M2PM1: Real Analysis

Dr. Claude Warnick

August 24, 2017

Abstract

In first year analysis courses, you learned about the real numbers and were introduced to important concepts such as completeness; convergence of sequences and series; continuity; differentiability of functions of one real variable. In this course we will extend these ideas to higher dimensions, as well as introducing new fundamental tools for analysis, including uniform convergence and integration.

Contents

Notes on Exercises

A crucial feature of mathematics is that it is a *practical* subject. One learns to a large extent by *doing*: solving problems, checking details of proofs, finding counterexamples.

Throughout the notes are incorporated exercises to test your understanding, develop intuition and establish results that are required later on. You should not treat these exercises as optional[1]. **You won't understand the material in this course unless you spend the time trying to solve the problems!** You will have to think hard, and you may have to come back to some of the problems over more than one session to crack them. Persist.

Notes on the Appendix

Appendix A consists of an introduction to differentiation in one dimension, and will be covered in the JMC problem classes at the start of term. Students on the mathematics course who sat the course M1P1 Analysis 1 should have seen this material before, and can treat it as optional (but they may wish to study it for revision). Those students on the joint computing and mathematics course will not have seen this material before, so should attend the problem classes and study the appendix.

Notes on non-examinable material

Material marked $(*)$ is not examinable, but will be valuable in future courses, especially if you take (applied or pure) analysis courses in your third year and beyond. You should certainly at least read through the notes on these sections, even if you choose not to attempt the questions. I will try to indicate in lectures when I'm covering this material.

[1] The exception here are those questions marked $(*)$, which go beyond the scope of the course, but which are hopefully nevertheless interesting.

Introduction

In this chapter we're going to start off by introducing a few definitions so that we've agreed on notation. We're also going to extend some of the ideas that you saw last year (such as limits and continuity) to higher dimensions. The definitions are almost identical, so this should mostly feel like a review chapter to begin with, although some of the ideas we're going to approach from a different point of view.

Notation for $\mathbb{R}^n$

Numbers

First, we quickly recall the various sets of numbers that you were introduced to last year. I'm going to define the natural numbers to start from 0, and denote by $\mathbb{N}^*$ the set of positive natural numbers:

$$\mathbb{N} := \{0, 1, 2, \ldots\},$$
$$\mathbb{N}^* := \{1, 2, 3, \ldots\}.$$

Be careful: not all authors use this convention, for some $\mathbb{N}$ denotes the positive natural numbers. The set of integers is denoted $\mathbb{Z}$:

$$\mathbb{Z} := \{\ldots, -2, -1, 0, 1, 2, \ldots\}.$$

If needed, the set $\mathbb{Z}^*$ is the non-zero integers $\mathbb{Z}^* := \mathbb{Z} \setminus \{0\}$. We know how to add, subtract and multiply elements of $\mathbb{Z}$ (and hence $\mathbb{N}$, $\mathbb{N}^*$). The rational numbers are denoted $\mathbb{Q}$:

$$\mathbb{Q} := \left\{ \frac{p}{q} \;\middle|\; p \in \mathbb{Z}, \quad q \in \mathbb{Z}^* \right\},$$

where we make the identification:

$$\frac{p}{q} \sim \frac{mp}{mq}, \qquad \forall\, m \in \mathbb{Z}^*.$$

Again, if required $\mathbb{Q}^* := \mathbb{Q} \setminus \{0\}$ is the set of non-zero rational numbers. We can add multiply and subtract elements of $\mathbb{Q}$, and we also can divide by elements of $\mathbb{Q}^*$. On $\mathbb{Q}$

we have a notion of ordering $\leq$, so that we can say whether a rational number is greater than, less than or equal to another.

The set $\mathbb{R}$ of real numbers is obtained as the *completion* of $\mathbb{Q}$, and inherits the operations of arithmetic and the order relation from $\mathbb{Q}$. We have:

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}.$$

Moreover, $\mathbb{R}$ satisfies the *completeness axiom*:

> If $A \subset \mathbb{R}$ is non-empty and bounded above, then A has a least upper bound.

An important function defined on all real numbers is the *modulus function*, given by:

$$|x| := \begin{cases} x & x \geq 0, \\ -x & x < 0. \end{cases}$$

This has the properties:

i) For all $x \in \mathbb{R}$, we have $|x| \geq 0$, with $|x| = 0$ if and only if $x = 0$.

ii) For all $x \in \mathbb{R}$, if $a \in \mathbb{R}$, then $|ax| = |a|\,|x|$

iii) The modulus function satisfies the triangle inequality:

$$|x + y| \leq |x| + |y|, \qquad \forall\, x, y, \in \mathbb{R}$$

The set $\mathbb{R}^n$

We can now introduce the space $\mathbb{R}^n$, which consists of column vectors with n real components[2]:

$$\mathbb{R}^n = \left\{ \left(x^1, \ldots, x^n\right)^t \middle|\ x^i \in \mathbb{R}, i = 1, \ldots, n \right\}$$

If $x, y \in \mathbb{R}^n$ are two vectors in $\mathbb{R}^n$ with

$$x = \left(x^1, \ldots, x^n\right)^t, \qquad y = \left(y^1, \ldots, y^n\right)^t,$$

we can define the sum of the two by:

$$x + y := \left(x^1 + y^1, \ldots, x^n + y^n\right)^t.$$

If $\lambda \in \mathbb{R}$, we define:

$$\lambda x := \left(\lambda x^1, \ldots, \lambda x^n\right)^t.$$

With these definitions, $\mathbb{R}^n$ has the structure of a *vector space* over $\mathbb{R}$.

[2] I will try to stick to the convention of using superscripts to label components of vectors, and subscripts to label different vectors, so that $x_1, x_2 \in \mathbb{R}^n$ are two different vectors, while $x^1, x^2 \in \mathbb{R}$ are the component of one vector.

We can define the inner product of two vectors by:

$$\langle x, y \rangle = \sum_{i=1}^{n} x^i y^i,$$

and we define the length, or norm of a vector by:

$$||x|| = \sqrt{\langle x, x \rangle}.$$

This has the properties:

i) For all $x \in \mathbb{R}^n$, we have $||x|| \geq 0$, with $||x|| = 0$ if and only if $x = 0$.

ii) For all $x \in \mathbb{R}^n$, if $a \in \mathbb{R}$, then $||ax|| = |a| \, ||x||$

iii) The norm satisfies the triangle inequality:

$$||x + y|| \leq ||x|| + ||y|| \,, \qquad \forall \; x, y, \in \mathbb{R}^n. \tag{I.1}$$

These properties are often abstracted away to define more general *normed spaces* [See the Functional Analysis course M3P7]. The norm gives us a useful notion of *distance* between two points. The distance from x to y is given by $||x - y||$. Notice that if $n = 1$ we have $|\cdot| = ||\cdot||$, and we will use either interchangeably in this case.

A very important class of subsets of $\mathbb{R}^n$ are the *open balls.* For $x \in \mathbb{R}^n$, $r > 0$, these are defined as follows:

$$B_r(x) = \{ y \in \mathbb{R}^n : ||x - y|| < r \} \,.$$

That is, $B_r(x)$ consists of all points in $\mathbb{R}^n$ which are a distance less than r from x.

Exercise 1.1. a) Show that the inner product has the following properties:

$$\langle x, y \rangle = \langle y, x \rangle \,, \qquad \langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle \,, \qquad \langle ax, y \rangle = a \langle x, y \rangle \,.$$

for all $x, y, z \in \mathbb{R}^n$ and $a \in \mathbb{R}$.

b) For $t \in \mathbb{R}$ and $x, y \in \mathbb{R}^n$, show that:

$$||x + ty||^2 = ||x||^2 + 2t \langle x, y \rangle + t^2 \, ||y||^2 \geq 0 \tag{I.2}$$

c) By thinking of (I.2) as a quadratic in t, and considering its possible roots, deduce the *Cauchy-Schwartz* inequality:

$$|\langle x, y \rangle| \leq ||x|| \, ||y|| \,. \tag{I.3}$$

When does equality hold?

d) Deduce the triangle inequality (I.1).

e) Show the reverse triangle inequality:

$$\big| \, ||x|| - ||y|| \, \big| \leq ||x - y||$$

f) Suppose $x = (x^1, \ldots, x^n)^t \in \mathbb{R}^n$.

i) Show that:

$$\max_{k=1,\ldots,n} \left| x^k \right| \leq ||x|| \,.$$

ii) Show that:

$$||x|| \leq \sqrt{n} \max_{k=1,\ldots,n} \left| x^k \right| .$$

Sets and convergence

Now that we have a few definitions relating to $\mathbb{R}^n$, we're ready to revisit some concepts from first year analysis and see how they can be extended to higher dimensions.

Convergence of sequences in $\mathbb{R}^n$

A sequence in $\mathbb{R}^n$ is an ordered list

$$x_0, x_1, \ldots, x_i, \ldots$$

with each $x_i \in \mathbb{R}^n$ for $i = 0, 1, 2, \ldots$. This is often written $(x_i)_{i=0}^{\infty}$, or $(x_i)_{i \in \mathbb{N}}$. A very important concept relating to sequences is that of *convergence*.

Definition I.1. A sequence $(x_i)_{i=0}^{\infty}$ with $x_i \in \mathbb{R}^n$ converges to the vector $x \in \mathbb{R}^n$ if the following holds: given $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $i \geq N$ we have

$$||x_i - x|| < \epsilon.$$

We then write:

$$x_i \to x, \quad \text{as } i \to \infty,$$

or

$$\lim_{i \to \infty} x_i = x.$$

You should go back and compare this definition to the definition of convergence for a sequence of real numbers that you saw in your first year.

Lemma I.1. *The sequence of vectors $(x_i)_{i=0}^{\infty}$ with $x_i \in \mathbb{R}^n$ converges to the vector $x \in \mathbb{R}^n$ if and only if each component of x_i converges to the corresponding component of x. That is, if we write:*

$$x_i = (x_i^1, \ldots, x_i^n)^t, \quad \textit{and} \quad x = (x^1, \ldots x^n)^t,$$

then

$$x_i^k \to x^k, \quad \textit{as} \;\; i \to \infty, \quad \forall\, k = 1, \ldots n.$$

Proof. First suppose that:

$$x_i^k \to x^k, \quad \text{as} \quad i \to \infty, \quad \forall\, k = 1, \dots n.$$

Let $\epsilon > 0$. Then for each $k = 1, \dots n$, there exists N_k such that if $i \geq N_k$ we have:

$$\left| x_i^k - x^k \right| < \frac{\epsilon}{\sqrt{n}}.$$

Let $N = \max\{N_1 \dots, N_n\}$. Then for $i \geq N$, we have:

$$\max_{k=1,\dots,n} \left| x_i^k - x^k \right| < \frac{\epsilon}{\sqrt{n}}.$$

Now, recall from Exercise 1.1 f) ii) that:

$$||x|| \leq \sqrt{n} \max_{k=1,\dots,n} \left| x^k \right|,$$

so we deduce:

$$||x_i - x|| \leq \sqrt{n} \max_{k=1,\dots,n} \left| x_i^k - x^k \right| < \epsilon.$$

This establishes the result in one direction.

Now suppose that:

$$x_i \to x, \quad \text{as} \quad i \to \infty.$$

Let $\epsilon > 0$. Then there exists N such that for $i \geq N$ we have:

$$||x_i - x|| < \epsilon.$$

Recall from Exercise 1.1 f) i) that:

$$\max_{k=1,\dots,n} \left| x^k \right| \leq ||x|| \,.$$

Thus for any $k = 1, \dots n$ we have that if $i \geq N$:

$$\left| x_i^k - x^k \right| \leq \max_{k=1,\dots,n} \left| x_i^k - x^k \right| \leq ||x_i - x|| < \epsilon,$$

which implies $x_i^k \to x^k$, which completes the proof. □

Exercise 1.2. Suppose that $(x_i)_{i=0}^\infty$ and $(y_i)_{i=0}^\infty$ with $x_i, y_i \in \mathbb{R}^n$ are two sequences of vectors with

$$x_i \to x, \quad y_i \to y, \qquad \text{as } i \to \infty.$$

a) Show that

$$x_i + y_i \to x + y \qquad \text{as } i \to \infty.$$

b) Show that

$$\langle x_i, y_i \rangle \to \langle x, y \rangle \qquad \text{as } i \to \infty,$$

deduce that

$$||x_i|| \to ||x|| \qquad \text{as } i \to \infty.$$

c) Suppose that $(a_i)_{i=0}^\infty$ with $a_i \in \mathbb{R}$ is a sequence of real numbers with $a_i \to a$ as $i \to \infty$. Show that:

$$a_i x_i \to a x \qquad \text{as } i \to \infty.$$

Open, closed and compact sets

In the study of the real line we are familiar with objects such as the open interval $(0,1)$, the closed interval $[0,1]$ and other sets which are neither open nor closed such as $(0,1]$. In higher dimensions we will need to consider sets which have properties such as "openness" and "closedness". In fact, the study of open and closed sets on different spaces is an entire branch of mathematics in itself: *topology*. We will take a quick and dirty approach to the area, and defer more careful treatment to the course M2PM5 Metric Spaces and Topology.

Definition I.2. A subset $U \subset \mathbb{R}^n$ is *open* if for each $x \in U$, there exists $r > 0$ such that $B_r(x) \subset U$.

In other words, about any point in an open set we can find a small ball which is entirely contained in the set. This means that an open set has no edge or boundary points.

Example 1. The ball $B_1(0)$ is open. To see this, suppose $x \in B_1(0)$, so that $||x|| < 1$. Let $r = (1 - ||x||)/2$ and suppose $y \in B_r(x)$. Then:

$$||y|| = ||y - x + x|| \leq ||y - x|| + ||x|| < \frac{1 - ||x||}{2} + ||x|| < \frac{1 + ||x||}{2} < 1$$

so that $y \in B_1(0)$. Thus $B_r(x) \subset B_1(0)$.

Example 2. The set $A = \{x \in \mathbb{R}^n : ||x|| \leq 1\}$ is not open. Clearly $y := (1, 0, \ldots, 0)^t$ belongs to A. On the other hand, if $r > 0$ then $z := (1 + r/2, 0, \ldots, 0)^t$ belongs to $B_r(y)$ but not to A, so there is no $r > 0$ such that $B_r(y) \subset A$.

Exercise 1.3. Which of the following subsets of $\mathbb{R}^n$ is open:

a) $\mathbb{R}^n$?

b) $\emptyset$?

c) $\{x = (x^1, \ldots, x^n)^t \in \mathbb{R}^n : x^1 > 0\}$?

d) $\{x = (x^1, \ldots, x^n)^t \in \mathbb{R}^n : x^i \in [0, 1)\}$?

e) $\mathbb{Q}^n := \{x = (x^1, \ldots, x^n)^t \in \mathbb{R}^n : x^i \in \mathbb{Q}\}$?

Now we move on to discuss closed sets:

Definition I.3. A subset $E \subset \mathbb{R}^n$ is *closed* if for every convergent sequence $(x_i)_{i=0}^{\infty}$ with $x_i \in E$, we have $x = \lim_{i\to\infty} x_i \in E$.

In other words, E contains all of its *limit points*, that is points which can be approached by a sequence from inside E.

Example 3. The single point set $E = \{y\} \subset \mathbb{R}^n$ is closed. Any convergent sequence $(x_i)_{i=0}^{\infty}$ with $x_i \in E$ must in fact be the constant sequence $x_i = y$, for which we certainly have $y = \lim_{i\to\infty} x_i \in E$.

Example 4. The set $E = \{x \in \mathbb{R}^n : ||x|| \leq 1\}$ is closed. Any convergent sequence $(x_i)_{i=0}^{\infty}$ with $x_i \in E$ has $||x_i|| \leq 1$. Now, we know from Exercise 1.2 that if $x_i \to x$, then $||x_i|| \to ||x||$. It's also a result from last year that if we have a convergent sequence of real numbers $(a_i)_{i=0}^{\infty}$ with $a_i \in \mathbb{R}$ satisfying $a_i \leq M$, then $\lim_{i\to\infty} a_i \leq M$. Combining these two facts, we conclude that:

$$||x|| = \left|\left|\lim_{i\to\infty} x_i\right|\right| = \lim_{i\to\infty} ||x_i|| \leq 1,$$

so that $x \in E$. More generally, for $r > 0$ and $y \in \mathbb{R}^n$, the *closed ball*:

$$\overline{B_r(y)} := \{x \in \mathbb{R}^n : ||x - y|| \leq r\}$$

is closed.

Theorem I.2. *A set E is closed if and only if the set $E^c := \{x \in \mathbb{R}^n : x \notin E\}$ is open.*

Proof. First suppose that E is closed and let $y \in E^c$. We have to prove that there exists $r > 0$ such that $B_r(y) \subset E^c$. We proceed by contradiction. Suppose there is no such r, then for each $i \in \mathbb{N}$ we can find $x_i \in B_{2^{-i}}(y) \cap E$. We have:

$$||x_i - y|| < 2^{-i}$$

so $x_i \to y$ as $i \to \infty$. Since E is closed we must have $y \in E$, but this is a contradiction since $y \in E^c$. Thus if E is closed, E^c must be open.

