MAU22203/33203 - Analysis in Several Real Variables

Adam Keilthy

Trinity College Dublin

Contents

0 Recapping analysis in one real variable

Before trying to define notions of convergence, continuity, differentiation, and integration for functions $f : \mathbb{R}^m \to \mathbb{R}^n$, it will be helpful to remind ourselfs of some basic concepts in one dimensional analysis, both to refresh our knowledge, and to illustrate some important differences. We start by asking a painful question: what *is* a real number?

0.1 The real numbers

In our mathematical career, we encounter many objects with properties similar to the reals. We can do addition in $\mathbb{N}$, and subtraction in $\mathbb{Z}$. We can divide in $\mathbb{Q}$ and $\mathbb{C}$. We can even do all of these things in

$$\mathbb{R}(x) = \left\{ \frac{f(x)}{g(x)} \mid f, g \in \mathbb{R}[x],\, g \neq 0 \right\}.$$

So what makes the reals unique.

Fact 0.1. *The reals are the unique (up to isomorphism) Dedekind complete Archimedean ordered field.*

Remark 0.2. *Other than the above fact, and the definition of least upper bound, this subsection is mostly for flavour. Don't worry too much if you want to skip it.*

Let's break this down:

Definition 0.3. *A field is a set F equipped with two binary commutative, associative operations $+$ and $\cdot$, such that*

- *For every $x, y, z \in F$, $x \cdot (y + z) = x \cdot y + x \cdot z$,*
- *There exists an element 0 such that $0 + x = x + 0 = x$ for every $x \in F$,*
- *There exists an element 1 such that $1 \cdot x = x \cdot 1 = x$ for every $x \in F$,*
- *For every $x \in F$, there exists an element $-x \in F$ such that $x + (-x) = 0$, called an additive inverse,*
- *For every $x \in F \setminus \{0\}$, there exists an element x^{-1} such that $x \cdot x^{-1} = 1$, called a multiplicative inverse.*

The need for an additive inverse removes objects like $\mathbb{N}$, and the need for a multiplicative inverse cuts out objects like $\mathbb{Z}$.

Definition 0.4. *An ordered field is a field F with a binary relation $<$ such that*

- *For every x, y, exactly one of $x < y$, $y < x$ or $x = y$ holds,*
- *If $x < y$ and $y < z$, then $x < z$*
- *If $x < y$, then $x + z < y + z$ for every $z \in F$,*
- *If $x, y > 0$¡ then $x \cdot y > 0$.*

The rationals $\mathbb{Q}$ and the reals $\mathbb{R}$ form ordered fields with the usual order. The field of rational functions forms an ordered field with the ordering

$$\frac{f(x)}{g(x)} < \frac{p(x)}{q(x)} \text{ if and only if } \lim_{x \to \infty} (p(x)g(x) - f(x)q(x)) > 0$$

where this final inequality is considered in the usual order for $\mathbb{R}$. However $\mathbb{C}$ cannot be equipped with an order compatible with the field operations.

Definition 0.5. *An ordered field is Archimedean if for every $x \in F$ there exists $n \in \mathbb{Z}$ such that the sum of n copies of the multiplicative identity is greater than x:*

$$1 + 1 + \cdots + 1 > x$$

The field of rational functions fails to be Archimedean, as the function $f(x) = x$ is greater than every integer n. In order to distinguish between the rationals and the reals, however, we need to introduce the least upper bound.

Definition 0.6. *Let $S \subset F$ be a subset of an ordered field. We say S is bounded above if there exists $b \in F$ such that $x < b$ for every $x \in S$, and call such a b an upper bound. We say S is bounded below if there exists $a \in F$ such that $a < x$ for every $x \in S$, and call such an a a lower bound. We say S is bounded if it is bounded both above and below.*

Definition 0.7. *Let $S \subset F$ be a subset of an ordered field that is bounded above. We say that $s \in F$ is the least upper bound, or supremum, of S if s is an upper bound for S and $s \leq b$ for every upper bound b of S. We write $s = \sup S$.*

For $S \subset F$ a subset of an ordered field that is bounded below, we say that $\ell \in F$ is the greatest lower bound, or infimum, of S if ℓ is a lower bound for S and $\ell \geq a$ for every lower bound a of S. We write $\ell = \inf S$.

Example 0.8. *Consider the subset $S = \{x \in \mathbb{Q} \mid x^2 < 2\} \subset \mathbb{R}$. This set is bounded above, and has least upper bound $\sup S = \sqrt{2}$.*

Note that the existence of a least uppper bound depends on the field of definition. If we view $S = \{x \in \mathbb{Q} \mid x^2 < 2\}$ as a subset of the ordered field $\mathbb{Q}$, then $\sup S$ does not exist: for any rational number q such that $q > 0$ and $q^2 > 2$, we can find a smaller such rational number, better approximating $\sqrt{2}$.

Definition 0.9. *An ordered field F is called Dedekind complete, or is said to satisfy the Least Upper Bound Principle, if for any non-empty $S \subset F$ that is bounded above, $\sup S$ exists in F.*

Remark 0.10. *As*

$$\inf S = -\sup\{-x \mid x \in S\},$$

the existence of least upper bounds also guarantees the existence of greatest lower bounds. As such, we will often only prove results about one of them, and deduce the other by this symmetry.

And this is the key distinction between the rationals and the reals. The rational numbers do not contain all suprema, while the reals do.

The proof of Fact 0.1 consists of two components: one is to show that there is a unique Dedekind complete Archimedean ordered field, and the other is to explicitly construct the real numbers as such an object. There are two main approaches to defining the real numbers: one via Dedekind cuts, and one via Cauchy sequences. Either suffice for our purposes, but since the specific construction of the reals is not important for our purposes, we will not expand on these here.

0.2 Convergent sequences in $\mathbb{R}$

Let us now recap some important results and concepts about sequences of real numbers.

Definition 0.11. *A sequence of real numbers is a function $f : \mathbb{N} \to \mathbb{R}$, usually represented as $x_1, x_2, \ldots$ or $\{x_n\}_{n=1}^{\infty}$. We will often suppress the range of the index, and just write $\{x_n\}$.*

Definition 0.12. *A sequence $\{x_n\}$ of real numbers is said to converge to a limit $L \in \mathbb{R}$ if, for all $\varepsilon > 0$, there exists an integer $N > 0$ such that $|x_n - L| < \varepsilon$ for all $n \geq N$. We call a sequence $\{x_n\}$ convergent if it converges to a finite limit L, and write $x_n \to L$ or $\lim_{n\to\infty} x_n = L$ to denote this.*

Definition 0.13. *A sequence $\{x_n\}$ is bounded above if there exists $B \in \mathbb{R}$ such that $B \geq x_n$ for all $n \geq 1$. The sequence is bounded beow if there exists $A \in \mathbb{R}$ such that $A \leq x_n$ for all $n \geq 1$. We say the sequence is bounded if it is bounded both above and below.*

A sequence being bounded is essential to have hope of it converging. Indeed, we have the following basic result.

Lemma 0.14. *Every convergent sequence of real numbers is bounded.*

Proof. Suppose we have a convergent sequence $\{x_n\}$ with limit L. By definition, there exists some $N > 0$ such that, for all $n \geq N$, $|x_n - L| < 1$, and so

$$L - 1 < x_n < L + 1$$

for all $n \geq N$. Define

$$\begin{aligned} A &:= \min\{x_1, \ldots, x_N, L - 1\}, \\ B &:= \max\{x_1, \ldots, x_N, L + 1\}. \end{aligned}$$

These exist and are well defined since the sets involved are finite, and furthermore satisfy that $A < x_n < B$ for all $n \geq 1$. □

However, being bounded is not sufficient to guarantee convergence. The sequence $\{x_n = (-1)^n\}$ is bounded, but definitely does not converge. Boundedness only guarantees convergence for monotonic sequences.

Definition 0.15. *A sequence $\{x_n\}$ is called increasing if $x_n \leq x_{n+1}$ for all $n \geq 1$, and strictly increasing if $x_n < x_{n+1}$ for all $n \geq 1$. We define decreasing and strictly decreasing sequences similarly. A sequence is called monotonic if it is increasing or decreasing.*

Remark 0.16. *Note that a sequence $\{x_n\}$ is (strictly) increasing if and only if the sequence $\{-x_n\}$ is (strictly) decreasing. As such, to prove results about monotonic sequences, we can freely assume we are working with an increasing sequence.*

Theorem 0.17. *Any increasing sequence that is bounded above converges. Any decreasing sequence that is bounded below converges.*

Proof. By Remark 0.16, it suffices to consider only the increasing case. Let $\{x_n\}$ be an increasing sequence that is bounded above. Since $\mathbb{R}$ is Dedekind complete, there exists a least upper bound $p \in \mathbb{R}$. In particular, for any $\varepsilon > 0$, $p - \varepsilon$ is not an upper bound, and so there exists some $N > 0$ such that $x_N > p - \varepsilon$. Then, since the sequence is increasing, we have that

$$p \geq x_n \geq x_N > p - \varepsilon$$

for every $n \geq N$. In particular, $|x_n - p| < \varepsilon$ for every $n \geq N$. Thus, $\{x_n\}$ converges and $x_n \to p$. □

Without monotonicity, the best we can say about bounded sequences comes in the form of the Bolzano-Weierstrass theorem.

Theorem 0.18. *Let $\{x_n\}$ be a bounded sequence. Then there exists a subsequence $\{x_{j_k}\}_{k=1}^{\infty}$, with $j_k < j_{k+1}$ that converges to a finite limit.*

Proof. Let $\{x_n\}$ be a bounded sequence. Call $j \in \mathbb{N}$ a valley point if $x_j < x_k$ for all $k \geq j$, and let S be the (ordered) set of all valley points

$$S = \{j_1 < j_2 < \cdots\}.$$

By definition of a valley point,

$$x_{j_1} < x_{j_2} < x_{j_3} < \dots$$

so if S is infinite, we obtain an increasing sequence that is bounded above, and hence convergent. If S is finite, then there is a maximal valley point J. Then, since $j_1 := J + 1$ is not a valley point, there exists $j_2 > j_1$ such that $x_{j_1} \geq x_{j_2}$. Since j_2 is not a valley point, there exists $j_3 > j_2$ such that $x_{j_2} \geq x_{j_3}$. Continuing in this way, we construct an infinite decreasing sequence

$$x_{j_1} \geq x_{j_2} \geq x_{j_3} \geq \cdots$$

which is bounded below and hence converges. □

1 Convergent sequences in $\mathbb{R}^n$

We start our step into several variables by trying to extend the above results into higher dimensions. A sequence in $\mathbb{R}^n$ is a function $f : \mathbb{N} \to \mathbb{R}^n$, and can be thought of as an ordered collection of vectors or points $\{\vec{x}_n\}$ in $\mathbb{R}^n$. In order to talk about convergence, we need a notion of distance.

Remark 1.1. *It will be convenient to think of points in $\mathbb{R}^n$ as both points in a metric or topological space, and as vectors in a vector space. As such, we will somewhat abusively talk about adding points together, when we really mean their sum as vectors, and so on. This is not possible in a general metric or topological space, and it is important to keep this in mind in other analysis courses.*

Definition 1.2. *Given a point $\vec{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$, we define its Euclidean norm by*

$$\|\vec{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

Given a pair of points $\vec{x} = (x_1, \dots, x_n)$ and $\vec{y} = (y_1, \dots, y_n)$ in $\mathbb{R}^n$, we define their inner product by

$$\langle x, y \rangle = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

We refer to $\mathbb{R}^n$ equipped with the Euclidean norm as n-dimensional Euclidean space.

Using this norm, and the vector space structure on $\mathbb{R}^n$, we can define a notion of distance, and hence convergence.

Definition 1.3. *A sequence $\{\vec{x}_k\}$ of points in $\mathbb{R}^n$ is said to converge (with respect to the Euclidean norm) to a point $\vec{p} \in \mathbb{R}^n$ if, for any $\varepsilon > 0$, there exists an integer $N > 0$ such that $\|\vec{x}_k - \vec{p}\| < \varepsilon$ for all $k \geq N$.*

We call a sequence $\{\vec{x}_k\}$ bounded if there exists $R > 0$ such that $\|\vec{x}_n\| \leq R$ for all $n \geq 1$.

Note that it no longer makes sense to talk about a sequence being bounded above or below, as we no longer have an order structure. Similarly, monotonicity is no longer defined.

Lemma 1.4. *If $\vec{x}$, $\vec{y} \in \mathbb{R}^n$, then $|x_i - y_i| \leq \|\vec{x} - \vec{y}\|$ for all $1 \leq i \leq n$.*

Proof. Since all quantities involved are non-negative, it suffices to prove the squared version:

$$(x_i - y_i)^2 = |x_i - y_i|^2 \leq \|\vec{x} - \vec{y}\|^2 = \sum_{k=1}^{n} (x_k - y_k)^2.$$

This is clearly true, as the right hand side is equal to the left hand side plus a sum of non-negative numbers. □

With this lemma in mind, we can say something above converges of sequences points in $\mathbb{R}^n$ in terms of the sequences of their components.

Lemma 1.5. *Let $\vec{p} = (p_1, \ldots, p_n) \in \mathbb{R}^n$, and let $\{\vec{x}_k = (x_{k,1}, \ldots, x_{k,n})\}$ be a sequence of points in $\mathbb{R}^n$. Then $\{\vec{x}_k\}$ converges to $\vec{p}$ (with respect to the Euclidean norm) if and only if $\{x_{k,i}\}$ converges to p_i for every $1 \leq i \leq n$.*

Proof. Suppose $\vec{x}_k \to \vec{p}$. Then, for every $\varepsilon > 0$ there exists an $N > 0$ such that $\|\vec{x}_k - \vec{p}\| < \varepsilon$ for all $k \geq N$. By Lemma 1.4, we therefore have

$$|x_{k,i} - p_i| \leq \|\vec{x}_k - \vec{p}\| < \varepsilon$$

for every $k \geq N$ and each $1 \leq i \leq n$. Hence $x_{k,i} \to p_i$.

Now suppose $x_{k,i} \to p_i$ for each $1 \leq i \leq n$. Then, for every $\varepsilon > 0$, there exist $N_1, N_2, \ldots, N_n > 0$ such that

$$|x_{k,i} - p_i| < \frac{\varepsilon}{\sqrt{n}}$$

for every $k \geq N_i$. Let N be the maximum of $N_1, \ldots, N_n$. Then, for all $k \geq N$

$$\begin{aligned} \|\vec{x}_k - \vec{p}\| &= \sqrt{(x_{k,1} - p_1)^2 + \cdots + (x_{k,n} - p_n)^2} \\ &< \sqrt{\frac{\varepsilon^2}{n} + \cdots + \frac{\varepsilon^2}{n}} = \sqrt{\varepsilon^2} = \varepsilon, \end{aligned}$$

and hence $\vec{x}_k \to \vec{p}$. □

With this componentwise convergence in mind, we can still lift some results from the one dimensional case, even without monotonicity. A great example is the multivariable Bolzano-Weierstrauss theorem.

Theorem 1.6. *Every bounded sequences of points in $\mathbb{R}^n$ has a convergent subsequence.*

Proof. We will prove this by induction on n. We have already seen the case for $n = 1$, so let us assume the result holds in $\mathbb{R}^{n-1}$ for some $n > 1$. Let $\{\vec{x}_k\}$ be a bounded sequence in $\mathbb{R}^n$, such that $\|\vec{x}_k\| \leq R$ for every $k \geq 1$. Denote by $\{\vec{t}_k\}$ the sequence of points in $\mathbb{R}^{n-1}$ with components $t_{k,i} = x_{k,i}$ for $1 \leq i \leq n-1$.

Since

$$\|\vec{t}_k\|^2 = \|\vec{x}_k\|^2 - x_{k,n}^2 \leq \|\vec{x}_k\|^2 \leq R^2$$

the sequence $\{\vec{t}_k\}$ is bounded and hence has a convergent subsequence $\{\vec{t}_{j_k}\}$, converging to a point $\vec{q} \in \mathbb{R}^{n-1}$. Then, for any $\varepsilon > 0$ there exists an $N_1 > 0$ such that for all $j_k \geq N_1$, $\|\vec{t}_{j_k} - \vec{q}\| < \frac{1}{2}\varepsilon$.

By the one dimensional Bolzano-Weierstrauss theorem, the sequence $\{x_{j_k,n}\}$, which is bounded by Lemma 1.4, has a convergent subsequence $\{x_{j_{k_\ell},n}\}$, converging to a real number r. Then, there exists $N_2 > 0$ such that, for all $j_{k_\ell} \geq N_2$, $|x_{j_{k_\ell},n} - r| < \frac{1}{2}\varepsilon$. Define $\vec{p}$ to be the element of $\mathbb{R}^n$ such that $p_i = q_i$ for $1 \leq i \leq n-1$, and $p_n = r$. Then, for all $j_{k_\ell} \geq \max\{N_1, N_2\}$, we have that

$$\|\vec{x}_{j_{k_\ell}} - \vec{p}\|^2 = \|\vec{t}_{j_{k_\ell}} - \vec{q}\|^2 + (x_{j_{k_\ell},n} - r)^2 < \frac{1}{2}\varepsilon^2$$

and hence

$$\|\vec{x}_{j_{k_\ell}} - \vec{p}\| < \frac{\varepsilon}{\sqrt{2}} < \varepsilon.$$

Therefore, the subsequence $\{\vec{x}_{j_{k_\ell}}\}$ converges to the point $\vec{p}$. □

1.1 Cauchy-Schwarz and Cauchy sequences in $\mathbb{R}^m$

Definition 1.7. *A sequence $\{\vec{x}_n\}$ in $\mathbb{R}^m$ is called Cauchy if for every $\varepsilon > 0$ there exists $N > 0$ such that $\|\vec{x}_k - \vec{x}_\ell\| < \varepsilon$ for every $k, \ell \geq N$.*

In the one dimensional case, Cauchy sequences were exactly those sequences that converge. Our next goal will be to show that this equivalence holds in $\mathbb{R}^m$. In order to prove this, we will need an analogue of the triangle inequality

$$|x + y| \leq |x| + |y|$$

for the Euclidean norm. As an auxiliary step towards proving this, we will first show the Cauchy-Schwarz inequality.

Proposition 1.8. *Let $\vec{x}, \vec{y} \in \mathbb{R}^m$. Then $|\langle \vec{x}, \vec{y} \rangle| \leq \|\vec{x}\|\|\vec{y}\|$ with equality if and only if $\vec{x}$ is a scalar multiple of $\vec{y}$ or vice versa.*

Proof. First note that this statement is clearly true if $\vec{y} = 0$, so we will assume otherwise. Then note that, for all $t \in \mathbb{R}$ $\|t\vec{x} + \vec{y}\|^2 \geq 0$. But we can expand this out in terms of the inner product to conclude that

$$0 \leq \|t\vec{x} + \vec{y}\|^2 = t^2\|\vec{x}\|^2 + 2t\langle \vec{x}, \vec{y}\rangle + \|\vec{y}\|^2$$

for all real t. As such, the quadratic equation

$$t^2\|\vec{x}\|^2 + 2t\langle \vec{x}, \vec{y}\rangle + \|\vec{y}\|^2 = 0$$

has at most one real root, and hence the discriminant is non-positive, that is to say

$$(2\langle \vec{x}, \vec{y}\rangle)^2 - 4\|\vec{x}\|^2\|\vec{y}\|^2 \leq 0$$

which can be rearranged to give

$$|\langle \vec{x}, \vec{y}\rangle| \leq \|\vec{x}\|\|\vec{y}\|.$$

We get equality if and only if the quadratic polynomial has exactly one real root, which occurs if and only if $\|t\vec{x} + \vec{y}\| = 0$ for some real t, and since $\|\vec{v}\| = 0$ if and only if $\vec{v} = 0$, we must therefore have equality if and only if $\vec{y} = -t\vec{x}$ for some real t. □

Corollary 1.9. *Let $\vec{x}, \vec{y} \in \mathbb{R}^m$. Then $\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$.*

Proof. Note that

$$\begin{aligned}
\|\vec{x} + \vec{y}\|^2 &= \langle \vec{x} + \vec{y}, \vec{x} + \vec{y}\rangle \\
&= \|\vec{x}\|^2 + 2\langle \vec{x}, \vec{y}\rangle + \|\vec{y}\|^2 \\
&\leq \|\vec{x}\|^2 + 2|\langle \vec{x}, \vec{y}\rangle| + \|\vec{y}\|^2 \\
&\leq \|\vec{x}\|^2 + 2\|\vec{x}\|\|\vec{y}\| + \|\vec{y}\|^2 \\
&= (\|\vec{x}\| + \|\vec{y}\|)^2
\end{aligned}$$

where the first inequality follows from $z \leq |z|$, and the second is the Cauchy-Schwarz inequality. The triangle inequality follows upon taking the square root. □

Corollary 1.10. *Let $\vec{x}, \vec{y} \in \mathbb{R}^m$. Then $\|\vec{x} - \vec{y}\| \geq \|\vec{x}\| - \|\vec{y}\|$.*

Proof. Note that

$$\|\vec{x}\| = \|(\vec{x} - \vec{y}) + \vec{y}\| \leq \|\vec{x} - \vec{y}\| + \|\vec{y}\|,$$

from which the reverse triangle inequality follows. □

With these inequalities in mind, we can begin to say something about the convergence of Cauchy sequences.

Lemma 1.11. *Every Cauchy sequence in $\mathbb{R}^m$ is bounded.*

Proof. Let $\{\vec{x}_n\}$ be a Cauchy sequence. By the Cauchy property, there exists $N > 0$ such that $\|\vec{x}_k - \vec{x}_\ell\| < 1$ for every $k, \ell \geq N$. In particular, $\|\vec{x}_k - \vec{x}_N\| < 1$ for every $k \geq N$. The reverse triangle inequality gives that

$$\|\vec{x}_k\| - \|\vec{x}_N\| \leq \|\vec{x}_k - \vec{x}_N\| < 1$$

and so $\|\vec{x}_k\| < \|\vec{x}_N\| + 1$ for every $k \geq N$. Hence

$$\|\vec{x}_n\| \leq \max\{\|\vec{x}_1\|, \|\vec{x}_2\|, \ldots, \|\vec{x}_N\|, \|\vec{x}_N\| + 1\}$$

for every $n \geq 1$, and so the sequence is bounded. $\square$

And finally, we show that Cauchy and convergent sequences coincide.

Theorem 1.12. *A sequence converges in $\mathbb{R}^m$ if and only if it is Cauchy.*

Proof. Suppose $\{\vec{x}_n\}$ converges to a point $\vec{p} \in \mathbb{R}^m$. Then, for every $\varepsilon > 0$ there exists $N > 0$ such that $\|\vec{x}_n - \vec{p}\| < \frac{\varepsilon}{2}$ for every $n \geq N$. In particular, for $k, \ell \geq N$, we have that

$$\|\vec{x}_k - \vec{x}_\ell\| \leq \|\vec{x}_k - \vec{p}\| + \|\vec{p} - \vec{x}_\ell\| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

by the triangle inequality. Hence $\{\vec{x}_n\}$ is Cauchy.

Conversely, suppose that $\{\vec{x}_n\}$ is Cauchy. Then, by Lemma 1.11, $\{\vec{x}_n\}$ is bounded. Hence by Theorem 1.6, it contains a convergent subsequence $\{\vec{x}_{j_k}\}$ that converges to a point $\vec{p} \in \mathbb{R}^m$. We claim $\{\vec{x}_n\}$ converges to $\vec{p}$.

Let $\varepsilon > 0$. By the Cauchy property, there exists $N > 0$ such that for all $k, \ell \geq N$, the norm $\|\vec{x}_k - \vec{x}_\ell\| < \frac{\varepsilon}{2}$. Since $\vec{x}_{j_k} \to \vec{p}$, there exists some $j_n > N$ such that $\|\vec{x}_{j_n} - \vec{p}\| < \frac{\varepsilon}{2}$. Therefore, by the triangle inequality,

$$\|\vec{x}_k - \vec{p}\| \leq \|\vec{x}_k - \vec{x}_{j_n}\| + \|\vec{x}_{j_n} - \vec{p}\| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

for every $k \geq N$. Hence $\vec{x}_n \to \vec{p}$. $\square$

The next step in analysis in several real variables would be to talk about functions and continuity of functions in several real variables. But we have two perspectives on continuity in the one variable case: one in terms of $\varepsilon - \delta$, and one in terms of open sets. It will be important for us to be able to consider both, so we will spend a bit of time on the topology of Euclidean space first.

2 Open sets and topology on $\mathbb{R}^m$

When discussing open sets in $\mathbb{R}$, our basic open object was the open interval (a, b). The natural analogue of this in higher dimensions is an open ball. This will give us the structure we need, but we do need to be a bit more careful, as the union of two intersecting open balls is not an open ball, as it would have been for intervals. Still, essentially all the arguments will transfer over.

Definition 2.1. *Given a point* $\vec{p} \in \mathbb{R}^m$ *and a positive real number* $r > 0$*, we define the open ball of radius* r*, centred at* $\vec{p}$ *to be the set*

$$B(\vec{p}, r) := \{\vec{x} \in \mathbb{R}^m \mid \|\vec{x} - \vec{p}\| < r\}.$$

We similarly define the corresponding closed ball as the set

$$\overline{B}(\vec{p}, r) := \{\vec{x} \in \mathbb{R}^m \mid \|\vec{x} - \vec{p}\| \leq r\}.$$

Analogously to the case of $\mathbb{R}$, we will want unions of open sets to be open. As such, we make the following definition.

