,

$$\int_{f(S)} g(\mathbf{y})\, dy_1 \cdots dy_n = \int_S g(\mathbf{f}(\mathbf{x})) \det \mathbf{f}'(\mathbf{x})\, dx_1 \cdots dx_n \tag{4.24}$$

5 Day 2 exercises

Recommended exercises: Exercise 5.1 and Exercise 5.6

Exercise 5.1. ( ) Let $\mathbf{g} : \mathbb{R}^2 \to \mathbb{R}^2$ be $g(x_1, x_2) = (x_1^2 - x_2^2, 2x_1x_2)$. Compute $\mathbf{g}'(\mathbf{x})$.

Fun fact: Note that $\mathbf{g}'(\mathbf{x})$ is of the form

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} \tag{5.1}$$

This is related to Exercise 5.7.

Exercise 5.2. ( ) For $f : \mathbb{R}^m \to \mathbb{R}$ and any vector $\mathbf{v} \in \mathbb{R}^m$, define the directional derivative

$$\partial_{\mathbf{v}} f(\mathbf{x}) = \lim_{t \to 0} \frac{f(\mathbf{x} + t\mathbf{v}) - f(\mathbf{x})}{t}. \tag{5.2}$$

(Sometimes people restrict $\mathbf{v}$ to be a unit vector. We don't need to do that here.) Show that if f is differentiable at $\mathbf{x}$, then $\partial_{\mathbf{v}} f(\mathbf{x}) = \nabla f(\mathbf{x}) \cdot \mathbf{v}$.

Exercise 5.3. () Show that the function in Exercise 3.5 is not differentiable at $(0, 0)$. (Hint: Use Exercise 5.2.)

Exercise 5.4. () This exercise will be useful for tomorrow. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be

$$f(x,y) = \begin{cases} \dfrac{x^2 y}{x^4 + y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases} \tag{5.3}$$

(See a plot of the function here: https://www.wolframalpha.com/input/?i=plot+z%3Dx%5E2y%2F%28x%5E4%2By%5E2%29)

1. Let $\mathbf{v} = (v_1, v_2) \in \mathbb{R}^2$ be any vector. Compute

$$\lim_{h \to 0} \frac{f(h\mathbf{v}) - f(\mathbf{0})}{h} \tag{5.4}$$

2. Show that the function in Exercise 3.5 is not differentiable at $(0, 0)$. (Note that unlike in Exercise 5.3, here, Exercise 5.2 does not help.)

Exercise 5.5. () This problem asks you to prove the multivariable chain rule Theorem 4.14. As stated in Section 4.7, we can adapt the proof of the single-variable chain rule. We provide some ideas here.

One tool we need is the operator norm of a matrix. For $A \in \mathbb{R}^{n \times m}$, we define the *operator norm* of A to be

$$\|A\| = \max\{|A\mathbf{v}| : \mathbf{v} \in \mathbb{R}^m, |\mathbf{v}| = 1\} \tag{5.5}$$

We can show that

$$|A\mathbf{v}| \leq \|A\| \, |\mathbf{v}| \text{ for all } A \in \mathbb{R}^{n \times m}, \mathbf{v} \in \mathbb{R}^m \tag{5.6}$$

which is a suitable substitute of the fact $|ab| \leq |a| \, |b|$ for $a, b \in \mathbb{R}$.

We can use (6.9) to obtain multivariable analogues of (4.1), (4.2), and (4.3). For example, the analogue of line (4.1) is

$$\frac{|\mathbf{g}'(\mathbf{f}(\mathbf{x}))\mathsf{E}_{\mathbf{f}}(\mathbf{h})|}{|\mathbf{h}|} \leq \|\mathbf{g}'(\mathbf{f}(\mathbf{x}))\| \frac{|\mathsf{E}_{\mathbf{f}}(\mathbf{h})|}{|\mathbf{h}|} \tag{5.7}$$

and the analogue of line (4.2) is

$$|\mathbf{k}| = |\mathbf{f}'(\mathbf{x})\mathbf{h} + \mathsf{E}_{\mathbf{f}}(\mathbf{h})| \leq \|\mathbf{f}'(\mathbf{x})\| \, |\mathbf{h}| + \frac{|\mathsf{E}_{\mathbf{f}}(\mathbf{h})|}{|\mathbf{h}|}|\mathbf{h}| \qquad \text{so} \qquad \lim_{\mathbf{h} \to \mathbf{0}} \mathbf{k} = \mathbf{0} \tag{5.8}$$

Line (4.3) also needs an analogue. We leave that as part of the exercise.

Exercise 5.6. ( ) Suppose g is a function of x, y, z, and suppose x, y, z are all functions of s and t. Show that the multivariable chain rule Theorem 4.14 implies

$$\frac{\partial g}{\partial s} = \frac{\partial g}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial g}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial g}{\partial z}\frac{\partial z}{\partial s} \tag{5.9}$$

$$\frac{\partial g}{\partial t} = \frac{\partial g}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial g}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial g}{\partial z}\frac{\partial z}{\partial t} \tag{5.10}$$

Exercise 5.7. ( ) By identifying $\mathbb{C}$ with $\mathbb{R}^2$, we can define a bijection between the set of functions $\mathbb{C} \to \mathbb{C}$ and the set of functions $\mathbb{R}^2 \to \mathbb{R}^2$ as follows: Given a function $f : \mathbb{C} \to \mathbb{C}$, define $\mathbf{g} : \mathbb{R}^2 \to \mathbb{R}^2$ by $\mathbf{g}(x, y) = (\operatorname{Re} f(x + iy), \operatorname{Im} f(x + iy))$.