Now consider the converse. Suppose that E^c is open and that $(x_i)_{i=0}^{\infty}$ is a convergent sequence with $x_i \in E$. Let $x = \lim_{i\to\infty} x_i$. We wish to show that $x \in E$. Again we proceed by contradiction. Suppose $x \notin E$, i.e. $x \in E^c$. Since E^c is open, there exists $r > 0$ such that $B_r(x) \subset E^c$. In particular, this means that $||y - x|| \geq r$ for all $y \in E$. However, we know that since $x_i \to x$ there exists N such that if $i \geq N$ we have:

$$||x_i - x|| < \frac{r}{2}.$$

Since $x_i \in E$ this is a contradiction. Thus we have shown that if E^c is open, then E is closed. □

This result enables us to equivalently define closed sets as being those whose complement is open. The fact that these two definitions of closedness are equivalent relies on certain properties of $\mathbb{R}^n$ which do not hold for more general topological spaces, and it is the definition in terms of open complements that is usually taken to be the correct notion to generalise.

Exercise 1.4. Which of the following subsets of $\mathbb{R}^n$ is closed:

a) $\mathbb{R}^n$?

b) $\emptyset$?

c) $\left\{x = (x^1, \ldots, x^n)^t \in \mathbb{R}^n : x^1 \geq 0\right\}$?

d) $\{x = (x^1, \ldots, x^n)^t \in \mathbb{R}^n : x^i \in [0,1)\}$?

e) $\mathbb{Q}^n := \{x = (x^1, \ldots, x^n)^t \in \mathbb{R}^n : x^i \in \mathbb{Q}\}$?

Definition I.4. For any set $A \subset \mathbb{R}^n$, we define:

i) the *interior* of A, denoted A° is the set of points $x \in A$ such that there exists $r > 0$ with $B_r(x) \subset A$.

ii) the *closure* of A, denoted $\overline{A}$, is the set of points $x \in \mathbb{R}^n$ such that there exists a sequence $(x_i)_{i=0}^\infty$ with $x_i \in A$ and $x_i \to x$.

iii) the *boundary* of A, denoted ∂A is the set $\overline{A} \setminus A^\circ$.

From the definition we have that if U is open then $U = U^\circ$, while if E is closed $E = \overline{E}$. We also have that for any set the inclusions:

$$A^\circ \subset A \subset \overline{A},$$

hold.

Exercise 1.5. a) Let $(x_i)_{i=0}^\infty$ be a sequence of vectors $x_i \in \mathbb{R}^n$ with $x_i \to x$. Suppose that the x_i satisfy $||x_i|| < r$ for all i and some $r > 0$. Show that:

$$||x|| \leq r.$$

b) Show that the closure of the open ball $B_r(y) := \{x \in \mathbb{R}^n : ||x - y|| < r\}$ is the closed ball $\overline{B_r(y)} := \{x \in \mathbb{R}^n : ||x - y|| \leq r\}$.

Exercise 1.6. Find A°, $\overline{A}$ and ∂A for the following sets:

a) $A = \mathbb{R}^n$.

b) $A = \{0\}$.

c) $A = \{x = (x^1, \ldots, x^n)^t \in \mathbb{R}^n : x^1 \geq 0\}$?

d) $A = \{x = (x^1, \ldots, x^n)^t \in \mathbb{R}^n : x^i \in [0,1)\ \forall\, i = 1, \ldots, n\}$?

e) $A = \{x = (x^1, \ldots, x^n)^t \in \mathbb{R}^n : x^i \in \mathbb{Q}\ \forall\, i = 1, \ldots, n\}$?

Exercise 1.7. a) Show that if U_1, U_2 are open, then so are:

$$i)\ U_1 \cup U_2 \qquad ii)\ U_1 \cap U_2$$

b) Show that if E_1, E_2 are closed, then so are:

$$i)\ E_1 \cup E_2 \qquad ii)\ E_1 \cap E_2$$

*c) Suppose $\mathcal{U}$ is any collection of open sets.

i) Show that $\bigcup \mathcal{U}$ is open.

ii) Give an example showing that $\bigcap \mathcal{U}$ need not be open.

iii) What are the analogous statements for closed sets?

It is useful to also introduce the notion of *boundedness* for a subset of $\mathbb{R}^n$:

Definition I.5. A subset $A \subset \mathbb{R}^n$ is *bounded* if there exists $M \geq 0$ such that for all $x \in A$ we have:

$$||x|| \leq M.$$

A set which is both closed and bounded is called *compact.*

The Bolzano–Weierstrass theorem

The Bolzano–Weierstrass theorem is the the most basic of a family of important results in analysis that allow us to extract convergence from some sequence, provided that we have some compactness. These type of results can be very powerful, but often seem a bit mystical to begin with.

Theorem I.3 (Bolzano–Weierstrass)**.** *Suppose $E \subset \mathbb{R}^n$ is compact and that $(x_i)_{i=0}^{\infty}$ is a sequence with $x_i \in E$. Then $(x_i)_{i=0}^{\infty}$ has a subsequence which converges in E. That is, there exist integers:*

$$0 \leq i_1 < i_2 < \ldots < i_j < \ldots$$

such that the sequence $(x_{i_j})_{j=0}^{\infty}$ converges to some $x \in E$.

Before we attack the theorem in the generality it's stated, we'll first establish a simpler result, which you should have seen last year.

Lemma I.4. *Suppose $(a_i)_{i=0}^{\infty}$ is a bounded sequence of real numbers. Then$(a_i)_{i=0}^{\infty}$ has a convergent subsequence.*

Proof. We proceed by constructing a nested sequence of intervals, each containing an infinite number of elements of the sequence. First, since $(a_i)_{i=0}^{\infty}$ is bounded there exists some K such that $a_i \in [-K, K]$ for all $i = 0, 1, \ldots$. We call this interval $I_0 := [-K, K]$, and this certainly contains infinitely many elements of our sequence. Now, we can split the interval I_0 into two intervals $[-K, 0]$ and $[0, K]$. At least one of these intervals contains infinitely many elements of the sequence, so pick the leftmost and call it I_1. We proceed inductively. After j steps we have a closed interval I_j. We split this in half and choose the leftmost half which contains infinitely many elements of the sequence to call I_{j+1}.

We end up with a nested sequence:

$$I_0 \supset I_1 \supset I_2 \ldots \supset I_j \ldots$$

of closed intervals I_j of length $2^{1-j}K$, each containing infinitely many elements of the sequence (see Figure 1). We set $i_0 := 0$ and define i_j inductively by:

$$i_j := \min\{i : a_i \in I_j,\ i > i_{j-1}\}$$

for $j = 1, 2, \ldots$.

Let us denote by c_j the lower endpoint of I_j and d_j the upper endpoint, so that $I_j = [c_j, d_j]$. The sequence $(c_j)_{j=0}^{\infty}$ is monotone increasing and bounded above, so has a limit by the completeness axiom, say $c_j \to c$. Similarly, $(d_j)_{j=0}^{\infty}$ is monotone decreasing and bounded below so $d_j \to d$ for some d. Since $d_j - c_j = 2^{1-j}K$, we have that $c = d$. We also have $c_j \leq a_{i_j} \leq d_j$, so we conclude that

$$a_{i_j} \to c$$

as $j \to \infty$. Thus the subsequence $(a_{i_j})_{j=0}^{\infty}$ which we have constructed is convergent. □

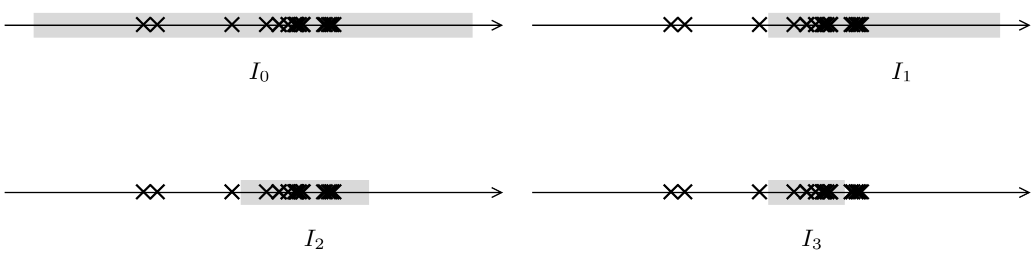

Figure 1 The nested intervals chosen in the proof of Lemma I.4

Now we can extend this result to $\mathbb{R}^n$:

Lemma I.5. *Suppose $(x_i)_{i=0}^{\infty}$, with $x_i \in \mathbb{R}^n$ is a bounded sequence. Then $(x_i)_{i=0}^{\infty}$ has a convergent subsequence.*

Proof. Since $(x_i)_{i=0}^{\infty}$ is bounded, there exists $K > 0$ such that $||x_i|| \leq K$. Recall from Exercise 1.1 f) i) that:

$$\max_{k=1,\ldots,n} \left| x^k \right| \leq ||x|| \,.$$

We conclude that in particular we have $x^1_i \in [-K, K]$ for all $i \in \mathbb{N}$. Now, we can apply Lemma I.4 to the sequence of real numbers $(x^1_i)_{i=0}^{\infty}$ to extract a subsequence $(x^1_{i_{j_1}})_{j_1=0}^{\infty}$ which converges to x^1. Consider the sequence $(x_{i_{j_1}})_{j_1=0}^{\infty}$. This is a subsequence of our original sequence, remains bounded and moreover $x^1_{i_{j_1}} \to x^1$. Now we know that $x^2_{i_{j_1}} \in [-K, K]$ for all $j_1 \in \mathbb{N}$. Again applying Lemma I.4, we can extract a subsequence $(x^2_{i_{j_{1} j_2}})_{j_2=0}^{\infty}$ which converges to x^2. The sequence $(x_{i_{j_{1} j_2}})_{j_2=0}^{\infty}$ is a subsequence of our original sequence, remains bounded and moreover $x^1_{i_{j_{1} j_2}} \to x^2$ and $x^2_{i_{j_{1} j_2}} \to x^1$. Continuing in this way we extract a subsequence of $(x_i)_{i=0}^{\infty}$ such that each component converges. By Lemma I.1 we conclude that the subsequence converges. □

This previous Lemma is sometimes also known as the Bolzano–Weierstrass theorem. With this result, we can finally conclude our proof of Theorem I.3.

Proof of Theorem I.3. Since $(x_i)_{i=0}^{\infty}$ is a sequence with $x_i \in E$ and E is bounded, the sequence is bounded. By Lemma I.5 we have that $(x_i)_{i=0}^{\infty}$ admits a convergent subsequence, $(x_{i_j})_{j=0}^{\infty}$. Since $x_{i_j} \in E$ for all j, and E is closed, we have that $\lim_{j\to\infty} x_{i_j} \in E$. □

Exercise 2.1. We say that a sequence $(x_i)_{i=0}^{\infty}$ with $x_i \in \mathbb{R}^n$ is a *Cauchy sequence* if the following holds: given $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $i, j \geq N$ we have:

$$||x_i - x_j|| < \epsilon.$$

a) Suppose the sequence $(x_i)_{i=0}^{\infty}$ with $x_i \in \mathbb{R}^n$ converges to x. Show that $(x_i)_{i=0}^{\infty}$ is a Cauchy sequence.

b) Suppose $(x_i)_{i=0}^{\infty}$ with $x_i \in \mathbb{R}^n$ is a Cauchy sequence. Show that $(x_i)_{i=0}^{\infty}$ is bounded.

c) Show that every Cauchy sequence in $\mathbb{R}^n$ converges.

Continuity

Last year, you learned about the notion of continuity for functions from $\mathbb{R}$ (or subsets thereof) to $\mathbb{R}$. In this section we'll revisit those definitions and upgrade them for functions from (subsets of) $\mathbb{R}^n$ to $\mathbb{R}^m$. In fact, the definitions we shall give are almost identical: the only thing that changes is that we use the appropriate norm for the spaces.

Continuity at a point

We will start with the simple definition:

Definition I.6. Let $A \subset \mathbb{R}^n$ and suppose $f : A \to \mathbb{R}^m$. We say that f is continuous at $p \in A$ if the following holds: given $\epsilon > 0$, there exists $\delta > 0$ such that for all $x \in A$ with $||x - p|| < \delta$ we have:

$$||f(x) - f(p)|| < \epsilon.$$

If f is continuous at p for all $p \in A$, we say f is continuous on A.

We can think of this as saying "f maps points in A close to p to points in $\mathbb{R}^m$ close to $f(p)$". Notice that in the definition above, the symbol $||\cdot||$ is playing two slightly different roles: as the norm on $\mathbb{R}^n$ and the norm on $\mathbb{R}^m$.

Example 5. The map:

$$\begin{array}{rccc} f & : & \mathbb{R}^n & \to & \mathbb{R} \\ & & x & \mapsto & ||x|| \end{array}$$

is continuous on $\mathbb{R}^n$. To see this, pick $p \in \mathbb{R}^n$. Suppose $||x - p|| < \delta$, then by the reverse triangle inequality (see Exercise 1.1) we have:

$$|f(x) - f(p)| = \big|\, ||x|| - ||p|| \,\big| \leq ||x - p|| < \delta.$$

Thus we can take $\delta = \epsilon$ and we have satisfied the condition for continuity.

Example 6. Suppose $\Lambda : \mathbb{R}^n \to \mathbb{R}^m$ is *linear*. Then Λ is continuous. To see this, we let $\{e_j\}_{j=1}^n$ be the canonical basis for $\mathbb{R}^n$ and we calculate:

$$\begin{aligned} ||\Lambda x - \Lambda p|| = ||\Lambda(x-p)|| &= \left|\left|\Lambda\left(\sum_{j=1}^n e_j (x-p)^j\right)\right|\right| \\ &= \left|\left|\sum_{j=1}^n \Lambda e_j (x-p)^j\right|\right| \leq \sum_{j=1}^n \left|\left|\Lambda e_j (x-p)^j\right|\right| \\ &\leq \sum_{j=1}^n \left|(x-p)^j\right| ||\Lambda e_j|| \end{aligned}$$

Setting $M = \max_{j=1,\ldots,n} ||\Lambda e_j||$, we have:

$$||\Lambda x - \Lambda p|| \leq Mn \max_{j=1,\ldots,n} \left|(x-p)^j\right| \leq Mn\,||x-p||\,,$$

using Exercise 1.1 f) i). Thus, if we take $\delta = \frac{\epsilon}{1+Mn}$, then for any x with $0 < ||x-p|| < \delta$, we have

$$||\Lambda x - \Lambda p|| < \frac{\epsilon}{1+Mn} Mn < \epsilon,$$

so Λ is continuous.

Example 7. The map

$$\begin{array}{cccc} f & : & \mathbb{R}^n & \to & \mathbb{R} \\ & & (x^1, \ldots, x^n)^t & \mapsto & x^1 \end{array}$$

is continuous on $\mathbb{R}^n$. To see this, pick $p \in \mathbb{R}^n$. Suppose $||x-p|| < \delta$, then by Exercise 1.1 f) i) we have:

$$|f(x) - f(p)| = \left|x^1 - p^1\right| \leq \max_{k=1,\ldots,n} \left|x^k - p^k\right| \leq ||x-p|| < \delta,$$

so we may take $\delta = \epsilon$ and we have satisfied the condition for continuity. Obviously the same argument shows that all of the coordinate maps (i.e. the map taking x to x^k) are continuous.

Theorem I.6. *Let $A \subset \mathbb{R}^n$, $B \subset \mathbb{R}^m$. Suppose $f : A \to B$ is continuous at p and $g : B \to \mathbb{R}^l$ is continuous at $f(p)$. Then $g \circ f : A \to \mathbb{R}^l$ is continuous at p.*

Proof. Let $\epsilon > 0$. Since g is continuous at $f(p)$, we know that there exists δ_1 such that for any $y \in B$ with $||y - f(p)|| < \delta_1$, we have $||g(y) - g(f(p))|| < \epsilon$. Similarly, Since f is continuous at p, we know that there exists δ such that for any $x \in A$ with $||x-p|| < \delta$, we have $||f(x) - f(p)|| < \delta_1$. Combining these two statements and taking $y = f(x)$, we deduce that if $x \in A$ with $||x-p|| < \delta$, we have $||g(f(x)) - g(f(p))|| < \epsilon$. □

It's sometimes useful to express the continuity of a function in a slightly different way, for which we need the following definition:

Definition I.7. Let $A \subset \mathbb{R}^n$ and suppose $f : A \to \mathbb{R}^m$. For $p \in A$ a limit point[3] of A, we say that $F \in \mathbb{R}^m$ is the limit of f as x tends to p if the following holds: given $\epsilon > 0$, there exists $\delta > 0$ such that for all $x \in A$ with $0 < ||x - p|| < \delta$ we have:

$$||f(x) - F|| < \epsilon.$$

In this case, we write

$$\lim_{x \to p} f(x) := F.$$

With this notion of a limit in hand, we can give the definition of continuity more compactly as:

> "f is continuous at p if $\lim_{x \to p} f(x) = f(p)$."