Definition 2.2. *A subset* $U \subset \mathbb{R}^m$ *is called open if for every* $\vec{x} \in U$*, there exists a real number* $r_x > 0$ *such that* $B(\vec{x}, r_x) \subset U$

Remark 2.3. *It is conventional and convenient to also define the empty set* $\emptyset$ *to be open.*

Example 2.4. *Let* $H = \{(x, y) \in \mathbb{R}^2 \mid y > c\}$ *for some fixed real* c*. I claim* H *is open: let* $\vec{q} = (s, t) \in H$*. Then, as* $t > c$*,* $\delta = t - c > 0$*. Suppose we have* $\vec{x} = (x, y) \in B(\vec{q}, \delta)$*. As*

$$|t - y| \leq \|\vec{q} - \vec{x}\| < \delta = t - c,$$

we have that

$$t - y < t - c \Rightarrow c < y$$

and so $\vec{x} \in H$*. Hence* $B(\vec{q}, \delta) \subset H$*. As* $\vec{q}$ *was arbitrary, we must have that* H *is open.*

Example 2.5. *Let* $A = \{\vec{x} \in \mathbb{R}^2 \mid 1 < \|x\| < 2\}$*. I claim* A *is open. Suppose we have* $\vec{q} \in A$ *and let* $r = \min(\|\vec{q}\| - 1, 2 - \|\vec{q}\|)$*, and consider the open ball* $B(\vec{q}, r)$*. If* $\vec{x} \in B(\vec{q}, r)$*, then*

$$\|\vec{x}\| \leq \|\vec{x} - \vec{q}\| + \|\vec{q}\| < r + \|\vec{q}\| \leq 2 - \|\vec{q}\| + \|\vec{q}\| = 2$$

and

$$\|\vec{x}\| \geq \|\vec{q}\| - \|\vec{x} - \vec{q}\| > \|\vec{q}\| - r \geq \|\vec{q}\| - \|\vec{q}\| + 1 = 1.$$

Hence $\vec{x} \in A$ *and so* $B(\vec{(q)}, r) \subset A$*. As* $\vec{q}$ *was arbitrary, we have that* A *is open.*

Example 2.6. *The closed ball* $\overline{B}(\vec{0}, 1)$ *is not open, as there is no open ball around* $(1, 0, \ldots, 0)$ *contained within* $\overline{B}(\vec{0}, 1)$*.*

Lemma 2.7. *The open ball is an open set.*

Proof. Let $\vec{p} \in \mathbb{R}^m$ and $r > 0$ be real. Suppose that $\vec{q} \in B(\vec{p}, r)$. Then $\|\vec{q} - \vec{p}\| < r$, and so $\delta = r - \|\vec{q} - \vec{p}\| > 0$. I claim $B(\vec{q}, \delta) \subset B(\vec{p}, r)$. As $\vec{q}$ is arbitrary, this will prove that $B(\vec{p}, r)$ is open. Suppose we have $\vec{x} \in B(\vec{q}, \delta)$. By the triangle inequality

$$\|\vec{x} - \vec{p}\| \leq \|\vec{x} - \vec{q}\| + \|\vec{q} - \vec{p}\| < r - \|\vec{q} - \vec{p}\| + \|\vec{q} - \vec{p}\| = r$$

and so $\vec{x} \in B(\vec{p}, r)$. Hence $B(\vec{q}, \delta) \subset B(\vec{p}, r)$, and so $B(\vec{p}, r)$ is open. □

2.1 Open sets in subsets of $\mathbb{R}^m$

Definition 2.8. *Let $X \subset \mathbb{R}^m$ be a subset of Euclidean space. Given a point $\vec{p} \in X$, and a real number $r > 0$, we define the open ball in X of radius r, centred at $\vec{p}$ to be the set*

$$B_X(\vec{p}, r) := \{\vec{x} \in X \mid \|\vec{x} - \vec{p}\| < r\}.$$

We define the closed ball in X of radius r, centred at $\vec{p}$ to be the set

$$\overline{B}_X(\vec{p}, r) := \{\vec{x} \in X \mid \|\vec{x} - \vec{p}\| \leq r\}.$$

We call a subset $U \subset X$ open in X if for every $\vec{x} \in U$, there exists $r_{\vec{x}} > 0$ such that $B_X(\vec{x}, r_{\vec{x}}) \subset U$. The empty set is open by convention.

Note that

$$B_X(\vec{x}, r) = B(\vec{x}, r) \cap X.$$

As such, we have an immediate useful observation.

Lemma 2.9. *If U is open in $\mathbb{R}^m$, and $X \subset \mathbb{R}^m$, then $U \cap X$ is open in X.*

Proof. This is by definition if $U \cap X = \emptyset$. Otherwise, suppose $\vec{x} \in U \cap X$. As U is open in $\mathbb{R}^m$, there exists $r_{\vec{x}} > 0$ such that $B(\vec{x}, r_{\vec{x}}) \subset U$. Hence

$$B_X(\vec{x}, r_{\vec{x}}) = B(\vec{x}, r_{\vec{x}}) \cap X \subset U \cap X$$

for every $\vec{x} \in U \cap X$, and so $U \cap X$ is open in X. □

Example 2.10. *Similarly to Example 2.4, the set*

$$H = \{(x, y, z) \in \mathbb{R}^3 \mid z > 0\}$$

is open in $\mathbb{R}^3$. Letting

$$S^2 = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}$$

be the unit sphere, we see that

$$V = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1 \text{ and } z > 0\} = H \cap S^2$$

is open in S^2. However V is not open in $\mathbb{R}^3$, as there is no open ball in $\mathbb{R}^3$ centred at $(0, 0, 1)$ contained in V. Openness is relative to the ambient space.

Corollary 2.11. *Let $X \subset \mathbb{R}^m$, and $\vec{p} \in X$. Then, for any $r > 0$, the open ball $B_X(\vec{p}, r)$ is open in X.*

Proof. As note previously, $B_X(\vec{p}, r) = B(\vec{p}, r) \cap X$. As $B(\vec{p}, r)$ is open in $\mathbb{R}^m$, Lemma 2.9 tells us that $B_X(\vec{p}, r)$ is open in X. □

Lemma 2.12. *Let $X \subset \mathbb{R}^m$, $\vec{p} \in X$ and $r > 0$. Then the set*

$$X \setminus \overline{B}_X(\vec{p}, r) = \{\vec{x} \in X \mid \|\vec{x} - \vec{p}\| > r\}$$

is open in X.

Proof. We will, for variety, prove this directly. Let $\vec{q} \in X \setminus \overline{B}_X(\vec{p}, r)$. As $\|\vec{q} - \vec{p}\| > r$, we have that $\delta = \|\vec{q} - \vec{p}\| - r > 0$. We claim that $B_X(\vec{q}, \delta) \subset X \setminus \overline{B}_X(\vec{p}, r)$, and hence that $X \setminus \overline{B}_X(\vec{p}, r)$ is open in X. Suppose $\vec{x} \in B_X(\vec{q}, \delta)$. Then, by the reverse triangle inequality,

$$\|\vec{x} - \vec{p}\| \geq \|\vec{p} - \vec{q}\| - \|\vec{x} - \vec{q}\| > \|\vec{q} - \vec{p}\| - \delta = r$$

and hence $\vec{x} \in X \setminus \overline{B}_X(\vec{p}, r)$, from which all other claims follow. □

2.2 The topology of Euclidean space

Definition 2.13. *A topology on a set X is a collection τ of subsets of X, called open sets in X, such that*

i) Both X and the empty set are open: $\emptyset, X \in \tau$,

ii) The union of any collection of open sets is open:

$$\{U_i\}_{i \in I} \subset \tau \Rightarrow \bigcup_{i \in I} U_i \in \tau,$$

iii) The intersection of any finite collection of open sets is open:

$$U_1, \ldots, U_k \in \tau \Rightarrow \bigcap_{i=1}^{n} U_i \in \tau.$$

Proposition 2.14. *Let $X \subset \mathbb{R}^m$. Then the collection of open sets in X is a topology on X.*

Proof. The empty set $\emptyset$ is open, by definition. Since $B_X(\vec{x}, r) \subset X$ for all $\vec{x} \in X$ and $r > 0$, X must be open.

Let $\{U_i\}_{i \in I}$ be any collection of open sets in X, and suppose $\vec{x} \in \bigcup_{i \in I} U_i$. Then $\vec{x} \in U_{i_0}$ for some $i_0 \in I$. Since U_{i_0} is open in X, there exists $r_{\vec{x}} > 0$ such that

$$B_X(\vec{x}, r_{\vec{x}}) \subset U_{i_0} \subset \bigcup_{i \in I} U_i.$$

Hence, $\bigcup_{i \in I} U_i$ is open in X.

Finally, let $U_1, \ldots, U_k$ be a finite collection of open sets in X, and suppose $\vec{x} \in \bigcap_{i=1}^{k} U_i$. Then $\vec{x} \in U_i$ for each $1 \leq i \leq k$. Since these are open in X, there exist $r_1, \ldots, r_k > 0$ such that

$$B(\vec{x}, r_i) \subset U_i$$

for each $1 \leq i \leq k$. Let $r = \min\{r_1, \ldots, r_k\}$. Then

$$B_X(\vec{x}, r) \subset B_X(\vec{x}, r_i) \subset U_i$$

for each $1 \leq i \leq k$, and so

$$B_X(\vec{x}, r) \subset \bigcap_{i=1}^{k} U_i$$

and so the intersection is open. □

Example 2.15. *The set*

$$\{(x,y) \in \mathbb{R}^2 \mid x^2+y^2<4,\ y>1\} = B(\vec{0},2) \cap \{(x,y)\in\mathbb{R}^2 \mid y>1\}$$

is open in $\mathbb{R}^2$*. The set*

$$\{(x,y) \in \mathbb{R}^2 \mid (x-n)^2+y^2<1,\ n\in\mathbb{Z}\} = \bigcup_{n\in\mathbb{Z}} B\left((n,0),1\right)$$

is open in $\mathbb{R}^2$*.*

Example 2.16. *For* $k \geq 1$*, the set*

$$\{(x,y) \in \mathbb{R}^2 \mid k^2(x^2+y^2)<1\} = \bigcap_{i=1}^{k} B(\vec{0},\frac{1}{k})$$

is open in $\mathbb{R}^2$*, but* $\bigcap_{i=1}^{\infty} B(\vec{0},\frac{1}{k}) = \{\vec{0}\}$ *is not. Similarly, the set*

$$\{(x,y) \in \mathbb{R}^2 \mid (x^2+y^2)<1+k^{-2}\} = \bigcap_{i=1}^{k} B(\vec{0},\sqrt{1+k^{-2}})$$

is open in $\mathbb{R}^2$*, but*

$$\{(x,y) \in \mathbb{R}^2 \mid (x^2+y^2)\leq 1\} = \bigcap_{i=1}^{\infty} B(\vec{0},\sqrt{1+k^{-2}})$$

is not.

This properties give us an easier proof of a complete description of open sets in $X \subset \mathbb{R}^m$

Proposition 2.17. *Suppose we have* $W \subset X \subset \mathbb{R}^m$*. Then* W *is open in* X *if and only if there exists* U *open in* $\mathbb{R}^m$ *such that* $W = U \cap X$*.*

Proof. We have seen in Lemma 2.9 that if U is open in $\mathbb{R}^m$, the set $W = U \cap X$ is open in X. Suppose we are given W open in X. Then for every $\vec{x} \in W$, there exists $r_{\vec{x}} > 0$ such that $B_X(\vec{x}, r_{\vec{x}}) \subset W$. Define $U = \bigcup_{\vec{x}\in W} B(\vec{x}, r_{\vec{x}}$ to be the union of the corresponding open balls in $\mathbb{R}^m$. As this is a union of open balls, which are open, this is open in $\mathbb{R}^m$. I claim that $U \cap X = W$. Suppose $\vec{p} \in U \cap X$. Then $\vec{p} \in X$, and $\vec{p} \in B(\vec{x}, r_{\vec{x}})$ for some $\vec{x} \in W$. Therefore

$$\vec{p} \in B(\vec{x}, r_{\vec{x}}) \cap X = B_X(\vec{x}, r_{\vec{x}}) \subset W$$

because of how we chose $r_{\vec{x}}$. Thus $U \cap X \subset W$. Clearly

$$\vec{x} \subset B_X(\vec{x}, r_{\vec{x}}) \subset B(\vec{x}, r_{\vec{x}}) \subset U$$

for every $\vec{x} \in W$, and so $W \subset U$ and hence $W \subset U \cap X$. Therefore $W = U \cap X$. □

2.3 Closed sets

Definition 2.18. *Let $X \subset \mathbb{R}^m$. We call a subset $F \subset X$ closed in X if and only if $X \setminus F$ is open in X.*

Example 2.19. *The closed ball*

$$\overline{B}(\vec{p}, r) = \{\vec{x} \in \mathbb{R}^m \mid \|\vec{x} - \vec{p}\| \leq r\} = \mathbb{R}^m \setminus \{\vec{x} \in \mathbb{R}^m \mid \|\vec{x} - \vec{p}\| > r\}$$

is closed in $\mathbb{R}^m$.

Example 2.20. *The set*

$$\{(x, y) \in \mathbb{R}^2 \mid y \leq c\} = \mathbb{R}^2 \setminus \{(x, y) \in \mathbb{R}^2 \mid y > c\}$$

is closed.

Example 2.21. *The singleton sets*

$$\begin{aligned} \{\vec{p}\} &= \mathbb{R}^m \setminus \{\vec{x} \in \mathbb{R}^m \mid \|\vec{x} - \vec{p}\| > 0\} \\ &= \mathbb{R}^m \setminus \bigcup_{n \geq 1} \{\vec{x} \in \mathbb{R}^m \mid \|\vec{x} - \vec{p}\| > \frac{1}{n}\} \end{aligned}$$

are closed.

Remark 2.22. *It is really important to note that not open is not the same as closed. Sets can be neither open nor closed, or both open and closed. For example, the set $[0, 1)$ is neither open nor closed in $\mathbb{R}$, while the set $(0, 1)$ is both open and closed in $(0, 1) \cup (2, 3)$.*

As closed sets are complementary, they satisfy dual properties to that of a topology, analogous to those proved in Proposition 2.14.

Proposition 2.23. *Let $X \subset \mathbb{R}^m$. The collection of closed sets in X satisfies the following properties:*

i) The empty set $\emptyset$ and X are closed in X,

ii) The intersection of any collection of closed sets in X is closed in X,

iii) The union of any finite collection of closed sets in X is closed in X.

Proof. The set $\emptyset = X \setminus X$ is open as X is open. Similarly, $X = X \setminus \emptyset$. Recall that, for any collection of sets $\{Y_i\}_{i \in I}$ of subsets of X

$$\begin{aligned} X \setminus \bigcap_{i \in I} Y_i &= \bigcup_{i \in I} (X \setminus Y_i)\,, \\ X \setminus \bigcup_{i \in I} Y_i &= \bigcap_{i \in I} (X \setminus Y_i)\,. \end{aligned}$$

(If you do not recall this, please prove it as an exercise)

Hence, given any collection of closed sets $\{F_i\}_{i\in I}$, we have that

$$X \setminus \bigcap_{i\in I} F_i = \bigcup_{i\in I} (X \setminus F_i)$$

is a union of open sets and therefore open. Hence $\bigcap_{i\in I} F_i$ is closed. Similarly, the finite union $\bigcup_{i=1}^k F_i$ of closed sets is the complement of a finite intersection of open sets, which is open, and therefore $\bigcup_{i=1}^k F_i$ is closed. □

2.4 Topology and convergence

Lemma 2.24. *A sequence $\{\vec{x}_n\}$ of points in $\mathbb{R}^m$ converges to a point $\vec{p}$ if and only if for any open set U containing $\vec{p}$, there exists $N > 0$ such that $\vec{x}_n \in U$ for every $n \geq N$.*

Proof. Suppose that for any open U containing $\vec{p}$, there exists $N > 0$ such that $\vec{x}_n \in U$ for every $n \geq N$. Then, in particular, for every $\varepsilon > 0$, there exists $N > 0$ such that $\vec{x}_n \in B(\vec{p}, \varepsilon)$ for every $n \geq N$, or equivalently, $\|\vec{x}_n - \vec{p}\| < \varepsilon$ for every $n \geq N$. That is to say that $\lim_{n\to\infty} \vec{x}_n = \vec{p}$.

Now suppose $\{\vec{x}_n\}$ converges to $\vec{p}$, and let U be an open set containing $\vec{p}$. Since U is open, there exists $r > 0$ such that $B(\vec{p}, r) \subset U$. Since $\{\vec{x}_n\}$ converges to $\vec{p}$, there exists $N > 0$ such that for all $n \geq N$, $\|\vec{x}_n - \vec{p}\| < r$, or equivalently $\vec{x}_n \in B(\vec{p}, r) \subset U$ for all $n \geq N$. □

Lemma 2.25. *Suppose $X \subset \mathbb{R}^m$, and let $F \subset X$ be closed in X. If $\{\vec{x}_n\}$ is a sequence of points in F converging to a point $\vec{p} \in X$, then $\vec{p} \in F$.*

Proof. Since F is closed, $X \setminus F$ is open. If the limit point $\vec{p}$ is in $X \setminus F$, then $X \setminus F$ is an open set containing $\vec{p}$ and hence there exists $N \geq 0$ such that $\vec{x}_n \in X \setminus F$ for all $n \geq N$. But $\vec{x}_n \in F$ for every $n \geq 1$, and so we must have $\vec{p} \in F$. □

Example 2.26. *The set $(0, 1)$ is open in $\mathbb{R}$ and contains $\{\frac{1}{n}\}$, but not the limit point 0. In contrast $[0, 1]$ is closed in $\mathbb{R}$ and contains the limit point 0.*

3 Continuous functions in several real variables

3.1 Continuity at a point

Definition 3.1. *Let $X \subset \mathbb{R}^m$ and $Y \subset \mathbb{R}^n$. A function $\varphi : X \to Y$ is called continuous at a point $\vec{p} \in X$ if for every $\varepsilon > 0$, there exists $\delta > 0$ such that $\|\varphi(\vec{x}) - \varphi(\vec{p})\| < \varepsilon$ for every $\vec{x} \in X$ with $\|\vec{x} - \vec{p}\| < \delta$.*

We say that $\varphi : X \to Y$ is continuous on X if φ is continuous at every point $\vec{p}$ of X.

All of the expected properties of continuity hold in the setting of several variables, and the proofs are largely similar to the one dimensional case. A very useful fact is that the composition of continuous functions is continuous.

Proposition 3.2. *Let $X \subset \mathbb{R}^m$, $Y \subset \mathbb{R}^n$, $Z \subset \mathbb{R}^s$, and suppose we have functions $\varphi : X \to Y$ and $\psi : Y \to Z$ such that φ is continuous at $\vec{p} \in X$ and ψ is continuous at $\phi(\vec{p}) \in Y$. Then the composition*

$$(\psi \circ \varphi) : X \to Z$$

is continuous at $\vec{p}$.

Proof. Let $\varepsilon > 0$ be given. Then, by continuity of ψ, there exists $\eta > 0$ such that $\|\psi(\vec{y}) - \psi(\varphi(\vec{p}))\| < \varepsilon$ for all $\vec{y} \in Y$ such that $\|\vec{y} - \varphi(\vec{p})\| < \eta$. By continuity of φ, there exists $\delta > 0$ such that $\|\varphi(\vec{x}) - \varphi(\vec{p})\| < \eta$ for all $\vec{x} \in X$ such that $\|\vec{x} - \vec{p}\| < \delta$. Thus, for all $\vec{x} \in X$, with $\|\vec{x} - \vec{p}\| < \delta$, we have that $\|\varphi(\vec{x}) - \varphi(\vec{p})\| < \eta$ and hence

$$\|(\psi \circ \varphi)(\vec{x}) - (\psi \circ \varphi)(\vec{p})\| = \|\psi(\varphi(\vec{x})) - \psi(\varphi(\vec{p}))\| < \epsilon.$$

Hence, $(\psi \circ \varphi)$ is continuous at $\vec{p}$. □

Continuity can also be classified in terms of limits, as described in the following two propositions.

Proposition 3.3. *Let $X \subset \mathbb{R}^m$ and $Y \subset \mathbb{R}^n$, and let $\varphi : X \to Y$ be a continuous function, continuous at a point $\vec{p} \in X$. Suppose $\{\vec{x}_k\}$ is a sequence in X converging to the point $\vec{p}$. Then the sequence $\{\varphi(\vec{x}_k)\}$ converges to $\varphi(\vec{p})$ in Y.*

Proof. Let $\varepsilon > 0$. By the continuity of φ, there exists $\delta > 0$ such that $\|\varphi(\vec{x}) - \varphi(\vec{p})$ for all $\vec{x} \in X$ such that $\|\vec{x} - \vec{p}\| < \delta$. Since $\vec{x}_k \to \vec{p}$, there exists $N > 0$ such that $\|\vec{x}_k - \vec{p}\| < \delta$ for all $k \geq N$. Hence, for all $k \geq N$, $\|\varphi(\vec{x}_k) - \varphi(\vec{p})\| < \varepsilon$, and so $\{\varphi(\vec{x}_k)\}$ converges to $\varphi(\vec{p})$. □

Proposition 3.4. *A function $\varphi : X \to Y$, where $X \subset \mathbb{R}^m$, $Y \subset \mathbb{R}^n$ is continuous at a point $\vec{p} \in X$ if and only if for every sequence $\{\vec{x}_k\}$ in X converging to $\vec{p}$, the sequence $\{\varphi(\vec{x}_k)\}$ converges to $\varphi(\vec{p})$.*

Proof. One direction of the equivalence is follows from Proposition 3.3. To prove the other, suppose that for every sequence $\{\vec{x}_k\}$ in X converging to $\vec{p}$, the sequence $\{\varphi(\vec{x}_k)\}$ converges to $\varphi(\vec{p})$, but φ is not continuous at $\vec{p}$. If φ is not continuous at $\vec{p}$, there must exist some $\varepsilon > 0$ such that, for every $\delta > 0$, there exists $\vec{x}_\delta \in X$ such that $\|\vec{x}_\delta - \vec{p}\| < \delta$, but $\|\varphi(\vec{x}_\delta) - \varphi(\vec{p})\| \geq \epsilon$. In particular, there will exist $\varepsilon > 0$ such that for every $k > 0$, there exists $\vec{x}_k$ with $\|\vec{x}_k - \vec{p}\| < \frac{1}{k}$, but $\|\varphi(\vec{x}_k) - \varphi(\vec{p})\| \geq \varepsilon$.

Clearly the sequence $\{\vec{x}_k\}$ must converge to $\vec{p}$, and so the sequence $\{\varphi(\vec{x}_k)\}$ must converge to $\varphi(\vec{p})$. But $\|\varphi(\vec{x}_k) - \varphi(\vec{p})\| \geq \varepsilon$, and so it cannot converge to $\varphi(\vec{p})$, a contradiction. Thus, φ must be continuous at $\vec{p}$. □

Remark 3.5. *As it gets a bit tedious to constantly state the assumption that $X \subset \mathbb{R}^m$ and $Y \subset \mathbb{R}^n$, we will usually omit it from this point onwards, unless there is any risk of ambiguity.*

As convergence of sequences can be considered component-wise, we must therefore be able to consider continuity in terms of components. Let $X \subset \mathbb{R}^m$ and $Y \subset \mathbb{R}^n$, and let $\varphi : X \to Y$ be a function between them. We can write

$$\varphi(\vec{x}) = (f_1(\vec{x}), f_2(\vec{x}), \dots, f_n(\vec{x}))$$

and we call the functions $f_i : X \to \mathbb{R}$ the components of φ.

Lemma 3.6. *Let $\pi_i : \mathbb{R}^m \to \mathbb{R}$ be the function given by projection onto the i^{th} component. Then π_i is continuous.*

Proof. Let $\varepsilon > 0$. By Lemma 1.4, $|x_i - p_i| \le \|\vec{x} - \vec{p}\|$ for each $1 \le i \le n$. Hence

$$|\pi_i(\vec{x}) - \pi_i(\vec{p})| = |x_i - p_i| \le \|\vec{x} - \vec{p}\| < \varepsilon$$

for all $\|\vec{x} - \vec{p}\| < \varepsilon$. This π_i is continuous at every $\vec{p} \in \mathbb{R}^m$, and is therefore continuous. □

Proposition 3.7. *Let $\varphi : X \to Y$ have components f_i, $1 \le i \le n$. Then φ is continuous at $\vec{p} \in X$ if and only if f_i is continuous at $\vec{p}$ for each $1 \le i \le n$.*

Proof. If φ is continuous at $\vec{p}$, then $f_i = \pi_i \circ \varphi$ is the composition of functions continuous at $\vec{p}$ and $\varphi(\vec{p})$, and is therefore continuous at $\vec{p}$.

Conversely, suppose f_i is continuous at $\vec{p}$ for each $1 \le i \le n$. Then, for any $\varepsilon > 0$ there exist $\delta_i > 0$ such that

$$|f_i(\vec{x}) - f_i(\vec{p})| < \frac{\varepsilon}{\sqrt{n}} \quad \text{for all} \|\vec{x} - \vec{p}\| < \delta_i.$$

Letting $\delta = \min\{\delta_1, \dots, \delta_n\}$, we must have that, for all $\vec{x} \in X$ with $\|\vec{x} - \vec{p}\| < \delta$, that

$$\|\varphi(\vec{x}) - \varphi(\vec{p})\|^2 = \sum_{i=1}^{n} (f_i(\vec{x}) - f_i(\vec{p}))^2 < \sum_{i=1}^{n} \frac{\varepsilon^2}{n} = \varepsilon^2$$

and hence $\|\varphi(\vec{x}) - \varphi(\vec{p})\| < \varepsilon$. Thus, φ is continuous at $\vec{p}$. □

3.2 Combining continuous functions

Projection onto a component is one important example of a continuous function that lets us deduce continuity properties of more general functions. Continuity of arithmetic functions allows us to further simplify considerations.

Lemma 3.8. *Let*

$$\begin{aligned} s : \mathbb{R}^2 &\to \mathbb{R} \\ (x, y) &\mapsto x + y \end{aligned}$$

be the sum function, and

$$\begin{aligned} m : \mathbb{R}^2 &\to \mathbb{R} \\ (x, y) &\mapsto xy \end{aligned}$$

be the multiplication function. Then both s and m are continuous.

Proof. Let $(u, v) \in \mathbb{R}^2$, and let $\varepsilon > 0$ and $\delta = \frac{\varepsilon}{2}$. Then, if $\|(x, y) - (u, v)\| < \delta$, Lemma 1.4 tells us that

$$|x - u| < \delta \quad \text{and} \quad |y - v| < \delta.$$

Hence

$$\begin{aligned} |s(x, y) - s(u, v)| &= |x + y - u - v| \\ &= |(x - u) + (y - v)| \\ &\leq |x - u| + |y - v| \\ &< 2\delta = \varepsilon \end{aligned}$$

and so s is continuous at (u, v). Since (u, v) was arbitrary, s is continuous on $\mathbb{R}^2$.

For multiplication, it is a bit more difficult to find δ. Suppose we have

$$\|(x, y) - (u, v)\| < \delta \quad \text{so} \quad |x - u|,\ |y - v| < \delta$$

for some δ. Then

$$\begin{aligned} |m(x, y) - m(u, v)| &= |xy - uv| \\ &= |(x - u)(y - v) + u(y - v) + v(x - u)| \\ &\leq |x - u||y - v| + |u||y - v| + |v||x - u| \\ &< \delta(\delta + |u| + |v|) \end{aligned}$$

Without loss of generality, we can choose $\delta \leq 1$, so that

$$|m(x, y) - m(u, v) < \delta(1 + |u| + |v|)$$

which, given $\varepsilon > 0$, will be less than ε if

$$\delta \leq \frac{\varepsilon}{1 + |u| + |v|}.$$

Hence, given $\varepsilon > 0$, let $\delta = \min(1, \frac{\varepsilon}{1+|u|+|v|})$ to fulfill the conditions of continuity at (u, v). □

Proposition 3.9. *Let $f, g : X \to \mathbb{R}$ be continuous functions. Then, for all real a, b, $af + bg$ and $f \cdot g$ are continuous functions. If $g(\vec{x}) \neq 0$ for any $\vec{x} \in X$, then $\frac{f}{g}$ is also continuous.*

Proof. First note that the constant functions $\underline{c}(\vec{x}) := c$ for any $c \in \mathbb{R}$ are continuous, and so the function $af = m \circ (\underline{a}, f)$ is continuous, as $(\underline{a}, f)$ has continuous components, and m is continuous. Similarly, as $af + bg = s \circ (af, bg)$ is a composition of continuous functions, it is continuous, as is $f \cdot g = m \circ (f, g)$. For the final case, note that it suffices to show that for such $g : X \to \mathbb{R}$, the reciprocal $\frac{1}{g}$ is continuous. But we can easily show that

$$\begin{aligned} r : \mathbb{R} \setminus 0 &\to \mathbb{R} \\ t &\mapsto \frac{1}{t} \end{aligned}$$

is a continuous function, and hence $\frac{1}{g} = r \circ g$ is continuous. □

Lemma 3.10. *Let $\varphi : X \to Y$ be continuous. Then the map*

$$\begin{aligned}|\varphi| : X &\to \mathbb{R}\\ \vec{x} &\mapsto \|\varphi(\vec{x})\|\end{aligned}$$

is continuous.