1. We say a function $f : \mathbb{C} \to \mathbb{C}$ is *complex linear* if there exists $a \in \mathbb{C}$ such that $f(z) = az$. Show that the corresponding $\mathbf{g} : \mathbb{R}^2 \to \mathbb{R}^2$ is linear, and find the 2×2 matrix for $\mathbf{g}$. Give a geometric interpretation for this matrix.

2. Let $f : \mathbb{C} \to \mathbb{C}$. We say f is *complex differentiable* at z if there exists a $a \in \mathbb{C}$ such that $f(z + h) = f(z) + ah + \mathsf{E}(h)$, where $\lim_{h\to 0} \frac{|\mathsf{E}(h)|}{|h|} = 0$. (Here, $h \in \mathbb{C}$.) The number a is called the derivative of f at z and is denoted $f'(z)$.

   If f is complex differentiable at $z = x_1 + ix_2 \in C$, show that the corresponding $\mathbf{g} : \mathbb{R}^2 \to \mathbb{R}^2$ satisfies the *Cauchy–Riemann equations* at (x_1, x_2):

$$\partial_1 g_1(x_1, x_2) = \partial_2 g_2(x_1, x_2) \tag{5.11}$$

$$\partial_1 g_2(x_1, x_2) = -\partial_2 g_1(x_1, x_2) \tag{5.12}$$

6 Day 3 (Inverse/Implicit, Day 1)

6.1 Continuously differentiable functions

In this class, we will focus exclusively on $\mathcal{C}^1$ functions.

Definition 6.1. A function $\mathbf{f} : \mathbb{R}^m \to \mathbb{R}^n$ is $\mathcal{C}^1$ (or *continuously differentiable*) if for every i and j, the partial derivative $\partial_i f_j$ exists and is a continuous function.

Recall from Theorem 4.11 that $\mathcal{C}^1$ functions are differentiable.

Definition 6.2. Let $U, V \subset \mathbb{R}^n$ be two open sets. We say $f : U \to V$ is a $\mathcal{C}^1$ *homeomorphism* (or $\mathcal{C}^1$ *diffeomorphism*) if f is a bijection, f is $\mathcal{C}^1$, and $f^{-1} : V \to U$ is $\mathcal{C}^1$.

Example 6.3. Here are some examples and non-examples.

1. For any $a, b \in \mathbb{R}$ with $b \neq 0$, the function $f(x) = a + bx$ is a $\mathcal{C}^1$ homeomorphism between $(0, 1)$ and $(a, a + b)$.

2. $f : (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R}$ given by $f(x) = \tan x$ is a $\mathcal{C}^1$ homeomorphism.

3. $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^3$: f is $\mathcal{C}^1$ and a bijection. However, f^{-1} is not $\mathcal{C}^1$. (The derivative of $f^{-1}(y) = y^{1/3}$ is not defined at $y = 0$.) Thus, f is not a $\mathcal{C}^1$ homeomorphism.

4. $f : (0, 1) \to (0, 1)$ given by $f(x) = x^3$ is a $\mathcal{C}^1$ homeomorphism. We removed the problematic point $x = 0$.

6.2 The inverse function theorem in one dimension

Here is a 1-dimensional inverse function theorem.

Theorem 6.4 (1-dimensional "global" inverse function theorem)**.** *Let $f : I \to \mathbb{R}$ be a $\mathcal{C}^1$ function. Suppose that $f'(x) \neq 0$ for all $x \in I$. Then $f(I)$ is an open interval, and f is a bijection between I and $f(I)$. Furthermore, the inverse map $g = f^{-1} : f(I) \to I$ satisfies*

$$g'(y) = \frac{1}{f'(g(y))} \tag{6.1}$$

Proof. We break the proof down into several steps. This theorem is not the main focus of the class, so we will leave some of the steps as exercises.
Step 1: First, we show f is injective. Because f' is continuous and $f'(x) \neq 0$ for all $x \in I$, we know either $f'(x) > 0$ for all x or $f'(x) < 0$ for all x. Without loss of generality, assume $f'(x) > 0$ for all x. This implies that f is strictly increasing on I, so f is injective.
Step 2: Next, we show $f(I)$ is open. Actually let's leave this as an exercise. See Exercise 7.1.
Step 3: Next, we show $g = f^{-1}$ is continuous on $f(I)$. Actually, let's leave this as an exercise too. See Exercise 7.1.
Step 4: Finally, we show (6.1). Fix $y \in f(I)$. Let $x = g(y)$.

We let h and k vary, but we relate the two by

$$h = g(y + k) - g(y) \tag{6.2}$$
$$k = f(x + h) - f(x) \tag{6.3}$$

(It helps to draw a picture. If we graph f, then k is the "rise" and h is the "run.") Note that

$$\lim_{k\to 0} h = \lim_{k\to 0}[g(y + k) - g(y)] = 0 \tag{6.4}$$

by the continuity of g. Thus,

$$g'(y) = \lim_{k\to 0} \frac{g(y + k) - g(y)}{k} = \lim_{k\to 0} \frac{h}{f(x + h) - f(x)} = \lim_{h\to 0} \frac{h}{f(x + h) - f(x)} = \frac{1}{f'(x)}. \tag{6.5}$$

□

Remark 6.5. The assumption that f' is continuous is unnecessary in Theorem 6.4. See Exercise 7.2

Here is a "local version" of Theorem 6.4.