Theorem I.7. *Suppose $A \subset \mathbb{R}^n$ and suppose $f, g : A \to \mathbb{R}$ and*

$$\lim_{x \to p} f(x) = F, \qquad \lim_{x \to p} g(x) = G,$$

then:

i) $\lim_{x \to p} f(x) + g(x) = F + G,$

ii) $\lim_{x \to p} f(x)g(x) = FG,$

iii) If, furthermore $G \neq 0$, then:

$$\lim_{x \to p} \frac{f(x)}{g(x)} = \frac{F}{G}.$$

Proof. i) Fix $\epsilon > 0$. Since $\lim_{x \to p} f(x) = F$, we know that there exists δ_1 such that if $x \in A$ with $0 < ||x - p|| < \delta_1$ then:

$$|f(x) - F| < \frac{\epsilon}{2}.$$

Similarly, there exists δ_2 such that if $x \in A$ with $0 < ||x - p|| < \delta_2$ then:

$$|g(x) - G| < \frac{\epsilon}{2}.$$

Take $\delta = \min\{\delta_1, \delta_2\}$. Then if $x \in A$ with $0 < ||x - p|| < \delta$, we have by the triangle inequality:

$$|f(x) + g(x) - (F + G)| \leq |f(x) - F| + |g(x) - G| < \epsilon.$$

[3] p is a limit point of A if $B_r(p) \cap A \neq \{p\}$ for any $r > 0$. That is, each open ball around p contains at least one other point of A. Note that every point in A° is a limit point.

ii) Fix $\epsilon > 0$, and assume without loss of generality that $\epsilon < 1$. Since $\lim_{x\to p} f(x) = F$, we know that there exists δ_1 such that if $x \in A$ with $0 < ||x - p|| < \delta_1$ then:

$$|f(x) - F| < \frac{\epsilon}{3(1 + |G|)}.$$

Similarly, there exists δ_2 such that if $x \in A$ with $0 < ||x - p|| < \delta_2$ then:

$$|g(x) - G| < \frac{\epsilon}{3(1 + |F|)}.$$

We note the following identity:

$$f(x)g(x) - FG = (f(x) - F)(g(x) - G) + (f(x) - F)G + (g(x) - G)F$$

Now, take $\delta = \min\{\delta_1, \delta_2\}$. If $x \in A$ with $0 < ||x - p|| < \delta$, we have by the triangle inequality:

$$\begin{aligned} |f(x)g(x) - FG| &\leq |f(x) - F|\,|g(x) - G| + |G|\,|f(x) - F| + |F|\,|g(x) - G| \\ &< \frac{\epsilon^2}{9(1 + |F|)(1 + |G|)} + \frac{\epsilon\,|G|}{3(1 + |G|)} + \frac{\epsilon\,|F|}{3(1 + |F|)} < \epsilon. \end{aligned}$$

iii) Given the previous part, it suffices to show that if $\lim_{x\to p} g(x) = G$ with $G \neq 0$, then

$$\lim_{x\to p} \frac{1}{g(x)} = \frac{1}{G}.$$

Fix $\epsilon > 0$. Since $\lim_{x\to p} g(x) = G$, we know that there exist δ such that if $x \in A$ with $0 < ||x - p|| < \delta$ then:

$$|g(x) - G| < \frac{\epsilon\,|G|^2}{2}.$$

Without loss of generality, we can assume that ϵ is sufficiently small that $\epsilon\,|G| < 1$, which implies $|g(x) - G| < \frac{|G|}{2}$ and so $|g(x)| > \frac{|G|}{2}$.
If $x \in A$ with $0 < ||x - p|| < \delta$, we can estimate:

$$\left|\frac{1}{g(x)} - \frac{1}{G}\right| = |G - g(x)| \cdot \frac{1}{|G|} \cdot \frac{1}{|g(x)|} < \frac{\epsilon\,|G|^2}{2} \cdot \frac{1}{|G|} \cdot \frac{2}{|G|} = \epsilon.$$

□

Corollary I.8. *Suppose $A \subset \mathbb{R}^n$ and $f, g : A \to \mathbb{R}$ are continuous at $p \in A$. Then:*

i) $f + g$ is continuous at p.

ii) fg is continuous at p.

iii) If, furthermore $g(p) \neq 0$, then $\dfrac{f}{g}$ is continuous at p.

Exercise 2.2 (*)**.** Suppose $A \subset \mathbb{R}^n$ and $f : A \to \mathbb{R}^m$. Show that $\lim_{x\to p} f(x) = F$ if and only if for any sequence $(x_i)_{i=0}^{\infty}$ with $x_i \in A, x_i \neq p$ and $x_i \to p$ we have:

$$f(x_i) \to F, \qquad \text{as } i \to \infty.$$

Exercise 2.3. a) Show that the map

$$\begin{array}{cccc} f & : & \mathbb{R} & \to & \mathbb{R}^n \\ & & x & \mapsto & (x, 0, \ldots, 0)^t \end{array}$$

is continuous on $\mathbb{R}$.

b) Let $A \subset \mathbb{R}^n$ and suppose we are given a map:

$$\begin{array}{cccc} f & : & A & \to & \mathbb{R}^m \\ & & (x^1, \ldots, x^n)^t & \mapsto & \left(f^1(x^1, \ldots, x^n), \ldots, f^m(x^1, \ldots, x^n)\right)^t. \end{array}$$

Show that f is continuous at $p \in A$ if and only if each map $f^k : A \to \mathbb{R}$ is continuous at p, for $k = 1, \ldots, m$.

c) Show that a map $f : \mathbb{R}^n \to \mathbb{R}$ which is polynomial in the coordinates is continuous on $\mathbb{R}^n$.

Extreme value theorem

The final result we shall establish in this chapter is the extreme value theorem. This is a valuable result which allows us to conclude something about the range of a continuous function by knowing something about the domain. More precisely, if the domain is compact, then the range is bounded.

Theorem I.9. *Let $E \subset \mathbb{R}^n$ be compact, and suppose $f : E \to \mathbb{R}$ is continuous on E. Then there exist $p_+, p_- \in E$ such that:*

$$f(p_-) \leq f(x) \leq f(p_+)$$

for all $x \in E$. That is to say that f is bounded and achieves its bounds.

Proof. We first argue that $f(x)$ is bounded above for $x \in E$. Suppose not, then for each $i \in \mathbb{N}$ there exists $x_i \in E$ such that $f(x_i) > i$. Since E is compact, by Bolzano-Weierstrass the sequence $(x_i)_{i=0}^{\infty}$ has a convergent subsequence $(x_{i_j})_{j=0}^{\infty}$ such that $x_{i_j} \to x \in E$. By the continuity of f we have:

$$\lim_{j\to\infty} f(x_{i_j}) = f(x),$$

but on the other hand $f(x_{i_j}) > i_j \to \infty$ as $j \to \infty$, a contradiction. Therefore, there exists M such that $f(x) \leq M$ for all $x \in E$. In other words, the set

$$f(E) := \{f(x) : x \in E\}$$

is bounded above. By the completeness axiom, $f(E)$ has a least upper bound, so without loss of generality we will assume $M = \sup_{x\in E} f(x)$.

Now since M is the least upper bound, for $i \geq 1$, we have that $M - 2^{-i}$ is not an upper bound. Therefore there exists $y_i \in E$ with $M - 2^{-i} < f(y_i) \leq M$. Clearly, we have that $f(y_i) \to M$ as $i \to \infty$. Now the sequence $(y_i)_{i=0}^{\infty}$ with $y_i \in E$ has a convergent subsequence $(y_{i_j})_{j=0}^{\infty}$ with $y_{i_j} \to p_+$ for some $p_+ \in E$ by the Bolzano-Weierstrass theorem. Since $f(y_{i_j}) \to M$, by the continuity of f we conclude that $f(p_+) = M$, Thus we have established that there exists $p_+ \in E$ such that $f(x) \leq f(p_+)$ for all $x \in E$. An identical argument deals with the lower bound. □

Exercise 2.4. Suppose that $E \subset \mathbb{R}^n$ is compact and $f : E \to \mathbb{R}^m$ is continuous on E.

a) Show that $f(E) := \{f(x) : x \in E\}$ is bounded.

b) Show that $f(E)$ is closed, hence compact.

c) Is it true that $f(E)$ is closed if we just assume E is closed (not necessarily compact)?

Exercise 2.5 (*). a) Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous on $\mathbb{R}^n$, and suppose $U \subset \mathbb{R}^m$ is open. Show that:

$$f^{-1}(U) := \{x \in \mathbb{R}^n : f(x) \in U\}$$

is open.

b) Suppose that $f : \mathbb{R}^n \to \mathbb{R}^m$ has the property that $f^{-1}(U) \subset \mathbb{R}^n$ is open for every open $U \subset \mathbb{R}^m$. Show that f is continuous on $\mathbb{R}^n$.

Chapter 1

Uniform continuity and convergence

So far, we have considered sequences of points in $\mathbb{R}^n$ and their limits. For many applications, we want to consider sequences of other objects, and to talk about their limits. For example it is often useful to consider sequences of *functions*. We might imagine that if we have a suitable limit of continuous (or differentiable) functions, then the limit should be a continuous (resp. differentiable) function. In certain circumstances this is true, and in this chapter we will develop some of the machinery required to establish such a result.

1.1 Uniform continuity

1.1.1 Definition and examples

In the previous section (and in last year's courses) we defined what it means for a function to be continuous at a point p. We said that the function is continuous on some set $A \subset \mathbb{R}^n$ if it is continuous at all $p \in A$. This is a very useful concept, however in certain circumstances we need a slightly stronger property, called *uniform continuity*.

Definition 1.1. Suppose $A \subset \mathbb{R}^n$ and $f : A \to \mathbb{R}^m$. We say that f is *uniformly continuous* on A if the following holds: given $\epsilon > 0$, there exists $\delta > 0$ such that for all $x, y \in A$ with $||x - y|| < \delta$ we have:

$$||f(x) - f(y)|| < \epsilon.$$

For comparison, let's recall the definition of continuity at a point:

Definition I.6. Let $A \subset \mathbb{R}^n$ and suppose $f : A \to \mathbb{R}^m$. We say that f is continuous at $p \in A$ if the following holds: given $\epsilon > 0$, there exists $\delta > 0$ such that for all $x \in A$ with $||x - p|| < \delta$ we have:

$$||f(x) - f(p)|| < \epsilon.$$

We see that the crucial difference between the definition for a function continuous on A and a function *uniformly* continuous on A is that in the first case δ is allowed to depend on the point p at which we're testing continuity. In the second case, one δ has to work at every point in A.

Lemma 1.1. *Let $A \subset \mathbb{R}^n$ and suppose $f : A \to \mathbb{R}^m$. If f is uniformly continuous on A, then f is continuous on A.*

Proof. Fix $p \in A$ and $\epsilon > 0$. By the uniform continuity of f, there exists $\delta > 0$ such that if $||x - p|| < \delta$, then:

$$||f(x) - f(p)|| < \epsilon.$$ □

Example 1.1. Suppose $\alpha \in \mathbb{R}$ and $z \in \mathbb{R}^n$. The function:

$$\begin{array}{rccc} f & : & \mathbb{R}^n & \to & \mathbb{R}^n \\ & & x & \mapsto & \alpha x + z \end{array}$$

is uniformly continuous. To see this, suppose that $||x - y|| < \delta$. Then we have:

$$||f(x) - f(y)|| = ||\alpha(x - y)|| = |\alpha| \, ||x - y|| \leq |\alpha| \, \delta.$$

Taking $\delta = \frac{\epsilon}{1+|\alpha|}$, we have the result. Note that our choice of δ does not depend on x, y.

Example 1.2. The function:

$$\begin{array}{rccc} f & : & (0, \infty) & \to & \mathbb{R} \\ & & x & \mapsto & \frac{1}{x} \end{array}$$

is continuous on $(0, \infty)$, but *not* uniformly continuous. The continuity of f at each $p > 0$ follows from Corollary I.8 *iii*). To see that f is not uniformly continuous, it is enough to show that for any $\delta > 0$ there exist $x, y \in (0, \infty)$ with $||x - y|| < \delta$ and $||f(x) - f(y)|| > 1$. Consider for some $0 < \kappa < 1$:

$$x = \frac{\kappa}{2}\delta, \quad y = \kappa\delta, \qquad \Longrightarrow \qquad ||x - y|| = \frac{\kappa}{2}\delta < \delta.$$

On the other hand,

$$||f(x) - f(y)|| = \frac{1}{\kappa\delta},$$

so if $\kappa = \frac{1}{1+\delta}$ we have $||f(x) - f(y)|| > 1$.

The key idea of uniform continuity is that points which are close together in the domain end up close together after they've been mapped to the range by f, and this should happen *independently of where in the range the points are.* In the second example above, we see that two points which are close together and near 0 are mapped much further apart than two points close together and near 1.

Exercise 2.6. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is differentiable (hence continuous) everywhere on $\mathbb{R}$ and the derivative satisfies:

$$\left| f'(x) \right| \leq M, \qquad \text{for all } x \in \mathbb{R}.$$

a) Show that for $x, y \in \mathbb{R}$ we have the estimate:

$$|f(x) - f(y)| \leq M \, |x - y| \, . \tag{1.1}$$

b) Deduce that f is uniformly continuous on $\mathbb{R}$.

The condition (1.1), which can be imposed whether or not f is differentiable, is known as the Lipschitz condition.

Exercise 2.7. Which of the following functions is i) continuous, ii) uniformly continuous on its domain?

a) $$\begin{array}{rcccc} f & : & \mathbb{R} & \to & \mathbb{R} \\ & & x & \mapsto & e^x \end{array}$$

b) $$\begin{array}{rcccc} f & : & (0,1) & \to & \mathbb{R} \\ & & x & \mapsto & e^x \end{array}$$

c) $$\begin{array}{rcccc} f & : & \mathbb{R} & \to & \mathbb{R} \\ & & x & \mapsto & \sin x \end{array}$$

d) $$\begin{array}{rcccc} f & : & [0,\infty) & \to & \mathbb{R} \\ & & x & \mapsto & \sqrt{x} \end{array}$$

*f) $$\begin{array}{rcccc} f & : & \mathbb{R}^n \setminus \{0\} & \to & \mathbb{R}^n \\ & & x & \mapsto & \frac{x}{||x||} \end{array}$$

Exercise 2.8. Let $A \subset \mathbb{R}^n$ and suppose that $f, g : A \to \mathbb{R}$.

a) Show that if f, g are uniformly continuous and bounded, then so is fg.

b) Show that if f, g are uniformly continuous but not necessarily bounded, then fg need not be uniformly continuous.

1.1.2 Continuous functions on a compact set

The following result is very useful, as it allows us to deduce that a continuous function is uniformly continuous, just based on knowledge about the domain of the function:

Theorem 1.2. *Suppose that $E \subset \mathbb{R}^n$ is compact and $f : E \to \mathbb{R}^m$ is continuous on E. Then f is* uniformly *continuous on E.*

Proof. Suppose f is not uniformly continuous. Then[1] there exists an $\epsilon > 0$ such that for any $i \in \mathbb{N}$ we can find x_i, y_i with:

$$||x_i - y_i|| < \frac{1}{2^i}, \qquad ||f(x_i) - f(y_i)|| \geq \epsilon.$$

Now $(x_i)_{i=0}^{\infty}$ is a sequence in a compact set, so by Bolzano–Weierstrass it has a convergent subsequence $(x_{i_j})_{j=0}^{\infty}$ with $x_{i_j} \to x \in E$. Next look at $(y_{i_j})_{j=0}^{\infty}$. Again this is a sequence in a compact set, so it has a convergent subsequence $(y_{i_{j_l}})_{l=0}^{\infty}$ with $y_{i_{j_l}} \to y$. Relabelling

[1] Check you understand why this is!

these sequences $u_l = x_{i_{j_l}}$, $w_l = y_{i_{j_l}}$, we have $(u_l)_{l=0}^{\infty}$ with $u_l \to x$ and $(w_l)_{l=0}^{\infty}$ with $w_l \to y$. Furthermore, since $i_{j_l} \geq j_l \geq l$, we have:

$$||u_l - w_l|| < \frac{1}{2^l}, \qquad ||f(u_l) - f(w_l)|| \geq \epsilon.$$

Sending $l \to \infty$, we conclude from the first inequality that $||x - y|| = 0$, so $x = y$. From the second inequality, using the continuity of f, we have $||f(x) - f(y)|| \geq \epsilon$, which is a contradiction. □

1.2 Uniform convergence

Now we want to think a bit about convergence of sequences of functions. When we say that a sequence $(x_i)_{i=0}^{\infty}$ with $x_i \in \mathbb{R}^n$ converges to $x \in \mathbb{R}^n$, crudely we're saying:

> "By going far enough along the sequence, we can ensure we're as close as we like to x".

For points in $\mathbb{R}^n$, it's pretty unambiguous what we mean by "close enough": we want the points in the sequence to eventually be a small distance away from x. For functions, it turns out that there are different plausible notions of "close enough". Which one we use depends to an extent on the problem that we wish to address.

1.2.1 Definition and examples

We will start with the most naïve definition possible for the convergence of a sequence of functions.

Definition 1.2. Let $A \subset \mathbb{R}^n$. Suppose $(f_i)_{i=0}^{\infty}$ is a sequence of functions $f_i : A \to \mathbb{R}^m$. We say that $f_i \to f$ *pointwise* if the following holds: given $\epsilon > 0$, for each $x \in A$ there exists $N \in \mathbb{N}$ such that for all $i \geq N$ we have:

$$||f_i(x) - f(x)|| \leq \epsilon.$$

Equivalently, we can write this as:

$$f_i(x) \to f(x) \qquad \text{for all } x \in A.$$

This definition has the advantage of being the simplest thing to write down, and it is usually relatively simple to check whether the condition holds.

Example 1.3. Consider the sequence of functions $(f_i)_{i=0}^{\infty}$ given by:

$$\begin{array}{rcccl} f_i & : & \mathbb{R} & \to & \mathbb{R} \\ & & x & \mapsto & x + 2^{-i}x^2. \end{array}$$

This converges pointwise to $f(x) = x$, since for any $x \in \mathbb{R}$ we have:

$$x + 2^{-i}x^2 \to x.$$

Example 1.4. For $i \in \mathbb{N}$, let $f_i : [-1,1] \to \mathbb{R}$ be given by:

$$f_i(x) = \begin{cases} -1 & x \leq -2^{-i}, \\ 2^i x & -2^{-i} < x < 2^{-i} \\ 1 & x \geq 2^{-i}, \end{cases}$$

Then $f_i \to f$ pointwise, where $f : [-1,1] \to \mathbb{R}$ is given by:

$$f(x) = \begin{cases} -1 & x < 0, \\ 0 & x = 0 \\ 1 & x > 0, \end{cases}$$

To see this, first suppose $x < 0$. Then for all sufficiently large i we have $x \leq -2^{-i}$, so $f_i(x) = -1$ and $f_i(x) \to f(x)$. Similarly, suppose $x > 0$. Then for all sufficiently large i we have $x \geq 2^{-i}$, so $f_i(x) = 1$ and $f_i(x) \to f(x)$. Finally, if $x = 0$ then $f_i(x) = 0$ for all i and so $f_i(x) \to f(x)$.