Proof. Let $\varepsilon > 0$. Then, note that by the reverse triangle inequality

$$\|\varphi(\vec{x}) - \varphi(\vec{p}\| \geq \|\varphi(\vec{x})\| - \|\varphi(\vec{p}\|,$$
$$\|\varphi(\vec{p}) - \varphi(\vec{x})\| \geq \|\varphi(\vec{p})\| - \|\varphi(\vec{x})\|,$$

and hence

$$\|\varphi(\vec{x}) - \varphi(\vec{p})\| = \|\varphi(\vec{p}) - \varphi(\vec{x})\| \geq \left| \|\varphi(\vec{x})\| - \|\varphi(\vec{p})\| \right|$$

for all $\vec{x}, \vec{p} \in X$. Now, since φ is continuous, there exists $\delta > 0$ such that $\|\varphi(\vec{x}) - \varphi(\vec{p})\| < \varepsilon$ for all $\|\vec{x} - \vec{p}\| < \delta$, and hence

$$\left| \|\varphi(\vec{x})\| - \|\varphi(\vec{p})\| \right| < \varepsilon$$

for all $\vec{x} \in X$ such that $\|\vec{x} - \vec{p}\| < \delta$. □

Example 3.11. *Consider the function $\varphi : \mathbb{R}^2 \setminus \{(0,0)\} \to \mathbb{R}$ given by*

$$\varphi(x,y) = \left(\frac{x}{x^2 + y^2}, \frac{x - y}{x^2 + y^2 + 3} \right).$$

This is continuous if and only if it has continuous components. All of $f(x,y) = x$, $f(x,y) = x - y$, $f(x,y) = x^2 + y^2$, and $f(x,y) = x^2 + y^2 + 3$ are continuous, and neither of the denominators are 0 on $\mathbb{R}^2 \setminus \{(0,0)\}$. Thus φ is continuous.

3.3 Continuity and topology

Continuity on a set has a very nice definition in terms of how it interacts with open sets, very similar to the situation with sequences. First, let us rephrase the definition of continuity in terms of open balls. The condition that, for every $\varepsilon > 0$ there exists a $\delta > 0$ such that $\|\varphi(\vec{x}) - \varphi(\vec{p})\| < \varepsilon$ whenever $\|\vec{x} - \vec{p}\| < \delta$ is equivalent to the statement that, for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\varphi(\vec{x}) \in B_Y(\varphi(\vec{p}), \varepsilon) \quad \text{for all} \vec{x} \in B_X(\vec{p}, \delta).$$

Recall also that for any map $\varphi : X \to Y$, we define the preimage of $V \subset Y$ as the set

$$\varphi^{-1}(V) := \{\vec{x} \in X \mid \varphi(\vec{x}) \in V\}.$$

Proposition 3.12. *A function $\varphi : X \to Y$ is continuous if and only if $\varphi^{-1}(V)$ is open in X for every V open in Y.*

Proof. Suppose φ is continuous, and let $V \subset Y$ be open in Y. If $\varphi^{-1}(V) = \emptyset$, we are done, as this is always open. Otherwise, suppose we have a point $\vec{p} \in \phi^{-1}(V)$. Since V is open, there exist $\varepsilon > 0$ such that $B_Y(\varphi(\vec{p}), \varepsilon) \subset V$. By continuity of φ, there exists $\delta > 0$ such that $\varphi(\vec{x}) \subset B_Y(\varphi(\vec{p}), \varepsilon)$ whenever $\vec{x} \in B_X(\vec{p}, \delta)$. In particular, $\varphi(\vec{x}) \in V$ whenever $\vec{x} \in B_X(\vec{p}, \delta)$, and so

$$B_X(\vec{p}, \delta) \subset \phi^{-1}(V).$$

We can find such a ball around every $\vec{p} \in \phi^{-1}(V)$, and so $\phi^{-1}(V)$ is open. Let $\vec{p} \in X$. Then, given any $\varepsilon > 0$, the set $V = B_Y(\varphi(\vec{p}), \varepsilon)$ is open in Y. Hence $\phi^{-1}(V)$ is open in X, and so there exists $\delta > 0$ such that

$$B_X(\vec{p}, \delta) \subset \phi^{-1}(V).$$

As such, for any $\vec{x} \in B_X(\vec{p}, \delta)$, we have that

$$\varphi(\vec{x}) \in V = B_Y(\varphi(\vec{p}, \varepsilon)$$

and hence φ is continuous at $\vec{p}$ and therefore continuous on X. □

Corollary 3.13. *A function $\varphi : X \to Y$ is continuous if and only if $\varphi^{-1}(F)$ is closed in X for every closed set F in Y.*

Corollary 3.14. *Let $f : X \to \mathbb{R}$ be continuous, and let $a, b, c \in \mathbb{R}$ with $a < b$. Then the sets*

$$\{\vec{x} \in X \mid f(\vec{x}) > c\}, \quad \{\vec{x} \in X \mid f(\vec{x}) < c\},$$
$$\{\vec{x} \in X \mid a < f(\vec{x}) < b\}, \quad \{\vec{x} \in X \mid f(\vec{x}) \neq c\},$$

are all open in X, and the sets

$$\{\vec{x} \in X \mid f(\vec{x}) \geq c\}, \quad \{\vec{x} \in X \mid f(\vec{x}) \leq c\},$$

$$\{\vec{x} \in X \mid a \leq f(\vec{x}) \leq b\}, \quad \{\vec{x} \in X \mid f(\vec{x}) = c\},$$

are closed in X.

3.4 Extrema of functions $f : X \to \mathbb{R}$

In one real variable, a continuous function on a closed and bounded interval $[a, b]$ achieves both its minimum and maximum value on $[a, b]$. Our next goal is to prove a higher dimensional analogue of this, via the following two lemmas.

Lemma 3.15. *Let X be a non-empty, closed and bounded subset of $\mathbb{R}^m$, and let $f : X \to \mathbb{R}$ be a continuous function. If the image $f(X)$ is bounded above (resp. below), then there exists $\vec{u} \in X$ such that $f(\vec{u}) \geq f(\vec{x})$ (resp. $f(\vec{u}) \leq f(\vec{x})$) for all $\vec{x} \in X$.*

Proof. First note that the bounded-below case follows from the bounded-above case, by considering the continuous function $-f$. As such, we will only prove the bounded-above case. Let $L = \sup\{f(\vec{x}) \mid \vec{x} \in X\}$. As this is the least upper bound for $f(X)$, there must be a sequence $\{\vec{x}_k\}$ of points in X such that

$$L - \frac{1}{k} < f(\vec{x}_k) \leq L$$

for all $k \geq 1$. Since X is closed and bounded, the multi-dimensional Bolzano-Weierstrass theorem tells us that this contains a convergent subsequence $\{\vec{x}_{j_k}\}$ converging to some point $\vec{u} \in \mathbb{R}^m$. In fact, as X is closed, Lemma 2.25 tells us that $\vec{u} \in X$. Furthermore, since

$$L - \frac{1}{j_k} < f(\vec{x}_{j_k}) \leq L,$$

we must have that

$$f(\vec{(u)}) = f(\lim_{k\to\infty} \vec{x}_{j_k}) = \lim_{k\to\infty} f(\vec{x}_{j_k}) = L,$$

where we have used the continuity of f to move the limit outside. Hence $f(\vec{u}) = L \geq f(\vec{x})$ for all $\vec{x} \in X$, as L is an upper bound. □

Lemma 3.16. *Let X be a non-empty, closed and bounded subset of $\mathbb{R}^m$, and let $f : X \to \mathbb{R}$ be a continuous function. Then there exists $M > 0$ such that $|f(\vec{x})| \leq M$ for all $\vec{x} \in X$, i.e $f(X) \subset [-M, M]$.*

Proof. Define $g : X \to \mathbb{R}$ by

$$g(\vec{x}) := \frac{1}{1 + |f(\vec{x})|}$$

and note that this is continuous, as $1+|f(\vec{x})| \geq 1$ is never zero and is continuous. Furthermore, $g(\vec{x}) \geq 0$ for all $\vec{x} \in X$. Thus, by Lemma 3.15, there exists $\vec{w} \in X$ such that $g(\vec{x}) \geq g(\vec{w})$ for all $\vec{x} \in X$. A bit of algebra gives that we must therefore have

$$|f(\vec{x})| \leq |f(\vec{w})| =: M$$

as claimed. □

Theorem 3.17 (Extreme value theorem)**.** *Let X be a non-empty, closed and bounded subset of $\mathbb{R}^m$, and let $f : X \to \mathbb{R}$ be a continuous function. Then there exist $\vec{u}, \vec{v} \in X$ such that*

$$f(\vec{u}) \leq f(\vec{x}) \leq f(\vec{v}) \text{ for all } \vec{x} \in X.$$

Proof. By Lemma 3.16, the set $f(X)$ is bounded both above and below. Hence, the existence of the points $\vec{u}, \vec{v} \in X$ as claimed follows from Lemma 3.15. □

3.5 Uniform continuity in several real variables

Definition 3.18. *Let $\varphi : X \to Y$ be a function between two Euclidean spaces. We call φ uniformly continuous if for any $\varepsilon > 0$, there exists $\delta > 0$ such that $\|\varphi(\vec{u}) - \varphi(\vec{v})\| < \varepsilon$ for all $\vec{u}, \vec{v} \in X$ such that $\|\vec{u} - \vec{v}\| < \delta$. Note that δ cannot depend on $\vec{u}$ or $\vec{v}$.*

Example 3.19. *For the sum function $s : \mathbb{R}^2 \to \mathbb{R}$, we took $\delta = \frac{\varepsilon}{2}$. As this works for any pair of points in $\mathbb{R}^2$, we can conclude that s is uniformly continuous.*

For the multiplication function $m : \mathbb{R}^2 \to \mathbb{R}$, we took

$$\delta = \min\left(1, \frac{\varepsilon}{1 + |u| + |v|}\right)$$

to establish continuity at the point (u, v). This does not imply uniform continuity, but maybe a universal δ exists. I claim otherwise. If m were uniformly continuous, there would exists $\delta > 0$ such that

$$|xy - uv| < \frac{1}{2}$$

for any pair of points (x, y), (u, v) such that

$$(x - u)^2 + (y - v)^2 < \delta.$$

Consider the points $(x, y) = (\frac{\delta}{2}, v)$ and $(u, v) = (0, v)$, for some $v \in \mathbb{R}$. These clearly satisfy the above condition, but

$$|xy - uv| = |\frac{\delta}{2}||v| < \frac{1}{2}$$

cannot possibly hold for all v. Hence m cannot be uniformly continuous.

Theorem 3.20. *For $\varphi : X \to Y$ a continuous function on a non-empty, closed and bounded set X, φ is uniformly continuous.*

Proof. Let $\varepsilon > 0$, and suppose there does not exist $\delta > 0$ such that $\|\varphi(\vec{u}) - \varphi(\vec{v})\| < \varepsilon$ for every $\vec{u}, \vec{v} \in X$ such that $\|\vec{u} - \vec{v}\| < \delta$. Then, for every $\delta > 0$, there exists a pair $(\vec{u}_\delta, \vec{v}_\delta) \in X^2$ such that $\|\vec{u}_\delta - \vec{v}_\delta\| < \delta$, but $\|\varphi(\vec{u}_\delta) - \varphi(\vec{v}_\delta)\| \geq \varepsilon$. In particular, we must have sequences $\{\vec{u}_k\}$ and $\{\vec{v}_k\}$ in X such that

$$\|\vec{u}_k - \vec{v}_k\| < \frac{1}{k}$$

for every $k \geq 0$, but

$$\|\varphi(\vec{u}_k) - \varphi(\vec{v}_k)\| \geq \varepsilon.$$

As X is closed and bounded, the sequence $\{\vec{u}_k\}$ must contain a subsequence $\{\vec{u}_{j_k}\}$ converging to a point $\vec{p} \in X$. Since

$$\|\vec{u}_{j_k} - \vec{v}_{j_k}\| < \frac{1}{j_k} \to 0$$

we must have that $\{\vec{v}_{j_k}\}$ also converges to $\vec{p}$. Thus, by continuity, the sequences

$$\{\varphi(\vec{u}_{j_k})\} \quad \text{and} \quad \{\varphi(\vec{v}_{j_k})\}$$

must both converge to $\varphi(\vec{p})$. But

$$\|\varphi(\vec{u}_{j_k}) - \varphi(\vec{v}_{j_k})\| \geq \varepsilon,$$

giving a contradiction. Therefore, there must exist a $\delta > 0$ for every $\varepsilon > 0$ such that

$$\|\vec{u} - \vec{v}\| < \delta \Rightarrow \|\varphi(\vec{u}) - \varphi(\vec{v})\| < \varepsilon$$

and so φ is uniformly continuous on X.

□

4 Limits of functions in several real variables

Definition 4.1. *A point $\vec{p} \in \mathbb{R}^m$ is called a limit point of a set $X \subset \mathbb{R}^m$ if for all $\varepsilon > 0$ $B(\vec{p}, \varepsilon) \cap X$ contains a point other than $\vec{p}$, i.e. there exists $\vec{x} \in X$ such that $0 < \|\vec{x} - \vec{p}\| < \varepsilon$.*

Example 4.2. *The point 1 is a limit point of $[0, 1)$, as is $\frac{1}{2}$. The point 0 is a limit point of $\{\frac{1}{n}\}$. The point $(0, 0)$ is a limit point of*

$$\{(x, y) \in \mathbb{R}^2 \mid x, y > 0\}.$$

Remark 4.3. *Not every point in X is a limit point of X. Despite the fact that $0 \in \mathbb{Z}$ 0 is not a limit point of $\mathbb{Z}$. Points of X which are not limit points of X are called isolated points.*

Proposition 4.4. *A set $X \in \mathbb{R}^m$ is closed if and only if X contains all its limit points.*

Proof. Suppose $X \subset \mathbb{R}^m$ is closed ad let $\vec{p}$ be a limit point of X that is not contained in X. Then $\vec{p} \in \mathbb{R}^m \setminus X$ is a point contained in an open set, and hence there exists $\varepsilon > 0$ such that $B(\vec{p}, \varepsilon) \subset \mathbb{R}^m \setminus X$. But this means that

$$B(\vec{p}, \varepsilon) \cap X = \emptyset$$

and so $\vec{p}$ cannot be a limit point. This gives a contradiction and hence, X must contain all its limit points.

Conversely, suppose that X contains all its limit points, but X is not closed. Then $\mathbb{R}^m \setminus X$ is not open, and so there exists a point $\vec{p} \in \mathbb{R}^m \setminus X$ with no open ball around it contained within $\mathbb{R}^m \setminus X$, i.e. for all $\varepsilon > 0$ $B(\vec{p}, \varepsilon) \cap X$ is non-empty. Since $\vec{p} \notin X$, $B(\vec{p}, \varepsilon) \cap X$ contains a point other than $\vec{p}$, and so $\vec{p}$ is a limit point of X. This is a contradiction to our assumption, and we must have that X is closed. □

Definition 4.5. *Let $X \subset \mathbb{R}^m$ and let $\vec{p} \in \mathbb{R}^m$ be a limit point of X. Suppose we have a function $\phi : X \to \mathbb{R}^n$ and a point $\vec{v} \in \mathbb{R}^n$. The point $\vec{v}$ is called the limit of $\phi(\vec{x})$ as $\vec{x}$ tends to $\vec{p}$ if and only if for all $\varepsilon > 0$, there exists $\delta > 0$ such that $\|\phi(\vec{x}) - \vec{v}\| < \varepsilon$ for all $\vec{x} \in X$ satisfying $0 < \|\vec{x} - \vec{p}\| < \delta$. We write*

$$\lim_{\vec{x} \to \vec{p}} \phi(\vec{x}) = \vec{v}$$

in this scenario.

Remark 4.6. *Note that, if $\vec{p} \in X$, we do not care about the behaviour of ϕ at $\vec{p}$ when talking about the limit. The classical example is*

$$\begin{aligned} f : [0, \infty) &\to \mathbb{R} \\ x &\mapsto \begin{cases} 1 \text{ if } x > 0, \\ 0 \text{ if } x = 0 \end{cases} \end{aligned}$$

which has $f(0) = 0$, but $\lim_{x \to 0} f(x) = 1$.

As with every other property we have considered so far in this course, limits of functions can be considered entirely componentwise.

Proposition 4.7. *Let $X \subset \mathbb{R}^m$ have limit point $\vec{p}$, and let $\vec{v} = (v_1, \dots, v_n) \in \mathbb{R}^n$. A function*

$$\begin{aligned} \phi : X &\to \mathbb{R}^n, \\ \vec{x} &\mapsto (f_1(\vec{x}), f_2(\vec{x}), \dots, f_n(\vec{x})), \end{aligned}$$

has the property that

$$\lim_{\vec{x} \to \vec{p}} \phi(\vec{x}) = \vec{v}$$

if and only if

$$\lim_{\vec{x} \to \vec{p}} f_i(\vec{x}) = v_i$$

for each $1 \leq i \leq n$.

Proof. Suppose $\lim_{\vec{x} \to \vec{p}} \phi(\vec{x}) = \vec{v}$, and let $\varepsilon > 0$. Then there exists $\delta > 0$ such that $\|\phi(\vec{x}) - \vec{v}\| < \varepsilon$ for every $\vec{x} \in X$ such that $0 < \|\vec{x} - \vec{p}\| < \delta$. Then, by Lemma 1.4,

$$|f_i(\vec{x}) - v_i| \leq \|\phi(\vec{x}) - \vec{v}\| < \varepsilon$$

for all $\vec{x} \in X$ such that $0 < \|\vec{x} - \vec{p}\| < \delta$. Hence $\lim_{\vec{x} \to \vec{p}} f_i(\vec{x}) = v_i$ for each $1 \leq i \leq n$.

Conversely, if $\lim_{\vec{x} \to \vec{p}} f_i(\vec{x}) = v_i$ for each $1 \leq i \leq n$, then for any ε, there exist $\delta_1, \dots, \delta_n$ such that

$$|f_i(\vec{x}) - v_i| < \frac{\varepsilon}{\sqrt{n}}$$

for all $\vec{x} \in X$ such that $0 < \|\vec{x} - \vec{p}\| < \delta_i$, for each $1 \leq i \leq n$. Thus, for all $\vec{x} \in X$ such that

$$0 < \|\vec{x} - \vec{p}\| < \min\{\delta_1, \dots, \delta_n\}$$

we have that

$$\|\phi(\vec{x}) - \vec{v}\| = \left(\sum_{i=1}^{n}(f_i(\vec{x}) - v_i)^2\right)^{\frac{1}{2}} < \left(\sum_{i=1}^{n}\frac{\varepsilon^2}{n}\right)^{\frac{1}{2}} = \varepsilon$$

and so $\lim_{\vec{x}\to\vec{p}} \phi(\vec{x}) = \vec{v}$. □

Limits behave well with respect to linear combinations of functions, which can make evaluation of limits much easier.

Proposition 4.8. *Let $X \subset \mathbb{R}^m$ have limit point $\vec{p} \in \mathbb{R}^m$. Suppose we have two functions $\phi, \psi : X \to \mathbb{R}^n$ such that*

$$\lim_{\vec{x}\to\vec{p}} \phi(\vec{x}) = \vec{v}, \quad \textit{and} \quad \lim_{\vec{x}\to\vec{p}} \psi(\vec{p}) = \vec{w}$$

for some $\vec{v}, \vec{w} \in \mathbb{R}^n$. Then, for any $c \in \mathbb{R}$, we have that

$$\lim_{\vec{x}\to\vec{p}} (\phi(\vec{x}) + \psi(\vec{x})) = \vec{v} + \vec{w}, \quad \textit{and} \quad \lim_{\vec{x}\to\vec{p}} (c\phi(\vec{x})) = c\vec{v}.$$

Proof. Let $\varepsilon > 0$. Then there exist δ_1, δ_2 such that

$$\|\phi(\vec{x}) - \vec{v}\| < \frac{\varepsilon}{2} \text{ for all } \vec{x} \in X \text{ such that } 0 < \|\vec{x} - \vec{p}\| < \delta_1$$

and

$$\|\psi(\vec{x}) - \vec{w}\| < \frac{\varepsilon}{2} \text{ for all } \vec{x} \in X \text{ such that } 0 < \|\vec{x} - \vec{p}\| < \delta_2.$$

Hence, for all $\vec{x} \in X$ such that $0 < \|\vec{x} - \vec{p}\| < \min\{\delta_1, \delta_2\}$, we have that

$$\begin{aligned}\|\phi(\vec{x}) + \psi(\vec{x}) - \vec{v} - \vec{w}\| &\leq \|\phi(\vec{x}) - \vec{v}\| + \|\psi(\vec{x}) - \vec{w}\| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon\end{aligned}$$

and hence

$$\lim_{\vec{x}\to\vec{p}} (\phi(\vec{x}) + \psi(\vec{x})) = \vec{v} + \vec{w}.$$

If $c = 0$, there is nothing to prove, so we will assume $c \neq 0$. Then, for all $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\|\phi(\vec{x}) - \vec{v}\| < \frac{\varepsilon}{|c|}$$

for all $\vec{x} \in X$ with $0 < \|\vec{x} - \vec{p}\| < \delta$. Hence, for all $\vec{x} \in X$ with $0 < \|\vec{x} - \vec{p}\| < \delta$, we have that

$$\|c\phi(\vec{x}) - c\vec{v}\| = |c|\|\phi(\vec{x}) - \vec{p}\| < |c|\frac{\varepsilon}{|c|} = \varepsilon.$$

Thus, $\lim_{\vec{x}\to\vec{p}}(c\phi(\vec{x})) = c\vec{v}$. □

Limits and continuity are extremely closely related. As in the one dimensional case, a function if continuous at a point if and only if limits at that point behave nicely.

Lemma 4.9. *Let $X \subset \mathbb{R}^m$ have a limit point $\vec{p}$ contained within X. Then a function $\phi : X \to \mathbb{R}^n$ is continuous at $\vec{p}$ if and only if $\lim_{\vec{x}\to\vec{p}} \phi(\vec{x}) = \phi(\vec{p})$.*

Proof. Suppose ϕ is continuous at $\vec{p}$. Then for every $\varepsilon > 0$, there exists $\delta > 0$ such that $\|\phi(\vec{x}) - \phi(\vec{p})\| < \varepsilon$ for every $\vec{x} \in X$ such that $\|\vec{x} - \vec{p}\| < \delta$. In particular, this inequality hold for every $\vec{x} \in X$ such that $0 < \|\vec{x} - \vec{p}\| < \delta$, and hence $\lim_{\vec{x}\to\vec{p}} \phi(\vec{x}) = \phi(\vec{p})$.

Conversely, if $\lim_{\vec{x}\to\vec{p}} \phi(\vec{x}) = \phi(\vec{p})$, the only thing we need to check to see that ϕ is continuous at $\vec{p}$ is that for every $\varepsilon > 0$, $\|\phi(\vec{x}) - \phi(\vec{p})\| < \varepsilon$ for $\|\vec{x} - \vec{p}\| = 0$, i.e. for $\vec{x} = \vec{p}$, which clearly holds. □

Lemma 4.10. *Let $X \subset \mathbb{R}^m$ have a limit point $\vec{p}$, and let $Y \subset \mathbb{R}^n$ contain a point $\vec{v}$. If we have a function $\phi : X \to Y$ such that*

$$\lim_{\vec{x}\to\vec{p}} \phi(\vec{x}) = \vec{v}$$

and a function $\psi : Y \to \mathbb{R}^s$ continuous at $\vec{v}$, then

$$\lim_{\vec{x}\to\vec{p}} \psi(\phi(\vec{x})) = \psi(\vec{v}).$$

Proof. Let $\varepsilon > 0$. Then by continuity of ψ, there exists $\eta > 0$ such that $\|\psi(\vec{y}) - \psi(\vec{v})\| < \varepsilon$ for all $\vec{y} \in Y$ such that $\|\vec{y} - \vec{v}\| < \eta$. From the limit property, there exists $\delta > 0$ such that $\|\phi(\vec{x}) - \vec{v}\| < \eta$ for all $\vec{x} \in X$ such that $0 < \|\vec{x} - \vec{p}\| < \delta$. Hence, for all $\vec{x} \in X$ with $0 < \|\vec{x} - \vec{p}\| < \delta$, we must have $\|\psi(\phi(\vec{x})) - \psi(\vec{v})\| < \varepsilon$, and so

$$\lim_{\vec{x}\to\vec{p}} \psi(\phi(\vec{x})) = \psi(\vec{v}).$$

□

Remark 4.11. *This lemma follows from Lemma 4.9 in the event that $\vec{v}$ is a limit point of Y.*

Finally, we can give a very neat collection of properties we can use to compute componentwise limits.