Corollary 6.6 (1-dimensional "local" inverse function theorem)**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be a $\mathcal{C}^1$ function and $x_0 \in \mathbb{R}$. Suppose that $f'(x_0) \neq 0$. Then there exists an open interval $I \subset \mathbb{R}$ containing x_0 such that f is a $\mathcal{C}^1$ homeomorphism between I and $f(I)$.*

Proof. If $f'(x_0) \neq 0$ and f' is continuous, there exists an open interval I containing x_0 such that $f'(x) \neq 0$ for all $x \in I$. Now we can apply Theorem 6.4. □

Here is another way to state Corollary 6.6: If $f : \mathbb{R} \to \mathbb{R}$ is $\mathcal{C}^1$, then $f'(x_0) \neq 0$ implies that f is a *local $\mathcal{C}^1$ homeomorphism near x_0*. Here, the key word is "local."

Remark 6.7. Unlike Remark 6.5, in Corollary 6.6, it is important to require f to be $\mathcal{C}^1$ and not only differentiable. Exercise 7.3 asks you for a counterexample.

6.3 The statement of the inverse function theorem in higher dimensions

We will generalize Corollary 6.6 to higher dimensions.

Example 6.8. Consider $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x, y) = (e^x \cos y, e^x \sin y)$. Then we compute

$$\mathbf{f}'(x,y) = \begin{pmatrix} \partial_x f_1 & \partial_y f_1 \\ \partial_x f_2 & \partial_y f_2 \end{pmatrix} = \begin{pmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{pmatrix}, \tag{6.6}$$

so $\det \mathbf{f}'(x, y) = e^{2x} \neq 0$ for all $(x, y) \in \mathbb{R}^2$.

The inverse function theorem is another statement along the ideas of "functions behave like their affine approximations" (2.1). In this case, it is "If the affine approximation to $\mathbf{f}$ at $\mathbf{a}$ is invertible, then $\mathbf{f}$ is locally invertible near $\mathbf{a}$."

Theorem 6.9 (Inverse function theorem)**.** *Let $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ be a $\mathcal{C}^1$ function, and suppose $\det \mathbf{f}'(\mathbf{a}) \neq 0$. Then there exists an open set $U \subset \mathbb{R}^n$ such that $\mathbf{a} \in U$, and such that $\mathbf{f}$ is a $\mathcal{C}^1$ homeomorphism between U and $\mathbf{f}(U)$. Furthermore, the inverse map $\mathbf{g} = \mathbf{f}^{-1} : \mathbf{f}(U) \to U$ satisfies*

$$\mathbf{g}'(\mathbf{y}) = \mathbf{f}'(\mathbf{g}(\mathbf{y}))^{-1}. \tag{6.7}$$

Before we begin the proof of the inverse function theorem, we'll start by presenting some useful tools in the proof. Knowing these tools may even more helpful than knowing the proof of the theorem. These tools are:

1. Operator norm of a matrix
2. Mean value inequality
3. Banach fixed point theorem

6.4 Operator norm

For $A \in \mathbb{R}^{n\times m}$, we define the *operator norm* of A to be

$$\|A\| = \max\{|A\mathbf{v}| : \mathbf{v} \in \mathbb{R}^m, |\mathbf{v}| = 1\} \tag{6.8}$$

Note that for any $A \in \mathbb{R}^{n\times m}$, $\|A\|$ is finite by the extreme value theorem. (The linear function $A : \mathbb{R}^m \to \mathbb{R}^n$ is continuous, and the set $\{\mathbf{v} \in \mathbb{R}^m : |\mathbf{v}| = 1\}$ is compact.) The extreme value theorem also explains why we can write max instead of sup.

The key property of the operator norm we need is the following

$$|A\mathbf{v}| \leq \|A\| \, |\mathbf{v}| \text{ for all } A \in \mathbb{R}^{n\times m}, \mathbf{v} \in \mathbb{R}^m \tag{6.9}$$

You are asked to prove this in Exercise 7.4. This is a useful way to bound expressions of the form $|A\mathbf{v}|$.

7 Day 3 exercises

Recommended exercises: Exercise 7.4, Exercise 7.8

Exercise 7.3 is fun if you like counterexamples in analysis.

Exercise 7.1. ( ) Fill in steps 2 and 3 of the proof of Theorem 6.4. (Hints: For step 2, show that $f(I) = (\inf_I f, \sup_I f)$. For step 3, use the fact that f is continuous and strictly increasing.)

Exercise 7.2. ( )

1. Let $f : \mathbb{R} \to \mathbb{R}$ be a differentiable function. (We do not assume that f is $\mathcal{C}^1$.) Show that f' has the intermediate value property: If $a < b$, and s is between $f'(a)$ and $f'(b)$, then there exists $c \in [a, b]$ such that $f'(c) = s$.

2. Use the previous part to show that Theorem 6.4 is still true if we replace "f is a $\mathcal{C}^1$ function" with "f is a differentiable function."

Exercise 7.3. Give an example of a differentiable function $f : \mathbb{R} \to \mathbb{R}$ such that $f'(0) \neq 0$ but there does *not* exist an open interval I containing 0 such that f is a $\mathcal{C}^1$ homeomorphism between I and $f(I)$.

Exercise 7.4. Prove the inequality involving the operator norm (6.9).

Exercise 7.5. Suppose $A \in \mathbb{R}^{1\times n}$ is a row vector, i.e., A is the transpose of some column vector $\mathbf{u} \in \mathbb{R}^n$. Show that $\|A\| = |u|$. (Hint: This is related to a well-known inequality involving the dot product of two vectors.)