Example 1.4 shows that pointwise convergence has some drawbacks. The functions f_i are all continuous (indeed uniformly continuous), however the limiting function f is discontinuous. We often want to argue that the limit of a sequence of continuous functions is continuous and it's clear that in order to do this, we need something stronger than pointwise continuity. This is the motivation for the idea of *uniform continuity.*

Definition 1.3. Let $A \subset \mathbb{R}^n$. Suppose $(f_i)_{i=0}^{\infty}$ is a sequence of functions $f_i : A \to \mathbb{R}^m$. We say that $f_i \to f$ *uniformly* if the following holds: given $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $i \geq N$ we have:

$$||f_i(x) - f(x)|| \leq \epsilon, \qquad \text{for all } x \in A.$$

Notice that there is very little difference in the way that uniform convergence and pointwise convergence are defined! The key difference is in where in the statement we have "for all $x \in A$". Ultimately, the issue is that in the case of pointwise convergence, we are allowed to choose N depending on x. In the case of uniform convergence we are only allowed to choose one N and it has to work for all $x \in A$.

Notice that uniform convergence implies pointwise convergence.

Example 1.5. The sequence:

$$\begin{array}{rcrcl} f_i & : & [-1,1] & \to & \mathbb{R} \\ & & x & \mapsto & x + 2^{-i}x^2. \end{array}$$

converges uniformly to $f(x) = x$. To see this, note that for any $x \in [-1,1]$

$$|f_i(x) - f(x)| = \left|2^{-i}x\right| \leq 2^{-i}.$$

For any $\epsilon > 0$ we can find N such that $2^{-i} < \epsilon$ for all $i \geq N$, so we indeed have that the convergence is uniform.

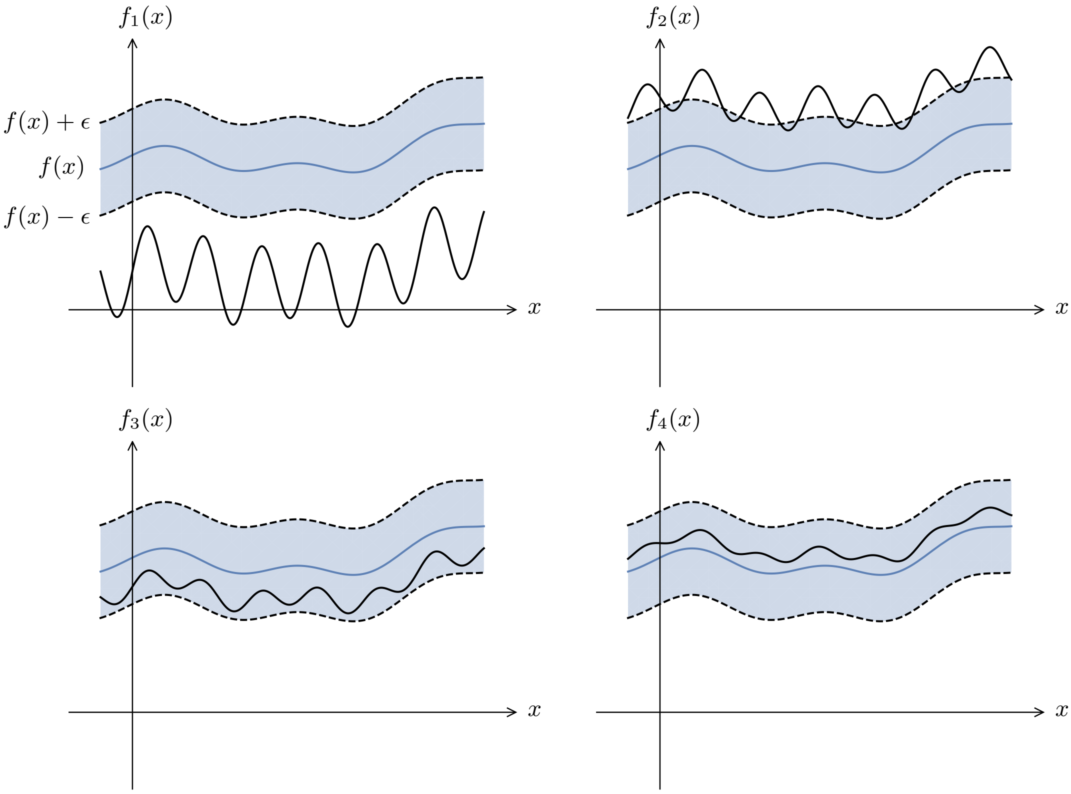

Figure 1.1 A sequence of functions converging uniformly

Example 1.6. The sequence in Example 1.4 does *not* converge uniformly. To see this, note that

$$f_i\left(2^{-(i+1)}\right) - f\left(2^{-(i+1)}\right) = \frac{1}{2} - 1 = -\frac{1}{2},$$

so that for $0 < \epsilon < \frac{1}{2}$ there can exist no $N \in \mathbb{N}$ such that for all $i \geq N$ and $x \in [-1, 1]$we have:

$$||f_i(x) - f(x)|| < \epsilon.$$

The key feature of uniform convergence is the idea that $f_i(x) \to f(x)$ at (roughly) the same rate everywhere on A. We show this for a few terms of a sequence of functions converging uniformly in Figure 1.1

Exercise 3.1. For $i \in \mathbb{N}$, let $f_i : [0, 1] \to \mathbb{R}$ be given by:

$$f_i(x) = \begin{cases} 2^i x & 0 \leq x < 2^{-i}, \\ 2 - 2^i x & 2^{-i} \leq x < 2^{-(i-1)} \\ 0 & x \geq 2^{-(i-1)}, \end{cases}$$

a) Sketch f_i.

b) Show that $f_i \to 0$ pointwise.

c) Is it true that:

$$\lim_{i\to\infty} \sup_{x\in[0,1]} f_i(x) = \sup_{x\in[0,1]} \lim_{i\to\infty} f_i(x)?$$

Exercise 3.2. State whether the sequence $(f_i)_{i=0}^{\infty}$ of functions converges pointwise or uniformly, when:

a) $\begin{array}{cccc} f_i & : & \mathbb{R} & \to & \mathbb{R} \\ & & x & \mapsto & e^x + \frac{1}{i+1} \end{array}$

b) $\begin{array}{cccc} f_i & : & \mathbb{R} & \to & \mathbb{R} \\ & & x & \mapsto & x + 2^{-i}x^2 \end{array}$

c) $f_i : [0,1] \to \mathbb{R}$, where:

$$f_i(x) = \begin{cases} 2^i x & 0 \le x < 2^{-i}, \\ 2 - 2^i x & 2^{-i} \le x < 2^{-(i-1)} \\ 0 & x \ge 2^{-(i-1)}, \end{cases}$$

Exercise 3.3. Suppose $A \subset \mathbb{R}^n$ and let $(f_i)_{i=0}^{\infty}$, where $f_i : A \to \mathbb{R}^m$ be a sequence of functions. Classify whether the following statements are equivalent to

i) $f_i \to f$ pointwise,

ii) $f_i \to f$ uniformly,

iii) neither.

a) Given $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $i \ge N$:

$$\sup_{x\in A} ||f_i(x) - f(x)|| < \epsilon.$$

b) For all $x \in A$, for all $i \in \mathbb{N}$ there exists $\epsilon > 0$ such that:

$$||f_i(x) - f(x)|| < \epsilon.$$

c) There exists $N \in \mathbb{N}$ such that for all $x \in A$, $i \ge N$, $\epsilon > 0$:

$$||f_i(x) - f(x)|| < \epsilon.$$

*d) $\forall \epsilon > 0 \, \forall x \in A \, \exists N \in \mathbb{N} \, (i \ge N \implies ||f_i(x) - f(x)|| < \epsilon)$

*e) $\forall \epsilon > 0 \, \exists N \in \mathbb{N} \, \forall x \in A \, (i \ge N \implies ||f_i(x) - f(x)|| < \epsilon)$

*f) $\forall \epsilon > 0 \, \exists N \in \mathbb{N} \, \forall x \in A \, (i \ge N \iff ||f_i(x) - f(x)|| < \epsilon)$

Exercise 3.4. Suppose $A \subset \mathbb{R}^n$ and let $(f_i)_{i=0}^{\infty}$, where $f_i : A \to \mathbb{R}^m$ be a sequence of functions. Suppose that $f_i \to f$ uniformly on A. Show that if $B \subset A$, then $f_i \to f$ uniformly on B.

1.2.2 Uniform limits of uniformly continuous functions

One of the main motivations for introducing the idea of uniform continuity and uniform convergence is the result we shall prove in this section. It tells us that if we have a sequence of functions which are uniformly continuous and they converge uniformly, then the function they converge to must also be uniformly continuous. We shall see that this is a very useful result in many situations.

Theorem 1.3. *Let $A \subset \mathbb{R}^n$ and suppose $(f_i)_{i=0}^{\infty}$ is a sequence of functions $f_i : A \to \mathbb{R}^m$ such that:*

i) $f_i \to f$ uniformly on A.

ii) Each f_i is uniformly continuous on A,

Then $f : A \to \mathbb{R}^m$ is uniformly continuous on A.

Proof. Fix $\epsilon > 0$. Firstly, by property i) above, we know that there exists $N \in \mathbb{N}$ such that for all $i \geq N$ and all $x \in A$ we have:

$$||f_i(x) - f(x)|| < \frac{\epsilon}{3}.$$

Fix $k \geq N$. By property ii) above, we know that there exists δ such that for all $x, y \in A$ with $||x - y|| < \delta$, we have:

$$||f_k(x) - f_k(y)|| < \frac{\epsilon}{3}.$$

Now, suppose $x, y \in A$ with $||x - y|| < \delta$. We estimate:

$$\begin{aligned} ||f(x) - f(y)|| &= ||f(x) - f_k(x) + f_k(x) - f_k(y) + f_k(y) - f(y)|| \\ &\leq ||f(x) - f_k(x)|| + ||f_k(x) - f_k(y)|| + ||f_k(y) - f(y)|| \\ &\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon. \end{aligned}$$

This is precisely the statement that f is uniformly continuous on A. □

Note that this theorem gives us an alternative proof that the sequence in Example 1.4 does not converge uniformly. The functions f_i are all continuous functions on a compact set (hence uniformly continuous by Theorem 1.2). If $f_i \to f$ uniformly, then f is continuous, however we know that the limit f is not continuous and thus the convergence cannot be uniform.

We can in fact relax the requirement of uniform continuity on the sequence f_i.

Corollary 1.4. *Let $U \subset \mathbb{R}^n$ be open and suppose $(f_i)_{i=0}^{\infty}$ is a sequence of functions $f_i : U \to \mathbb{R}^m$ such that:*

i) $f_i \to f$ uniformly on U.

ii) Each f_i is continuous on U,

Then $f : U \to \mathbb{R}^m$ is continuous on U.

Proof. Fix $p \in U$. Since U is open, there exists r such that $B_{2r}(p) \subset U$, so in particular $\overline{B_r(p)} \subset U$. By Theorem 1.2, each f_i restricted to $\overline{B_r(p)}$ is uniformly continuous, and moreover we have that $f_i \to f$ uniformly on $\overline{B_r(p)}$. By Theorem 1.3, we have that f is uniformly continuous on $\overline{B_r(p)}$, hence continuous at p. Since p was arbitrary, we're done. □

1.2.3 (*) The space $C^0([a,b])$

Let us restrict our attention for a moment to the case of functions defined on a closed interval in $\mathbb{R}$. Let us introduce some new notation. We will say that $C^0([a,b])$ is the set of all continuous functions mapping $[a,b] \to \mathbb{R}$. In other words:

$$f \in C^0([a,b]) \iff f : [a,b] \to \mathbb{R}, \text{ and } f \text{ is continuous on } [a,b].$$

Note that by Theorem 1.2, we know that $f \in C^0([a,b])$ implies that f is uniformly continuous. We also know by Theorem I.9 that f achieves its maximum and minimum on $[a,b]$.

We can define a function on $C^0([a,b])$ as follows:

$$\begin{array}{rccl} ||\cdot||_{C^0} & : & C^0 & \to & [0,\infty) \\ & & f & \mapsto & ||f||_{C^0} := \sup_{x\in[a,b]} |f(x)| \,. \end{array}$$

In other words, $||\cdot||_{C^0}$ assigns to each element f of $C^0([a,b])$ the largest value achieved by $|f(x)|$ on the interval. We call $||\cdot||_{C^0}$ the norm on $C^0([a,b])$.

Lemma 1.5. *The the norm on $C^0([a,b])$ has the following properties:*

i) For all $f \in C^0([a,b])$, we have $||f||_{C^0} \geq 0$, with $||f||_{C^0} = 0$ if and only if $f \equiv 0$.

ii) For all $f \in C^0([a,b])$, if $a \in \mathbb{R}$, then $||af||_{C^0} = |a|\,||f||_{C^0}$

iii) The norm satisfies the triangle inequality:

$$||f+g||_{C^0} \leq ||f||_{C^0} + ||g||_{C^0}\,, \qquad \forall\ f,g, \in C^0. \tag{1.2}$$

Proof. i) It is clear that since $||f||_{C^0}$ is the supremum of a set of non-negative numbers, it must be non-negative. If $||f||_{C^0} = 0$, then we have for all $x \in [a,b]$:

$$0 \leq |f(x)| \leq \sup_{y\in[a,b]} |f(y)| = 0$$

so that $f(x) = 0$ for all $x \in [a,b]$.

ii) If $a \in \mathbb{R}$, then:

$$||af||_{C^0} = \sup_{x\in[a,b]} |af(x)| = |a| \sup_{x\in[a,b]} |f(x)| = |a|\,||f||_{C^0}\,.$$

iii) We estimate:

$$||f+g||_{C^0} = \sup_{x\in[a,b]} |f(x)+g(x)| \le \sup_{x\in[a,b]} (|f(x)|+|g(x)|)$$
$$\le \sup_{x\in[a,b]} |f(x)| + \sup_{y\in[a,b]} |g(y)| = ||f||_{C^0} + ||g||_{C^0}.$$

Here we have used the triangle inequality for $|\cdot|$, as well as the fact that $\sup(a+b) \le \sup a + \sup b$.

□

With the norm on $C^0([a,b])$, we can give the definition of uniform convergence very succinctly. A sequence $(f_i)_{i=0}^{\infty}$ converges uniformly to f if: given $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $i \ge N$ we have:

$$||f_i - f||_{C^0} < \epsilon.$$

If we compare this with the definition of convergence in $\mathbb{R}^n$, Definition I.1, we see there are some very close parallels. This is because $\mathbb{R}^n$ with its standard norm $||\cdot||$ and $C^0([a,b])$ with the norm $||\cdot||_{C^0}$ are both examples of a more general structure, a *normed space.*

Recall from Exercise 2.1 that a sequence $(x_i)_{i=0}^{\infty}$ with $x_i \in \mathbb{R}^n$ converges if and only if it is a Cauchy sequence. A Cauchy sequence satisfies the following condition: given $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $i, j \ge N$ we have:

$$||x_i - x_j|| < \epsilon.$$

In other words a Cauchy sequence is one for which all of the terms are eventually close together. This fact, that Cauchy sequences converge, is a fundamental property of $\mathbb{R}^n$ called *completeness.* In fact, one can replace the least upper bound axiom of $\mathbb{R}$ with the assumption that Cauchy sequences converge: the two statements are equivalent.

We make the obvious definition of a Cauchy sequence in $C^0([a,b])$.

Definition 1.4. A sequence $(f_i)_{i=0}^{\infty}$ with $f_i \in C^0([a,b])$ is *Cauchy* if the following holds: given $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $i, j \ge N$ we have:

$$||f_i - f_j||_{C^0} < \epsilon.$$

Remarkably, $C^0([a,b])$ also has the completeness property enjoyed by $\mathbb{R}^n$:

Theorem 1.6. *A sequence $(f_i)_{i=0}^{\infty}$ with $f_i \in C^0([a,b])$ converges uniformly to some $f \in C^0([a,b])$ if and only if it is Cauchy.*

Proof. First, suppose $(f_i)_{i=0}^{\infty}$ converges uniformly to f. Let $\epsilon > 0$. By the convergence of f_i, we know that there exists $N \in \mathbb{N}$ such that for all $i \ge N$ we have:

$$||f_i - f||_{C^0} < \frac{\epsilon}{2}.$$

Now suppose $i, j \ge N$. Then:

$$||f_i - f_j||_{C^0} = ||f_i - f + f - f_j||_{C^0} \le ||f_i - f||_{C^0} + ||f - f_j||_{C^0} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Next, suppose that $(f_i)_{i=0}^{\infty}$ is Cauchy. This means that given $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for any $y \in [a,b]$ and $i, j \geq N$:

$$|f_i(y) - f_j(y)| \leq \sup_{x\in[a,b]} |f_i(x) - f_j(x)| = ||f_i - f_j||_{C^0} < \epsilon. \tag{1.3}$$

Thus, for any $y \in [a,b]$, the sequence $(f_i(y))_{i=0}^{\infty}$ is a Cauchy sequence in $\mathbb{R}$, so by the results of Exercise 2.1, $f_i(y)$ converges to some $f(y) \in \mathbb{R}$. Thus we have established that f_i converges to some f pointwise.