Proposition 4.12. *Let $X \subset \mathbb{R}^m$ have a limit point $\vec{p}$, and let $f, g : X \to \mathbb{R}$ be two functions such that*

$$\lim_{\vec{x}\to\vec{p}} f(\vec{x}) \quad \text{and} \quad \lim_{\vec{x}\to\vec{p}} g(\vec{x})$$

exist. Then both

$$\lim_{\vec{x}\to\vec{p}} (f(\vec{x}) + g(\vec{x})) \quad \text{and} \quad \lim_{\vec{x}\to\vec{p}} (f(\vec{x})g(\vec{x}))$$

exist and

$$\lim_{\vec{x}\to\vec{p}} (f(\vec{x}) + g(\vec{x})) = \lim_{\vec{x}\to\vec{p}} f(\vec{x}) + \lim_{\vec{x}\to\vec{p}} g(\vec{x}),$$
$$\lim_{\vec{x}\to\vec{p}} (f(\vec{x})g(\vec{x})) = \left(\lim_{\vec{x}\to\vec{p}} f(\vec{x})\right)\left(\lim_{\vec{x}\to\vec{p}} g(\vec{x})\right).$$

If furthermore $g(\vec{x}) \neq 0$ *for all* $\vec{x} \in X$ *and* $\lim_{\vec{x}\to\vec{p}} g(\vec{x}) \neq 0$*, then*

$$\lim_{\vec{x}\to\vec{p}} \frac{f(\vec{x})}{g(\vec{x})} \text{ exists and } \lim_{\vec{x}\to\vec{p}} \frac{f(\vec{x})}{g(\vec{x})} = \frac{\lim_{\vec{x}\to\vec{p}} f(\vec{x})}{\lim_{\vec{x}\to\vec{p}} g(\vec{x})}.$$

Proof. Let $a = \lim_{\vec{x}\to\vec{p}} f(\vec{x})$ and $b = \lim_{\vec{x}\to\vec{p}} g(\vec{x})$. Define the function

$$\begin{aligned} \psi : X &\to \mathbb{R}^2, \\ \vec{x} &\mapsto (f(\vec{x}), g(\vec{x})). \end{aligned}$$

Since limits are defined componentwise, $\lim_{\vec{x}\to\vec{p}} \psi(\vec{x})$ exists and is equal to (a, b). Recall the addition function $s : \mathbb{R}^2 \to \mathbb{R}$ and the multiplication function $m : \mathbb{R}^2 \to \mathbb{R}$:

$$\begin{aligned} f(\vec{x}) + g(\vec{x}) &= (s \circ \psi)(\vec{x}), \\ f(\vec{x})g(\vec{x}) &= (m \circ \psi)(\vec{x}). \end{aligned}$$

Thus, the claim follows from continuity of s and m, and Lemma 4.10. Similarly, the quotient limit follows from continuity of the reciprocal function $r : \mathbb{R}\setminus\{0\} \to \mathbb{R}$, and so

$$\lim_{\vec{x}\to\vec{p}} \frac{1}{g(\vec{x})} = \frac{1}{b}$$

and the claim follows from the multiplication limit result. □

Remark 4.13. *We can actually reduce the restrictions on* $g(\vec{x})$ *in order for the quotient limit to exist to* $\lim_{\vec{x}\to\vec{p}} g(\vec{x}) \neq 0$*, as this would imply that* $g(\vec{x}) \neq 0$ *in some small open ball* B *in* X *centred at* $\vec{p}$*, and we can then apply Proposition 4.12 to* $\phi|_B$*.*

5 Recapping one dimensional derivatives

Derivatives capture local information about "nice enough" functions. We see this locality in the fact that almost every result about derivatives only needs a function to be defined in some small open set around a point of interest. As such, we cannot define derivatives for points too close to the boundary of the domain of definition of a function.

Definition 5.1. *Let* $D \subset \mathbb{R}^m$*. A point* $\vec{s} \in D$ *is called an interior point of* D *if there exists* $\varepsilon > 0$ *such that* $B(\vec{s}, \varepsilon) \subset D$*.*

We say that a function $\phi : D \to \mathbb{R}^n$ *is defined around* $\overline{s}$ *if* $\vec{s}$ *is an interior point of* D*.*

More generally, we talk about a function f *being defined around* $\vec{s}$ *if* $\vec{s}$ *is an interior point of the domain of* f*.*

Remark 5.2. *A function* f *is defined around a point* $\vec{s}$ *if and only if* f *is defined on some open ball centred at* $\vec{s}$*.*

In $\mathbb{R}$, a function f is defined around a real number s if and only if there is some $\delta > 0$ such that f is defined on $(s - \delta, s + \delta)$. This is precisely the space we need to define a derivative.

Definition 5.3. *Let $x_0 \in \mathbb{R}$, and let f be a real-valued function defined around x_0. We say that f is differentiable at x_0 with derivative $f'(x_0)$ if and only if the limit*

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

exists and is well defined.

If the derivative is defined for $f'(x_0 + h)$ is defined for small enough h, then f' defines a function around x_0. In this scenario, we can attempt to take a second derivative, and beyond.

Definition 5.4. *We say that f is k-times differentiable for $k \geq 1$ if the $(k-1)^{st}$ derivative exists around x_0, and the limit*

$$f^{(k)} = \lim_{h \to 0} \frac{f^{(k-1)}(x_0 + h) - f^{(k-1)}(x_0)}{h}$$

exists and is well defined.

Remark 5.5. *It is notationally convenient to define the zero-th derivative of f by $f^{(0)} = f$.*

It is very straightforward to show that derivatives are linear: if f and g are defined around x_0 and have derivatives at x_0, then the derivative of $f + g$ and cf exist for any $c \in \mathbb{R}$, and are given by

$$(f + g)'(x_0)) = f'(x_0) + g'(x_0), \quad \text{and} \quad (cf)'(x_0) = cf'(x_0).$$

Some other very important properties of derivatives for f and g as described include:

Proposition 5.6 (Product Rule)**.** *The product function $f \cdot g$ is differentiable at x_0 with derivative*

$$(f \cdot g)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0).$$

Proof. We only consider h small enough that $f(x_0+h)$ and $g(x_0+h)$ are defined. We can then write

$$\begin{aligned}&\frac{f(x_0 + h)g(x_0 + h) - f(x_0)g(x_0)}{h} \\ &= \left(\frac{f(x_0 + h) - f(x)}{h}\right) g(x_0 + h) + f(x_0)\left(\frac{g(x_0 + h) - g(x_0)}{h}\right).\end{aligned}$$

Since a function differentiable at x_0 is continuous at x_0, our existing results on limits give that

$$\begin{aligned}&\lim_{h\to 0}\frac{f(x_0+h)g(x_0+h)-f(x_0)g(x_0)}{h}\\&=\lim_{h\to 0}\left(\frac{f(x_0+h)-f(x)}{h}\right)\lim_{h\to 0}g(x_0+h)+f(x_0)\lim_{h\to 0}\left(\frac{g(x_0+h)-g(x_0)}{h}\right)\end{aligned}$$

and hence

$$(f\cdot g)'(x_0)=f'(x_0)g(x_0)+f(x_0)g'(x_0).$$

□

Proposition 5.7 (Quotient Rule)**.** *For $g(x)\neq 0$ for all x sufficiently close to x_0, and $g(x_0)\neq 0$, the quotient function $\frac{f}{g}$ is differentiable at x_0, with derivative*

$$\left(\frac{f}{g}\right)'(x_0)=\frac{f'(x_0)g(x_0)-f(x_0)g'(x_0)}{g(x_0)^2}$$

Proof. For h sufficiently small, we can write

$$\begin{aligned}&\frac{1}{h}\left(\frac{f(x_0+h)}{g(x_0+h)}-\frac{f(x_0)}{g(x_0)}\right)=\frac{1}{h}\left(\frac{f(x_0+h)g(x_0)-f(x_0)g(x_0+h)}{g(x_0+h)g(x_0)}\right)\\&=\left(\frac{f(x_0+h)-f(x)}{h}\right)\frac{g(x_0)}{g(x_0+h)g(x_0)}-\frac{f(x_0)}{g(x_0+h)g(x_0)}\left(\frac{g(x_0+h)-g(x_0)}{h}\right).\end{aligned}$$

Taking limits, we get that

$$\lim_{h\to 0}\frac{1}{h}\left(\frac{f(x_0+h)}{g(x_0+h)}-\frac{f(x_0)}{g(x_0)}\right)=\frac{f'(x_0)g(x_0)-f(x_0)g'(x_0)}{g(x_0)^2}.$$

□

Proposition 5.8 (Chain Rule)**.** *Let $x_0\in\mathbb{R}$, f be a function defined around and differentiable at x_0, and let g be a function defined around and differentiable at $f(x_0)$. Then $g\circ f$ is differentiable at x_0, and*

$$(g\circ f)'(x_0)=g'(f(x_0))f'(x_0).$$

Proof. Let $y_0=f(x_0)$, and let

$$G(y)=\begin{cases}\frac{g(y)-g(y_0)}{y-y_0} \text{ if } y\neq y_0,\\ g'(y_0) \text{ if } y=y_0.\end{cases}$$

We can easily check that

$$g(f(x_0+h))-g(f(x_0))=G(f(x_0+h))\left(f(x_0+h)-f(x_0)\right).$$

As $G(y)$ is continuous at y_0, $\lim_{h\to 0} G(f(x_0+h)) = g'(f(x_0))$, and hence

$$\lim_{h\to 0} \frac{g(f(x_0+h)) - g(f(x_0))}{h} = \lim_{h\to 0} G(f(x_0+h)) \lim_{h\to 0} \left(\frac{f(x_0+h) - f(x_0)}{h} \right)$$

and hence

$$(g \circ f)'(x_0) = g'(f(x_0)) f'(x_0).$$

□

5.1 Rolle and the Mean Value Theorem

Theorem 5.9 (Rolle's Theorem)**.** *Let $f : [a,b] \to \mathbb{R}$ be continuous on $[a,b]$ and differentiable on (a,b). Suppose further that $f(a) = f(b)$. Then there exists $c \in (a,b)$ such that $f'(c) = 0$.*

Proof. If $f(x) = 0$ for all $x \in [a,b]$, there is nothing to prove. So suppose there exists $x \in (a,b)$ such that $f(x) \neq 0$. Without loss of generality, we can assume that $f(x) > 0$ for some $x \in (a,b)$. By the extreme value theorem, f therefore achieves its maximum value on $[a,b]$ at some point $c \in (a,b)$. We claim that $f'(c) = 0$. From the maximality of $f(c)$,

$$\begin{aligned} f(c+h) - f(c) \leq 0 \quad &\text{for} \quad h < 0, \\ f(c+h) - f(c) \leq 0 \quad &\text{for} \quad h > 0, \end{aligned}$$

and so

$$\begin{aligned} \frac{f(c+h) - f(c)}{h} \geq 0 \quad &\text{for} \quad h < 0, \\ \frac{f(c+h) - f(c)}{h} \leq 0 \quad &\text{for} \quad h > 0. \end{aligned}$$

As such, we must have that

$$f'(c) \leq 0 \quad \text{and} \quad f'(c) \geq 0,$$

and so $f'(c) = 0$. □

The mean value theorem is a quick corollary of this.

Theorem 5.10 (Mean Value Theorem)**.** *Let $f : [a,b] \to \mathbb{R}$ be continuous on $[a,b]$ and differentiable on (a,b). Then there exists $c \in (a,b)$ such that*

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

Proof. Define

$$g(x) = f(x) - \frac{b-x}{b-a} f(a) - \frac{x-a}{b-a} f(b).$$

The function g satisfies the conditions of Rolle's theorem and so there exists $c \in (a, b)$ such that $g'(c) = 0$, and so

$$0 = f'(c) + \frac{1}{b-a} f(a) - \frac{1}{b-a} f(b)$$

and so

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

□

Corollary 5.11. *Let $f : [a, b] \to \mathbb{R}$ be as above. If $f'(x) > 0$ for all $x \in (a, b)$, then $f(b) > f(a)$.*

Corollary 5.12. *Let $f : [a, b] \to \mathbb{R}$ be as above. If $f'(x) = 0$ for all $x \in (a, b)$, then $f(x) = f(a)$ for all $x \in [a, b]$.*

Corollary 5.13. *Let $f : [a, b] \to \mathbb{R}$ be as above, and let $M \in \mathbb{R}$. If $f'(x) \leq M$ for all $x \in (a, b)$, then $f(x) \leq f(a) + M(x - a)$ for all $x \in [a, b]$.*

Corollary 5.14. *Let $f : [a, b] \to \mathbb{R}$ be as above, and let $M \in \mathbb{R}$. If $|f'(x)| \leq M$ for all $x \in (a, b)$, then $|f(b) - f(a)| \leq M|b - a|$.*

5.2 Real Analytic Functions

Definition 5.15. *We call a function $f : (a, b) \to \mathbb{R}$ real analytic on (a, b) if the n^{th} derivative $f^{(n)}(x)$ exists for all $x \in (a, b)$ and $n \geq 0$, and for every $x \in (a, b)$ there exists $\delta > 0$ such that*

$$f(x + h) = \sum_{n=0}^{\infty} f^{(n)}(x) \frac{h^n}{n!}$$

for all $|h| < \delta$. This power series is called the Taylor series of f around x.

Most functions we care about are real analytic on their domain of definition. This includes polynomials, trigonometric functions, exponentials, the natural logarithm, and rational functions. But not every function is real analytic, even if it has derivatives of every order. The classical counter example is

$$f(x) = \begin{cases} e^{-\frac{1}{x^2}} \text{ for } x \neq 0, \\ 0 \text{ otherwise.} \end{cases}$$

This has $f^{(n)}(0) = 0$ for every $n \geq 0$, and so the Taylor series around 0 converges everywhere and is uniformly 0, while $f(x)$ is decidedly non-zero away from 0.

5.2.1 Taylor's theorem and extrema

Lemma 5.16. *Let $x_0, h \in \mathbb{R}$ and let f be a k-times differentiable function defined on an open interval containing both x_0 and $x_0 + h$. Let $c_0, \ldots, c_{k-1} \in \mathbb{R}$ and define*

$$p(t) = f(x_0 + th) - \sum_{n=0}^{k-1} c_n t^n$$

for all t in some open interval D containing $[0, 1]$ such that $f(x_0 + th)$ is defined for all $t \in D$. Then $p^{(n)} = 0$ for all $0 \le n < k$ if and only if $c_n = f^{(n)}(x_0)\frac{h^n}{n!}$ for all $0 \le n < k$.

Proof. Differentiating $p(t)$ s times, we see that

$$p^{(s)}(t) = h^s f^{(s)}(x_0 + th) - \sum_{n=s}^{k-1} \frac{n!}{(n-s)!} c_n t^{n-s}.$$

Evaluating this at $t = 0$, we see that $p^{(s)}(0) = 0$ if and only if

$$h^s f^{(s)}(x_0) - s! c_s = 0$$

which is to say that $c_s = f^{(s)}(x_0)\frac{h^s}{s!}$. Hence, the claim follows. □

Theorem 5.17. *Let $x_0, h \in \mathbb{R}$ and let f be a k-times differentiable function defined on an open interval containing both x_0 and $x_0 + h$. Then*

$$f(x_0 + h) = \sum_{n=0}^{k-1} f^{(n)}(x_0)\frac{h^n}{n!} + f^{(k)}(x_0 + \theta h)\frac{h^k}{k!}$$

for some $\theta \in (0, 1)$.

Proof. Let D be an open interval containing $[0, 1]$ such that $f(x_0 + th)$ is defined for all $t \in D$, and define

$$p(t) = f(x_0 + th) - \sum_{n=0}^{k-1} f^{(n)}(x_0)\frac{(th)^n}{n!}$$

for all $t \in D$. By Lemma 5.16, $p^{(n)}(0) = 0$ for $n = 0, 1, \ldots, k-1$. Also define for $t \in D$

$$q(t) = p(t) - p(1)t^k.$$

Then $q^{(n)}(0) = p^{(n)}(0) = 0$ for each $n = 0, 1, \ldots, k-1$ and $q(1) = 0$. By Rolle's theorem there exists $\theta_1 \in (0, 1)$ such that $q'(\theta_1) = 0$. Hence, by Rolle's theorem, there exists $\theta_2 \in (0, \theta_1)$ such that $q^{(2)}(\theta_2) = 0$. Repeating this process, we construct a sequence

$$0 < \theta_k < \theta_{k-1} < \cdots < \theta_1 < 1$$

such that $q^{(n)}(\theta_n) = 0$ for each $n = 1, 2, \ldots, k$. In particular, for $n = k$, we get that

$$0 = q^{(k)}(\theta_k) = p^{(k)}(\theta_k) - k!p(1) = f^{(k)}(x_0 + \theta_k h)h^k - k!p(1).$$

Thus, evaluating $p(1)$ in terms of derivatives of f, we get that

$$f(x_0 + h) = p(1) + \sum_{n=0}^{k-1} f^{(n)}(x_0)\frac{h^n}{n!} = \sum_{n=0}^{k-1} f^{(n)}(x_0)\frac{h^n}{n!} + f^{(k)}(x_0 + \theta_k h)\frac{h^k}{k!}.$$

□

Corollary 5.18. *Let $f : (a, b) \to \mathbb{R}$ be a k-times continuously differentiable function, and let $x_0 \in (a, b)$. Then, for any $\varepsilon > 0$, there exist $\delta > 0$ such that*

$$\left| f(x_0 + h) - \sum_{n=0}^{k} f^{(n)}(x_0)\frac{h^n}{n!} \right| < \varepsilon |h|^k$$

for all h such that $x_0 + h \in (a, b)$ and $|h| < \delta$.

Proof. The k^{th} derivative is continuous at x_0, and so for all $\varepsilon > 0$ there exists $\delta > 0$ such that $x_0 + h \in (a, b)$ and

$$|f^{(k)}(x_0 + h) - f^{(k)}(x_0)| < k!\varepsilon$$

for all h such that $|h| < \delta$. In particular, if $|h| < \delta$ and $0 < \theta < 1$, we have that

$$|f^{(k)}(x_0 + \theta h) - f^{(k)}(x_0)| < k!\varepsilon.$$

Choosing θ to be the value from Taylor's Theorem, we have that

$$\begin{aligned} \left| f(x_0 + h) - \sum_{n=0}^{k} f^{(n)}(x_0)\frac{h^n}{n!} \right| &= \left| \sum_{n=0}^{k-1} f^{(n)}(x_0)\frac{h^n}{n!} + f^{(k)}(x_0 + \theta h)\frac{h^k}{k!} - \sum_{n=0}^{k} f^{(n)}(x_0)\frac{h^n}{n!} \right| \\ &= \frac{|h|^k}{k|} |f^{(k)}(x_0 + \theta h) - f^{(k)}(x_0)| < \varepsilon |h|^k. \end{aligned}$$

□

Taylor's theorem applied to twice-differentiable functions gives us a useful criteria for determining whether a point is a local maximum or minimum.

Lemma 5.19. *Let $f : (x_0 - c, x_0 + c) :\to \mathbb{R}$ be a twice differentiable function, and suppose that $f^{(2)}(x_0 + h) > 0$ for all $h \in (-c, c)$. Then*

$$f(x_0 + h) \geq f(x_0) + f'(x_0)h$$

for all $h \in (-c, c)$.

Proof. Taylor's theorem tells us that

$$f(x_0 + h) = f(x_0) + f'(x_0)h + f^{(2)}(x_0 + \theta h)\frac{h^2}{2}$$

for some $\theta \in (0, 1)$. By the hypotheses of the lemma, $f^{(2)}(x_0 + \theta h) \geq 0$, and $\frac{h^2}{2} > 0$, and so the claim follows. □

Corollary 5.20. *Let $f : (a, b) \to \mathbb{R}$ be a twice differentiable function and let $x_0 \in (a, b)$ be such that $f'(x_0) = 0$ and $f^{(2)}(x) > 0$ ($f^{(2)}(x) < 0$) for all x in a small interval around x_0. Then x_0 is a local minimum (maximum respectively).*

Proof. We consider only the case of the positive second derivative. From Lemma 5.19, we have that

$$f(x_0 + h) \geq f(x_0) + f'(x_0)h = f(x_0)$$

for all small enough h, and hence $f(x_0)$ is a local minimum. □

6 Derivatives of functions in several real variables

6.1 Directional derivatives and partial derivatives

Before we define a notion of a total derivative for a function $\varphi : X \to Y$ for $X \subset \mathbb{R}^m$ and $Y \subset \mathbb{R}^n$, we should spend a bit of time defining derivatives in terms of rates of changes in a certain direction, and the implications for the growth and continuity of the function.

Definition 6.1. *Let $X \subset \mathbb{R}^m$ and let $f : X \to \mathbb{R}$ be a function defined around a point $\vec{x}_0 \in X$. Let $\vec{v} \in \mathbb{R}^m$ be a unit vector. The directional derivative of f (in the direction of $\vec{v}$) at $\vec{x}_0$ is the limit*

$$\partial_{\vec{v}} f(\vec{x}_0) := \lim_{h \to 0} \frac{f(\vec{x}_0 + h\vec{v}) - f(\vec{x}_0)}{h}$$

assuming the limit exists.

Letting $\{\vec{e}_1, \ldots, \vec{e}_m\}$ be the standard basis of $\mathbb{R}^m$. We define the partial derivatives of the function

$$f(\vec{x}) = f(x_1, \ldots, x_m)$$

with respect to x_i to be the directional derivative

$$\partial_i f(\vec{x}_0) := \partial_{\vec{e}_i} f(\vec{x}_0).$$

Written in terms of the coordinates $(x_1, \ldots, x_m)$, the partial derivatives are

$$\frac{\partial f}{\partial x_i}(\vec{x}_0) = \lim_{h \to 0} \frac{f(x_{0,1}, \ldots, x_{0,i} + h, \ldots, x_{0,m}) - f(x_{0,1}, \ldots, x_{0,i}, \ldots, x_{0,m})}{h}.$$

Essentially every major result about derivatives of one variable has an analogue for partial derivatives, including the product, quotient, and chain rules.

While partial derivatives provide a substantial amount of information about the local behaviour of functions of several real functions, knowledge of the partial derivatives is weaker than knowledge of derivatives of functions in one variable. For example, for functions of one real variable, being differentiable at a point implies being continuous at that point. This is not the case for partial derivatives!

Example 6.2. *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by*

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

If $(x,y) \neq (0,0)$, then the partial derivatives of f are well defined at (x,y), and are given by

$$\frac{\partial f}{\partial x} = \frac{2y(x^2-y^2)}{(x^2+y^2)^2}, \quad \text{and} \quad \frac{\partial f}{\partial y} = \frac{-2x(x^2-y^2)}{(x^2+y^2)^2}.$$

We can also compute the partial derivatives at $(0,0)$:

$$\frac{\partial f}{\partial x}(0,0) = \lim_{h\to 0} \frac{0-0}{h} = 0,$$
$$\frac{\partial f}{\partial y}(0,0) = \lim_{h\to 0} \frac{0-0}{h} = 0,$$

as $f(x,0) = f(0,y) = 0$. Thus, the partial derivatives of f exist everywhere, but note that $f(x,x) = 1$ for all $x \neq 0$, and so f is not continuous at $(0,0)$.

We can even have that the restriction of a function to any line is continuous, and that the partial derivatives exist and are well defined everywhere, but that the function fails to be continuous everywhere.

Example 6.3. *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by*

$$f(x,y) = \begin{cases} \frac{2x^2y}{x^4+y^2} \text{ if } (x,y) \neq (0,0), \\ 0 \text{ if } (x,y) = (0,0). \end{cases}$$

We can again check that this is differentiable at every point of $\mathbb{R}^2$. Also note that for $(u,v) \neq (0,0)$ and $t \neq 0$

$$f(tu,tv) = \frac{2tu^2v}{t^2u^4+v^2}$$

and

$$\lim_{t\to 0} f(tu,tv) = 0,$$

so the limit as (x,y) approaches the origin along any line is 0. But $f(t,t^2) = 1$ for all $t \neq 0$, and so f cannot be continuous at $(0,0)$.

6.2 Growth of functions with bounded partial derivatives

For functions defined on sets containing straight lines, it is possible to give analogues of results such as Rolle's theorem and the mean value theorem, and in particular the results on the growth of functions.

Lemma 6.4. *Let $\vec{u}$ and $\vec{v}$ be points of $\mathbb{R}^m$. Then*

$$\sum_{i=1}^{m} |u_i - v_i| \leq \sqrt{m}\|\vec{u} - \vec{v}\|.$$

Proof. Define a vector $\vec{s} \in \mathbb{R}^m$ in terms of its components by

$$s_i := \begin{cases} \frac{|u_i - v_i|}{u_i - v_i} \text{ if } u_i \neq v_i, \\ 1 \text{ if } u_i = v_i. \end{cases}$$

Then

$$\langle \vec{u} - \vec{v}, \vec{s} \rangle = \sum_{i=1}^{m} |u_i - v_i|$$

and $\|\vec{s}\| = \sqrt{m}$. Hence, by the Cauchy-Schwarz inequality,

$$\sum_{i=1}^{m} |u_i - v_i| = \langle \vec{u} - \vec{v}, \vec{s} \rangle \leq \|\vec{u} - \vec{v}\| \cdot \|\vec{s}\| = \sqrt{m}\|\vec{u} - \vec{v}\|$$

□

Remark 6.5. *In the following discussion, whenever a derivative appears, there is the implicit assumption that it exists and is well defined at all points that we evaluate it at. We will not usually spell this out explicitly.*

Proposition 6.6. *Let X be a open subset of $\mathbb{R}^m$, and let $f : X \to \mathbb{R}$. Let $\vec{u} \neq \vec{v} \in X$ be such that*

$$(1-t)\vec{u} + t\vec{v} \in X$$

for all $t \in [0,1]$. Suppose there exists a positive constant $M > 0$ such that

$$|\partial_{\vec{n}} f((1-t)\vec{u} + t\vec{v})| \leq M$$

for every $t \in [0,1]$ where $\vec{n}$ is the unit vector in the direction $\vec{v} - \vec{u}$. Then

$$|f(\vec{u}) - f(\vec{v})| \leq M\|\vec{u} - \vec{v}\|.$$

Proof. Define a function $F : [0,1] \to \mathbb{R}$ by

$$F(t) = f(\vec{u} + t\|\vec{v} - \vec{u}\|\vec{n}).$$

It follows immediately that

$$F'(t) = \partial_{\vec{n}} f(\vec{u} + t\|\vec{v} - \vec{u}\|\vec{n})\|\vec{u} - \vec{v}\|.$$

From the hypothesis of the theorem, this is bounded in absolute value, and so by Corollary 5.14, we have that

$$|F(0) - F(1)| \leq M\|\vec{u} - \vec{v}\|$$

and thus

$$|f(\vec{u}) - f(\vec{v})| \leq M\|\vec{u} - \vec{v}\|.$$

□

Corollary 6.7. *Let X be a product of open intervals in $\mathbb{R}^m$, and let $f : X \to \mathbb{R}$ be a function which has partial derivatives defined everywhere in X. Suppose there exists a positive constant $M > 0$ such that*

$$\left|\frac{\partial f}{\partial x_i}(\vec{x})\right| \leq M$$

for every $\vec{x} \in X$ and each $1 \leq i \leq m$. Then

$$|f(\vec{u}) - f(\vec{v})| \leq \sqrt{m}M\|\vec{u} - \vec{v}\|$$

for every $\vec{u}$, $\vec{v} \in X$.