Exercise 7.6. Let $c \in \mathbb{R}$ and let

$$A = \begin{pmatrix} 0 & c \\ 0 & 0 \end{pmatrix} \tag{7.1}$$

Determine $\|A\|$. (Note that the eigenvalues of A are 0 and 0. This exercise shows that the eigenvalues do not determine $\|A\|$.)

Exercise 7.7. (This problem requires some linear algebra.) Let $A \in \mathbb{R}^{n\times m}$. (Note that A does not need to be a square matrix.) The *singular values* of A are the square roots of the eigenvalues of $A^T A$, where A^T denotes the transpose of A. Show that $\|A\|$ is equal to the largest singular value of A.

See https://en.wikipedia.org/wiki/Singular_value_decomposition for more information on singular values.

Exercise 7.8. (This problem is relevant for tomorrow.) Here is a version of the mean value theorem from single-variable calculus:

- **Single-variable mean value theorem.** Let $f : \mathbb{R} \to \mathbb{R}$ be a $\mathcal{C}^1$ function. For any $a, b \in \mathbb{R}$ with $a < b$, there exists a $c \in (a, b)$ such that

$$f(b) - f(a) = f'(c)(b - a). \tag{7.2}$$

Show that the theorem above is false if you replace $f : \mathbb{R} \to \mathbb{R}$ with $\mathbf{f} : \mathbb{R} \to \mathbb{R}^2$.

8 Day 4 (Inverse/Implicit, Day 2)

8.1 Mean value inequality

For $\mathbb{R}^m \to \mathbb{R}$, we have the mean value theorem (see Exercise 9.1). However, for $\mathbb{R}^m \to \mathbb{R}^n$ and $n \geq 2$, the mean value theorem does not hold (see Exercise 7.8). That is why we use the following instead.

Theorem 8.1 (Mean value inequality)**.** *Let* $\mathbf{f} : \mathbb{R}^m \to \mathbb{R}^n$ *be a* $\mathcal{C}^1$ *function. Let* $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$*. Suppose that for all* $\mathbf{c} \in \mathbb{R}^m$ *on the line segment joining* $\mathbf{a}$ *and* $\mathbf{b}$*, we have* $\|\mathbf{f}'(\mathbf{c})\| \leq M$*. Then*

$$|\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a})| \leq M|\mathbf{b} - \mathbf{a}| \tag{8.1}$$

Proof. Define $\gamma : \mathbb{R} \to \mathbb{R}^m$ by $\gamma(t) = (1-t)\mathbf{a}+t\mathbf{b}$. Let $g(t) = \mathbf{f}(\gamma(t))$, so by the (multivariable) chain rule Theorem 4.14, $g'(t) = \mathbf{f}'(\gamma(t))\gamma'(t) = \mathbf{f}'(\gamma(t))(y - x)$. Then

$$\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a}) = g(1) - g(0) = \int_0^1 g'(t)\, dt = \int_0^1 \mathbf{f}'(\gamma(t))(y - x)\, dt \tag{8.2}$$

By the triangle inequality (for vector-valued integrals) followed by the inequality for the operator norm (6.9),

$$\begin{aligned}
|\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a})| &= \left|\int_0^1 \mathbf{f}'(\gamma(t))(\mathbf{b} - \mathbf{a})\, dt\right| && (8.3)\\
&\leq \int_0^1 |\mathbf{f}'(\gamma(t))(\mathbf{b} - \mathbf{a})|\, dt && (8.4)\\
&\leq \int_0^1 \|\mathbf{f}'(\gamma(t))\||\mathbf{b} - \mathbf{a}|\, dt && (8.5)\\
&\leq \int_0^1 M|\mathbf{b} - \mathbf{a}|\, dt && (8.6)\\
&= M|\mathbf{b} - \mathbf{a}| && (8.7)
\end{aligned}$$

□

8.2 Banach fixed-point theorem

Definition 8.2. A subset $X \subset \mathbb{R}^n$ is *closed* if for every sequence $(x_n)_{n=1}^\infty$ in X, if the sequence converges to some $a \in \mathbb{R}^n$, then $a \in X$.

Remark 8.3. There are other ways to defined closed sets. For example, another definition is "a set $X \subset \mathbb{R}^n$ is closed if its complement is open." This turns out to be equivalent to the definition above.

We only need the Banach fixed-point theorem for closed subsets of $\mathbb{R}^n$, but the proof generalizes to all complete metric spaces, so we state it in this general form.

Definition 8.4. A metric space (X, d) is *complete* if every Cauchy sequence in X converges (to something in X).

Example 8.5. Closed subsets of $\mathbb{R}^n$ (with the standard metric $d(x, y) = |x-y|$) are complete metric spaces. In contrast, the open interval $(0, 1) \subset \mathbb{R}$ is not a complete metric space.

Definition 8.6. For a metric space (X, d), we say that a function $\phi : X \to X$ is a *contraction of X into itself* if there exists a $c < 1$ such that $d(\phi(x), \phi(y)) \leq cd(x, y)$.

Example 8.7. The function $\phi : \mathbb{R} \to \mathbb{R}$ given by $\phi(x) = \frac{1}{2}x$ is a contraction of $\mathbb{R}$ into itself.

Theorem 8.8 (Banach fixed point theorem, a.k.a. contraction mapping principle)**.** *Let (X, d) be a complete metric space and let $\phi : X \to X$ be a contraction of X into itself. Then there exists a unique point $a \in X$ such that $\phi(a) = a$.*

Furthermore, given any $x_0 \in X$, define a sequence $(x_n)_{n=0}^{\infty}$ recursively by $x_{n+1} = \phi(x_n)$. Then $x_n \to a$.