To complete the proof, it is enough to show that $f_i \to f$ uniformly. Repeating the argument that led to (1.3), we know that for any $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for any $y \in [a,b]$ and $i, j \geq N$:

$$|f_i(y) - f_j(y)| < \frac{\epsilon}{2}.$$

Taking the limit $j \to \infty$, we conclude that for any $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for $i \geq N$ we have:

$$|f_i(y) - f(y)| \leq \frac{\epsilon}{2} < \epsilon,$$

for any $y \in [a,b]$. This is precisely the statement of uniform convergence. Since we know that a uniformly convergent sequence of uniformly continuous functions is uniformly continuous, we have that $f \in C^0([a,b])$. □

A normed space with the completeness property is called a *Banach space*. These are very important structures, and you will learn much more about them if you take the course M3P7 Functional Analysis.

Exercise 3.5 (*)**.** Let $E \subset \mathbb{R}^n$ be compact. We define $C^0(E, \mathbb{R}^m)$ to be the set of all continuous functions $f : E \to \mathbb{R}^m$, and equip $C^0(E, \mathbb{R}^m)$ with the norm:

$$\begin{array}{rccl} ||\cdot||_{C^0} & : & C^0 & \to & [0,\infty) \\ & & f & \mapsto & ||f||_{C^0} := \sup_{x\in E} ||f(x)|| \,. \end{array}$$

Show that $C^0(E, \mathbb{R}^m)$ is complete.

Exercise 3.6 (*)**.** Suppose that $f : (a,b) \to \mathbb{R}$ is uniformly continuous, and suppose $(x_j)_{j=0}^{\infty}$ and $(y_j)_{j=0}^{\infty}$ are two sequences of points $x_j, y_j \in (a,b)$ with $x_j \to a$ and $y_j \to b$.

a) Show that the sequences $(f(x_j))_{j=0}^{\infty}$ and $(f(y_j))_{j=0}^{\infty}$ are Cauchy.

b) Setting $f_L := \lim_{j\to\infty} f(x_j)$, $f_R := \lim_{j\to\infty} f(y_j)$, show that the function $\tilde{f} : [a,b] \to \mathbb{R}$ given by:

$$\tilde{f}(x) = \begin{cases} f_L & x = a, \\ f(x) & a < x < b \\ f_R & x = b, \end{cases}$$

is continuous on $[a,b]$.

c) Deduce that $f : (a,b) \to \mathbb{R}$ is uniformly continuous if and only if there exists a continuous function $\tilde{f} : [a,b] \to \mathbb{R}$ such that $f(x) = \tilde{f}(x)$ for all $x \in (a,b)$. We say that $\tilde{f}$ is an extension of f.

1.2.4 (*) Weierstrass M−test

In your first year analysis courses, you looked at series. Given a sequence $(a_i)_{i=0}^{\infty}$ with $a_i \in \mathbb{R}$, we can form the partial sum:

$$S_j = \sum_{i=0}^{j} a_i.$$

If the sequence of partial sums $(S_j)_{j=0}^{\infty}$ converges, then we say that the series converges and we write:

$$\sum_{i=0}^{\infty} a_i := \lim_{j\to\infty} S_j.$$

The series is said to be *absolutely convergent* if the sum:

$$\sum_{i=0}^{\infty} |a_i|$$

converges. You may have studied various criteria for a series to converge (ratio test, root test, comparison test ...).

Often, one wishes to consider an infinite sum of *functions* rather than numbers. Let $A \subset \mathbb{R}^n$. Suppose we're given a sequence $(f_i)_{i=0}^{\infty}$ of functions $f_i : A \to \mathbb{R}^m$. If the sequence of partial sums $(S_j)_{j=0}^{\infty}$ with $S_j : A \to \mathbb{R}^m$ given by:

$$S_j(x) = \sum_{i=0}^{j} f_i(x)$$

converges (pointwise or uniformly) to some function $S : A \to \mathbb{R}^m$, then we write:

$$\sum_{i=0}^{\infty} f_i(x) := S(x),$$

with the sum converging pointwise or uniformly as appropriate.

An important and useful result for confirming that a series converges uniformly is the Weierstrass M−test.

Theorem 1.7 (Weierstrass M−test)**.** *Let $A \subset \mathbb{R}^n$. Suppose $(f_i)_{i=0}^{\infty}$ is a sequence of functions $f_i : A \to \mathbb{R}^m$. Suppose further that there exists a sequence $(M_i)_{i=0}^{\infty}$ with $M_i \in \mathbb{R}$ such that:*

i) $\sum_{i=0}^{\infty} M_i$ *converges.*

ii) $||f_i(x)|| \leq M_i$ *for all* $x \in A$, $i \in \mathbb{N}$.

Then $\sum_{i=0}^{\infty} f_i(x)$ *converges uniformly on* A.

Proof. Since $\sum_i M_i$ converges, the partial sums form a Cauchy sequence by Exercise 2.1. As a result, given $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that:

$$\left| \sum_{i=0}^{l} M_i - \sum_{i=0}^{m} M_i \right| = \sum_{i=m+1}^{l} M_i < \epsilon$$

for $N \leq m \leq l$. Letting $S_j(x) = \sum_{i=0}^{j} f_i(x)$, we can estimate for any $x \in A$:

$$||S_l(x) - S_m(x)|| = \left|\left| \sum_{i=m+1}^{l} f_i(x) \right|\right| \leq \sum_{i=m+1}^{l} ||f_i(x)|| \leq \sum_{i=m+1}^{l} M_i < \epsilon.$$

Thus $(S_j(x))_{j=0}^{\infty}$ is a Cauchy sequence in $\mathbb{R}^m$. Again invoking Exercise 2.1, we have that $S_j(x) \to S(x)$ for some $S(x) \in \mathbb{R}^m$. Repeating the arguments above, we have that given $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for $N \leq m \leq l$ and for any $x \in A$ we have:

$$||S_l(x) - S_m(x)|| < \frac{\epsilon}{2}.$$

Sending $l \to \infty$, we conclude:

$$||S(x) - S_m(x)|| \leq \frac{\epsilon}{2} < \epsilon,$$

so that $S_j \to S$ uniformly. □

As a useful example, we can show that functions defined by power series are continuous within their radius of convergence.

Corollary 1.8. *Suppose that $(a_i)_{i=0}^{\infty}$ is a sequence of real numbers $a_i \in \mathbb{R}$, and suppose there exists $r > 0$ such that the sum*

$$\sum_{i=0}^{\infty} |a_i| \, r^i$$

converges. Then the power series:

$$f(x) = \sum_{i=0}^{\infty} a_i x^i$$

converges uniformly to a continuous function on the interval $[-r, r]$.

Proof. We apply the Weierstrass $M-$test with $f_i(x) = a_i x^i$, and $M_i = |a_i| \, r^i$. □

Exercise 3.7 (*)**.** A Fourier series is a series of the form:

$$f(x) = a_0 + \sum_{k=1}^{\infty} \left(a_k \cos kx + b_k \sin kx \right),$$

where $a_k, b_k \in \mathbb{R}$. Show that if $\sum_{k=1}^{\infty} (|a_k| + |b_k|)$ converges, then the Fourier series converges uniformly on $\mathbb{R}$ to a uniformly continuous function $f : \mathbb{R} \to \mathbb{R}$ satisfying:

$$f(x + 2\pi) = f(x), \quad \text{for all } x.$$

Exercise 3.8 (*)**.** Let $h : \mathbb{R} \to \mathbb{R}$ be the periodic function satisfying:

$$h(x) = |x| \qquad x \in \left[-\frac{1}{2}, \frac{1}{2}\right],$$

and $h(x+1) = h(x)$.

a) Sketch $h(x)$.

b) Show that $h(x)$ is uniformly continuous on $\mathbb{R}$ and differentiable at each point $x \in \mathbb{R}$ such that $2x \notin \mathbb{Z}$.

Define the Takagi function[2] $T : \mathbb{R} \to \mathbb{R}$ by:

$$T(x) = \sum_{i=0}^{\infty} \frac{1}{2^i} h\left(2^i x\right) \tag{1.4}$$

c) Show that the sum in (1.4) converges uniformly on $\mathbb{R}$ to a uniformly continuous function $T : \mathbb{R} \to \mathbb{R}$.

d) Let $x \in \mathbb{R}$. Show that there exist unique u_n, v_n such that:

   i) $2^n v_n, 2^n u_n \in \mathbb{Z}$,

   ii) $v_n - u_n = 2^{-n}$ and,

   iii) $u_n \leq x < v_n$.

e) Let $h_i(x) = \frac{1}{2^i} h\left(2^i x\right)$. Show that

   i) $h_i(u_n) = h_i(v_n) = 0$ for $i \geq n$.

   ii) $\dfrac{h_i(v_n) - h_i(u_n)}{v_n - u_n} = \pm 1$ for $i < n$.

f) Deduce that:

$$\frac{T(v_n) - T(u_n)}{v_n - u_n} = \sum_{i=0}^{n-1} \frac{h_i(v_n) - h_i(u_n)}{v_n - u_n},$$

   and conclude that the sum on the right does *not* converge as $n \to \infty$.

g) Show that T is not differentiable at any x.

1.3 (*) Arzelà-Acscoli theorem

The final result we shall prove before we move on to consider integration is the Arzelà-Acscoli theorem. This can be thought of loosely as a Bolzano–Weierstrass theorem, but for sequences of functions rather than sequences of points in $\mathbb{R}^n$. It turns out that this theorem is surprisingly powerful, and is important in several branches of analysis.

Before we state the theorem, we first need a definition.

[2]Sometimes called the blancmange function.

Definition 1.5. Let $F = (f_i)_{i=0}^{\infty}$ be a sequence of continuous functions $f_i : [a,b] \to \mathbb{R}$. We say that F is *uniformly equicontinuous* if the following holds: given $\epsilon > 0$, there exists $\delta > 0$ such that:

$$|f_i(x) - f_i(y)| < \epsilon,$$

for all $i \in \mathbb{N}$ and all $x, y \in [a,b]$ with $|x-y| < \delta$.

Notice that the key point here (and the way in which this differs from uniform continuity of each f_i) is that we can choose δ *independently* of i.

Exercise 3.9 (*)**.** Suppose $F = (f_i)_{i=0}^{\infty}$ is a sequence of continuous functions $f_i : [a,b] \to \mathbb{R}$, which are differentiable on (a,b) and satisfy:

$$\sup_{x\in(a,b)} \left|f_i'(x)\right| \leq M$$

for some M independent of i. Show that F is uniformly equicontinuous.

Theorem 1.9 (Arzelà-Ascoli)**.** *Let $F = (f_i)_{i=0}^{\infty}$ be a uniformly equicontinuous sequence of functions $f_i : [a,b] \to \mathbb{R}$ satisfying:*

$$\sup_{x\in[a,b]} |f_i(x)| \leq K$$

for some K independent of i. Then F has a subsequence which converges uniformly.

Proof. First of all, note that the set $[a,b] \cap \mathbb{Q}$ is countable, so let us write $[a,b] \cap \mathbb{Q} = \{x_1, x_2, \ldots\}$. Now consider the sequence of numbers $(f_i(x_1))_{i=0}^{\infty}$. This is bounded, so by Bolzano–Weierstrass has a convergent subsequence $(f_{i_j}(x))_{j=0}^{\infty}$. To keep notation a bit easier, let's denote $f_{0,j} = f_j$ and $f_{1,j} = f_{i_j}$, so that $f_{1,j}$ is a subsequence of F with $f_{1,j}(x_1)$ convergent. Now let's consider $(f_{1,i}(x_2))_{i=0}^{\infty}$, which again is a bounded sequence of real numbers, with a convergent subsequence $f_{1,i_j}(x_2)$. Let us set $f_{2,j} = f_{1,i_j}$. This is a subsequence of $f_{1,j}$ such that $f_{2,j}(x_1)$ and $f_{2,j}(x_2)$ both converge. We can repeat this process of extracting subsequences to obtain functions $f_{k,i}$ for $k, i \in \mathbb{N}$ with the property that $(f_{k+1,i})_{i=0}^{\infty}$ is a subsequence of $(f_{k,i})_{i=0}^{\infty}$ and moreover $f_{k,i}(x_l)$ converges for $l \leq k$.

Let us define a new subsequence $G = (g_i)_{i=0}^{\infty}$ by $g_i = f_{i,i}$. This has the property that G is a subsequence of F, and moreover $g_i(x_l)$ converges for each $l \in \mathbb{N}$, let us say $g_i(x_l) \to g(x_l)$. It remains to show that in fact g_i converges uniformly. To see this, let us fix $\epsilon > 0$. Since F is equicontinuous, we may find $\delta > 0$ such that $|g_i(x) - g_i(y)| < \frac{1}{3}\epsilon$ for all $|x-y| < \delta$. Since $[a,b]$ is finite, there exist a finite number of rational points $\{x_1, \ldots x_K\} \in [a,b] \cap \mathbb{Q}$ such that $[a,b] \subset \cup_{m=1}^{K}(x_m - \delta, x_m + \delta)$. Since g_i converges at each rational point, for each $m = 1, \ldots K$ there exists N_m such that for all $i, j \geq N_m$ we have $|g_i(x_m) - g_j(x_m)| < \frac{\epsilon}{3}$. Let $N = \max\{N_1, \ldots N_K\}$ and suppose $i, j \geq N$. Fix $x \in [a,b]$. By construction, there is an $m \in \{1, \ldots, K\}$ such that $|x - x_m| < \delta$. We estimate:

$$\begin{aligned}
|g_i(x) - g_j(x)| &= |g_i(x) - g_i(x_m) + g_i(x_m) - g_j(x_m) + g_j(x_m) - g_j(x)| \\
&\leq |g_i(x) - g_i(x_m)| + |g_i(x_m) - g_j(x_m)| + |g_j(x_m) - g_j(x)| \\
&\leq \frac{1}{3}\epsilon + \frac{1}{3}\epsilon + \frac{1}{3}\epsilon = \epsilon.
\end{aligned}$$

Thus, g_i is a Cauchy sequence in $C^0([a,b])$, so by Theorem 1.6, we deduce that $g_i \to g$ uniformly for some uniformly continuous $g : [a,b] \to \mathbb{R}$. □

Chapter 2

Integration

At school, and in your methods courses last year and this, you will have been introduced to the integral. Typically, it is introduced as 'an inverse operation of differentiation'. In practice, to integrate a function $f : \mathbb{R} \to \mathbb{R}$, we try and find an $F : \mathbb{R} \to \mathbb{R}$ such that $F'(x) = f(x)$. We call F the antiderivative (or indefinite integral) of f, and write:

$$\int_a^b f(t)dt := F(b) - F(a).$$

For many of the functions that you are familiar with, this approach is fine. We have a catalogue of antiderivatives, and can follow this algorithm. When we want to make statements about *general* functions f, things become more difficult. We need a better definition of the integral, one which we can use even if we're not able to find an antiderivative. The basic idea that we shall pursue is the following:

$$\int_a^b f(t)dt := \text{'The area under the graph of } f \text{ between } a \text{ and } b\text{'}.$$

It will turn out that we can make sense of the area under the graph of f for a reasonably large class of functions. The approach we shall take will lead us to a definition of the integral known as the *Riemann* integral (strictly we shall use the Darboux integral, but the two are equivalent). It is not the only possible definition of an integral. If you take the course M3P19 Measure and Integration you will be introduced to the Lebesgue integral, which for many problems in analysis is more powerful.

The basic idea will be to split the region under the graph of f up and approximate the area by sufficiently narrow rectangles, as shown in Figures 2.1, 2.2. We will estimate the area from above by considering rectangles whose tops lie above the graph, and we will estimate the area from below by considering rectangles whose tops lie below the graph. For a 'reasonable' function, these two estimates should get closer and closer as the size of the rectangles shrinks to zero. Of course, since this is an analysis course, we will need to be precise about notions like 'closer and closer' and 'reasonable' in this context.

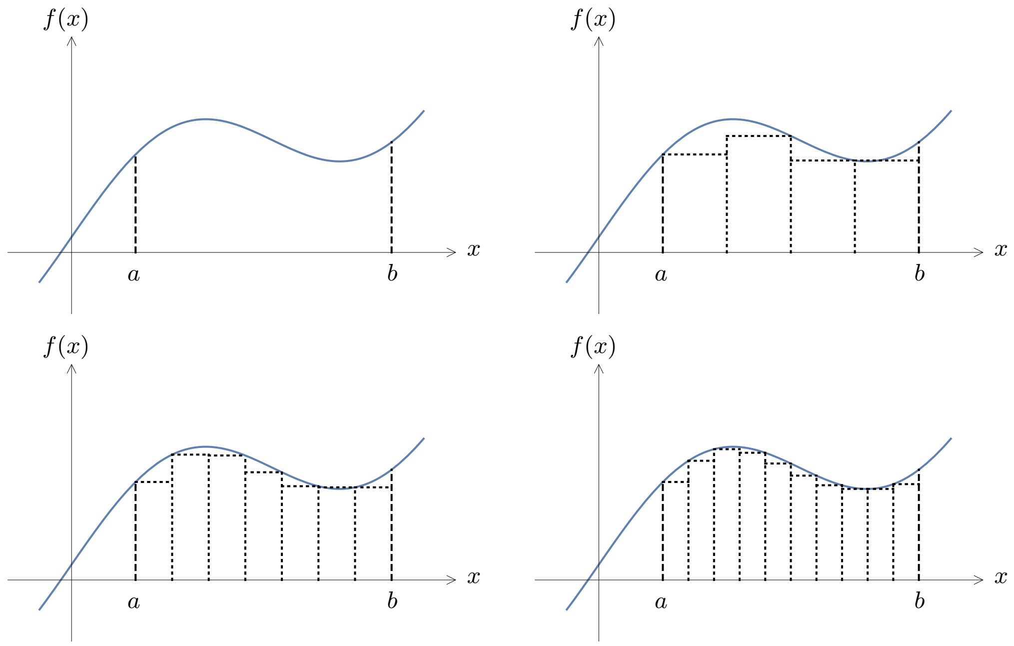

Figure 2.1 Estimating the area under a graph from below with rectangles

2.1 The Darboux integral on $\mathbb{R}$

The basic idea of our definition of the integral has been spelled out above, but we will need to spend a little bit of time setting up some machinery first in order to make our definition precise.