Proof. It follows immediately if $\vec{u} = \vec{v}$. Otherwise, we can find $\{(a_i, b_i)\}_{i=1}^m$ such that $a_i < u_i$, $v_i < b_i$ for each i. For each $0 \leq k \leq m$, define

$$\vec{w}_k = (w_{k,1}, \ldots, w_{k,m})$$

by

$$w_{k,i} = \begin{cases} u_i \text{ if } i > k, \\ v_i \text{ if } i \leq k. \end{cases}$$

This gives a collection of points in X such that $\vec{w}_0 = \vec{u}$, $\vec{w}_m = \vec{v}$ and $\vec{w}_{k-1}$ and $\vec{w}_k$ differ only in the k^{th} coordinate. The line segment joining each $\vec{w}_{k-1}$ and $\vec{w}_k$ is contained entirely within X, so by Proposition 6.6

$$|f(\vec{w}_{k-1}) - f(\vec{w}_k)| \leq M\|\vec{w}_{k-1} - \vec{w}_k\| = M|u_k - v_k|$$

for each $k = 1, \ldots, m$. The triangle inequality gives

$$|f(\vec{u}) - f(\vec{v})| = |f(\vec{w}_0) - f(\vec{w}_m)| \leq \sum_{k=1}^{m} |f(\vec{w}_{k-1}) - f(\vec{w}_k)|$$

and the result then follows on applying Lemma 6.4. □

Example 6.8. *This bound is the best we can do in general! For example, take*

$$f : \mathbb{R}^m \to \mathbb{R},$$
$$(x_1, \ldots, x_m) \mapsto x_1 + \cdots + x_m.$$

This has partial derivatives $\frac{\partial f}{\partial x_i} = 1$ *for each* i *at every point of* $\mathbb{R}^m$*. Thus, for every* $\vec{u}$, $\vec{v}$, *we have that*

$$|f(\vec{u}) - f(\vec{v})| \leq \sqrt{m}\|\vec{u} - \vec{v}\|.$$

Taking $\vec{u} = (0, 0, \dots, 0)$ *and* $\vec{v} = (1, 1, \dots, 1)$*, we see that this bound is achieved.*

This corollary can be extended to a result about functions mapping to $\mathbb{R}^n$ and gives us our first conditions for continuity in terms of partial derivatives.

Corollary 6.9. *Let* $X \subset \mathbb{R}^m$ *be a product of open intervals, and let* $\varphi : X \to \mathbb{R}^n$ *have components* $\varphi(\vec{x}) = (f_1(\vec{x}), \dots, f_n(\vec{x}))$*. Suppose there exists a positive constant* $M > 0$ *such that*

$$\left|\frac{\partial f_j}{\partial x_i}\right| \leq M$$

at every point of X*, for each* $1 \leq i \leq m$ *and* $1 \leq j \leq n$*. Then*

$$\|\varphi(\vec{u}) - \varphi(\vec{v})\| \leq \sqrt{mn}M\|\vec{u} - \vec{v}\|$$

for every $\vec{u}$, $\vec{v} \in X$.

Proof. This follows immediately from applying Corollary 6.7 to each term of

$$\|\varphi(\vec{u}) - \varphi(\vec{v})\|^2 = \sum_{j=1}^{n}(f_j(\vec{u}) - f_j(\vec{v}))^2.$$

□

Corollary 6.10. *Let* $X \subset \mathbb{R}^m$ *be an open set and let* $\varphi(\vec{x}) = (f_1(\vec{x}), \dots, f_n(\vec{x}))$ *be a function* $\varphi : X \to \mathbb{R}^n$*. Suppose there exists a positive constant* $M > 0$ *such that*

$$\left|\frac{\partial f_j}{\partial x_i}\right| \leq M$$

at every point of X*, for each* $1 \leq i \leq m$ *and* $1 \leq j \leq n$*. Then* φ *is continuous on* X*.*

Proof. Let $\vec{p} \in X$. Since X is open, there exists $c > 0$ such that $B(\vec{p}, c) \subset X$. Let V be the product of the open intervals

$$(p_1 - \frac{c}{\sqrt{2}}, p_1 + \frac{c}{\sqrt{2}}) \times (p_2 - \frac{c}{\sqrt{2}}, p_2 + \frac{c}{\sqrt{2}}) \times \cdots \times (p_m - \frac{c}{\sqrt{2}}, p_m + \frac{c}{\sqrt{2}})$$

and note that

$$B(\vec{p}, \frac{c}{2}) \subset V \subset B(\vec{p}, c).$$

Then, for all $\vec{x} \in B(\vec{p}, \frac{c}{2}) \subset V$, by Corollary 6.9, we have that

$$\|\varphi(\vec{x}) - \varphi(\vec{p})\| \leq \sqrt{mn}M\|\vec{x} - \vec{p}\|$$

and hence, for any $\varepsilon > 0$, we can take

$$\delta = \frac{\min(\frac{c}{2}, \frac{\varepsilon}{2})}{\sqrt{mn}M}$$

to prove continuity at $\vec{p}$. □

Remark 6.11. *The results of Corollary 6.7, and hence Corollaries 6.9 and 6.10, can be extended to convex open sets by covering the line segment joining two points with products of open intervals. I recommend drawing some pictures in two dimensions to illustrate this argument.*

6.3 Functions with continuous partial derivatives and first order approximation

As we have seen, just having partial derivatives is not sufficient to guarantee continuity of a function at a point. This is reflective of the failure of the gradient to give a first order approximation of the function around a point. In contrast, when we have continuous partial derivatives, we obtain quite a strong approximation result.

Definition 6.12. *Let X be an open set in $\mathbb{R}^m$, and let $f : X \to \mathbb{R}$ be a function whose first order partial derivatives are defined at a point $\vec{p} \in X$. The gradient of f at $\vec{p}$ is the element of $\mathbb{R}^m$ given by*

$$(\nabla f)_{\vec{p}} = (\partial_1 f(\vec{p}), \partial_2 f(\vec{p}), \ldots \partial_m f(\vec{p}))\,.$$

Proposition 6.13. *Let $X \subset \mathbb{R}^m$ be open, and let $f : X \to \mathbb{R}$ be such that all its first order partial derivatives exist at all points of X and are continuous at a point $\vec{p} \in X$. Then, for any $\varepsilon > 0$, there exists $\delta > 0$ such that*

$$|f(\vec{u}) - f(\vec{v}) - \langle (\nabla f)_{\vec{p}}, \vec{u} - \vec{v}\rangle| \leq \varepsilon \|\vec{u} - \vec{v}\|$$

for all $\vec{u}$, $\vec{v} \in X$ such that $\|\vec{u} - \vec{p}\| < \delta$ and $\|\vec{v} - \vec{p}\| < \delta$.

Proof. Let $\vec{p} = (p_1, \ldots, p_m)$ and define $g : X \to \mathbb{R}$ by

$$g(\vec{x}) = f(\vec{x}) - f(\vec{p}) - \langle (\nabla f)_{\vec{p}}, \vec{x} - \vec{p}\rangle.$$

Then, the function g has first order partial derivatives

$$\partial_i g(\vec{x}) = \partial_i f(\vec{x}) - \langle (\nabla f)_{\vec{p}}, \vec{e}_i\rangle$$

that are continuous at $\vec{p}$. Furthermore, $g(\vec{p}) = \partial_i g(\vec{p}) = 0$ for each $1 \leq i \leq m$.

Given $\varepsilon > 0$, the continuity of $\partial_i g$ at $\vec{p}$ and openness of X ensures that we can find δ small enough such that the product of open intervals

$$\{\vec{x} \in \mathbb{R}^m \mid |x_i - p_i| < \delta \text{ for each } 1 \leq i \leq m\}$$

is a subset of X, and such that for all $\vec{x} \in X$ satisfying $\|\vec{x} - \vec{p}\| < \delta$,

$$|\partial_i g(\vec{x})| \leq \frac{\varepsilon}{\sqrt{m}}$$

for each $1 \leq i \leq m$. As $\|\vec{u} - \vec{p}\| < \delta$ implies (by Lemma 1.4) $|u_i - p_i| < \delta$, and similarly $\|\vec{v} - \vec{p}\| < \delta$ implies that $|v_i - p_i| < \delta$, we can apply Corollary 6.7 to obtain that

$$|g(\vec{u}) - g(\vec{v})| < \varepsilon \|\vec{u} - \vec{v}\|$$

for all $\vec{u}$, $\vec{v} \in X$ such that $\|\vec{u} - \vec{p}\| < \delta$ and $\|\vec{v} - \vec{p}\| < \delta$. The result follows immediately. ∎

Corollary 6.14. *Let $X \subset \mathbb{R}^m$ be open, and let $f : X \to \mathbb{R}$ be such that all its first order partial derivatives exist at all points of X and are continuous at a point $\vec{p} \in X$. Then*

$$\lim_{\vec{x}\to\vec{p}} \frac{|f(\vec{x}) - f(\vec{p}) - \langle(\nabla f)_{\vec{p}}, \vec{x} - \vec{p}\rangle|}{\|\vec{x} - \vec{p}\|} = 0$$

We extend the notion of a gradient to functions valued in $\mathbb{R}^n$ by defining the output as a matrix rather than a vector.

Definition 6.15. *Let $X \subset \mathbb{R}^m$ be open, and suppose $\varphi : X \to \mathbb{R}^n$ is a function with components*

$$\varphi(\vec{x}) = (f_1(\vec{x}), \ldots, f_n(\vec{x}))\,.$$

Suppose further that the partial derivatives $\partial_j f_i$ exist at a point $\vec{p} \in X$ for each $1 \le i \le n$ and $1 \le j \le m$. Define the Jacobian of φ at $\vec{p}$ to be the $(n \times m)$-matrix with components

$$\Big((\mathrm{J}\,\varphi)_{\vec{p}}\Big)_{i,j} := \partial_j f_i(\vec{p}).$$

Corollary 6.16. *Let $X \subset \mathbb{R}^m$ be open and let*

$$\varphi(\vec{x}) = (f_1(\vec{x}), \ldots, f_n(\vec{x}))$$

be a function $\varphi : X \to \mathbb{R}^n$ such that the partial derivatives $\partial_j f_i$ exist in X and are continuous at a point $\vec{p} \in X$ for each $1 \le i \le n$ and $1 \le j \le m$.

Treating the Jacobian matrix $(\mathrm{J}\,\varphi)_{\vec{p}}$ as a linear transformation $\mathbb{R}^m \to \mathbb{R}^n$, then for any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\Big\|\varphi(\vec{u}) - \varphi(\vec{v}) - (\mathrm{J}\,\varphi)_{\vec{p}}\,(\vec{u} - \vec{v})\Big\| \le \varepsilon\|\vec{u} - \vec{v}\|$$

for all $\vec{u}$, $\vec{v} \in X$ such that $\|\vec{u} - \vec{p}\| < \delta$ and $\|\vec{v} - \vec{p}\| < \delta$.

Proof. Note that

$$\begin{aligned}\|\varphi(\vec{u}) - \varphi(\vec{v}) - (\mathrm{J}\,\varphi)_{\vec{p}}\,(\vec{u} - \vec{v})\|^2 \\ = \sum_{j=1}^{n} (f_j(\vec{u}) - f_j(\vec{v}) - \langle(\nabla f_j)_{\vec{p}}, \vec{u} - \vec{v}\rangle)^2.\end{aligned}$$

The result then follows immediately from applying Proposition 6.13 for $\frac{\varepsilon}{\sqrt{n}}$. □

Corollary 6.17. *Let $X \subset \mathbb{R}^m$ be open, and let*

$$\varphi(\vec{x}) = (f_1(\vec{x}), \ldots, f_n(\vec{x}))$$

be a function $\varphi : X \to \mathbb{R}^n$ such that the partial derivatives $\partial_j f_i$ exist in X and are continuous at a point $\vec{p} \in X$ for each $1 \le i \le n$ and $1 \le j \le m$. Then

$$\lim_{\vec{x}\to\vec{p}} \frac{\Big\|\varphi(\vec{u}) - \varphi(\vec{v}) - (\mathrm{J}\,\varphi)_{\vec{p}}\,(\vec{u} - \vec{v})\Big\|}{\|\vec{x} - \vec{p}\|} = 0$$

6.4 Derivatives of functions of several real variables

It turns out that these inequalities, and their associated limits, are precisely what we need to best define a good notion of differentiability of a function $\varphi : X \to \mathbb{R}^n$ – a function will be differentiable at a point if it can be well approximated by an (affine-)linear function at that point.

Definition 6.18. *Let $X \subset \mathbb{R}^m$ be an open set, and let $\varphi : X \to \mathbb{R}^n$ be a function. Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation, and let $\vec{p} \in X$ be a point. We say that φ is differentiable at $\vec{p}$ with derivative T if and only if*

$$\lim_{\vec{x}\to\vec{p}} \frac{\varphi(\vec{x}) - \varphi(\vec{p}) - T(\vec{x}-\vec{p})}{\|\vec{x}-\vec{p}\|} = 0.$$

Given a map $\varphi : X \to \mathbb{R}^n$ that is differentiable at $\vec{p} \in X$, we usually denote its derivative at $\vec{p}$ by $(\mathrm{D}\,\varphi)_{\vec{p}}$, and call this the total derivative of φ at $\vec{p}$. If φ is differentiable at every point of X, we say φ is differentiable on X.

Lemma 6.19. *Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation. Then T is differentiable at every point $\vec{p} \in \mathbb{R}^m$ with derivative $(\mathrm{D}\,T)_{\vec{p}} = T$.*

Proof. This is immediate as

$$T\vec{x} - T\vec{p} - T(\vec{x}-\vec{p}) = 0$$

for all $\vec{x}, \vec{p} \in \mathbb{R}^m$. □

Lemma 6.20. *Let $X \subset \mathbb{R}^m$ be an open set, and let $\varphi : X \to \mathbb{R}^n$ be a function on X. Then φ is differentiable at a point $\vec{p} \in X$, with derivative T if and only if for every $\varepsilon > 0$, there exists $\delta > 0$ such that*

$$\|\varphi(\vec{x}) - \varphi(\vec{p}) - T(\vec{x}-\vec{p})\| \le \varepsilon\|\vec{x}-\vec{p}\|$$

for all $\vec{x} \in X$ such that $\|\vec{x}-\vec{p}\| < \delta$.

Remark 6.21. *Observe that this is one of the few times we do not demand a strict inequality. This is basically to allow us to include the case $\vec{x} = \vec{p}$, without having to consider it as a special case.*

Proof. Suppose that for every $\varepsilon > 0$, there exists such a δ. Then, for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\|\varphi(\vec{x}) - \varphi(\vec{p}) - T(\vec{x}-\vec{p})\| \le \frac{\varepsilon}{2} < \varepsilon$$

for all $\vec{x} \in X$ such that $\|\vec{x}-\vec{p}\| < \delta$, and in particular for all $\vec{x} \in X$ such that $0 < \|\vec{x}-\vec{p}\| < \delta$. Hence

$$\frac{\|\varphi(\vec{x}) - \varphi(\vec{p}) - T(\vec{x}-\vec{p})\|}{\|\vec{x}-\vec{p}\|} < \varepsilon$$

for every $\vec{x} \in X$ such that $0 < \|\vec{x} - \vec{p}\| < \delta$, which is precisely to say that

$$\lim_{\vec{x} \to \vec{p}} \frac{\|\varphi(\vec{x}) - \varphi(\vec{p}) - T(\vec{x} - \vec{p})\|}{\|\vec{x} - \vec{p}\|} = 0$$

and hence

$$\lim_{\vec{x} \to \vec{p}} \frac{\varphi(\vec{x}) - \varphi(\vec{p}) - T(\vec{x} - \vec{p})}{\|\vec{x} - \vec{p}\|} = 0.$$

Conversely, if φ is differentiable at $\vec{p}$ with derivative T, then for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\frac{\|\varphi(\vec{x}) - \varphi(\vec{p}) - T(\vec{x} - \vec{p})\|}{\|\vec{x} - \vec{p}\|} = \left| \frac{\|\varphi(\vec{x}) - \varphi(\vec{p}) - T(\vec{x} - \vec{p})\|}{\|\vec{x} - \vec{p}\|} - 0 \right| < \varepsilon$$

for all $\vec{x} \in X$ such that $0 < \|\vec{x} - \vec{p}\| < \delta$. Hence

$$\|\varphi(\vec{x}) - \varphi(\vec{p}) - T(\vec{x} - \vec{p})\| < \varepsilon \|\vec{x} - \vec{p}\|$$

for all $\vec{x} \in X$ such that $0 < \|\vec{x} - \vec{p}\| < \delta$ and

$$\|\varphi(\vec{x}) - \varphi(\vec{p}) - T(\vec{x} - \vec{p})\| \leq \varepsilon \|\vec{x} - \vec{p}\|$$

for all $\vec{x} \in X$ such that $\|\vec{x} - \vec{p}\| < \delta$ $\square$

Lemma 6.22. *Let $X \subset \mathbb{R}^m$ be open, $\vec{p} \in X$, and $\varphi : X \to \mathbb{R}^n$ a function differentiable at $\vec{p}$. Then φ is continuous at $\vec{p}$.*

Proof. Denote the derivative of φ at $\vec{p}$ by $(\mathrm{D}\,\varphi)_{\vec{p}}$. From Lemma 6.20, there exists $c > 0$ such that

$$\|\varphi(\vec{x}) - \varphi(\vec{p}) - T(\vec{x} - \vec{p})\| \leq \varepsilon \|\vec{x} - \vec{p}\| \leq \|\vec{x} - \vec{p}\|$$

for all $0 < \|\vec{x} - \vec{p}\| < c$. Thus, for any $\varepsilon > 0$, let $\delta = \min(c, \varepsilon)$ to obtain that

$$[\|\varphi(\vec{x}) - \varphi(\vec{p}) - T(\vec{x} - \vec{p})\| \leq \varepsilon \|\vec{x} - \vec{p}\| \leq \frac{1}{2} \|\vec{x} - \vec{p}\| < \delta \leq \varepsilon$$

for all $\vec{x} \in X$ such that $\|\vec{x} - \vec{p}\| < \delta$. Hence

$$\lim_{\vec{x} \to \vec{p}} \varphi(\vec{x}) - \varphi(\vec{p}) - (\mathrm{D}\,\varphi)_{\vec{p}} = 0$$

and so

$$\lim_{\vec{x} \to \vec{p}} \varphi(\vec{x}) = \varphi(\vec{p}) + \lim_{\vec{x} \to \vec{p}} (\mathrm{D}\,\varphi)_{\vec{p}}(\vec{x} - \vec{p}).$$

From tutorial sheet 3, we know that linear transformations are continuous, so

$$\lim_{\vec{x} \to \vec{p}} (\mathrm{D}\,\varphi)_{\vec{p}}(\vec{x} - \vec{p}) = (\mathrm{D}\,\varphi)_{\vec{p}} \left(\lim_{\vec{x} \to \vec{p}} \vec{x} - \vec{p} \right) = \vec{0}.$$

By Lemma 4.9, φ is therefore continuous at $\vec{p}$. $\square$

6.4.1 Determining the derivative

From results such that Corollary 6.17, it is clear that for nice enough φ, the derivative must be given by Jacobian. We will show this formally, and then go on to show that this is essentially the only possibility.

Proposition 6.23. *Let $X \subset \mathbb{R}^m$ be open, let $\vec{p} \in X$ be a point, and let $\varphi : X \to \mathbb{R}^n$ be the function*

$$\varphi(\vec{x}) = (f_1(\vec{x}), f_2(\vec{x}), \ldots, f_n(\vec{x})).$$

Suppose that the partial derivatives $\partial_j f_i(\vec{x})$ are defined around $\vec{p}$ and continuous at $\vec{p}$. Then, φ is differentiable at $\vec{p}$ with derivative $(\mathrm{D}\,\varphi)_{\vec{p}} = (\mathrm{J}\,\varphi)_{\vec{p}}$.

Proof. From Corollary 6.17, we know that for φ as in the proposition statement

$$\lim_{\vec{x}\to\vec{p}} \frac{\left\| \varphi(\vec{u}) - \varphi(\vec{v}) - (\mathrm{J}\,\varphi)_{\vec{p}}\,(\vec{u} - \vec{v}) \right\|}{\|\vec{x} - \vec{p}\|} = 0$$

and hence

$$\lim_{\vec{x}\to\vec{p}} \frac{\varphi(\vec{u}) - \varphi(\vec{v}) - (\mathrm{J}\,\varphi)_{\vec{p}}\,(\vec{u} - \vec{v})}{\|\vec{x} - \vec{p}\|} = 0.$$

Therefore, φ is differentiable at $\vec{p}$ with derivative $(\mathrm{J}\,\varphi)_{\vec{p}}$. □

An important result that we have yet to show, and have implicitly been assuming is that the derivative of a differentiable function is uniquely defined. Fortunately, all our results still hold is we replace "the derivative" with "a derivative", but it will be better to show that the derivative is indeed uniquely determined by φ.

Proposition 6.24. *Let $X \subset \mathbb{R}^m$ be open, $\vec{p} \in X$, and let $\varphi : X \to \mathbb{R}^n$ be a function differentiable at $\vec{p}$. Then, for every $\vec{v} \in \mathbb{R}^m$*

$$(\mathrm{D}\,\varphi)_{\vec{p}}\vec{v} = \lim_{t\to 0} \frac{\varphi(\vec{p} + t\vec{v}) - \varphi(\vec{p})}{t},$$

and the derivative $(\mathrm{D}\,\varphi)_{\vec{p}}$ is uniquely determined by φ.

Proof. Let $\vec{v} \in \mathbb{R}^m$, and suppose we are given $\varepsilon > 0$. We should (and totally could) now separate out the case where $\|\vec{v}\| = 0$, but to get around this, we will introduce an auxiliary $\varepsilon_1 > 0$, chosen such that

$$\varepsilon_1 \|\vec{v}\| < \varepsilon.$$

Differentiability of φ at $\vec{p}$, and openness of X, ensures that there exists $\delta_1 > 0$ such that

$$\|\varphi(\vec{x}) - \varphi(\vec{p}) - (\mathrm{D}\,\varphi)_{\vec{p}}(\vec{x} - \vec{p})\| \le \varepsilon_1 \|\vec{x} - \vec{p}\|$$

and $\vec{x} \in X$ for all $\vec{x}$ such that $\|\vec{x} - \vec{p}\| < \delta_1$. We then choose $\delta > 0$ such that $\delta\|\vec{v}\| < \delta_1$. Then, for $0 < |t| < \delta$, $\|(\vec{p} + t\vec{v}) - \vec{p}\| < \delta_1$ and so

$$\|\varphi(\vec{p} + t\vec{v}) - \varphi(\vec{p}) - t(\mathrm{D}\,\varphi)_{\vec{p}}\vec{v}\| \le \varepsilon_1 |t| \|\vec{v}\| < \varepsilon |t|$$

and hence
$$\frac{\|\varphi(\vec{p}+t\vec{v})-\varphi(\vec{p})-t(\mathrm{D}\,\varphi)_{\vec{p}}\vec{v}\|}{|t|}<\varepsilon$$
for all $0<|t|<\delta$. This implies that
$$\lim_{t\to 0}\frac{\|\varphi(\vec{p}+t\vec{v})-\varphi(\vec{p})-t(\mathrm{D}\,\varphi)_{\vec{p}}\vec{v}\|}{|t|}=0$$
and hence
$$\lim_{t\to 0}\frac{\varphi(\vec{p}+t\vec{v})-\varphi(\vec{p})}{t}=(\mathrm{D}\,\varphi)_{\vec{p}}\vec{v}.$$
□

With this result, we can give an adjacent result to Proposition 6.23, where we trade the assumption of continuous partial derivatives near $\vec{p}$ for an assumption of differentiability at $\vec{p}$.

Corollary 6.25. *Let $X\subset\mathbb{R}^m$ be open, $\vec{p}\in X$, and let $\varphi:X\to\mathbb{R}^n$ be differentiable at $\vec{p}$. Suppose further that*
$$\varphi(\vec{x})=(f_1(\vec{x}),\ldots,f_n(\vec{x}))$$
has components such that $\partial_j f_i(\vec{p})$ exists. Then $(\mathrm{D}\,\varphi)_{\vec{p}}=(\mathrm{J}\,\varphi)_{\vec{p}}$.

Proof. Denote by $\vec{e}_1,\ldots,\vec{e}_m$ the standard basis of $\mathbb{R}^m$. The linear transformation $(\mathrm{D}\,\varphi)_{\vec{p}}$ is uniquely determined by its action on the elements of this basis. Writing it as a matrix with respect to the standard bases of $\mathbb{R}^m$ and $\mathbb{R}^n$, we see that the (i,j)-entry of $(\mathrm{D}\,\varphi)_{\vec{p}}$ is the i^{th} component of
$$(\mathrm{D}\,\varphi)_{\vec{p}}\vec{e}_j=\lim_{t\to 0}\frac{\varphi(\vec{p}+t\vec{e}_j)-\varphi(\vec{p})}{t}.$$
As limits can be computed componentwise, the i^{th} component of this is
$$\lim_{t\to 0}\frac{f_i(\vec{p}+t\vec{e}_j)-f_i(\vec{p})}{t}=\partial_j f_i(\vec{p})$$
which is precisely the (i,j)-entry of $(\mathrm{J}\,\varphi)_{\vec{p}}$. □

Example 6.26. *We will compute the derivative at a point $\vec{p}=\left(\begin{smallmatrix}p\\q\end{smallmatrix}\right)$ of*
$$\varphi\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}x^2-y^2\\2xy\end{pmatrix}$$
via limits, and verify that it is equal to the Jacobian
$$(\mathrm{J}\,\varphi)_{\vec{p}}=\begin{pmatrix}2p&-2q\\2q&2p\end{pmatrix}$$

in the standard basis. Since the Jacobian has continuous entries everywhere, this follows immediately from Proposition 6.23, but it is good practice to do it the old fashioned way. The goal will essentially to be to write

$$\varphi\begin{pmatrix}x\\y\end{pmatrix}-\varphi\begin{pmatrix}p\\q\end{pmatrix}$$

in the form

$$T\begin{pmatrix}x-p\\y-q\end{pmatrix}+E(x,y)$$

where $E(x,y)$ is some function such that of quadratic order in the difference.

$$\begin{aligned}\varphi\begin{pmatrix}x\\y\end{pmatrix}-\varphi\begin{pmatrix}p\\q\end{pmatrix} &= \begin{pmatrix}x^2-p^2-y^2+q^2\\2xy-2pq\end{pmatrix}\\ &= \begin{pmatrix}(x-p)(x+p)-(y-q)(y+q)\\2(x-p)(y-q)+2q(x-p)+2p(y-q)\end{pmatrix}\\ &= \begin{pmatrix}(x-p)(x-p+2p)-(y-q)(y-q+2q)\\2(x-p)(y-q)+2q(x-p)+2p(y-q)\end{pmatrix}\\ &\begin{pmatrix}(x-p)^2+2p(x-p)-(y-q)^2-2q(y-q)\\2(x-p)(y-q)+2q(x-p)+2p(y-q)\end{pmatrix}\\ &= \begin{pmatrix}2p&-2q\\2q&2p\end{pmatrix}\begin{pmatrix}x-p\\y-q\end{pmatrix}+\begin{pmatrix}(x-p)^2-(y-q)^2\\2(x-p)(y-q)\end{pmatrix}.\end{aligned}$$

Hence

$$\left\|\varphi\begin{pmatrix}x\\y\end{pmatrix}-\varphi\begin{pmatrix}p\\q\end{pmatrix}-\begin{pmatrix}2p&-2q\\2q&2p\end{pmatrix}\begin{pmatrix}x-p\\y-q\end{pmatrix}\right\|=\left\|\begin{pmatrix}(x-p)^2-(y-q)^2\\2(x-p)(y-q)\end{pmatrix}\right\|.$$

The right hand side is equal to

$$\begin{aligned}\sqrt{((x-p)^2-(y-q)^2)^2+4(x-p)^2(y-q)^2} &= \sqrt{(x-p)^4+2(x-p)^2(y-q)^2+(y-q)^4}\\ &= \sqrt{((x-p)^2+(y-q)^2)^2}=(x-p)^2+(y-q)^2\\ &= \|\vec{x}-\vec{p}\|^2.\end{aligned}$$

Thus, letting $T=\left(\begin{smallmatrix}2p&-2q\\2q&2p\end{smallmatrix}\right)$, we see that

$$\lim_{\vec{x}\to\vec{p}}\frac{\|\varphi(\vec{x})-\varphi(\vec{p})-T(\vec{x}-\vec{p})\|}{\|\vec{x}-\vec{p}\|}=\lim_{\vec{x}\to\vec{p}}\|\vec{x}-\vec{p}\|=0$$

and hence $(\mathrm{D}\,\varphi)_{\vec{p}}=T=(\mathrm{J}\,\varphi)_{\vec{p}}$.