Proof. We break the proof into three steps.
Step 1: We'll show any sequence defined as above converges.

Take a starting point $x_0 \in X$. Then $d(x_{n+1}, x_n) = d(\phi(x_n), \phi(x_{n-1})) \leq cd(x_n, x_{n-1})$, which implies $d(x_{n+1}, x_n) \leq c^n d(x_1, x_0)$. Thus if $m \leq n$, then by the triangle inequality,

$$d(x_m, x_n) \leq \sum_{i=m}^{n-1} d(x_{i+1}, x_i) \leq d(x_1, x_0) \sum_{i=m}^{n-1} c^i \leq d(x_1, x_0) \sum_{i=m}^{\infty} c^i \leq d(x_1, x_0) \cdot \frac{c^m}{1-c} \tag{8.8}$$

which shows that $(x_n)_n$ is a Cauchy sequence. Since the space (X, d) is complete, the sequence converges.
Step 2: We'll show that any two sequences defined as above converges to the same limit $a \in X$.

Let $x_0, y_0 \in X$. Then $d(x_{n+1}, y_{n+1}) \leq cd(x_n, y_n)$, so $d(x_n, y_n) \leq c^n d(x_0, y_0)$. This implies that the sequences $(x_n)_n$ and $(y_n)_n$ converges to the same element of X. Let us call this element a.
Step 3: We'll show $\phi(a) = a$.

Let $(x_n)_n$ be a sequence defined as above. By Step 2, $x_n \to a$. Note that contractions are continuous, so

$$a = \lim_{n\to\infty} x_n = \lim_{n\to\infty} \phi(x_{n-1}) = \phi\left(\lim_{n\to\infty} x_{n-1}\right) = \phi(a). \tag{8.9}$$

□

8.3 Proof of the inverse function theorem, preliminary

Let us now discuss how the proof of Theorem 6.9 begins.

By a change of variables, we may assume without loss of generality that $\mathbf{a} = \mathbf{0}$, $\mathbf{f}(\mathbf{0}) = \mathbf{0}$, and $\mathbf{f}'(\mathbf{0}) = I$. (Here, I denotes the $n \times n$ identity matrix.) To justify this more, see Exercise 9.5

Since $\mathbf{f}(\mathbf{0}) = \mathbf{0}$ and $\mathbf{f}'(\mathbf{0}) = I$, this means $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ is well-approximated by the identity function $\mathbf{x} \mapsto \mathbf{x}$. If $\mathsf{E}(\mathbf{x}) = \mathbf{f}(\mathbf{x}) - \mathbf{x}$ denotes the "error" in this approximation, then $\mathsf{E}'(\mathbf{x}) = \mathbf{f}'(\mathbf{x}) - I$. Since $\mathsf{E}'(\mathbf{0})$ is the zero matrix, and E' is continuous,

$$\exists \delta > 0 \text{ such that} \qquad \text{if } x \in \mathbb{R}^n \text{ and } |x| \leq \delta, \text{ then} \qquad \|\mathsf{E}'(\mathbf{x})\| \leq \frac{1}{2} \tag{8.10}$$

This allows us to use the mean value inequality (Theorem 8.1) on E. Furthermore, because $\frac{1}{2} < 1$, we will then be able to apply the Banach fixed-point theorem. The details will come tomorrow.

9 Day 4 exercises

Highly recommended: Exercise 9.3

Exercise 9.1. () In contrast to Exercise 7.8, we do have the following multivariable version of the mean value theorem when the range of the function is $\mathbb{R}$. (We will not need this in the class.)

- **Multivariable mean value theorem.** Let $f : \mathbb{R}^m \to \mathbb{R}$ be a $\mathcal{C}^1$ function. For any $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$, there exists a $\mathbf{c} \in \mathbb{R}^m$ on the line segment joining $\mathbf{a}$ and $\mathbf{b}$ such that

$$f(\mathbf{b}) - f(\mathbf{a}) = \nabla f(\mathbf{c}) \cdot (\mathbf{b} - \mathbf{a}) \tag{9.1}$$

Using the single-variable mean value theorem (as stated in Exercise 7.8), prove the multivariable mean value theorem.

Hint: Some of the ideas from the proof of Theorem 8.1 work here too.

Exercise 9.2. () Consider $\phi(x) = x^2$. Show that ϕ is a contraction of $[0.4, 0.4]$ into itself. Then show that ϕ is *not* a contraction of $[-0.5, 0.5]$ into itself.

Exercise 9.3. () Suppose $f : \mathbb{R} \to \mathbb{R}$ is $\mathcal{C}^1$ and $f(0) = 0$. Furthermore, suppose that $\frac{1}{2} \leq f'(x) \leq \frac{3}{2}$ for all $x \in [-2, 2]$. Fix $z \in [-1, 1]$.

1. Show that there exists a unique $c \in [-2, 2]$ such that $f(c) = z$.
2. Use the Banach fixed-point theorem to construct a sequence $(x_n)_{n=0}^{\infty}$ such that $x_n \to c$. (Hint: Consider $\phi(x) = x - f(x) + z$.)

Exercise 9.4. () Suppose $f : \mathbb{R} \to \mathbb{R}$ is $\mathcal{C}^1$. Fix $a < b$. Suppose $f(a) < 0 < f(b)$. Furthermore, suppose that there exist $0 < k_1 \leq k_2$ such that $k_1 \leq f'(x) \leq k_2$ for all $x \in [a, b]$.