2.1.1 Partitions of an interval

Throughout this section, we will be working on a finite interval $I := [a, b] \subset \mathbb{R}$. A *partition*, $\mathcal{P}$, of the interval I is a finite sequence of numbers $\mathcal{P} := (x_0, \ldots, x_k)$ satisfying:

$$a = x_0 < x_1 < \ldots < x_{k-1} < x_k = b.$$

Given a partition, we can break the interval I up into segments $[x_j, x_{j+1}]$ which overlap only at their endpoints. We call these segments *subintervals*. We denote the length of the n^{th} subinterval by Δx_j, i.e.

$$\Delta x_j = x_{j+1} - x_j, \qquad j = 0, \ldots, k-1.$$

Notice that we don't require the subintervals to all have the same length. The *mesh* of a partition $\mathcal{P} = (x_0 \ldots, x_k)$ is the length of the longest subinterval:

$$\operatorname{mesh} \mathcal{P} := \max_{j=0,\ldots,k-1} \Delta x_j = \max_{j=0,\ldots,k-1} (x_{j+1} - x_j).$$

Given two partitions, $\mathcal{P}$, $\mathcal{Q}$, we say that $\mathcal{Q}$ is a *refinement* of $\mathcal{P}$, or $\mathcal{Q}$ *refines* $\mathcal{P}$ if every point of $\mathcal{P}$ is also a point of $\mathcal{Q}$. In this case, we write $\mathcal{P} \preceq \mathcal{Q}$.

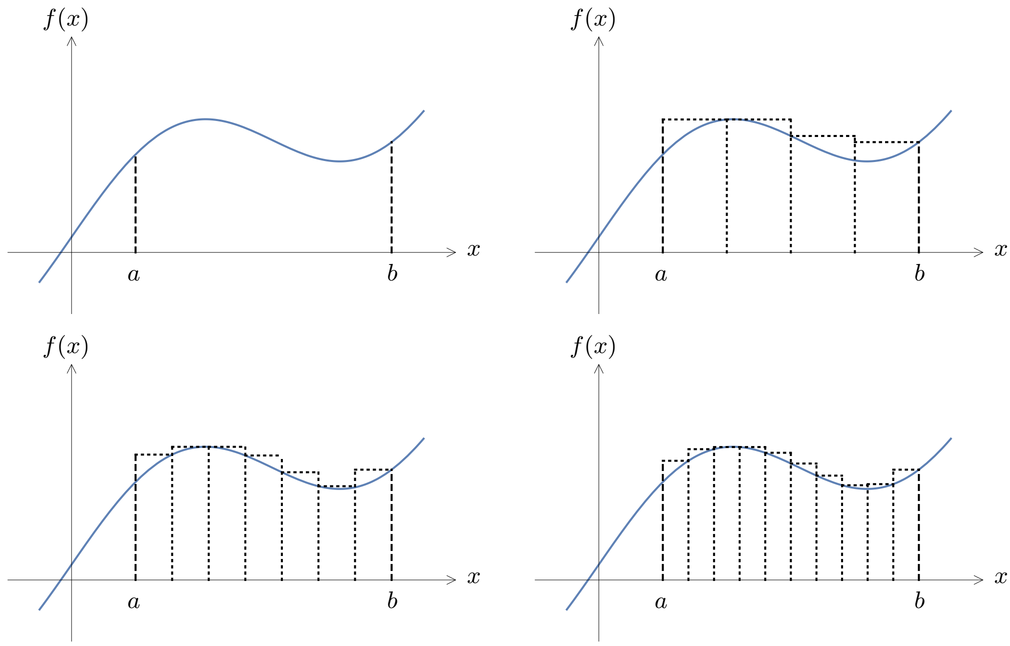

Figure 2.2 Estimating the area under a graph from above with rectangles

Lemma 2.1. *Given two partitions, $\mathcal{P}$, $\mathcal{Q}$, there is a unique minimal partition $\mathcal{R}$ such that $\mathcal{P} \preceq \mathcal{R}$* and *$\mathcal{Q} \preceq \mathcal{R}$. $\mathcal{R}$ is minimal in the sense that if $\mathcal{R}'$ satisfies $\mathcal{P} \preceq \mathcal{R}'$* and *$\mathcal{Q} \preceq \mathcal{R}'$, then $\mathcal{R} \preceq \mathcal{R}'$.*

Proof. Given $\mathcal{P} := (x_0, \ldots, x_k)$ and $\mathcal{Q} := (y_0, \ldots, y_l)$, we simply take $\mathcal{R}$ to be the set $\{x_0, \ldots, x_k, y_0, \ldots, y_l\}$ with any duplicates deleted and re-ordered as appropriate to give $\mathcal{R} = (z_0, \ldots, z_m)$. This certainly gives a refinement of both $\mathcal{P}$ and $\mathcal{Q}$, and moreover it is minimal, as any other partition which refines both $\mathcal{P}$ and $\mathcal{Q}$ must contain all the points of $\mathcal{R}$ and hence must refine $\mathcal{R}$. □

We call $\mathcal{R}$ as constructed above the *common refinement* of $\mathcal{P}$ and $\mathcal{Q}$. Figure 2.3 shows a sketch of the common refinement of two partitions.

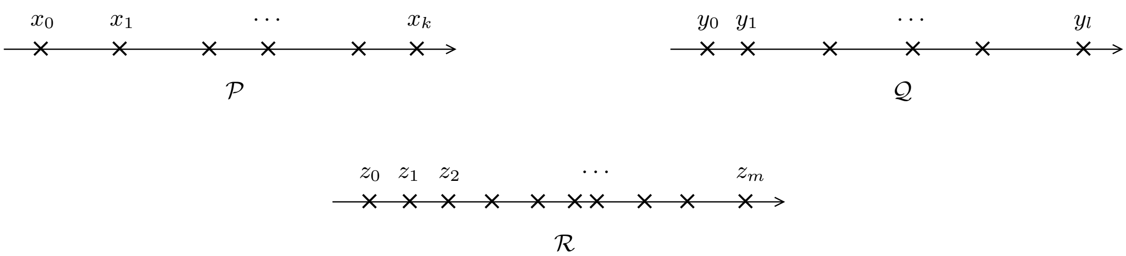

Figure 2.3 The common refinement $\mathcal{R}$ of two partitions $\mathcal{P}, \mathcal{Q}$.

2.1.2 Upper and lower Darboux integrals

Now, let us consider a bounded function $f : [a, b] \to \mathbb{R}$. It is helpful for visualising things to think of f as continuous to begin with, but we will not make that assumption. Given a partition $\mathcal{P} = (x_0, \ldots, x_k)$ of the interval $[a, b]$, we introduce the numbers:

$$m_j := \inf_{x \in [x_j, x_{j+1}]} f(x),$$
$$M_j := \sup_{x \in [x_j, x_{j+1}]} f(x),$$

for $j = 0, \ldots, k-1$, which are respectively the greatest lower and least upper bounds on f in the interval $[x_j, x_{j+1}]$. Since f is bounded, these are all finite and well-defined. Next, we introduce the *lower* and *upper Darboux sums* of f with respect to $\mathcal{P}$:

$$L(f, \mathcal{P}) := \sum_{j=0}^{k-1} m_j \Delta x_j$$
$$U(f, \mathcal{P}) := \sum_{j=0}^{k-1} M_j \Delta x_j.$$

Recall that $\Delta x_j = (x_{j+1} - x_j)$ is the length of the interval $[x_j, x_{j+1}]$, so each term in the sum represents the area of a rectangle with base Δx_j and height m_j (resp. M_j). The geometrical picture describing these sums is given in Figures 2.1, 2.2. In both cases, the sums represent the area of the rectangles, which in the case of the lower sum *under* estimates the area under the graph and in the case of the upper sum *over* estimates the area under the graph. In this discussion, we have assumed that the graph lies above the axis. In the case where it doesn't, we understand 'area' to be negative when it's below the axis.

Since $m_j \leq M_j$ for all $j = 0, k-1$, we certainly have:

$$L(f, \mathcal{P}) \leq U(f, \mathcal{P})$$

We expect that the area under the graph of f should lie somewhere between these two bounds. We also expect that a more refined partition should estimate the area more accurately. We can make this precise with the following theorem, which is really the crucial result of this section.

Theorem 2.2. *Let $f : [a, b] \to \mathbb{R}$ be bounded. Suppose that $\mathcal{P}$, $\mathcal{Q}$ are partitions of $[a, b]$ with $\mathcal{Q}$ a refinement of $\mathcal{P}$. Then:*

$$L(f, \mathcal{P}) \leq L(f, \mathcal{Q}) \leq U(f, \mathcal{Q}) \leq U(f, \mathcal{P}).$$

Proof. Suppose first that $\mathcal{P} = (x_0, \ldots, x_k)$ and $\mathcal{Q} = (x_0, \ldots, x_l, c, x_{l+1}, \ldots x_k)$, so that $\mathcal{P}$ and $\mathcal{Q}$ differ only in that a single subinterval $[x_l, x_{l+1}]$ has been split into $[x_l, c]$ and $[c, x_{l+1}]$. From basic properties of the infimum, we have:

$$\inf_{x \in [x_l, c]} f(x) \geq \inf_{x \in [x_l, x_{l+1}]} f(x), \quad \text{and} \quad \inf_{x \in [c, x_{l+1}]} f(x) \geq \inf_{x \in [x_l, x_{l+1}]} f(x).$$

Now, if we take the difference of the lower sums associated to $\mathcal{P}$ and $\mathcal{Q}$, then the only terms which will not cancel will be those involving the subinterval $[x_l, x_{l+1}]$:

$$\begin{aligned} L(f, \mathcal{Q}) - L(f, \mathcal{P}) &= (c - x_l) \inf_{x \in [x_l, c]} f(x) + (x_{l+1} - c) \inf_{x \in [c, x_{l+1}]} f(x) \\ &\quad - (x_{l+1} - x_l) \inf_{x \in [x_l, x_{l+1}]} f(x) \\ &\geq [(c - x_l) + (x_{l+1} - c) - (x_{l+1} - x_l)] \inf_{x \in [x_l, x_{l+1}]} f(x), \\ &\geq 0. \end{aligned}$$

Thus we have:

$$L(f, \mathcal{P}) \leq L(f, \mathcal{Q}).$$

Now, note that by Exercise 4.2 *a*), we have $U(f, \mathcal{P}) = -L(-f, \mathcal{P})$, so that:

$$U(f, \mathcal{Q}) = -L(-f, \mathcal{Q}) \leq -L(-f, \mathcal{P}) = U(f, \mathcal{P}),$$

so that we have established

$$L(f, \mathcal{P}) \leq L(f, \mathcal{Q}) \leq U(f, \mathcal{Q}) \leq U(f, \mathcal{P}).$$

in the special case where $\mathcal{P} = (x_0, \ldots, x_k)$ and $\mathcal{Q} = (x_0, \ldots, x_l, c, x_{l+1}, \ldots x_k)$. For the general result, we observe that if $\mathcal{P} \preceq \mathcal{Q}$ then we can obtain $\mathcal{Q}$ by inserting a finite number of points into $\mathcal{P}$, and we can apply our special case result at each stage to obtain the general result. □

Corollary 2.3. *Let $f : [a, b] \to \mathbb{R}$ be bounded. Suppose $\mathcal{P}, \mathcal{P}'$ are any two partitions of $[a, b]$. Then:*

$$L(f, \mathcal{P}) \leq U(f, \mathcal{P}').$$

Proof. Let $\mathcal{Q}$ be the common refinement of $\mathcal{P}, \mathcal{P}'$. Then by Theorem 2.2 we have:

$$L(f, \mathcal{P}) \leq L(f, \mathcal{Q}) \leq U(f, \mathcal{Q}) \leq U(f, \mathcal{P}'),$$

which gives the result. □

Exercise 4.1. Let $A \subset \mathbb{R}^n$, suppose $f, g : A \to \mathbb{R}$ are bounded, and let $\lambda \geq 0$. Show that:

a) $\displaystyle\sup_{x \in A]} -f(x) = -\inf_{x \in A} f(x)$,

b) $\displaystyle\inf_{x \in A} -f(x) = -\sup_{x \in A} f(x)$,

c) $\displaystyle\sup_{x \in A} \lambda f(x) = \lambda \sup_{x \in A} f(x)$,

d) $\displaystyle\sup_{x \in A} (f(x) + g(x)) \leq \sup_{x \in A} f(x) + \sup_{x \in A} g(x)$,

e) $\displaystyle\inf_{x \in A} (f(x) + g(x)) \geq \inf_{x \in A} f(x) + \inf_{x \in A} g(x)$,

f) $\displaystyle\inf_{x \in A} \lambda f(x) = \lambda \inf_{x \in A} f(x)$,

g) $\displaystyle\left|\sup_{x \in A} f(x)\right| \leq \sup_{x \in A} |f(x)|$,

h) $\displaystyle\left|\inf_{x \in A} f(x)\right| \leq \sup_{x \in A} |f(x)|$.

Exercise 4.2. Let $[a,b] \subset \mathbb{R}$ be any finite interval and $\mathcal{P}$ be any partition of $[a,b]$. Suppose $f, g : [a,b] \to \mathbb{R}$ are bounded functions, and that $\lambda \geq 0$. Show that:

a) $U(-f,\mathcal{P}) = -L(f,\mathcal{P})$,

b) $L(-f,\mathcal{P}) = -U(f,\mathcal{P})$,

c) $U(\lambda f,\mathcal{P}) = \lambda U(f,\mathcal{P})$,

d) $L(\lambda f,\mathcal{P}) = \lambda L(f,\mathcal{P})$,

e) $U(f+g,\mathcal{P}) \leq U(f,\mathcal{P}) + U(g,\mathcal{P})$,

f) $L(f+g,\mathcal{P}) \geq L(f,\mathcal{P}) + L(g,\mathcal{P})$.

Exercise 4.3. Suppose that $\mathcal{P} \preceq \mathcal{Q}$. Show that:

$$0 \leq U(f,\mathcal{Q}) - L(f,\mathcal{Q}) \leq U(f,\mathcal{P}) - L(f,\mathcal{P})$$

2.1.3 The Darboux Integral

We are now ready to define the integral:

Definition 2.1. The *lower* and *upper Darboux integrals* are defined as follows:

$$\underline{\int_a^b} f(x)dx := \sup_{\mathcal{P}} L(f,\mathcal{P}) \qquad \overline{\int_a^b} f(x)dx := \inf_{\mathcal{P}} U(f,\mathcal{P})$$

where the supremum and infimum are taken over all partitions of $[a,b]$. By Corollary 2.3 we have that:

$$\underline{\int_a^b} f(x)dx \leq \overline{\int_a^b} f(x)dx.$$

We say that a bounded function $f : [a,b] \to \mathbb{R}$ is *Darboux integrable*, or simply *integrable*,[1] if it is bounded, and the upper and lower Darboux integrals are equal. In this case, we write:

$$\int_a^b f(x)dx := \underline{\int_a^b} f(x)dx = \overline{\int_a^b} f(x)dx.$$

Notice that the lower and upper Darboux integrals are always well defined for a bounded function f.