6.5 Properties of total derivatives

Proposition 6.27. *Let $X\subset\mathbb{R}^m$ be open, and let $\varphi: X\to\mathbb{R}^n$ and $\psi: X\to\mathbb{R}^n$ be functions differentiable at a point $\vec{p}\in X$. Then, for any $c\in\mathbb{R}$, $\varphi+\psi$ and $c\varphi$ are differentiable at $\vec{p}$ with total derivatives*

$$(\mathrm{D}(\varphi+\psi))_{\vec{p}}=(\mathrm{D}\,\varphi)_{\vec{p}}+(\mathrm{D}\,\psi)_{\vec{p}}$$

and

$$(\mathrm{D}(c\varphi))_{\vec{p}} = c(\mathrm{D}\,\varphi)_p.$$

Proof. Since limits are linear, assuming everything involved converges, we have that

$$\begin{aligned}
&\lim_{\vec{x}\to\vec{p}} \frac{\varphi(\vec{x}) + \psi(\vec{x}) - \varphi(\vec{p}) - \psi(\vec{p}) - \left((\mathrm{D}\,\varphi)_{\vec{p}} + (\mathrm{D}\,\varphi)_{\vec{p}}\right)(\vec{x}-\vec{p})}{\|\vec{x}-\vec{p}\|} \\
=&\lim_{\vec{x}\to\vec{p}} \frac{\varphi(\vec{x}) - \varphi(\vec{p}) - (\mathrm{D}\,\varphi)_{\vec{p}}(\vec{x}-\vec{p})}{\|\vec{x}-\vec{p}\|} + \lim_{\vec{x}\to\vec{p}} \frac{\psi(\vec{x}) - \psi(\vec{p}) - (\mathrm{D}\,\varphi)_{\vec{p}}(\vec{x}-\vec{p})}{\|\vec{x}-\vec{p}\|} \\
=&0
\end{aligned}$$

from which the first claim follows. The second is similar. □

There are also analogues of the product and chain rule, at least where such things make sense, but in order to prove that we need some auxiliary results.

6.5.1 Operators norms and growth of differential operators

In the first homework, we saw a norm $\|\cdot\|_{HS}$ on matrices, called the Hilbert-Schmidt norm. Absolutely everything in this second can be done using the Hilbert-Schmidt norm, as all we need is a norm on linear operators such that

$$\|T\vec{x}\| \leq \|T\|_{op}\|\vec{x}\|$$

for every vector $\vec{x}$ on which T acts. However, in order to keep these notes mostly self contained, we will introduced an additional norm that can be used in more general contexts and is independent of a choice of basis.

Definition 6.28. *Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation. Define*

$$\|T\|_{op} = \sup_{\|\vec{x}\|=1} \{\|T\vec{x}\|\}$$

This is well defined, as the unit sphere is closed and bounded and linear transformations are continuous. Thus, by the extreme value theorem, the supremum exists and is attained.

Note that, for non-zero $\vec{x}$

$$\|T\vec{x}\| = \|\vec{x}\| \left\| T\frac{\vec{x}}{\|\vec{x}\|} \right\| \leq \|T\|_{op}\|\vec{x}\|.$$

This inequality also clearly holds for $\vec{x} = \vec{0}$, so the norm of T bounds the growth its image.

Lemma 6.29. *Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation, which we will identify with its matrix with respect to the standard bases of $\mathbb{R}^m$ and $\mathbb{R}^n$. Then $A = T^T T$ is a real symmetric $(m \times m)$-matrix with real eigenvalues. Denote by λ_{max} the maximal eigenvalue of A. Then $\lambda_{max} \geq 0$ and $\|T\|_{op} = \sqrt{\lambda_{max}}$.*

Proof. Since A is real symmetric, there exists an orthonormal change of basis of $\mathbb{R}^m$ diagonalising A, i.e. there exists an orthonormal matrix R such that $R^T AR$ is diagonal. Let $\vec{b}_1, \dots, \vec{b}_m$ be the associated basis with eigenvalues $\lambda_1, \dots, \lambda_m$, so that

$$\langle \vec{b}_j, A\vec{b}_i \rangle = \begin{cases} 0 \text{ if } i \neq j, \\ \lambda_i \text{ if } i = j. \end{cases}$$

Then, for any $\vec{x} = \sum_{i=1}^m u_i \vec{b}_i$ of unit norm, we have that

$$\begin{aligned} \|T\vec{x}\|^2 &= \langle T\vec{x}, T\vec{x} \rangle \\ &= \langle \vec{x}, T^T T\vec{x} \rangle \\ &= \sum_{i,j=1}^m u_i u_j \langle \vec{b}_j, A\vec{b}_i \rangle \\ &= \sum_{i=1}^m \lambda_i u_i^2 \leq \lambda_{max} \sum_{i=1}^m u_i^2 = \lambda_{max}, \end{aligned}$$

Thus $\|T\vec{x}\|^2 \leq \lambda_{max}$ for all unit vectors $\vec{x}$. Taking $\vec{x}$ to be a unit eigenvector associated to λ_{max}, we see that this upper bound is attained, and hence we must have

$$T\|_{op}^2 = \sup_{\|\vec{x}\|=1} \{\|T\vec{x}\|^2\} = \lambda_{max}.$$

□

Proposition 6.30. *Let $X \subset \mathbb{R}^m$ be open, let $\vec{p} \in X$ and let $\varphi : X \to \mathbb{R}^n$ be a function differentiable at $\vec{p}$. Choose $M > \|(\mathrm{D}\,\varphi)_{\vec{p}}\|_{op}$. Then there exists $\delta > 0$ such that*

$$\|\varphi(\vec{x}) - \varphi(\vec{p})\| \leq M\|\vec{x} - \vec{p}\|$$

for all $\vec{x} \in X$ such that $\|\vec{x} - \vec{p}\| < \delta$.

Proof. Differentiability implies that there exists $\delta > 0$ such that

$$\|\varphi(\vec{x}) - \varphi(\vec{p}) - (\mathrm{D}\,\varphi)_{\vec{p}}(\vec{x} - \vec{p}) \leq (M - \|(\mathrm{D}\,\varphi)_{\vec{p}}\|_{op})\, \|\vec{x} - \vec{p}\|$$

for all $\vec{x} \in X$ such that $\|\vec{x} - \vec{p}\| < \delta$. By the reverse triangle inequality, we therefore have

$$\|\varphi(\vec{x}) - \varphi(\vec{p})\| - \|(\mathrm{D}\,\varphi)_{\vec{p}}(\vec{x} - \vec{p})\| \leq (M - \|(\mathrm{D}\,\varphi)_{\vec{p}}\|_{op})\, \|\vec{x} - \vec{p}\|$$

and hence

$$\begin{aligned} \|\varphi(\vec{x}) - \varphi(\vec{p})\| &\leq (M - \|(\mathrm{D}\,\varphi)_{\vec{p}}\|_{op})\, \|\vec{x} - \vec{p}\| + \|(\mathrm{D}\,\varphi)_{\vec{p}}(\vec{x} - \vec{p})\| \\ &\leq (M - \|(\mathrm{D}\,\varphi)_{\vec{p}}\|_{op})\, \|\vec{x} - \vec{p}\| + \|(\mathrm{D}\,\varphi)_{\vec{p}}\|_{op}\|(\vec{x} - \vec{p})\| \\ &\leq M\|\vec{x} - \vec{p}\| \end{aligned}$$

for all $\|\vec{x} - \vec{p}\| < \delta$. □

Remark 6.31. *As mention, an operator norm such as the Hilbert-Schmidt norm would work here, but will generally give a worse bound, and requires both finite dimensionality and a choice of basis. In contrast, the operator norm here is independent of a choice of basis, and can be generalised to infinite dimensional vector spaces with relatively few changes.*

We can now prove versions of the product and chain rule.

6.5.2 Product rule

Proposition 6.32. *Let $X \subset \mathbb{R}^m$ be open, and let $\vec{p} \in X$. If $f, g : X \to \mathbb{R}$ are differentiable at $\vec{p}$, then so is fg, with derivative*

$$(\mathrm{D}\, fg)_{\vec{p}} = g(\vec{p})(\mathrm{D}\, f)_{\vec{p}} + f(\vec{p})(\mathrm{D}\, g)_{\vec{p}}.$$

Proof. By Proposition 6.30, there exist $M, N, \delta > 0$ such that

$$|f(\vec{x}) - f(\vec{p})| \leq M\|\vec{x} - \vec{p}\|,$$
$$|g(\vec{x}) - g(\vec{p})| \leq N\|\vec{x} - \vec{p}\|$$

for all $\vec{x} \in X$ such that $\|\vec{x} - \vec{p}\| < \delta$.

We can write

$$f(\vec{x})g(\vec{x}) = f(\vec{p})g(\vec{x}) + g(\vec{p})f(\vec{x}) - f(\vec{p})g(\vec{p}) + \left(f(\vec{x}) - f(\vec{p})\right)\left(g(\vec{x}) - g(\vec{p})\right),$$

so by linearity of derivatives, it suffices to show that

$$h(\vec{x}) := \left(f(\vec{x}) - f(\vec{p})\right)\left(g(\vec{x}) - g(\vec{p})\right)$$

is differentiable at $\vec{p}$ with derivative 0, as if h is differentiable at $\vec{p}$, then

$$(\mathrm{D}\, fg)_{\vec{p}} = g(\vec{p})(\mathrm{D}\, f)_{\vec{p}} + f(\vec{p})(\mathrm{D}\, g)_{\vec{p}} + (\mathrm{D}\, h)_{\vec{p}}.$$

Noting that $h(\vec{p} = 0$, and that for all $0 < \|\vec{x} - \vec{p}\| < \delta$,

$$\frac{|h(\vec{x})|}{\|\vec{x} - \vec{p}\|} = \frac{|f(\vec{x}) - f(\vec{p})||g(\vec{x}) - g(\vec{p})|}{\|\vec{x} - \vec{p}\|} \leq MN\|\vec{x} - \vec{p}\|$$

we conclude that

$$0 \leq \lim_{\vec{x} \to \vec{p}} \frac{h(\vec{x}) - h(\vec{p})|}{\|\vec{x} - \vec{p}\|} \leq MN \lim_{\vec{x} \to \vec{p}} \|\vec{x} - \vec{p}\| = 0$$

and hence $(\mathrm{D}\, h)_{\vec{p}} = 0$. □

6.5.3 Chain rule

Proposition 6.33. *Let $X \subset \mathbb{R}^m$, $Y \subset \mathbb{R}^n$ be open sets, and let $\vec{p} \in X$. Let $\varphi : X \to Y$ be differentiable at $\vec{p}$ and let $\psi : Y \to \mathbb{R}^s$ be differentiable at $\varphi(\vec{p})$. Then the composition*

$$(\psi \circ \varphi) : X \to \mathbb{R}^s$$

is differentiable at $\vec{p}$ with derivative

$$(\mathrm{D}(\psi \circ \varphi))_{\vec{p}} = (\mathrm{D}\,\psi)_{\varphi(\vec{p})} \circ (\mathrm{D}\,\phi)_{\vec{p}}.$$

The derivative of a composition is the composition of the derivatives.

Proof. From Proposition 6.30, there exist $M, N, \delta_1, \eta_1 > 0$ such that for all $\vec{x} \in X$ such that $\|\vec{x} - \vec{p}\| < \delta_1$

$$\|\varphi(\vec{x}) - \varphi(\vec{p})\| \leq M\|\vec{x} - \vec{p}\|$$

and for all $\vec{y} \in Y$ such that $\|vecy - \varphi(\vec{p})\| < \eta_1$,

$$\|\psi(\vec{y}) - \psi(\varphi(\vec{p}))\| \leq N\|\vec{y} - \varphi(\vec{p})\|.$$

More precisely, we have that $M > \|(\mathrm{D}\,\varphi)_{\vec{p}}\|_{op}$ and $N > \|(\mathrm{D}\,\psi)_{\varphi(\vec{p})}\|_{op}$.

Suppose we are given $\varepsilon > 0$. Then there exists η_2, which we can take to be such that $\eta_2 \leq \eta_1$, such that

$$\|\psi(\vec{y}) - \psi(\varphi(\vec{p})) - (\mathrm{D}\,\psi)_{\varphi(\vec{p})}(\vec{y} - \varphi(\vec{p})\| \leq \frac{\varepsilon}{2M}\|\vec{y} - \varphi(\vec{p})\|$$

for all $\vec{y} \in Y$ such that $\|\vec{y} - \varphi(\vec{p})\| < \eta_2$. We can also find $0 < \delta_2 \leq \delta_1$ such that $M\delta_2 \leq \eta_2$. Then, for $\vec{x} \in X$ satisfying $\|\vec{x} - \vec{p}\| < \delta_2$, then

$$\|\varphi(\vec{x}) - \varphi(\vec{p})\| \leq M\|\vec{x} - \vec{p}\| < M\delta_2 \leq \eta_1.$$

Thus, if $\|\vec{x} - \vec{p}\| < \delta_2$,

$$\begin{aligned}\|\psi(\varphi(\vec{x})) - \psi(\varphi(\vec{p})) - (\mathrm{D}\,\psi)_{\varphi(\vec{p})}(\varphi(\vec{x}) - \varphi(\vec{p}))\| &\leq \frac{\varepsilon}{2M}\|\varphi(\vec{x}) - \varphi(\vec{p})\| \\ &\leq \frac{\varepsilon}{2}\|\vec{x} - \vec{p}\|.\end{aligned}$$

We can also find $0 < \delta < \delta_2$ such that

$$\|\varphi(\vec{x}) - \varphi(\vec{p}) - (\mathrm{D}\,\varphi)_{\vec{p}}(\vec{x} - \vec{p})\| \leq \frac{\varepsilon}{2N}\|\vec{x} - \vec{p}\|$$

for all $\vec{x} \in X$ such that $\|\vec{x} - \vec{p}\| < \delta$.

Now, since

$$\|(\mathrm{D}\,\psi)_{\varphi(\vec{p})}\vec{w}\| \leq \|(\mathrm{D}\,\psi)_{\varphi(\vec{p})}\|_{op}\|\vec{w}\| \leq N\|\vec{w}\|$$

for all $\vec{w} \in \mathbb{R}^n$, we have that

$$\begin{aligned}&\|(\mathrm{D}\,\psi)_{\varphi(\vec{p})}\varphi(\vec{x}) - (\mathrm{D}\,\psi)_{\varphi(\vec{p})}\varphi(\vec{p}) - (\mathrm{D}\,\psi)_{\varphi(\vec{p})} \circ (\mathrm{D}\,\varphi)_{\vec{p}}(\vec{x} - \vec{p})\| \\ &\leq N\|\varphi(\vec{x}) - \varphi(\vec{p}) - (\mathrm{D}\,\varphi)_{\vec{p}}(\vec{x} - \vec{p})\| \\ &\leq \frac{\varepsilon}{2}\|\vec{x} - \vec{p}\|\end{aligned}$$

for all $\|\vec{x}-\vec{p}\|<\delta$. Thus, for all $\vec{x}\in X$ such that $\|\vec{x}-\vec{p}\|<\delta$, we have that

$$\begin{aligned}
&\|\psi(\varphi(\vec{x}))-\psi(\varphi(\vec{p}))-(\mathrm{D}\,\psi)_{\varphi(\vec{p})}\circ(\mathrm{D}\,\varphi)_{\vec{p}}(\vec{x}-\vec{p})\| \\
&=\|\psi(\varphi(\vec{x}))-\psi(\varphi(\vec{p}))-(\mathrm{D}\,\psi)_{\varphi(\vec{p})}(\varphi(\vec{x})-\varphi(\vec{p})) \\
&\qquad\qquad +(\mathrm{D}\,\psi)_{\varphi(\vec{p})}\varphi(\vec{x})-(\mathrm{D}\,\psi)_{\varphi(\vec{p})}\varphi(\vec{p})-(\mathrm{D}\,\psi)_{\varphi(\vec{p})}\circ(\mathrm{D}\,\varphi)_{\vec{p}}(\vec{x}-\vec{p})\| \\
&\le\|\psi(\varphi(\vec{x}))-\psi(\varphi(\vec{p}))-(\mathrm{D}\,\psi)_{\varphi(\vec{p})}(\varphi(\vec{x})-\varphi(\vec{p}))\| \\
&\qquad\qquad +\|(\mathrm{D}\,\psi)_{\varphi(\vec{p})}\varphi(\vec{x})-(\mathrm{D}\,\psi)_{\varphi(\vec{p})}\varphi(\vec{p})-(\mathrm{D}\,\psi)_{\varphi(\vec{p})}\circ(\mathrm{D}\,\varphi)_{\vec{p}}(\vec{x}-\vec{p})\| \\
&\le\frac{\varepsilon}{2}\|\vec{x}-\vec{p}\|+\frac{\varepsilon}{2}\|\vec{x}-\vec{p}\|=\varepsilon\|\vec{x}-\vec{p}\|.
\end{aligned}$$

Thus, by Lemma 6.20, $(\psi\circ\phi)$ is differentiable at $\vec{p}$ with derivative

$$(\mathrm{D}(\psi\circ\varphi))_{\vec{p}}=(\mathrm{D}\,\psi)_{\varphi(\vec{p})}\circ(\mathrm{D}\,\phi)_{\vec{p}}.$$

□

Example 6.34. *The majority of classical differential rules can be viewed as a consequence of chain rule for total derivatives. Lets see a few examples, where we will take our functions to be differentiable everywhere, although differentiability at appropriate points is sufficient.*

Let $f:\mathbb{R}^m\to\mathbb{R}$ be a differentiable function, and let $g_1,\dots,g_m:\mathbb{R}\to\mathbb{R}$ be a collection of differentiable functions. Then the function

$$F(t):=f(g_1(t),\dots,g_m(t))$$

is differentiable with derivative

$$\frac{dF}{dt}(t)=\sum_{i=1}^{m}\frac{\partial f}{\partial x_i}(g_1(t),\dots,g_m(t))\times\frac{dg_i}{dt}(t).$$

To see this, note that

$$F=f\circ\begin{pmatrix}g_1\\ g_2\\ \vdots\\ g_m\end{pmatrix}=:f\circ G.$$

Since limits, and hence derivatives, can be computed componentwise, we compute the derivatives

$$(\mathrm{D}\,f)_{\vec{x}}=\begin{pmatrix}\frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_m}\end{pmatrix}$$

and

$$(\mathrm{D}\,G)_t=\begin{pmatrix}\frac{dg_1}{dt}\\ \frac{dg_2}{dt}\\ \vdots\\ \frac{dg_m}{dt}\end{pmatrix}.$$

Hence, by chain rule

$$\frac{dF}{dt}(t) = (\mathrm{D}\,F)_t = (\mathrm{D}\,f)_{G(t)}(\mathrm{D}\,G)_t = \sum_{i=1}^{m} \frac{\partial f}{\partial x_i}(g_1(t), \ldots, g_m(t)) \times \frac{dg_i}{dt}(t).$$

A particular choice of f includes

$$f(x_1, \ldots, x_m) = x_1 + \cdots + x_m$$

which is differentiable everywhere, and hence

$$\frac{d}{dt}(g_1 + \cdots + g_m) = \begin{pmatrix} 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \frac{dg_1}{dt} \\ \vdots \\ \frac{dg_m}{dt} \end{pmatrix} = \frac{dg_1}{dt} + \cdots + \frac{dg_m}{dt}$$

and so linearity of differentiation follows from chain rule.

Similarly, product rule is a consequence of chain rule. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be

$$f(x_1, x_2) = x_1 x_2.$$

Chain rule implies that

$$\frac{d}{dt}(g_1 g_2) = \begin{pmatrix} g_2 & g_1 \end{pmatrix} \begin{pmatrix} \frac{dg_1}{dt} \\ \frac{dg_2}{dt} \end{pmatrix} = g_2 \frac{dg_1}{dt} + g_1 \frac{dg_2}{dt}.$$

7 The Inverse and Implicit Function Theorems

7.1 Contraction mappings and a fixed point theorem

Definition 7.1. *Let $X \subset \mathbb{R}^m$. We call a function $\varphi : X \to \mathbb{R}^m$ a contraction if there exists $0 \leq \lambda < 1$ such that*

$$\|\varphi(\vec{u}) - \varphi(\vec{v})\| \leq \lambda \|\vec{u} - \vec{v}\|$$

for all $\vec{u}, \vec{v} \in X$.

Remark 7.2. *A contraction is automatically continuous. It is a good exercise to prove why this is the case!*

Theorem 7.3 ((Weak) Banach Fixed Point Theorem)**.** *Let $F \subset \mathbb{R}^m$ be a closed subset, and let $\varphi : F \to F$ be a contraction. Then there is a unique point $\vec{p} \in F$ such that $\varphi(\vec{p}) = \vec{p}$.*

Proof. As φ is a contraction, there exists $\lambda \in [0, 1)$ such that

$$\|\varphi(\vec{u}) - \varphi(\vec{v})\| \leq \lambda \|\vec{u} - \vec{v}\|$$

for all $\vec{u}, \vec{v} \in F$. Choose any $\vec{x}_0 \in F$ and define a sequence by $\vec{x}_{n+1} = \varphi(\vec{x}_n)$ for every $n \geq 0$. Then

$$\|\vec{x}_{n+1} - \vec{x}_n\| = \|\varphi(\vec{x}_n) - \varphi(\vec{x}_{n-1}\| \leq \lambda \|\vec{x}_n - \vec{x}_{n-1}\| \leq \cdots \leq \lambda^n \|\vec{x}_1 - \vec{x}_0\|$$

for all $n \geq 0$. Then, for any $k > j > 0$

$$\begin{aligned}\|\vec{x}_k - \vec{x}_j\| &= \|\vec{x}_k - \vec{x}_{k-1} + \vec{x}_{k-1} - \vec{x}_{k-2} + \cdots + \vec{x}_{j+1} - \vec{x}_j\| \\ &\leq \sum_{m=j}^{k-1} \|\vec{x}_{m+1} - \vec{x}_m\| \leq \sum_{m=j}^{k-1} \lambda^m \|\vec{x}_1 - \vec{x}_0\| \\ &= \frac{\lambda^j - \lambda^k}{1-\lambda}\|\vec{x}_1 - \vec{x}_0\| \leq \frac{\lambda^j}{1-\lambda}\|\vec{x}_1 - \vec{x}_0\|.\end{aligned}$$

As $\lim_{j\to 0} \lambda^j = 0$ for $\lambda \in [0,1)$, for any $\varepsilon > 0$, there exists $N > 0$ such that

$$\frac{\lambda^j}{1-\lambda}\|\vec{x}_1 - \vec{x}_0\| < \varepsilon$$

for all $j \geq N$. Hence, for all $k > j \geq N$,

$$\|\vec{x}_k - \vec{x}_j\| < \varepsilon$$

and so $\{\vec{x}_n\}$ is a Cauchy sequence. Therefore $\{\vec{x}_n\}$ converges to a point $\vec{p} \in \mathbb{R}^m$. As F is closed and $\vec{x}_n \in F$ for every $n \geq 0$, the limit $\vec{p} \in F$. Thus

$$\vec{p} = \lim_{n\to\infty} \vec{x}_{n+1} = \lim_{n\to\infty} \varphi(\vec{x}_{n+1}) = \varphi(\lim_{n\to\infty} \vec{x}_n) = \varphi(\vec{p})$$

as φ is continuous on F. SO φ has a fixed point in F. To see that this point is unique, suppose there exists $\vec{q} \in F$ such that $\varphi(\vec{q}) = \vec{q}$. Then

$$\|\vec{q} - \vec{p}\| = \|\varphi(\vec{q}) - \varphi(\vec{p})\| \leq \lambda\|\vec{q} - \vec{p}\| < \|\vec{q} - \vec{p}\|$$

unless $\|\vec{q} - \vec{p}\| = 0$, i.e. $\vec{p} = \vec{q}$. Thus the fixed point is unique. □

Remark 7.4. *This holds much more generally in any complete metric space: a set with a sensible notion of distance, in which every Cauchy sequence converges.*

7.2 A pre-inverse function theorem

Lemma 7.5. *Let $X \subset \mathbb{R}^m$ be open, let $\varphi : X \to \mathbb{R}^n$ be differentiable at a point $\vec{p} \in X$. Suppose there exists a constant $K > 0$ such that*

$$\|\vec{x} - \vec{p}\| \leq K\|\varphi(\vec{x}) - \varphi(\vec{p})\|$$

for all $\vec{x} \in X$. Then, for all $\vec{w} \in \mathbb{R}^m$

$$\|vecw\| \leq K\|(\mathrm{D}\,\varphi)_{\vec{p}}\vec{w}\|$$

Proof. Take $\vec{w} \in \mathbb{R}^m$. Then, for all $t \neq 0$ small enough such that $\vec{p} + t\vec{w} \in X$, we have that

$$|t|\|\vec{w}\| = \|\vec{p} + t\vec{w} - \vec{p}\| \leq K\|\varphi(\vec{p} + t\vec{w}) - \varphi(\vec{p})\|.$$

Thus, for all sufficiently small $0 < |t|$, we have that

$$\|\vec{w}\| \leq K \left\| \frac{\varphi(\vec{p} + t\vec{w}) - \varphi(\vec{p})}{t} \right\|$$

and hence

$$\|\vec{w}\| \leq \lim_{t \to 0} K \left\| \frac{\varphi(\vec{p} + t\vec{w}) - \varphi(\vec{p})}{t} \right\| = K\|(\mathrm{D}\,\varphi)_{\vec{p}}\vec{w}\|$$

using Proposition 6.24 and continuity of the norm. □

Proposition 7.6. *Let $X \subset \mathbb{R}^m$ be an open set and let $\varphi : X \to \mathbb{R}^m$ be a differentiable function. Let $Y \subset \varphi(X)$ be a non-empty open set in $\mathbb{R}^m$, and suppose there exists a constant $K > 0$ such that*

$$\|\vec{u} - \vec{v}\| \leq K\|\varphi(\vec{u}) - \varphi(\vec{v})\|$$

for all $\vec{u}, \vec{v} \in X$. Then there is a differentiable function $\mu : Y \to \mathbb{R}^m$ such that $\varphi(\mu(\vec{y})) = \vec{y}$ for all $\vec{y} \in Y$. Furthermore $\mu(Y)$ is open in $\mathbb{R}^m$ and $(\mathrm{D}\,\varphi)_{\vec{p}}$ is invertible with inverse $(\mathrm{D}\,\mu)_{\varphi(\vec{p}}$ for all $\vec{p} \in \mu(Y)$.