1. Show that there exists a unique $c \in [a, b]$ such that $f(c) = 0$.

2. Use the Banach fixed-point theorem to construct a sequence $(x_n)_{n=0}^{\infty}$ such that $x_n \to c$.

Exercise 9.5. ( ) Let $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^n$ be a $\mathcal{C}^1$ function and let $\mathbf{a} \in \mathbb{R}^n$. Suppose $\det \mathbf{g}'(\mathbf{a}) \neq 0$. Show that there exists function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ such that $\mathbf{f}(\mathbf{0}) = \mathbf{0}$, $\mathbf{f}'(0) = I$, and most importantly,

$$\exists \text{invertible } A \in \mathbb{R}^{n \times n}, \exists \mathbf{b} \in \mathbb{R}^n \text{ such that} \qquad \mathbf{g}(\mathbf{x}) = A\mathbf{f}(\mathbf{x} - \mathbf{a}) + \mathbf{b} \tag{9.2}$$

This shows that if we prove the conclusion of the inverse function theorem holds for $\mathbf{f}$ at $\mathbf{0}$, then we also prove the conclusion of the inverse function holds for $\mathbf{g}$ at $\mathbf{a}$.

10 Day 5 (Inverse/Implicit, Day 3)

10.1 Proof of the inverse function theorem

Finally we get to the proof of Theorem 6.9. This is a difficult proof.

Proof of Theorem 6.9. As explained in Exercise 9.5, we may assume without loss of generality that $\mathbf{a} = \mathbf{0}$, $\mathbf{f}(\mathbf{0}) = \mathbf{0}$, and $\mathbf{f}'(\mathbf{0}) = I$. (Here, I denotes the $n \times n$ identity matrix.)
Step 1: The key inequality for this proof.

Let $\mathsf{E}(\mathbf{x}) = \mathbf{f}(\mathbf{x}) - \mathbf{x}$. As explained in Section 8.3 there exists a $\delta > 0$ such that

$$\forall \mathbf{x} \in \overline{B}(\mathbf{0}, \delta), \qquad \|\mathsf{E}'(\mathbf{x})\| \leq \frac{1}{2}, \tag{10.1}$$

where $\overline{B}(\mathbf{0}, \delta) = \{x \in \mathbb{R}^n : |x| \leq \delta\}$. This is the closed ball in $\mathbb{R}^n$ centered at $\mathbf{0}$ and with radius δ. (Note that in this step, we also use the fact that the operator norm $\|\cdot\|$, viewed as a function $\mathbb{R}^{n \times m} \to \mathbb{R}$ is a continuous function.) Then by the mean value inequality (Theorem 8.1),

$$\forall \mathbf{x}, \mathbf{y} \in \overline{B}(\mathbf{0}, \delta), \qquad |\mathsf{E}(\mathbf{y}) - \mathsf{E}(\mathbf{x})| \leq \frac{1}{2}|\mathbf{y} - \mathbf{x}| \tag{10.2}$$

We will use (10.2) over and over again in this proof, so we will refer to it as the "key inequality."
Step 2: We'll show that $\mathbf{f}$ is injective on $\overline{B}(\mathbf{0}, \delta)$.

If $\mathbf{x}, \mathbf{y} \in \overline{B}(\mathbf{0}, \delta)$, then by the key inequality (10.2) and the triangle inequality

$$\frac{1}{2}|\mathbf{y} - \mathbf{x}| \geq |\mathsf{E}(\mathbf{y}) - \mathsf{E}(\mathbf{x})| = |(f(\mathbf{y}) - f(\mathbf{x})) - (\mathbf{y} - \mathbf{x})| \geq |\mathbf{y} - \mathbf{x}| - |f(\mathbf{y}) - f(\mathbf{x})|. \tag{10.3}$$

This shows that

$$\forall \mathbf{x}, \mathbf{y} \in \overline{B}(\mathbf{0}, \delta), \qquad |f(\mathbf{y}) - f(\mathbf{x})| \geq \frac{1}{2}|\mathbf{y} - \mathbf{x}| \tag{10.4}$$

which implies that $\mathbf{f}$ is injective on $\overline{B}(\mathbf{0}, \delta)$.

Step 3: We'll show that $\mathbf{f}(\overline{B}(\mathbf{0}, \delta))$ contains $\overline{B}(\mathbf{0}, \frac{1}{2}\delta)$.

For $\mathbf{z} \in \mathbb{R}^n$, we define a function $\phi_{\mathbf{z}} : \mathbb{R}^n \to \mathbb{R}^n$ by

$$\phi_{\mathbf{z}}(\mathbf{x}) = \mathbf{x} - \mathbf{f}(\mathbf{x}) + \mathbf{z} = \mathbf{z} - \mathsf{E}(\mathbf{x}). \tag{10.5}$$

Note that

$$\mathbf{x}_0 \text{ is a fixed point of } \phi_{\mathbf{z}} \iff \phi_{\mathbf{z}}(\mathbf{x}_0) = \mathbf{x}_0 \iff \mathbf{f}(\mathbf{x}_0) = \mathbf{z} \tag{10.6}$$

Thus, we would like to apply the Banach fixed point theorem (Theorem 8.8) with $\phi_{\mathbf{z}}$. To do that, we will show two things about $\phi_{\mathbf{z}}$:

$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \qquad |\phi_{\mathbf{z}}(\mathbf{y}) - \phi_{\mathbf{z}}(\mathbf{x})| \leq \frac{1}{2}|\mathbf{y} - \mathbf{x}| \tag{10.7}$$

$$\forall \mathbf{z} \in \overline{B}(\mathbf{0}, \tfrac{1}{2}\delta), \qquad \phi_{\mathbf{z}}(\overline{B}(\mathbf{0}, \delta)) \subset \overline{B}(\mathbf{0}, \delta). \tag{10.8}$$