As a notational convention, x appears in the integral as a 'dummy variable'. We can replace x with any other letter and the meaning is unchanged:

$$\int_a^b f(x)dx = \int_a^b f(t)dt = \int_a^b f(\alpha)d\alpha, \quad \text{etc.}$$

Example 2.1. Suppose $f : [a,b] \to \mathbb{R}$ is the constant function $f(x) = c$. Then we find, for any partition $\mathcal{P}$ that $m_j = M_j = c$ and so:

$$L(f,\mathcal{P}) = U(f,\mathcal{P}) = \sum_{j=0}^{k-1} c\Delta x_j = c\sum_{j=0}^{k-1}(x_{j+1} - x_j) = c(a-b).$$

[1] If you continue with your study of analysis you will discover that there are other inequivalent definitions of integrability, in which case one has to distinguish what type of integrability one refers to.

as the sum is a telescoping sum. Thus f is integrable and:

$$\int_a^b f(x)dx = c(a-b).$$

Example 2.2. Suppose $f : [-1, 1] \to \mathbb{R}$ is given by:

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

For $l \in \mathbb{N}$, let us consider the partition of $[-1, 1]$ given by $\mathcal{P}_l = (-1, -2^{-l}, 2^{-l}, 1)$. For the intervals I= $[-1, -2^{-l}], [2^{-l}, 1]$ we have that:

$$\sup_{x \in I} f(x) = \inf_{x \in I} f(x) = 0,$$

so these intervals do not contribute to either the upper or lower sums. For the interval $I = [-2^{-l}, 2^{-l}]$, we have:

$$\inf_{x \in I} f(x) = 0, \quad \sup_{x \in I} f(x) = 1, \quad \Delta x = 2^{-l+1},$$

so that:

$$L(f, \mathcal{P}_l) = 0, \qquad U(f, \mathcal{P}_l) = 2^{-l+1}.$$

We conclude that[2]:

$$\overline{\int_{-1}^{1}} f(x)dx = \inf_{\mathcal{P}} U(f, \mathcal{P}) \leq 0, \qquad \underline{\int_{-1}^{1}} f(x)dx = \sup_{\mathcal{P}} L(f, \mathcal{P}) \geq 0,$$

from which, since:

$$\underline{\int_{-1}^{1}} f(x)dx \leq \overline{\int_{a}^{b}} f(x)dx,$$

we deduce that f is integrable, and

$$\int_{-1}^{1} f(x)dx = 0.$$

Example 2.3. Consider the Dirichlet function $f : [0, 1] \to \mathbb{R}$:

$$f(x) = \begin{cases} 1 & x \in \mathbb{Q}, \\ 0 & x \notin \mathbb{Q}. \end{cases}$$

If $I \subset [0, 1]$ is any interval of positive length, then $\sup_{x \in I} f(x) = 1$ and $\inf_{x \in I} f(x) = 0$, so that for any partition $\mathcal{P}$ we have $M_j = 1$, $m_j = 0$. We conclude that:

$$L(f, \mathcal{P}) = 0, \qquad U(f, \mathcal{P}) = 1.$$

Thus, we see that this function is *not* integrable.

[2] Note carefully that we cannot immediately conclude that both are zero.

An important property of the integral is that it is linear:

Theorem 2.4 (Linearity of the integral)**.** *Suppose that $f : [a, b] \to \mathbb{R}$ and $g : [a, b] \to \mathbb{R}$ are both integrable, and that $\lambda \in \mathbb{R}$. Then $f + \lambda g$ is integrable and we have:*

$$\int_a^b (f(x) + \lambda g(x))\, dx = \int_a^b f(x)dx + \lambda \int_a^b g(x)dx.x$$

Proof. We shall proceed in two stages: first we will show that if g is integrable, then so is $-g$. Then we will establish that $f + \lambda g$ is integrable for $\lambda \geq 0$. The full result will follow from combining these two facts.

1. Note that for any partition, $L(-g, \mathcal{P}) = -U(g, \mathcal{P})$, which implies:

$$\underline{\int_a^b} - g(x)dx = \sup_{\mathcal{P}} L(-g, \mathcal{P}) = \sup_{\mathcal{P}} -U(g, \mathcal{P}) = -\inf_{\mathcal{P}} U(g, \mathcal{P}) = -\int_a^b g(x)dx$$

   Similarly,

$$\overline{\int_a^b} - g(x)dx = \inf_{\mathcal{P}} U(-g, \mathcal{P}) = \inf_{\mathcal{P}} -L(g, \mathcal{P}) = -\sup_{\mathcal{P}} U(g, \mathcal{P}) = -\int_a^b g(x)dx,$$

   so that $-g$ is integrable and:

$$\int_a^b -g(x)dx = -\int_a^b g(x)dx.$$

2. Now, suppose $\lambda \geq 0$, then for any partition $\mathcal{P}$, we have by Exercise 4.2 that:

$$L(f, \mathcal{P}) + \lambda L(g, \mathcal{P}) \leq L(f + \lambda g, \mathcal{P}),$$

   From the definition of the supremum we of course have:

$$L(f + \lambda g, \mathcal{P}) \leq \sup_{\mathcal{Q}} L(f + \lambda g, \mathcal{Q}),$$

   so that:

$$L(f, \mathcal{P}) + \lambda L(g, \mathcal{P}) \leq \sup_{\mathcal{Q}} L(f + \lambda g, \mathcal{Q}).$$

   Now, suppose that $\mathcal{P}$, $\mathcal{P}'$ are any two partitions and $\mathcal{R}$ is their common refinement. Then

$$L(f, \mathcal{P}) + \lambda L(g, \mathcal{P}') \leq L(f, \mathcal{R}) + \lambda L(g, \mathcal{R}).$$

   Putting these estimates together gives:

$$L(f, \mathcal{P}) + \lambda L(g, \mathcal{P}') \leq \sup_{\mathcal{Q}} L(f + \lambda g, \mathcal{Q}),$$

   for any pair of partitions $\mathcal{P}$, $\mathcal{P}'$. Taking the supremum over $\mathcal{P}, \mathcal{P}'$ we have:

$$\sup_{\mathcal{P}} L(f, \mathcal{P}) + \lambda \sup_{\mathcal{P}'} L(g, \mathcal{P}') \leq \sup_{\mathcal{Q}} L(f + \lambda g, \mathcal{Q}),$$

or equivalently:

$$\int_a^b f(x)dx + \lambda \int_a^b g(x)dx \leq \underline{\int_a^b} [f(x) + \lambda g(x)]\, dx.$$

Applying this estimate to $-f$ and $-g$ we have:

$$\sup_{\mathcal{P}} L(-f, \mathcal{P}) + \lambda \sup_{\mathcal{P}'} L(-g, \mathcal{P}') \leq \sup_{\mathcal{Q}} L(-\left(f + \lambda g\right), \mathcal{Q})$$

which implies:

$$\sup_{\mathcal{P}} -U(f, \mathcal{P}) + \lambda \sup_{\mathcal{P}'} -U(g, \mathcal{P}') \leq \sup_{\mathcal{Q}} -U(f + \lambda g, \mathcal{Q})$$

and thus:

$$-\inf_{\mathcal{P}} U(f, \mathcal{P}) - \lambda \inf_{\mathcal{P}'} U(g, \mathcal{P}') \leq -\inf_{\mathcal{Q}} U(f + \lambda g, \mathcal{Q})$$

so that:

$$\overline{\int_a^b} [f(x) + \lambda g(x)]\, dx \leq \int_a^b f(x)dx + \lambda \int_a^b g(x)dx.$$

Since by definition:

$$\underline{\int_a^b} [f(x) + \lambda g(x)]\, dx \leq \overline{\int_a^b} [f(x) + \lambda g(x)]\, dx,$$

we conclude that

$$\underline{\int_a^b} [f(x) + \lambda g(x)]\, dx = \overline{\int_a^b} [f(x) + \lambda g(x)]\, dx = \int_a^b f(x)dx + \lambda \int_a^b g(x)dx,$$

which is the result we require. Combining these two cases gives the general result.

□

Before we establish some further results concerning integrable functions, we need a useful lemma.

Lemma 2.5. *A function $f : [a, b] \to \mathbb{R}$ is integrable if and only if for any $\epsilon > 0$ there exists a partition $\mathcal{P}$ such that:*

$$U(f, \mathcal{P}) - L(f, \mathcal{P}) < \epsilon. \tag{2.1}$$

Proof. Note that if $\mathcal{P}$ is any partition, from the definition of the supremum and infimum we have:

$$\sup_{\mathcal{Q}} L(f, \mathcal{Q}) \geq L(f, \mathcal{P}), \qquad \inf_{\mathcal{Q}} U(f, \mathcal{Q}) \leq U(f, \mathcal{P}).$$

Combining these two inequalities, we have that for any $\mathcal{P}$:

$$0 \leq \inf_{\mathcal{Q}} U(f, \mathcal{Q}) - \sup_{\mathcal{Q}} L(f, \mathcal{Q}) < U(f, \mathcal{P}) - L(f, \mathcal{P})$$

Now suppose that for every $\epsilon > 0$, there exists a partition $\mathcal{P}$ such that (2.1) holds. Then for any ϵ we have:

$$0 \leq \inf_{\mathcal{Q}} U(f, \mathcal{Q}) - \sup_{\mathcal{Q}} L(f, \mathcal{Q}) < \epsilon,$$

which implies (since ϵ is arbitrary):

$$0 = \inf_{\mathcal{Q}} U(f, \mathcal{Q}) - \sup_{\mathcal{Q}} L(f, \mathcal{Q}) = \overline{\int_a^b} f(x)dx - \underline{\int_a^b} f(x)dx$$

so that f is integrable.

Conversely, suppose that f is integrable, so that $\inf_{\mathcal{Q}} U(f, \mathcal{Q}) = \sup_{\mathcal{Q}} L(f, \mathcal{Q}) = I$. Then from the properties of the supremum and infimum we can find $\mathcal{P}, \mathcal{P}'$ such that:

$$U(f, \mathcal{P}) < I + \frac{\epsilon}{2}, \qquad L(f, \mathcal{P}') > I - \frac{\epsilon}{2}$$

Taking $\mathcal{R}$ to be the common refinement of $\mathcal{P}, \mathcal{P}'$, we deduce:

$$U(f, \mathcal{R}) < I + \frac{\epsilon}{2}, \qquad L(f, \mathcal{R}) > I - \frac{\epsilon}{2},$$

which implies:

$$U(f, \mathcal{R}) - L(f, \mathcal{R}) < \epsilon.$$

□

Note that by Exercise 4.3, if $\mathcal{P}$ satisfies (2.1), then so does any partition $\mathcal{Q}$ with $\mathcal{P} \preceq \mathcal{Q}$. This formulation is often more useful to actually work with, as we shall now see, when we consider additivity of the integral. It is important to know that if we break up an interval into two or more subintervals, then the integral behaves as we would expect.

Theorem 2.6 (Additivity of the integral)**.** *Suppose $f : [a, b] \to \mathbb{R}$, and let $c \in (a, b)$. Then f is integrable if and only if $f|_{[a,c]}$ and $f|_{[c,b]}$ are integrable, and:*

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx.$$

Proof. 1. First suppose $f : [a, b] \to \mathbb{R}$ is integrable. Fix $\epsilon > 0$. By Lemma 2.5 there exists a partition of $[a, b]$, $\mathcal{P}$ such that:

$$U(f, \mathcal{P}) - L(f, \mathcal{P}) < \epsilon.$$

By refining $\mathcal{P}$ by adding the point c if necessary, we can assume that c is an endpoint of $\mathcal{P}$, so that:

$$\mathcal{P} = (x_0, \ldots, x_l, c, x_{l+1}, \ldots, x_k).$$

We have, by Exercise 4.5

$$L(f, \mathcal{P}) = L(f|_{[a,c]}, \mathcal{P}_L) + L(f|_{[c,b]}, \mathcal{P}_R),$$
$$U(f, \mathcal{P}) = U(f|_{[a,c]}, \mathcal{P}_L) + U(f|_{[c,b]}, \mathcal{P}_R),$$

where

$$\mathcal{P}_L = (x_0, \ldots, x_l, c), \qquad \mathcal{P}_R = (c, x_{l+1}, \ldots, x_k).$$

We obtain:

$$\begin{aligned}&\Big[U(f|_{[a,c]}, \mathcal{P}_L) - L(f|_{[a,c]}, \mathcal{P}_L)\Big] + \Big[U(f|_{[c,b]}, \mathcal{P}_R) - L(f|_{[c,b]}, \mathcal{P}_R)\Big]\\ &\quad = U(f, \mathcal{P}) - L(f, \mathcal{P}) < \epsilon.\end{aligned}$$

Now, both terms in square brackets are non-negative, so we conclude that:

$$U(f|_{[a,c]}, \mathcal{P}_L) - L(f|_{[a,c]}, \mathcal{P}_L) < \epsilon, \qquad U(f|_{[c,b]}, \mathcal{P}_R) - L(f|_{[c,b]}, \mathcal{P}_R) < \epsilon,$$

which implies that indeed $f|_{[a,c]}$ and $f|_{[c,b]}$ are integrable.

2. Now conversely, suppose $f|_{[a,c]}$ and $f|_{[c,b]}$ are integrable. Fix $\epsilon > 0$. Then there exist partitions $\mathcal{P}_\mathcal{L}$ of $[a, c]$ and $\mathcal{P}_\mathcal{R}$ of $[c, b]$ such that:

$$U(f|_{[a,c]}, \mathcal{P}_L) - L(f|_{[a,c]}, \mathcal{P}_L) < \frac{\epsilon}{2}, \qquad U(f|_{[c,b]}, \mathcal{P}_R) - L(f|_{[c,b]}, \mathcal{P}_R) < \frac{\epsilon}{2}$$

Now, we form a partition $\mathcal{P}$ of $[a, b]$ by concatenating $\mathcal{P}_L$ and $\mathcal{P}_R$. We again have that:

$$\begin{aligned}L(f, \mathcal{P}) &= L(f|_{[a,c]}, \mathcal{P}_L) + L(f|_{[c,b]}, \mathcal{P}_R),\\ U(f, \mathcal{P}) &= U(f|_{[a,c]}, \mathcal{P}_L) + U(f|_{[c,b]}, \mathcal{P}_R),\end{aligned}$$

so

$$\begin{aligned}U(f, \mathcal{P}) - L(f, \mathcal{P}) = &\Big[U(f|_{[a,c]}, \mathcal{P}_L) - L(f|_{[a,c]}, \mathcal{P}_L)\Big]\\ &+ \Big[U(f|_{[c,b]}, \mathcal{P}_R) - L(f|_{[c,b]}, \mathcal{P}_R)\Big] < \epsilon.\end{aligned}$$

Which implies f is integrable. We further deduce that:

$$\begin{aligned}\int_a^b f(x)dx &\leq U(f, \mathcal{P}) = U(f|_{[a,c]}, \mathcal{P}_L) + U(f|_{[c,b]}, \mathcal{P}_R)\\ &< \int_a^c f(x)dx + \int_c^b f(x)dx + \epsilon,\end{aligned}$$

Similarly, we have:

$$\begin{aligned}\int_a^b f(x)dx &\geq L(f, \mathcal{P}) = L(f|_{[a,c]}, \mathcal{P}_L) + L(f|_{[c,b]}, \mathcal{P}_R)\\ &> \int_a^c f(x)dx + \int_c^b f(x)dx - \epsilon.\end{aligned}$$

Since $\epsilon > 0$ is arbitrary, we conclude that:

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx. \qquad \square$$

It is useful to extend slightly our definition of the integral to allow for the case where $a \geq b$. This is purely a convention, but we define:

$$\int_a^b f(x)dx := -\int_b^a f(x)dx \quad \text{for } a > b, \qquad \int_a^a f(x)dx := 0.$$

The next Lemma we shall require is slightly technical, but important.

Lemma 2.7. *If $A \subset \mathbb{R}^n$, and $f : A \to \mathbb{R}$, then:*

$$\sup_{x \in A} |f(x)| - \inf_{x \in A} |f(x)| \leq \sup_{x \in A} f(x) - \inf_{x \in A} f(x).$$

Proof. We consider three possible cases:

a) $f(x) \geq 0$ for all $x \in A$. In this case, $|f(x)| = f(x)$ and we have

$$\sup_{x \in A} |f(x)| - \inf_{x \in A} |f(x)| = \sup_{x \in A} f(x) - \inf_{x \in A} f(x).$$

b) $f(x) \leq 0$ for all $x \in A$. In this case, $|f(x)| = -f(x)$ and $\sup_{x \in A} |f(x)| = -\inf_{x \in A} f(x)$, $\inf_{x \in A} |f(x)| = -\sup_{x \in A} f(x)$ so again:

$$\sup_{x \in A} |f(x)| - \inf_{x \in A} |f(x)| = \sup_{x \in A} f(x) - \inf_{x \in A} f(x).$$

c) Finally, suppose $\inf_{x \in A} f(x) < 0 < \sup_{x \in I} f(x)$. Then

$$\sup_{x \in A} |f(x)| = \max\{\sup_{x \in A} f(x), -\inf_{x \in A} f(x)\}$$

where both terms are positive, and we have:

$$\sup_{x \in A} |f(x)| - \inf_{x \in A} |f(x)| \leq \sup_{x \in A} |f(x)| \leq \sup_{x \in A} f(x) - \inf_{x \in A} f(x)$$

□

Theorem 2.8. *Suppose $f : [a, b] \to \mathbb{R}$ and $g : [a, b] \to \mathbb{R}$ are both integrable. Then so is:*

i) $|f|$,

ii) f^2,

iii) fg.