Proof. As $Y \subset \varphi(X)$, given $\vec{y} \in Y$, there exists at least one $\vec{x} \in X$ such that $\varphi(\vec{x}) = \vec{y}$. I claim there can be at most one such point: if $\vec{x}_1, \vec{x}_2 \in X$ are such that $\varphi(\vec{x}_1) = \varphi(\vec{x}_2) = \vec{y}$, then

$$\|\vec{x}_1 - \vec{x}_2\| \leq K\|\varphi(\vec{x}_1) - \varphi(\vec{x}_2) = K\|\vec{y} - \vec{y}\| = 0$$

so $\vec{x}_1 = \vec{x}_2$. Thus, there is a unique function $\mu : Y \to \mathbb{R}^m$ defined by $\mu(\vec{y})$ is the unique $\vec{x} \in X$ such that $\varphi(\vec{x}) = \vec{y}$. Clearly $\varphi(\mu(\vec{y})) = \vec{y}$ for every $\vec{y} \in Y$, so it remains to show that $\mu(Y)$ is open in $\mathbb{R}^m$, μ is differentiable, and to determine its derivative.

To see that $\mu(Y)$ is open, we first note that $\vec{p} \in \mu(Y)$ implies that there exists $\vec{y} \in Y$ such that $\vec{y} = \varphi(\vec{p})$, and hence $\mu(Y) \subset \varphi^{-1}(Y)$. And if $\vec{p} \in \varphi^{-1}(Y)$, then $\varphi(\vec{p}) \in Y$, and so $\vec{p} = \mu(\varphi(\vec{p}))$ by definition. Hence $\vec{p} \in \mu(Y)$, and so we must have that $\mu(Y) = \varphi^{-1}(Y)$.

As φ is differentiable on X, it is continuous on X, and so $\mu(Y) = \varphi^{-1}(Y)$ is open in X. Hence, for all $\vec{p} \in \mu(Y)$, there exists $\delta_1 > 0$ such that $B_X(\vec{p}, \delta_1) \subset \mu(Y)$. As X is open in $\mathbb{R}^m$, there exists $\delta_2 > 0$ such that $B(\vec{p}, \delta_2) \subset X$. Thus, letting $\delta = \min(\delta_1, \delta_2)$, we must have that $B(\vec{p}, \delta) \subset \mu(Y)$. Therefore $\mu(Y)$ is open in $\mathbb{R}^m$.

To see that $(\mathrm{D}\,\varphi)_{\vec{p}}$ is invertible for all $\vec{p} \in \mu(Y)$, note that Lemma 7.5 implies that

$$\|\vec{w}\| \leq K\|(\mathrm{D}\,\varphi)_{\vec{p}}\vec{w}\|$$

for all $\vec{w} \in \mathbb{R}^m$. Thus, if $(\mathrm{D}\,\varphi)_{\vec{p}}\vec{w} = \vec{0}$, we must have $\vec{w} = \vec{0}$. Hence $(\mathrm{D}\,\varphi)_{\vec{p}}$ is an injective map $\mathbb{R}^m \to \mathbb{R}^m$ and is therefore invertible.

Finally, we want to show that μ is differentiable with derivative

$$(\mathrm{D}\,\mu)_{\varphi(\vec{p})} = (\mathrm{D}\,\varphi)_{\vec{p}}^{-1}$$

for all $\vec{p} \in \mu(Y)$.

Let $\vec{q} \in Y$ and let $\vec{p} = \mu(\vec{q})$ so that $\vec{q} = \varphi(\vec{p})$. Suppose we are given $\varepsilon > 0$. As φ is differentiable and X is open, we can find $\delta > 0$ such that $\vec{x} \in X$ and

$$\|\varphi(\vec{x}) - \varphi(\vec{p}) - (\mathrm{D}\,\varphi)_{\vec{p}}(\vec{x} - \vec{p})\| \leq \frac{\varepsilon}{K^2}\|\vec{x} - \vec{p}\|$$

for all $\|\vec{x} - \vec{p}\| < K\delta$. By reducing δ as needed, we can assume that $B(\vec{q}, \delta) \subset Y$. Let $\vec{y} \in B(\vec{q}, \delta)$, and let $\vec{x} = \mu(\vec{y})$. Then

$$\|\vec{x} - \vec{p}\| \leq K\|\varphi(\vec{x}) - \varphi(\vec{p})\| = K\|\vec{y} - \vec{q}| < K\delta.$$

Hence

$$\begin{aligned}\|\vec{y} - \vec{q} - (\mathrm{D}\,\varphi)_{\vec{p}}(\mu(\vec{y}) - \mu(\vec{p}))\| &= \|\varphi(\vec{x}) - \varphi(\vec{p}) - (\mathrm{D}\,\varphi)_{\vec{p}}(\vec{x} - \vec{p})\| \\ &\leq \frac{\varepsilon}{K^2}\|\vec{x} - \vec{p}\| \leq \frac{\varepsilon}{K}\|\vec{y} - \vec{q}\|.\end{aligned}$$

This implies, via Lemma 7.5, that

$$\begin{aligned}\|\mu(\vec{y}) - \mu(\vec{q}) - (\mathrm{D}\,\varphi)_{\vec{p}}^{-1}(\vec{y} - \vec{p})\| &\leq K\|(\mathrm{D}\,\varphi)_{\vec{p}}\left(\mu(\vec{y}) - \mu(\vec{q}) - (\mathrm{D}\,\varphi)_{\vec{p}}^{-1}(\vec{y} - \vec{p})\right)\| \\ &= K\|\vec{y} - \vec{p} - (\mathrm{D}\,\varphi)_{\vec{p}}(\mu(\vec{y}) - \mu(\vec{q}))\| \\ &\leq \varepsilon\|\vec{y} - \vec{q}\|\end{aligned}$$

for all $\vec{y} \in Y$ such that $\|\vec{y} - \vec{q}\| < \delta$.

Hence μ is differentiable at $\vec{q}$ with derivative

$$(\mathrm{D}\,\mu)_{\vec{q}} = (\mathrm{D}\,\mu)_{\varphi(\vec{p})} = (\mathrm{D}\,\varphi)_{\vec{p}}^{-1}.$$

□

7.3 The inverse function theorem

Definition 7.7. *A function $\varphi : X \to \mathbb{R}^n$ defined on a open subset $X \subset RR^m$ is called continuously differentiable if it is differentiable with continuous first order partial derivatives throughout X.*

Remark 7.8. *By Proposition 6.23, any function with continuous first order partial derivatives throughout X is continuously differentiable on X.*

Theorem 7.9 (Inverse function theorem)**.** *Let $X \subset \mathbb{R}^m$ be open, and let $\varphi : X \to \mathbb{R}^m$ be a continuously differentiable function. Let $\vec{p} \in X$ be a point such that $(\mathrm{D}\,\varphi)_{\vec{p}}$ is invertible. Then there exists an open $Y \subset \mathbb{R}^m$ containing $\varphi\vec{p}$ and a continuously differentiable function $\mu : Y \to \mathbb{R}^m$ such that $\mu(Y)$ is open in $\mathbb{R}^m$, $\vec{p} \in \mu(Y)$ and $\varphi(\mu(\vec{y})) = \vec{y}$ for every $\vec{y} \in Y$.*

Proof. We essentially need to prove that $(\mathrm{D}\,\varphi)_{\vec{p}}$ being invertible implies the inequality assumed in Proposition 7.6, but this will require us restricting to small open sets. Let $\vec{q} = \varphi(\vec{p})$ and let $T + (\mathrm{D}\,\varphi)_{\vec{p}}^{-1}$. Define a constant $K = 2\|T\|_{op} > 0$, so that

$$\|T\vec{w}\| \leq \frac{1}{2}K\|\vec{w}\|$$

for all $\vec{w} \in \mathbb{R}^m$.

We first need to find an open set $U \subset X$ on which the inequality of Proposition 7.6 holds. In order to achieve this, we will introduce some auxiliary functions.

Define $\psi : X \to \mathbb{R}^m$ by

$$\psi(\vec{x}) = \vec{x} - T(\varphi(\vec{x}) - \vec{q}).$$

Via chain rule, ψ is differentiable at all points of X, with derivative

$$(\mathrm{D}\,\psi)_{\vec{x}} = I - T(\mathrm{D}\,\varphi)_{\vec{x}}.$$

In particular, we note that $(\mathrm{D}\,\psi)_{\vec{p}} = 0$.

As φ, and therefore ψ, have continuous first order derivatives $\vec{p}$, there exists $r > 0$ such that, for all $\|\vec{x} - \vec{p}\| < 2r$, $\vec{x} \in X$ and

$$\|\psi(\vec{u}) - \psi(\vec{v})\| = \|\psi(\vec{u}) - \psi(\vec{v}) - (\mathrm{D}\,\psi)_{\vec{p}}(\vec{u} - \vec{v})\| \leq \frac{1}{2}\|\vec{u} - \vec{v}\|$$

for all $\vec{u}, \vec{v} \in B(\vec{p}, 2r) \subset X$. Thus, for all $\vec{u}, \vec{v} \in B(\vec{p}, 2r)$,

$$\|\vec{u} - \vec{v} - T(\varphi(\vec{u}) - \varphi(\vec{v}))\| \leq \frac{1}{2}\|\vec{u} - \vec{v}\|.$$

By the reverse triangle inequality, we therefore have that

$$\|\vec{u} - \vec{v}\| \leq 2\|T(\varphi(\vec{u}) - \varphi(\vec{v})\| \leq K\|\varphi(\vec{u}) - \varphi(\vec{v})\|$$

for all $\vec{u}, \vec{v} \in B(\vec{p}, 2r)$. If the image of this open ball contained an open set containing $\varphi(\vec{p})$, we would be done. We don't yet know such an open set exists.

Pick a point $\vec{y} \in B(\vec{q}, \frac{r}{K})$. We will show that there exists $\vec{x} \in B(\vec{p}, r)$ such that $\varphi(\vec{x}) = \vec{y}$, and hence $B(\vec{q}, \frac{r}{K}) \subset \varphi(B(\vec{p}, r))$.

Consider the function $\theta_{\vec{y}} : \overline{B}(\vec{p}, r) \to \mathbb{R}^m$ defined by

$$\theta_{\vec{y}}(\vec{x}) = \psi(\vec{x}) + T(\vec{y} - \vec{q}) = \vec{x} + T(\vec{y} - \varphi(\vec{x})).$$

Note that $\theta_{\vec{y}}(\vec{x}) = \vec{x}$ if and only if $T(\varphi(\vec{x}) - \vec{y}) = \vec{0}$, which occurs if and only if $\varphi(\vec{x}) = \vec{y}$. So if we can show that $\theta_{\vec{y}}$ has a fixed point in $B(\vec{p}, r)$, we are essentially done.

Note that

$$\begin{aligned}
\|\theta_{\vec{y}}(\vec{x}) - \vec{p}\| &= \|\theta_{\vec{y}}(\vec{x}) - \vec{p} - T(\vec{y} - \vec{q}) + T(\vec{y} - \vec{q})\| \\
&\leq \|\theta_{\vec{y}}(\vec{x}) - \vec{p} - T(\vec{y} - \vec{q})\| + \|T(\vec{y} - \vec{q})\| \\
&\leq \|\vec{x} - \vec{p} - T(\varphi(\vec{x}) - \varphi(\vec{p}))\| + \frac{1}{2}K\|\vec{y} - \vec{q}\| \\
&= \|\psi(\vec{x}) - \psi(\vec{p})\| + \frac{1}{2}K\|\vec{y} - \vec{q}\| \\
&\leq \frac{1}{2}\|\vec{x} - \vec{p}\| + \frac{1}{2}K\|\vec{y} - \vec{q}\| < r
\end{aligned}$$

for $\vec{x} \in \overline{B}(\vec{p}, r)$, $\vec{y} \in B(\vec{q}, \frac{r}{K})$.

Thus, $\theta_{\vec{y}} : \overline{B}(\vec{p}, r) \to B(\vec{p}, r)$ is a map from the closed ball to the open ball. We also note that

$$\|\theta_{\vec{y}}(\vec{u}) - \theta_{\vec{y}}(\vec{v})\| = \|\psi(\vec{u}) - \psi(\vec{v})\| \leq \frac{1}{2}\|\vec{u} - \vec{v}\|$$

and so $\theta_{\vec{y}}$ is a contraction on a closed set. Thus, there exists $\vec{x} \in \overline{B}(\vec{p}, r)$ such that $\theta_{\vec{y}}(\vec{x}) = \vec{x}$. Since $\theta_{\vec{y}}$ maps into the open ball, $\vec{x}$ is contained in the open ball, as needed.

Thus, we have a continuously differentiable map $\varphi : B(\vec{p}, r) \to \mathbb{R}^m$ and non-empty open $Y = B(\vec{q}, \frac{r}{K}) \subset \varphi(B(\vec{p}, r))$ such that

$$\|\vec{u} - \vec{v}\| \leq K\|\varphi(\vec{u}) - \varphi(\vec{v})\|$$

for all $\vec{u}, \vec{v} \in B(\vec{p}, r)$. Therefore, by Proposition 7.6, there exists differentiable $\mu : Y : \mathbb{R}^m$ such that $\mu(Y)$ is open, $\varphi(\mu(\vec{y})) = \vec{y}$, and $(\mathrm{D}\,\mu)_{\vec{y}} = (\mathrm{D}\,\varphi)^{-1}_{\mu(\vec{y})}$ for all $\vec{y} \in Y$.

Furthermore, as φ is continuously differentiable, $(\mathrm{D}\,\varphi)_{\mu(\vec{y})} = (\mathrm{J}\,\varphi)_{\mu(\vec{y})}$ has entries that are continuous functions on Y. Therefore

$$(\mathrm{J}\,\mu)_{\vec{y}} = (\mathrm{D}\,\mu)_{\vec{y}} = (\mathrm{D}\,\varphi)^{-1}_{\mu(\vec{y})}$$

has entries that are continuous functions on Y, and so μ is continuously differentiable on Y. ☐

Example 7.10. *Consider the function $F : \mathbb{R}^2 \to \mathbb{R}^2$ defined by*

$$F(x, y) = \begin{pmatrix} x^2 - y^2 \\ x^2 + y^2 \end{pmatrix}$$

This is continuously differentiable, with derivative

$$\begin{pmatrix} 2x & -2y \\ 2x & 2y \end{pmatrix}$$

which has determinant $8xy$. Hence, this derivative is invertible for all (x, y) for which $x \neq 0$ and $y \neq 0$. This implies that for any $x, y \neq 0$, there is some open set containing (x, y) on which F has a continuously differentiable inverse. For example, near $(1, 1)$, an inverse is given by

$$F^{-1}(u, v) = \begin{pmatrix} \sqrt{\frac{u+v}{2}} \\ \sqrt{\frac{v-u}{2}} \end{pmatrix}$$

This does not extend to having an inverse on

$$\mathbb{R}^2 \setminus \{(x, y) \in \mathbb{R}^2 \mid x = 0 \text{ or } y = 0\}$$

as $F(x, y) = F(\pm x, \pm y)$, so F is not injective.

Example 7.11. *Consider the function* $F : \mathbb{R}^2 \to \mathbb{R}^2$ *defined by*

$$F(x, y) = \begin{pmatrix} e^x \cos y \\ e^x \sin y \end{pmatrix}$$

This is continuously differentiable, with derivative

$$\begin{pmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{pmatrix}$$

which has determinant $e^x \neq 0$ *for any* $(x, y) \in \mathbb{R}^2$*. Hence* F *has a continuously differentiable inverse on some open set around any point of* $\mathbb{R}^2$*.*

Note that this does not extend to having an inverse on $\mathbb{R}^2$*:* $F(x, y) = F(x, y+2\pi)$*, so* F *is very much not injective and hence not invertible. The existence of an inverse is strictly local.*

7.4 The implicit function theorem

Consider the equation $x^2 + y^2 - 1 = 0$ defining a circle. Near to the point $(0, 1)$, this equation defines y as an implicit function of x: every point (x, y) satisfying $x^2 + y^2 - 1 = 0$ near $(0, 1)$ satisfies $y = \sqrt{1 - x^2}$. We can exploit the inverse function theorem to generalise this to describe when we can express the set of points satisfying a set of equations as functions of a subset of our variables, at least locally.

Lemma 7.12. *Let* $U \subset \mathbb{R}^m$ *be open, and let* $(p_1, \ldots, p_m)$ *be a point of* U*. Then, for each* $1 \leq k < m$*, the set*

$$U_k = \{(x_{k+1}, \ldots, x_m) \in \mathbb{R}^{m-k} \mid (p_1, \ldots, p_k, x_{k+1}, \ldots, x_m) \in U\}$$

is open.

Proof. Suppose $\vec{q} \in U_k$. Then, as $(p_1, \ldots, p_k, \vec{q}) \in U$, there exists $\delta > 0$ such that $B((p_1, \ldots, p_k, \vec{q}), \delta) \subset U$. It is then easy to check that $B(\vec{q}, \delta) \subset U_k$, as $\|\vec{x} - \vec{q}\| < \delta$ implies that

$$\|(p_1, \ldots, p_k, \vec{x}) - (p_1, \ldots, p_k), \vec{q}\| < \delta$$

and so $(p_1, \ldots, p_k, \vec{x}) \in U$, implying that $\vec{x} \in U_k$. □

Theorem 7.13 (Implicit function theorem)**.** *Let* $X \subset \mathbb{R}^m$ *be open,* $f_1, \ldots, f_k : X \to \mathbb{R}$ *be continuously differentiable functions on* X*, with* $k < m$*. Define*

$$S = \{\vec{x} \in X \mid f_i(\vec{x}) = 0 \text{ for all } 1 \leq i \leq k\}$$

and let $\vec{p} \in S$*. Suppose that the matrix*

$$J(\vec{x}) = (\partial_j f_i(\vec{x}))$$

has rank at least k at $\vec{p}$. Without loss of generality, we assume that the first k columns are independent, reordering the variables $x_1, \ldots, x_m$ if needed. Then there exists open $V \subset X$ containing $\vec{p}$, and continuously differentiable functions $g_1, \ldots, g_k : U \to \mathbb{R}$ defined on some open set $U \subset \mathbb{R}^{m-k}$ containing $(p_{k+1}, p_{k+2} \ldots, p_m)$ such that

$$S \cap V = \{(g_1(\vec{x}')), \ldots, g_k(\vec{x}'), x_{k+1}, \ldots, x_m) \mid \vec{x}' = (x_{k+1}, \ldots x_m) \in U\}$$

Proof. Define $F : \mathbb{R}^m \to \mathbb{R}^m$ by

$$F(\vec{x}) = (f_1(\vec{x}), \ldots, f_k(\vec{x}), x_{k+1}, \ldots, x_m).$$

The function F is continuously differentiable, with derivative

$$(\mathrm{D}\, F)_{\vec{p}} = \begin{pmatrix} \partial_1 f_1(\vec{p}) & \cdots & \partial_k f_1(\vec{p}) & \partial_{k+1} f_1(\vec{p}) & \cdots & \partial_m f_1(\vec{p}) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \partial_1 f_k(\vec{p}) & \cdots & \partial_k f_k(\vec{p}) & \partial_{k+1} f_k(\vec{p}) & \cdots & \partial_m f_k(\vec{p}) \\ 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1 \end{pmatrix}$$

where the lower right is the $(m-k) \times (m-k)$ identity matrix. By the assumption of the theorem, the upper left hand matrix has rank k, and so this matrix has rank m and is therefore invertible.

Thus, by the inverse function theorem, there exists an open set $Y \subset \mathbb{R}^m$ containing

$$F(\vec{p}) = (f_1(\vec{p}), \ldots, f_k(\vec{p}), p_{k+1}, \ldots, p_m) = (0, \ldots, 0, p_{k+1}, \ldots, p_m)$$

and a continuously differentiable function $\mu : Y \to \mathbb{R}^m$ such that $\mu(Y)$ is an open set $\vec{p} \in \mu(Y)$ and $F(\mu(\vec{y}) = \vec{y})$ for all $\vec{y} \in Y$.

By looking at the last $m-k$ components, we see that μ must be of the form

$$\mu(\vec{y}) = (G_1(\vec{y}), G_2(\vec{y}), \ldots, G_k(\vec{y}), y_{k+1}, \ldots, y_m)$$

for some continuously differentiable functions $G_1, \ldots, G_k : Y \to \mathbb{R}$. Furthermore, we have that

$$f_i(G_1(\vec{y}), \ldots G_k(\vec{y}), y_{k+1}, \ldots, y_m) = y_i$$

for each $1 \leq i \leq k$. Thus

$$\begin{aligned} S \cap \mu(Y) &= \{\vec{x} \in \mu(Y) \mid f_1(\vec{x}) = \cdots = f_k(\vec{x}) = 0\} \\ &= \{\vec{x} = (G_1(\vec{y}), \ldots, G_k(\vec{y}), y_{k+1}, \ldots, y_m) \mid \vec{y} \in Y,\, y_1 = \cdots = y_k = 0, \} \end{aligned}$$

Let $V = \mu(Y)$. Applying Lemma 7.12, we have an open set $U \subset \mathbb{R}^{m-k}$

$$U = \{(y_{k+1}, \ldots, y_m) \mid (0, \ldots, 0, y_{k+1}, \ldots, y_m) \in Y\}.$$

For $1 \leq i \leq k$, define

$$g_i(y_{k+1}, \ldots, y_m) = G_i(0, \ldots, 0, y_{k+1}, \ldots, y_m).$$

Then

$$S \cap V = \{(g_1(\vec{y}'), \ldots, g_k(\vec{y}'), y_{k+1}, \ldots, y_m) \mid \vec{y}' = (y_{k+1}, \ldots, y_m) \in U\}$$

□

Remark 7.14. *Note that, when the implicit function theorem applies, the map*

$$\pi : V \cap S \to U$$
$$(x_1, \ldots, x_k, x_{k+1}, \ldots, x_m) \mapsto (x_{k+1}, \ldots, x_m)$$

is a bijection with inverse

$$\pi^{-1}(x_{k+1}, \ldots, x_m) = (g_1(x_{k+1}, \ldots, x_m), \ldots, g_k(x_{k+1}, \ldots, x_m), x_{k+1}, \ldots, x_m).$$

The map π is continuous, and in fact continuously differentiable, and the implicit function theorem tells us that π^{-1} is continuously differentiable. Hence π is a homeomorphism $V \cap S \cong U$. In fact, it is a diffeomorphism, which is a continuously differentiable map with continuously differentiable inverse.

Example 7.15. *Consider the function $f(x, y) = x^2 + y^2 - 1$. This satisfies all the differentiability requirements, and has $J(x, y) = (2x, 2y)$, which has rank 1 at every point on the unit circle. Let (x_0, y_0) be a point on the circle. When $x_0 \neq 0$, the first column has rank 1, and so the set of (x, y) satisfying $x^2 + y^2 - 1 = 0$ can be given as a continuously differentiable function of y in a small open around y_0 and similarly, when $y_0 \neq 0$, we can define the set of (x, y) satisfying $x^2 + y^2 - 1 = 0$ can be given as a continuously differentiable function of x in a small open around y_0.*

8 Multiple integrals and Fubini's theorem

Integration will play a much larger role next term, no analysis course would be complete without at least touching upon integrals. In our case, we will not discuss integrals over general domains or Lebesgue integration. Instead we shall focus on integration over rectangular domains, and the conditions under which such integrals can be computed using iterated one dimensional integrals.

8.1 An informal recap of one dimensional Riemann integrals

Definition 8.1. *A partition of an interval $[a, b]$ is a finite sequence $P = (x_0, \ldots, x_n)$ such that*

$$a = x_0 < x_1 < \cdots < x_n = b.$$

A tagged partition of $[a,b]$ is a partition P of $[a,b]$ and a choice of $t_i \in [x_i, x_{i+1}]$ for each $0 \le i \le n$. We denote by $\mathcal{P}_n([a,b])$ the set of tagged partitions of $[a,b]$ into n parts.

Definition 8.2. *A function $f : [a,b] \to \mathbb{R}$ is called Riemann integrable if*

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}_n([a,b])} \sum_{i=0}^{n-1} f(t_i)(x_{i+1} - x_i) = \lim_{n\to\infty} \inf_{P\in\mathcal{P}_n([a,b])} \sum_{i=0}^{n-1} f(t_i)(x_{i+1} - x_i).$$

The integral $\int_a^b f(x)\,dx$ is given by this common value.

Fact 8.3. *If $f : [a,b] \to \mathbb{R}$ is continuous, it is Riemann integrable.*

Fact 8.4. *Whenever the integrals involved are well defined, integration satisfies*

$$\int_a^b f(x) + g(x)\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx$$

$$\int_a^b f(x)\,dx = \int_a^b f(x)\,dx + \int_c^b f(x)\,dx$$

$$\left|\int_a^b f(x)\,dx\right| \le \int_a^b |f(x)|\,dx$$

and if $f(x) \le g(x)$ for all $x \in [a,b]$, then

$$\int_a^b f(x)\,dx \le \int_a^b g(x)\,dx.$$

8.2 Integration in several variables

In order to integrate a function of several variables over a rectangular domain, we have two main options:

1. Define a partition of a product of intervals, construct m-dimensional analogues of Riemann sums, and define the integral as the common value of the supremum and infinimum of these Riemann sums over the set of tagged partitions.,
2. Define it as an iterated integral of one dimensional functions, as follows

$$\int_{[a_1,b_1]\times\cdots\times[a_m,b_m]} g(x_1,\ldots,g_m)\,dx_m \ldots dx_1$$
$$:= \int_{[a_1,b_1]\times\cdots\times[a_{m-1},b_{m-1}]} \left(\int_{a_m}^{b_m} g(x_1,\ldots,x_{m-1},t)\,dt\right) dx_{m-1}\ldots dx_1.$$

The first approach is arguably more useful theoretically, as we can easily derive the m-dimensional analogues of Fact 8.4, and this common value is what is normally meant by the Riemann integral of a function f. However, it is not particularly well suited to computation. The second is much more practical for actually calculating things, but *a priori* depends on the order of integration.