To prove (10.7), by the key inequality (10.2) again,

$$|\phi_{\mathbf{z}}(\mathbf{y}) - \phi_{\mathbf{z}}(\mathbf{x})| = |\mathsf{E}(\mathbf{y}) - \mathsf{E}(\mathbf{x})| \leq \frac{1}{2}|\mathbf{y} - \mathbf{x}|. \tag{10.9}$$

To prove (10.8), suppose $|\mathbf{z}| \leq \frac{1}{2}\delta$ and $|\mathbf{x}| \leq \delta$. We need to show $|\phi_{\mathbf{z}}(\mathbf{x})| \leq \delta$. By the fact that $\mathsf{E}(\mathbf{0}) = \mathbf{0}$ and the key inequality (10.2),

$$|\mathsf{E}(\mathbf{x})| = |\mathsf{E}(\mathbf{x}) - \mathsf{E}(\mathbf{0})| \leq \frac{1}{2}|\mathbf{x}| \tag{10.10}$$

Then

$$|\phi_{\mathbf{z}}(\mathbf{x})| = |\mathbf{z} - \mathsf{E}(\mathbf{x})| \leq |\mathbf{z}| + |\mathsf{E}(\mathbf{x})| = |\mathbf{z}| + \frac{1}{2}|\mathbf{x}| \leq \frac{1}{2}\delta + \frac{1}{2}\delta = \delta \tag{10.11}$$

which proves (10.8).

Suppose $\mathbf{z} \in \overline{B}(\mathbf{0}, \frac{1}{2}\delta)$. As a consequence of (10.8) and (10.7), we have shown that $\phi_{\mathbf{z}}$ is a contraction of $\overline{B}(\mathbf{0}, \delta)$ into itself. Furthermore, $\overline{B}(\mathbf{0}, \delta)$ is a complete metric space. By the Banach fixed point theorem (Theorem 8.8), there is a point $\mathbf{x}_0 \in \overline{B}(\mathbf{0}, \delta)$ such that $\phi_{\mathbf{z}}(\mathbf{x}_0) = \mathbf{x}_0$. As noted in (10.6), $\mathbf{f}(\mathbf{x}_0) = \mathbf{z}$, so $\mathbf{z} \in \mathbf{f}(\overline{B}(\mathbf{0}, \delta))$.

Step 4: The conclusion.

Let $V = B(\mathbf{0}, \frac{1}{2}\delta) = \{\mathbf{x} \in \mathbb{R}^n : |x| < \delta\}$. (Note that this is an open ball.) Let $U = \mathbf{f}^{-1}(V)$. Then by what we have shown above, $\mathbf{f} : U \to V$ is a bijection and it is $\mathcal{C}^1$. Let $\mathbf{g} = \mathbf{f}^{-1} : V \to U$. It only remains to show that

$$\mathbf{g}'(\mathbf{y}) = \mathbf{f}'(\mathbf{g}(\mathbf{y}))^{-1} \text{ for all } \mathbf{y} \in V. \tag{10.12}$$

Note that (10.12) implies that $\mathbf{g}$ is $\mathcal{C}^1$. (Actually, here we also use the fact that the map $A \mapsto A^{-1}$ is a continuous function $\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$. This is not hard to show.) Exercise 11.1 asks you to prove (10.12). □

Remark 10.1. This is a difficult proof. I remember that when I saw this proof for the first time, I did not feel like I understood anything.

Remark 10.2. Terry Tao gives a one-sentence summary of the proof in his blog post. https://terrytao.wordpress.com/2011/09/12/the-inverse-function-theorem-for-everywhere-differen The sentence is in the first paragraph, and begins "Indeed, one may normalize..."

See also Theorem 2 of that blog post for a generalization of the inverse function theorem.

10.2 Implicit function theorem

As we have seen, the inverse function theorem is a statement that can be thought in terms of the theme "functions behave like their affine approximations" (2.1). The inverse function theorem only applies to functions $\mathbb{R}^m \to \mathbb{R}^n$ with $m = n$.

The *implicit function theorem*, which we will state below, applies to functions $\mathbb{R}^m \to \mathbb{R}^n$ with $m \geq n$. It can also be thought of as another instance of the theme (2.1). Let us start with linear functions $A : \mathbb{R}^m \to \mathbb{R}^n$ with $m \geq n$. In other words, A is a matrix such that the number of columns is at least the number of rows.

Example 10.3. Suppose

$$A = \begin{pmatrix} 2 & -6 \end{pmatrix} \tag{10.13}$$

A point in $\ker A = \{\mathbf{x} \in \mathbb{R}^2 : A\mathbf{x} = 0\}$ (the kernel of A) is completely determined by its first coordinate x_1. In other words, for points in $\ker A$, we can write x_1 as a function of x_2. (It is just $x_1 = 3x_1$.) Furthermore, this is true for any other 1×2 matrix as long as the first entry is nonzero.

Example 10.4. Suppose

$$A = \begin{pmatrix} 3 & 1 & 4 & 1 & 5 \\ 0 & 0 & 9 & 2 & 6 \\ 0 & 0 & 0 & 5 & 3 \end{pmatrix} \tag{10.14}$$

A point in $\ker A = \{\mathbf{x} \in \mathbb{R}^5 : A\mathbf{x} = 0\}$ is completely determined by the coordinates x_2 and x_5. In other words, for points in $\ker A$, we can write x_1, x_3, x_4 as functions of x_2 and x_5.