Example 8.5. *Define a function $f : \mathbb{R}^2 \to \mathbb{R}$ be*

$$f(x,y) := \begin{cases} \frac{4xy(x^2-y^2)}{(x^2+y^2)^3} \text{ if } (x,y) \neq (0,0), \\ 0 \text{ if } (x,y) = (0,0). \end{cases}$$

Note that this is not continuous on $[0,1] \times [0,1]$, even though it is continuous on $(0,1] \times (0,1]$. Let us compute its integral using the second method for both orders.

Letting $u = x^2 + y^2$, we find that

$$\begin{aligned} \int_0^1 \left(\int_0^1 f(x,y)\, dy \right) dx &= \int_0^1 \left(\int_{x^2}^{x^2+1} \frac{2x(2x^2 - u^2)}{u^3}\, du \right) dx \\ &= \int_0^1 \left(\int_{x^2}^{x^2+1} \frac{4x^3}{u^3} - \frac{2x}{u^2}\, du \right) dx \\ &= \int_0^1 \frac{2x}{(x^2+1)^2}\, dx \\ &= \int_1^2 \frac{1}{v^2}\, dv = 1 - \frac{1}{2} = \frac{1}{2}. \end{aligned}$$

Noting that $f(x,y) = -f(y,x)$, we must have that

$$\begin{aligned} \int_0^1 \left(\int_0^1 f(x,y)\, dx \right) dy &= -\int_0^1 \left(\int_0^1 f(y,x)\, dx \right) dy \\ &= -\int_0^1 \left(\int_0^1 f(u,v)\, dv \right) du \\ &= -\frac{1}{2} \end{aligned}$$

where we may the substitutions $v = x$ and $u = y$. Thus, the two integrals do not agree!

Thus, using iterated integrals can only define the integral for a certain class of functions, for which the order does not matter. Otherwise $f(x,y) = xy$ and $g(x,y) = f(y,x) = xy$ could have different integrals! This class of functions turns out to be functions continuous on the entire rectangular domain of integration.

Proposition 8.6. *Let $m > 1$ be an integer, $a_1, \ldots, a_m, b_1, \ldots, b_m \in \mathbb{R}$ be real numbers such that $a_i < b_i$ for each $1 \leq i \leq m$, and let*

$$f : [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_m, b_m] \to \mathbb{R}$$

be a continuous function. Then

$$\begin{aligned} g : [a_1, b_1] \times \cdots \times [a_{m-1}, b_{m-1}] &\to \mathbb{R}, \\ (x_1, \ldots, x_{m-1}) &\mapsto \int_{a_m}^{b_m} f(x_1, \ldots, x_{m-1}, t)\, dt \end{aligned}$$

is continuous.

Proof. Suppose we are given $\varepsilon > 0$, and choose $\varepsilon_0 > 0$ such that $(b_n - a_n)\varepsilon_0 < \varepsilon$. Since f is continuous on a closed and bounded set, it is uniformly continuous by Theorem 3.20. Hence, there exists $\delta > 0$ such that

$$|f(x_1, \ldots, x_{m-1}, t) - f(u_1, \ldots, u_{m-1}, t)| < \varepsilon_0$$

for all reals with $a_i \leq x_i$, $u_i \leq b_i$ for $1 \leq i \leq m-1$ satisfying

$$\sqrt{\sum_{i=1}^{m-1} (x_i - u_i)^2} < \delta,$$

and all $t \in [a_m, b_m]$. Then, for such $(x_1, \ldots, x_{m-1})$ and $(u_1, \ldots, u_{m-1})$,

$$\begin{aligned} |g(x_1, \ldots, x_{m-1}) - g(u_1, \ldots, u_{m-1})| &= \left| \int_{a_m}^{b_m} f(x_1, \ldots, x_{m-1}, t) - f(u_1, \ldots, u_{m-1}, t)\, dt \right| \\ &\leq \int_{a_m}^{b_m} |f(x_1, \ldots, x_{m-1}, t) - f(u_1, \ldots, u_{m-1}, t)|\, dt \\ &\leq \int_{a_m}^{b_m} \varepsilon_0 \, dt = (b_m - a_m)\varepsilon_0 < \varepsilon. \end{aligned}$$

Hence, g is continuous. □

We will only prove that the order of our integrating variables does not matter in the two dimensional case. The m-dimensional case follows via repeated applications of the two dimensional case, since transpositions generated the symmetric group.

Theorem 8.7 (Fubini's Theorem)**.** *Let $f : [a_1, b_1] \times [a_2, b_2] \to \mathbb{R}$ be continuous. Then*

$$\int_{a_2}^{b_2} \left(\int_{a_1}^{b_1} f(x, y)\, dx \right) dy = \int_{a_1}^{b_1} \left(\int_{a_2}^{b_2} f(x, y)\, dy \right) dx.$$

Furthermore, if f is Riemann integrable, in the sense of limiting values of double Riemann sums, then its Riemann integral equal to the common value of the iterated integrals.

Proof. We first consider the iterated integrals. Since f is continuous on a closed and bounded set, f is uniformly continuous. Thus, given any $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|f(x, y) - f(u, v)| < \varepsilon$$

for all $a_1 \leq x, u \leq b_1$, $a_2 \leq y, v \leq b_2$ such that

$$|x - u| < \delta \text{ and } |y - v| < \delta.$$

This follows by choosing δ such that

$$\sqrt{(x-u)^2 + (y-v)^2} < \sqrt{2}\delta$$

implies our ε-bound.

Let $P = (u_0, \ldots, u_p)$ be a partition of $[a_1, b_1]$ and $Q = (v_0, \ldots, v_q)$ be a partition of $[a_2, b_2]$ such that

$$u_{j+1} - u_j < \delta \text{ and } v_{k+1} - v_k < \delta$$

for all $0 \le j \le p-1$, $0 \le k \le q-1$. Then, whenever $x \in [u_j, u_{j+1}]$ and $y \in [v_k, v_{k+1}]$, our δ-bound is satisfied and so

$$|f(x,y) - f(u_j, v_k)| < \varepsilon.$$

Now, by linearity of the integral

$$\begin{aligned}
\int_{a_2}^{b_2} \left(\int_{a_1}^{b_1} f(x,y)\,dx \right) dy &= \sum_{j=0}^{p-1} \sum_{k=0}^{q-1} \int_{v_k}^{v_{k+1}} \left(\int_{u_j}^{u_{j+1}} f(x,y)\,dx \right) dy \\
&\le \sum_{j,k} \int_{v_k}^{v_{k+1}} \left(\int_{u_j}^{u_{j+1}} f(u_j, v_k) + \varepsilon\,dx \right) dy \\
&= \sum_{j,k} (f(u_j, v_k) + \varepsilon)(u_{j+1} - u_j)(v_{k+1} - v_k) \\
&= S + \varepsilon(b_1 - a_1)(b_2 - a_2)
\end{aligned}$$

where

$$S = \sum_{j=0}^{p-1} \sum_{k=0}^{q-1} f(u_j, v_k)(u_{j+1} - u_j)(v_{k+1} - v_k).$$

Similarly

$$\int_{a_2}^{b_2} \left(\int_{a_1}^{b_1} f(x,y)\,dx \right) dy \ge S - \varepsilon(b_1 - a_1)(b_2 - a_2)$$

and so

$$\left| \int_{a_2}^{b_2} \left(\int_{a_1}^{b_1} f(x,y)\,dx \right) dy - S \right| \le \varepsilon(b_1 - a_1)(b_2 - a_2)$$

As p and q are finite, it does not matter whether we sum over j first or k first:

$$\sum_{j=0}^{p-1} \sum_{k=0}^{q-1} = \sum_{k=0}^{q-1} \sum_{j=0}^{p-1},$$

and so the same argument shows that

$$\left| \int_{a_1}^{b_1} \left(\int_{a_2}^{b_2} f(x,y)\,dy \right) dx - S \right| \le \varepsilon(b_1 - a_1)(b_2 - a_2)$$

and hence, by the triangle inequality

$$\left| \int_{a_2}^{b_2} \left(\int_{a_1}^{b_1} f(x,y)\,dx \right) dy - \int_{a_1}^{b_1} \left(\int_{a_2}^{b_2} f(x,y)\,dy \right) dx \right| \leq 2\varepsilon(b_1 - a_1)(b_2 - a_2).$$

This holds for every $\varepsilon > 0$, but the left hand side is constant and has no dependence on ε or δ. As we can take ε arbitrarily small, this can only hold if

$$\int_{a_2}^{b_2} \left(\int_{a_1}^{b_1} f(x,y)\,dx \right) dy = \int_{a_1}^{b_1} \left(\int_{a_2}^{b_2} f(x,y)\,dy \right) dx.$$

Now, suppose that f is Riemann integrable, and consider a partition

$$P = ((x_0, x_1, \ldots, x_p); (y_0, y_1, \ldots, y_q))$$

of $[a_1, b_1] \times [a_2, b_2]$ into rectangles $R_{j,k} = [x_j, x_{j+1}] \times [y_k, y_{k+1}]$. As each $R_{j,k}$ is closed and bounded, and f is continuous, f achieves a minimum and maximum value in $R_{j,k}$. Let $(s_{j,k}, t_{j,k})$ be a point in $R_{j,k}$ for which f is minimal, and $(S_{j,k}, T_{j,k})$ be a point in $R_{j,k}$ for which f is maximal. Then we have that

$$f(s_{j,k}, t_{j,k})(x_{j+1} - x_j)(y_{k+1} y_k) \leq \int_{y_k}^{y_{k+1}} \int_{x_j}^{x_{j+1}} f(x,y)\,dx\,dy$$

and

$$\int_{y_k}^{y_{k+1}} \int_{x_j}^{x_{j+1}} f(x,y)\,dx\,dy \leq f(S_{j,k}, T_{j,k})(x_{j+1} - x_j)(y_{k+1} y_k)$$

for every $0 \leq j \leq p-1$, $0 \leq k \leq q-1$. Summing over all possible j and k, we get that

$$\sum_{j,k} f(s_{j,k}, t_{j,k})(x_{j+1} - x_j)(y_{k+1} y_k) \leq \int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x,y)\,dx\,dy$$

and

$$\int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x,y)\,dx\,dy \leq \sum_{j,k} f(S_{j,k}, T_{j,k})(x_{j+1} - x_j)(y_{k+1} y_k).$$

In particular, we must have that

$$\inf_{\mathcal{P}_{p,q}([a_1,b_1]\times[a_2,b_2])} \sum_{j,k} f(z_{j,k}, w_{j,k})(x_{j+1} - x_j)(y_{k+1} y_k) \leq \int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x,y)\,dx\,dy$$

and

$$\int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x,y)\,dx\,dy \leq \sup_{\mathcal{P}_{p,q}([a_1,b_1]\times[a_2,b_2]} \sum_{j,k} f(z_{j,k}, w_{j,k})(x_{j+1} - x_j)(y_{k+1} y_k).$$

where we take the supremum/infimum over tagged partitions. But, by the assumption that f is Riemann integrable, the limit of these as we consider partitions with increasingly many parts are both equal to some real number $\int_{[a_1,b_1]\times[a_2,b_2]} f(x,y)\,dA$. Thus

$$\int_{[a_1,b_1]\times[a_2,b_2]} f(x,y)\,dA \le \int_{a_2}^{b_2}\int_{a_1}^{b_1} f(x,y)\,dx\,dy \le \int_{[a_1,b_1]\times[a_2,b_2]} f(x,y)\,dA$$

and so we must have equality. □

9 Counterexamples in Analysis

The following is just a collection of functions that display counter-intuitive behaviour, or otherwise demonstrate the importance of certain assumptions in analysis. We begin with some examples illustrating the difference between two dimensional limits and iterated limits.

Example 9.1. *Let*

$$f(x,y) = \begin{cases} \frac{x^2-y^2}{x^2+y^2} & \text{if } (x,y) \neq (0,0), \\ 0 & \text{otherwise.} \end{cases}$$

Then, for $x \neq 0$

$$\lim_{y\to 0} f(x,y) = \frac{x^2}{x^2} = 1$$

and hence

$$\lim_{x\to 0}\lim_{y\to 0} f(x,y) = \lim_{x\to 0} 1 = 1$$

as limits depend only on the values away from the limit point.

Similarly, for $y \neq 0$

$$\lim_{x\to 0} f(x,y) = \frac{-y^2}{y^2} = -1$$

and hence

$$\lim_{y\to 0}\lim_{x\to 0} = \lim_{y\to 0} -1 = -1.$$

Thus, the two ordered limits do not agree, and we cannot possibly have a two dimensional limit.

Example 9.2. *Even if the two ordered limits do agree, this does not guarantee that the two dimensional limit will exist. Consider*

$$f(x,y) = \begin{cases} 1 & \text{if } xy \neq 0, \\ 0 & \text{if } x = 0 \text{ or } y = 0. \end{cases}$$

Then

$$\lim_{y\to 0} f(x,y) = \begin{cases} 1 & \text{if } x \neq 0, \\ 0 & \text{if } x = 0 \end{cases}$$

and

$$\lim_{x\to 0} f(x,y) = \begin{cases} 1 \text{ if } y \neq 0, \\ 0 \text{ if } y = 0 \end{cases}$$

Hence

$$\lim_{x\to 0}\lim_{y\to 0} f(x,y) = 1 = \lim_{y\to 0} \lim x \to 0 f(x,y)$$

but for all $\delta > 0$ *there exists points* (x,y) *within* δ *of the origin such that* $f(x,y) = 0$ *and points such that* $f(x,y) = 1$. *Thus,* $f(x,y)$ *cannot approach a meaningful limit as* $(x,y) \to (0,0)$. *Thus, the two dimensional limit does not exist.*

Next we will give some examples of unusual behaviour with partial derivatives and total derivatives.

Example 9.3. *Let*

$$f(x,y) = \begin{cases} x^2 \sin\left(\frac{1}{x}\right) + y^2 \sin\left(\frac{1}{y}\right) \text{ if } x, y \neq 0, \\ x^2 \sin\left(\frac{1}{x}\right) \text{ if } x \neq 0, y = 0, \\ y^2 \sin\left(\frac{1}{y}\right) \text{ if } x = 0, y \neq 0, \\ 0 \text{ otherwise.} \end{cases}$$

We can easily compute the partial derivatives

$$\frac{\partial f}{\partial x}(x,y) = \begin{cases} 2x \sin\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right) \text{ if } x \neq 0, \\ 0 \text{ otherwise} \end{cases}$$

and

$$\frac{\partial f}{\partial y}(x,y) = \begin{cases} 2y \sin\left(\frac{1}{y}\right) - \cos\left(\frac{1}{y}\right) \text{ if } y \neq 0, \\ 0 \text{ otherwise.} \end{cases}$$

Neither of these are continuous at $(0,0)$, *but* f *is differentiable at* $(0,0)$ *with* $(\mathrm{D}\, f)_{(0,0)} = 0$. *We can see that*

$$f(h,k) - f(0,0) = \begin{cases} h^2 \sin\left(\frac{1}{h}\right) + k^2 \sin\left(\frac{1}{k}\right) \text{ if } h, k \neq 0, \\ h^2 \sin\left(\frac{1}{h}\right) \text{ if } h \neq 0, k = 0, \\ k^2 \sin\left(\frac{1}{k}\right) \text{ if } h = 0, k \neq 0. \end{cases}$$

Since $|\sin(\alpha)| \leq 1$ *for every* $\alpha \in \mathbb{R}$, *we can see that in each of these cases, we must have that*

$$|f(h,k) - f(0,0)| \leq h^2 + k^2 = \|(h,k)\|^2.$$

Thus

$$\left| \frac{f(h,k) - f(0,0)}{\|(h,k)\|} \right| \leq \|(h,k)\|$$

which tends to 0 *as* $(h,k) \to (0,0)$, *as needed.*

Example 9.4. *Let*

$$X = \mathbb{R}^2 \setminus \{(x,0) \in \mathbb{R}^2 \mid x \geq 0\}$$

be the open set obtained by removing the non-negative real axis. Define $f : X \to \mathbb{R}$ by

$$f(x,y) = \begin{cases} x^2 \text{ if } x,y > 0, \\ 0 \text{ otherwise.} \end{cases}$$

Then

$$\frac{\partial f}{\partial x}(x,y) = \begin{cases} 2x \text{ if } x,y > 0, \\ 0 \text{ otherwise} \end{cases}$$

and $\frac{\partial f}{\partial y}(x,y) = 0$. These are both continuous on X, so f is continuously differentiable on X with everywhere vanishing y-derivative, but f is not independent of y:

$$f(1,1) = 1 \neq 0 = f(1,-1).$$

Next, we consider some functions with interesting relationships to continuity and differentiability.

Example 9.5. *The function*

$$f(x) = \begin{cases} x \text{ if } x \in \mathbb{Q}, \\ -x \text{ if } x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

is continuous only at 0. For $p \in \mathbb{Q}$, $p \neq 0$, and any $\delta > 0$, we can find $x \in \mathbb{R} \setminus \mathbb{Q}$ such that $|x - p| < \delta$, but

$$|f(x) - f(p)| = p + x \geq \max(2p - \delta, p)$$

which cannot be made arbitrarily small by making δ smaller. Similarly, f is not continuous at any $p \in \mathbb{R} \setminus \mathbb{Q}$. But at $p = 0$,

$$|f(x) - f(p)| = |x|$$

which can be made arbitrarily small, by considering $|x| < \varepsilon$ for any given $\varepsilon > 0$.

Example 9.6. *The function*

$$f(x) = \begin{cases} x^2 \sin\left(\frac{1}{x}\right) \text{ if } x \neq 0, \\ 0 \text{ if } x = 0 \end{cases}$$

is continuous and differentiable everywhere, but its derivative is not continuous. The function f is clearly continuous away from 0, and it is continuous at 0, as we can squeeze it between $\pm x^2$:

$$0 = \lim_{x\to 0} -x^2 \leq \lim_{x \to 0} f(x) \leq \lim_{x \to 0} x^2 = 0.$$

The derivative is given by

$$f'(x) = \begin{cases} 2x\sin\left(\frac{1}{x}\right) - \cos x & \textit{if } x \neq 0, \\ \lim_{h\to 0} \frac{h^2 \sin\left(\frac{1}{x}\right)}{h} = 0 & \textit{if } x = 0. \end{cases}$$

But $\lim_{x\to 0} f'(x) = -1 \neq f'(0)$*, so* f' *is not continuous at* 0*.*

Example 9.7. *The function*

$$f(x) = \begin{cases} e^{\frac{-1}{x}} & \textit{if } x > 0, \\ 0 & \textit{if } x \leq 0 \end{cases}$$

if infinitely differentiable (smooth) at $x = 0$*, but* $f^{(k)}(0) = 0$ *for every* $k \geq 0$*, so the associated Taylor series gives the* 0 *function!*

To see that the derivatives are 0 *at* 0*, we first show that for all* $k \geq 0$*,*

$$f^{(k)}(x) = \frac{p_k(x)}{x^{2k}} e^{\frac{-1}{x}}$$

for all $x > 0$*. This is true for* $k = 0$*, so we will proceed by induction and assume it holds true for some* $k \geq 0$*. Then*

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

which simplifies to

$$f^{(k+1)}(x) = \frac{p_{k+1}(x)}{x^{2k+2}} e^{\frac{-1}{x}}$$

where

$$p_{k+1}(x) = x^2 p_k'(x) - (2kx + 1)p_k(x)$$

is a polynomial.

Next we show that $f^{(k)}(0) = 0$ *for all* $k \geq 0$*. It is easy to check that this holds for* $k = 0$*, so suppose it holds for some* $k \geq 0$*.*

$$\begin{aligned} f^{(k+1)}(0) &= \lim_{h\to 0^+} \frac{f^{(k)}(h) - f^{(k)}(0)}{h} \\ &= \lim_{h\to 0^+} \frac{p_k(h)}{h^{2k+1}} e^{\frac{-1}{h}}. \end{aligned}$$

We have restricted to the limit from the positive direction, as the negative direction is clearly 0*. As* $\lim_{h\to 0^+} p_k(h) = p_k(0)$ *definitely exists, it suffices to show that*

$$\lim_{h\to 0^+} h^{-2k-1} e^{\frac{-1}{h}} = 0$$

or, even better, that there exists M_{2k+1} *such that*

$$h^{-2k-1} e^{\frac{-1}{h}} \leq M_{2k+1} h \quad \Leftrightarrow h^{-2k-2} e^{\frac{-1}{h}} \leq M_{2k+1}$$

for all $h > 0$, as then the result would follow from the squeeze theorem. Let $g_k(x) = e^{-x}x^{2k+2}$. This has derivative

$$g_k'(x) = (2k+2)e^{-x}x^{2k+1} - e^{-x}x^{2k+2} = e^{-x}x^{2k+1}(2k+2-x)$$

which is positive for $x < 2k+2$ and negative for $x > 2k+2$. Hence g_k has a global maximum M_{2k+1} at $x = 2k+2$, which is to say that

$$g_k(x) = x^{2k+2}e^{-x} \leq M_{2k+1}$$

for all $x \in \mathbb{R}$. In particular, $g_k(\frac{1}{h}) \leq M_{2k+1}$ for all $h > 0$, as needed.

Remark 9.8. *Having functions like this is actually a good thing! The related function*

$$f(x) = \begin{cases} e^{\frac{-1}{1-x^2}} & \text{if } x \in (-1,1), \\ 0 & \text{otherwise} \end{cases}$$

gives a really good smooth approximation to the indicator function

$$\chi_{(-1,1)}(x) = \begin{cases} 1 & \text{if } x \in (-1,1), \\ 0 & \text{otherwise} \end{cases}$$

which allows us to extend many results about convergence of Fourier series and transforms to "nice" discontinuous functions, a result with finds many applications in both physics and analytic number theory!

Example 9.9. *The Weierstrass function*

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

can be shown to be continuous everywhere using the Weierstrass M-test, which is a combination of the comparison test (to establish both pointwise and uniform convergence) and a result about uniform convergence of sequences of continuous functions. However, it is too bumpy, exhibiting fractal-ish behaviour, to be differentiable at any point.

The last two examples also illustrate some unexpected or unpleasant behaviour involving derivatives.

Example 9.10. *The sequence of functions $\{f_n(x) = \frac{\sin(nx)}{n}\}$ is a sequence of continuously differentiable functions. Furthermore this sequence convergences uniformly to the zero function. However, the sequence of derivatives*

$$\{f_n'(x) = \cos(nx)\}$$

does not even converge pointwise to a function.

Example 9.11. *Cantor's function,*

$$c : [0, 1] \to \mathbb{R}$$

sometimes called the Devil's Staircase, is an example of a non-constant monotonically increasing continuous function, which is differentiable almost everywhere. Almost-everywhere differentiability is something that we would formally define using measures, but for now it is sufficient to think of it as only failing to be differentiable on a subset of $[0, 1]$ *containing no open intervals as a subset - the points are "almost discrete".*

What is bizarre about this function is that $c(0) = 0$*,* $c(1) = 1$*, but* $c'(x) = 0$ *everywhere it is defined. How do we define* c*? Well, to define* $c(x)$*, we begin by writing* x *in base* 3*. If the base* 3 *expansion of* x *contains a* 1*, we discard all digits after the first* 1*. If the base* 3 *expansion contains only* 2*s and* 0*s, we replace all the* 2*s with* 1*s. We then read the obtained sequence of* 0*s and* 1*s as though it were a binary number, giving* $c(x)$*.*

10 Summary of main results

As a possible study aid, we summarise here the major results from each section of the course. It is not a comprehensive list, but should provide a good start point. It notably does not include definitions or examples, or every useful corollary.

Chapter 0

- Theorem 0.17 - Bounded monotonic sequences converge
- Bolzano-Weierstrass Theorem - Theorem 0.18 - Bounded sequences contain monotonic subsequences

Chapter 1

- Lemma 1.4 - Components are bounded by norm
- Lemma 1.5 - Sequences converge iff their components do
- Bolzano-Weierstrass Theorem - Theorem 1.6 - Bounded sequences have convergent subsequences
- Proposition 1.8 - The Cauchy-Schwarz inequality
- Corollary 1.9 - The triangle inequality
- Theorem 1.12 - Cauchy sequences converge

Chapter 2

- Corollary 2.11 - Open balls are open
- Proposition 2.14 - Open sets form a topology
- Lemma 2.25 - Convergent sequences in closed sets converge in the closed set

Chapter 3

- Proposition 3.2 - Composition of continuous functions is continuous
- Proposition 3.3 - Continuous functions commute with limits
- Lemma 3.6 - Coordinate maps are continuous
- Proposition 3.7 - A function is continuous iff its components are.
- Lemma 3.8 - Sums and products are continuous - You might be asked to prove that sums of continuous functions are continuous without use of this lemma

- Proposition 3.12 - The inverse image of open sets is open for continuous functions
- Corollary 3.14 - Special cases of Proposition 3.12
- Extreme Value Theorem - Theorem 3.17 - Continuous functions on closed and bounded sets achieve their extrema
- Theorem 3.20 - Continuous functions on closed and bounded sets are uniformly continuous

Chapter 4

- Proposition 4.4 - A closed set contains all its limit points
- Proposition 4.7 - Limits can be computed in terms of components
- Proposition 4.8 - Linear combinations of limits are limits of linear combinations
- Lemma 4.10 - Limits commute with continuous functions

Chapter 5

- Rolle's Theorem - Theorem 5.9 - Derivatives vanish if a function turns around
- Mean Value Theorem - Theorem 5.10 - The average slope is attained
- Taylor's Theorem - Theorem 5.17 - Taylor series mostly work!

Chapter 6

- Corollary 6.10 - Bounded partial derivaitves imply continuity
- Lemma 6.19 - The derivative of a linear map is linear
- Lemma 6.20 - Differentiable functions satisfy inequalities
- Lemma 6.22 - Differentiable functions are continuous
- Proposition 6.23 - Continuous partial derivatives implies differentiable with derivative given by the Jacobian
- Proposition 6.24 - Directional derivatives are determined by total derivatives
- Corollary 6.25 - If a function is differentiable and has partial derivatives, the derivatives is the Jacobian
- Proposition 6.30 - Differentiable functions have bounded growth
- Chain Rule - Proposition 6.33 - Derivatives of compositions of functions. Definitely a named theorem

Chapter 7

- Banach Fixed Point Theorem - Theorem 7.3 - Contractions have fixed points
- The Inverse Function Theorem - Theorem 7.9 - Continuously differentiable functions with non-vanishing derivative have local inverses
- The Implicit Function Theorem - Theorem 7.13 - Nice level sets can be parametrised

Chapter 8

- Example 8.5 - A counterexample in which order of integration matters
- Fubini's Theorem - Theorem 8.7 - Order of integration does not matter for continuous functions

Chapter 9

Chapter 9 contains no examinable material, just interesting examples.