Example 10.5. As a generalization of Example 10.4, suppose $A \in \mathbb{R}^{n \times m}$ and $n \geq m$. If columns $i_1, \ldots, i_m$ of A are linearly independent, then points in $\ker A$ are completely determined by the coordinates $x_{i_1}, \ldots, x_{i_m}$

Now let us look at some nonlinear examples.

Example 10.6. Consider $f(x, y) = x^2 + y^2 - 1$. The zero set $Z = \{(x, y) \in \mathbb{R}^2 : f(x, y) = 0\}$ is the unit circle.

1. At every point $(x_0, y_0) \in \mathbb{Z}$ with $x_0 \neq 0$, there is an open set $U \subset \mathbb{R}^2$ containing (x_0, y_0) such that $Z \cap U$ is the graph of a function $x = h(y)$. (Either $h(y) = (1 - y^2)^{1/2}$ or $h(y) = -(1 - y^2)^{1/2}$.)

2. At every point $(x_0, y_0) \in \mathbb{Z}$ with $y_0 \neq 0$, there is an open set $U \subset \mathbb{R}^2$ containing (x_0, y_0) such that $Z \cap U$ is the graph of a function $y = g(x)$. (Either $g(x) = (1 - x^2)^{1/2}$ or $g(x) = -(1 - x^2)^{1/2}$.)

What makes the above possible is the direction of the tangent line to Z at various points. Note that $f'(x, y) = \begin{pmatrix} 2x & 2y \end{pmatrix}$.

Here is the implicit function theorem, stated for maps $\mathbb{R}^2 \to \mathbb{R}$. The statement and proof generalize to $\mathbb{R}^m \to \mathbb{R}^n$ with $m \geq n$ with some small modifications. In the higher-dimensional version, you need to consider linearly independent columns of $\mathbf{f}'(\mathbf{x})$, as in Example 10.5. (See an analysis textbook for the statements.)

Theorem 10.7 (Implicit function theorem in $\mathbb{R}^2$)**.** *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a $\mathcal{C}^1$ function. Suppose that $f(0, 0) = (0, 0)$. $\partial_x f(0, 0) \neq 0$. Then there exist two open subsets $I_1, I_2 \subset \mathbb{R}$ such that $(0, 0) \in I_1 \times I_2$, and there exists a function $g : I_2 \to \mathbb{R}$ such that*

$$\text{for all } (x, y) \in I_1 \times I_2, \qquad f(x, y) = 0 \iff x = g(y). \tag{10.15}$$

The statement of the implicit function theorem looks complicated, so you might think the proof is as complicated as the proof of the inverse function theorem. However, we can actually use the inverse function theorem to prove the implicit function theorem, which saves us a lot of trouble.

Proof. See Exercise 11.3 for an outline of the proof. □

Informally, a set is a k-dimensional manifold if it "locally looks k-dimensional." For example, a circle is a 1-dimensional manifold and a sphere is a 2-dimensional manifold. The implicit function theorem gives sufficient condition for the zero set of a system of equations to be a k-dimensional manifold.

Corollary 10.8. *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a $\mathcal{C}^1$ function. Let $Z = \{(x, y) \in \mathbb{R}^2 : f(x, y) = 0\}$, and suppose that $f'(x, y) \neq 0$ for all $(x, y) \in Z$. Then Z is a 1-dimensional submanifold of $\mathbb{R}^2$.*

The result above generalizes to higher dimensions.

10.3 Conclusion of the class

Multivariable calculus is often taught without linear algebra as a prerequisite. The goal of this class was to show you how linear algebra actually plays a key role in multivariable calculus, starting with the definition of the derivative.

11 Day 5 exercises

Exercise 11.1. () Finish Step 4 in the proof of Theorem 6.9. Here are some ideas.

1. First show that $\mathbf{g}$ is continuous. Hint: Use (10.4).
2. Now show (10.12). To do this, let $\mathsf{E}_{\mathbf{g}}(\mathbf{k})$ be defined by

$$\mathbf{g}(\mathbf{y}+\mathbf{k}) = \mathbf{g}(\mathbf{y}) + \mathbf{f}'(\mathbf{g}(\mathbf{y}))^{-1}\mathbf{k} + \mathsf{E}_{\mathbf{g}}(\mathbf{k}) \tag{11.1}$$

and show

$$\lim_{\mathbf{k}\to\mathbf{0}} \frac{|\mathsf{E}_{\mathbf{g}}(\mathbf{k})|}{|\mathbf{k}|} = 0 \tag{11.2}$$

Hint: Some of the ideas in the proof of Theorem 6.4 may be useful. Also, use (10.4).

Exercise 11.2. () Suppose $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ satisfies $\det \mathbf{f}'(\mathbf{x}) \neq 0$ for all $\mathbf{x} \in \mathbb{R}^n$. Show that $\mathbf{f}$ is an open mapping, i.e., $\mathbf{f}(U)$ is open for all open sets $U \subset \mathbb{R}^n$.

Exercise 11.3. () Here is an outline of the proof of the implicit function theorem Theorem 10.7. Fill in the details.

1. Define $\mathbf{F} : \mathbb{R}^2 \to \mathbb{R}^2$ by $\mathbf{F}(x,y) = (f(x,y), y)$ and show that $\det \mathbf{F}'(0,0) \neq 0$.
2. Apply the inverse function theorem to $\mathbf{F}$ to get an inverse map $\mathbf{G}(s,t) = (G_1(s,t), G_2(s,t))$
3. Show that $g(y) = G_1(0,y)$ satisfies (10.15).