Proof. i) Fix $\epsilon > 0$. Since f is integrable, there exists a partition $\mathcal{P} = (x_0, \ldots, x_k)$ such that:

$$U(f, \mathcal{P}) - L(f, \mathcal{P}) = \sum_{j=0}^{k-1} \left[\sup_{x \in [x_j, x_{j+1}]} f(x) - \inf_{x \in [x_j, x_{j+1}]} f(x) \right] \Delta x_j < \epsilon$$

Now let us estimate:

$$\begin{aligned} U(|f|,\mathcal{P}) - L(|f|,\mathcal{P}) &= \sum_{j=0}^{k-1}\left[\sup_{x\in[x_j,x_{j+1}]}|f(x)| - \inf_{x\in[x_j,x_{j+1}]}|f(x)|\right]\Delta x_j \\ &\leq \sum_{j=0}^{k-1}\left[\sup_{x\in[x_j,x_{j+1}]} f(x) - \inf_{x\in[x_j,x_{j+1}]} f(x)\right]\Delta x_j \\ &< \epsilon, \end{aligned}$$

where we have used Lemma 2.7 to go from the first line to the second. We conclude that $|f|$ is integrable.

ii) Since f is integrable, it is bounded, say $|f(x)| \leq M$. Note that:

$$\left|f(x)^2 - f(y)^2\right| = |f(x) - f(y)|\,|f(x) + f(y)| \leq 2M\,|f(x) - f(y)|\,.$$

Since this holds for any x, y, in particular we conclude that for any interval $I \subset \mathbb{R}$:

$$\sup_{x\in I} f(x)^2 - \inf_{y\in I} f(y)^2 \leq 2M\left(\sup_{x\in I} f(x) - \inf_{y\in I} f(y)\right).$$

Now, fix $\epsilon > 0$. Since f is integrable, there exists a partition $\mathcal{P}$ such that:

$$U(f,\mathcal{P}) - L(f,\mathcal{P}) = \sum_{j=0}^{k-1}\left[\sup_{x\in[x_j,x_{j+1}]} f(x) - \inf_{x\in[x_j,x_{j+1}]} f(x)\right]\Delta x_j < \frac{\epsilon}{1+2M}.$$

We can estimate:

$$\begin{aligned} U(f^2,\mathcal{P}) - L(f^2,\mathcal{P}) &= \sum_{j=0}^{k-1}\left[\sup_{x\in[x_j,x_{j+1}]} f(x)^2 - \inf_{x\in[x_j,x_{j+1}]} f(x)^2\right]\Delta x_j \\ &\leq 2M\sum_{j=0}^{k-1}\left[\sup_{x\in[x_j,x_{j+1}]} f(x) - \inf_{x\in[x_j,x_{j+1}]} f(x)\right]\Delta x_j \\ &< \frac{2M}{1+2M}\epsilon < \epsilon. \end{aligned}$$

We conclude that f^2 is integrable.

iii) Note that:

$$fg = \frac{1}{4}\left[(f+g)^2 - (f-g)^2\right]$$

Applying previous results, we can see that the right-hand side is integrable, hence so is the left-hand side.

□

Exercise 4.4. Let $A \subset [-1,1]$ be a *finite* set, and let $\chi_A : [-1,1] \to \{0,1\}$ be the *characteristic function* of A. That is:

$$\chi_A(x) = \begin{cases} 1 & x \in A, \\ 0 & x \notin A. \end{cases}$$

Show that χ_A is integrable, and:

$$\int_{-1}^{1} \chi_A(x)dx = 0.$$

Exercise 4.5. Suppose that $a < c < b$ and suppose $f : [a,b] \to \mathbb{R}$ is a bounded function. Let $\mathcal{P}$ be a partition of $[a,b]$ of the form:

$$\mathcal{P} = (a, x_1, \ldots, x_{l-1}, x_l = c, x_{l+1}, \ldots, x_{k-1}, b).$$

and define $\mathcal{P}_L$ and $\mathcal{P}_R$ to be partitions of $[a,c]$ and $[c,b]$ respectively, given by:

$$\mathcal{P}_L = (a, x_1, \ldots, x_{l-1}, c), \qquad \mathcal{P}_R = (c, x_{l+1}, \ldots, x_{k-1}, b).$$

Show that:

$$\begin{aligned} L(f,\mathcal{P}) &= L(f|_{[a,c]}, \mathcal{P}_L) + L(f|_{[c,b]}, \mathcal{P}_R), \\ U(f,\mathcal{P}) &= U(f|_{[a,c]}, \mathcal{P}_L) + U(f|_{[c,b]}, \mathcal{P}_R). \end{aligned}$$

Exercise 4.6. Suppose $f : [a,b] \to \mathbb{R}$, $g : [a,b] \to \mathbb{R}$ are integrable.

a) Show that:

$$(b-a) \inf_{x \in [a,b]} f(x) \le \int_a^b f(x)dx \le (b-a) \sup_{x \in [a,b]} f(x).$$

b) Establish the estimate:

$$\left| \int_a^b f(x)dx \right| \le (b-a) \sup_{x \in [a,b]} |f(x)|$$

c) Show that if $0 \le f(x)$ for all $x \in [a,b]$ then:

$$0 \le \int_a^b f(x)dx$$

d) Show that if $f(x) \le g(x)$ for all $x \in [a,b]$ then:

$$\int_a^b f(x)dx \le \int_a^b g(x)dx$$

e) Prove that:

$$\left| \int_a^b f(x)dx \right| \le \int_a^b |f(x)|\, dx$$

Exercise 4.7 (*)**.** Consider Thomae's function $f : [0,1] \to \mathbb{R}$:

$$f(x) = \begin{cases} 1 & x = 0, \\ \frac{1}{q} & x = \frac{p}{q} \in \mathbb{Q}, \text{ where hcf}(p,q) = 1, \\ 0 & x \notin \mathbb{Q}. \end{cases}$$

Show that f is integrable.

2.1.4 Integrability of continuous functions

So far, we've only positively established that constant functions are integrable. Now we shall establish the important result that all continuous functions are integrable. This means that the set of all integrable functions is big enough to be interesting.

Theorem 2.9. *Suppose that $f : [a,b] \to \mathbb{R}$ is continuous on $[a,b]$. Then f is integrable.*

Proof. Fix $\epsilon > 0$. Since f is continuous on $[a,b]$, it is uniformly continuous. Thus there exists $\delta > 0$ such that for all $x, y \in [a,b]$ with $|x-y| < \delta$ we have $|f(x) - f(y)| < \frac{1}{2(b-a)}\epsilon$. Pick a partition $\mathcal{P} = (x_0, \ldots, x_k)$ such that $\Delta x_j < \delta$ for all $j = 0, \ldots k-1$. We can do this by (for example) setting:

$$x_j = a + \frac{b-a}{k} j, \qquad j = 0, \ldots, k,$$

where k is chosen sufficiently large that $(b-a) < k\delta$. Now, by construction, for $x, y \in [x_j, x_{j+1}]$ we have:

$$|f(x) - f(y)| < \frac{1}{2(b-a)}\epsilon,$$

which implies that:

$$\sup_{x \in [x_j, x_{j+1}]} f(x) - \inf_{x \in [x_j, x_{j+1}]} f(x) \leq \frac{1}{2(b-a)}\epsilon < \frac{1}{b-a}\epsilon.$$

Now, we can estimate:

$$\begin{aligned} U(f,\mathcal{P}) - L(f,\mathcal{P}) &= \sum_{j=0}^{k-1} \left[\sup_{x \in [x_j, x_{j+1}]} f(x) - \inf_{x \in [x_j, x_{j+1}]} f(x) \right] \Delta x_j \\ &< \sum_{j=0}^{k-1} \frac{\epsilon}{b-a}(x_{j+1} - x_j) = \epsilon, \end{aligned}$$

as we have a telescoping sum. Thus for any $\epsilon > 0$, we have found a partition $\mathcal{P}$ such that:

$$U(f,\mathcal{P}) - L(f,\mathcal{P}) < \epsilon,$$

which implies f is integrable. □

Remark. Note that for the proof to go through it is in fact sufficient for f to be uniformly continuous on (a,b). In view of Example 2.2 we can always modify f at the endpoints without affecting the integrability or the value of the integral.

Exercise 5.1. Suppose that $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$, and satisfies:

$$\int_a^b f(t)dt = 0, \qquad 0 \leq f(x), \text{ for all } x \in [a, b].$$

a) Show that $f(x) = 0$, for all $x \in [a, b]$.

b) Does the result hold if we do not assume f is continuous?

2.1.5 Riemann sums

While we've introduced a definition of what it means for a function to be integrable, and we've defined the integral, in practice it is useful to introduce a slightly different approach to find the integral, the Riemann sum. We will need to first refine slightly our criterion for integrability.

Lemma 2.10. *Suppose $f : [a, b] \to \mathbb{R}$ is integrable. Then given $\epsilon > 0$, there exists $\delta > 0$ such that for any partition $\mathcal{P}$ satisfying* $\operatorname{mesh}\mathcal{P} < \delta$ *we have:*

$$U(f, \mathcal{P}) - L(f, \mathcal{P}) < \epsilon.$$

(*) Proof. Since f is integrable, we have $|f(x)| \leq M$ for some M, and we know that there exists a partition $\mathcal{Q} = (y_0, \dots, y_m)$ such that:

$$U(f, \mathcal{Q}) - L(f, \mathcal{Q}) < \frac{\epsilon}{4}.$$

Suppose that $\mathcal{P}$ is any partition satisfying mesh $\mathcal{P} < \delta$ and consider $\mathcal{R}$ to be the common refinement of $\mathcal{P}$, $\mathcal{Q}$. Since a, b are common endpoints, there are at most $m - 1$ endpoints of $\mathcal{Q}$ that are distinct from the endpoints of $\mathcal{P}$. Therefore, at most $m - 1$ intervals in $\mathcal{P}$ contain endpoints in $\mathcal{Q}$. Those intervals of $\mathcal{P}$ which do not contain an endpoint in $\mathcal{Q}$ are left unchanged in $\mathcal{R}$. Let us consider:

$$L(f, \mathcal{R}) - L(f, \mathcal{P})$$

Those terms in the sum involving intervals which are common to $\mathcal{P}$, $\mathcal{R}$ will cancel. There will remain at most $3(m - 1)$ contributions to consider. Each interval has length at most δ, and we have $|f(x)| \leq M$, so that the contribution to the sums from these intervals is at most $3(m - 1)\delta M$, so that we find:

$$L(f, \mathcal{R}) - L(f, \mathcal{P}) \leq 3(m - 1)\delta M < \frac{\epsilon}{4},$$

where we take $\delta = \frac{\epsilon}{20m(M+1)}$. An identical argument leads to the estimate:

$$U(f, \mathcal{P}) - U(f, \mathcal{R}) < \frac{\epsilon}{4}.$$

Now, since $\mathcal{R}$ is a refinement of $\mathcal{Q}$, we have:

$$L(f, \mathcal{P}) > L(f, \mathcal{R}) - \frac{\epsilon}{4} > L(f, \mathcal{Q}) - \frac{\epsilon}{4} > U(f, \mathcal{Q}) - \frac{\epsilon}{2},$$

and similarly

$$U(f,\mathcal{P}) < U(f,\mathcal{R}) + \frac{\epsilon}{4} < U(f,\mathcal{Q}) + \frac{\epsilon}{4} < L(f,\mathcal{Q}) + \frac{\epsilon}{2},$$

so that we have:

$$U(f,\mathcal{P}) - L(f,\mathcal{P}) - \epsilon < L(f,\mathcal{Q}) - U(f,\mathcal{Q}) \leq 0$$

which gives us that:

$$U(f,\mathcal{P}) - L(f,\mathcal{P}) < \epsilon.$$

for any $\mathcal{P}$ with mesh $\mathcal{P} < \delta$. □

Definition 2.2. A *tagged partition* $(\mathcal{P},\tau)$ is a partition $\mathcal{P} = (x_0 \ldots, x_k)$, together with a choice of k points $\tau = (t_0, \ldots, t_{k-1})$ such that $t_j \in [x_j, x_{j+1}]$ for $j = 0, \ldots, k-1$. The *mesh* of the tagged partition is again the length of the longest subinterval, $\max_{j=0,\ldots,k-1} \Delta x_j$.

If $f : [a,b] \to \mathbb{R}$ is an integrable function, we introduce the Riemann sum of f with respect to the tagged partition:

$$R(f,\mathcal{P},\tau) = \sum_{j=0}^{k-1} f(t_j)\Delta x_j$$

Clearly, since for $t \in [x_j, x_{j+1}]$ we have:

$$m_j \leq f(t) \leq M_j$$

we deduce:

$$L(f,\mathcal{P}) \leq R(f,\mathcal{P},\tau) \leq U(f,\mathcal{P}).$$

Theorem 2.11. *Let $f : [a,b] \to \mathbb{R}$ be an integrable function. Suppose $((\mathcal{P}_l,\tau_l))_{l=0}^{\infty}$ is a sequence of tagged partitions with mesh $\mathcal{P}_l \to 0$ as $l \to \infty$. Then we have:*

$$R(f,\mathcal{P}_l,\tau_l) \to \int_a^b f(x)dx.$$

Proof. Fix $\epsilon > 0$. By Lemma 2.10 there exists $\delta > 0$ such that for any partition $\mathcal{P}$ satisfying mesh $\mathcal{P} < \delta$ we have:

$$U(f,\mathcal{P}) - L(f,\mathcal{P}) < \frac{\epsilon}{2}.$$

Since mesh $\mathcal{P}_l \to 0$ as $l \to \infty$, there exists $N \in \mathbb{N}$ such that for all $l \geq N$ we have mesh $\mathcal{P}_l < \delta$, so we have:

$$\int_a^b f(x) - \frac{\epsilon}{2} < L(f,\mathcal{P}_l) \leq R(f,\mathcal{P}_l,\tau_l) \leq U(f,\mathcal{P}_l) < \int_a^b f(x) + \frac{\epsilon}{2},$$

which implies:

$$\left| R(f,\mathcal{P}_l,\tau_l) - \int_a^b f(x) \right| < \epsilon.$$

□

Example 2.4. Let us use the Riemann sum approach to calculate:

$$\int_0^1 x dx.$$

We know that $f : x \mapsto x$ is integrable, because it is continuous. We will take as our sequence of tagged partitions the uniform partitions with tags at the left endpoint:

$$\mathcal{P}_l = \left(0, \frac{1}{l}, \frac{2}{l}, \ldots, \frac{l-1}{l}, 1\right), \qquad \mathcal{T}_l = \left(0, \frac{1}{l}, \frac{2}{l}, \ldots, \frac{l-1}{l}\right)$$

The Riemann sum is:

$$R(f, \mathcal{P}_l, \mathcal{T}_l) = \sum_{j=0}^{l-1} \frac{j}{l} \times \frac{1}{l} = \frac{1}{l^2} \sum_{j=0}^{l-1} j = \frac{1}{l^2} \times \frac{l(l-1)}{2} = \frac{1}{2} - \frac{1}{2l^2}.$$

We have that:

$$R(f, \mathcal{P}_l, \mathcal{T}_l) \to \frac{1}{2} = \int_0^1 x dx,$$

as $l \to \infty$, which agrees with what we would expect from considering the area under the graph.

Exercise 5.2. By choosing a suitable sequence of partitions to approximate the integral by a Riemann sum, show that:

$$\int_0^1 x^2 dx = \frac{1}{3}.$$

Exercise 5.3. By rewriting the left-hand side as a Riemann sum, show that:

$$\lim_{k \to \infty} \sum_{j=0}^{k-1} \frac{k}{j^2 + k^2} = \int_0^1 \frac{1}{1 + x^2} dx.$$

2.2 The fundamental theorem of calculus

Of course, when it comes to actually calculating integrals, you know that in practise the best approach is to find the anti-derivative of the integrand. The fact that this procedure works and gives the same answer as finding the area under the graph is non-trivial, and goes by the name of the *Fundamental Theorem of Calculus.*

Lemma 2.12. *Suppose that $f : [a, b] \to \mathbb{R}$ is integrable, and define the function $F : [a, b] \to \mathbb{R}$ by:*

$$F(x) = \int_a^x f(t) dt.$$

Then F is uniformly continuous on $[a, b]$.

Proof. Since f is integrable, by Theorem 2.6 the function F is well defined. As f is bounded, there exists M such that $|f(x)| \leq M$. For $x, y \in [a, b]$ we can estimate by Exercise 4.6 *b*):

$$|F(x) - F(y)| = \left| \int_x^y f(t)dt \right| \leq |x - y| \, M$$

so if we take $|x - y| < \frac{\epsilon}{1+M}$, we have $|F(x) - F(y)| < \epsilon$, which is the statement that $F : [a, b] \to \mathbb{R}$ is uniformly continuous. □

Theorem 2.13 (Fundamental Theorem of Calculus)**.** *Suppose that* $f : [a, b] \to \mathbb{R}$ *is continuous, and define the function* $F : [a, b] \to \mathbb{R}$ *by:*

$$F(x) = \int_a^x f(t)dt.$$

Then F *is differentiable on* (a, b)*, and:*

$$F'(x) = f(x).$$

Proof. First note the following two facts, which hold for $x, y \in [a, b]$ with $x \neq y$:

$$F(y) - F(x) = \int_x^y f(t)dt, \qquad \frac{1}{y - x} \int_x^y 1dt = 1.$$

These results follow from Theorem 2.6 and Example 2.1 respectively.

Now fix $\epsilon > 0$ and $x \in (a, b)$. Since f is continuous at x, there exists $\delta > 0$ such that if $|y - x| < \delta$ we have $|f(x) - f(y)| < \epsilon$. Suppose $y \neq x$ satisfies this condition. Let us estimate:

$$\begin{aligned} \left| \frac{F(y) - F(x)}{y - x} - f(x) \right| &= \left| \frac{1}{y - x} \int_x^y f(t)dt - f(x) \right| \\ &= \left| \frac{1}{y - x} \int_x^y \left(f(t) - f(x)\right) dt \right| \end{aligned}$$

Now, for t between x and y, we know that $|x - t| < \delta$, so $|f(t) - f(x)| < \epsilon$, and we can further estimate:

$$\begin{aligned} \left| \frac{1}{y - x} \int_x^y \left(f(t) - f(x)\right) dt \right| &= \frac{1}{|y - x|} \left| \int_x^y f(t) - f(x)dt \right| \\ &\leq \frac{1}{|y - x|} |y - x| \sup_{t \in [x,y]} |f(t) - f(x)| < \epsilon. \end{aligned}$$

This is precisely the condition that:

$$\lim_{y \to x} \frac{F(y) - F(x)}{y - x} = f(x).$$

□

Remark. Again, it is actually sufficient for $f : (a, b) \to \mathbb{R}$ to be uniformly continuous, rather than requiring continuity on $[a, b]$.

An immediate corollary of this result is the following:

Corollary 2.14. *Suppose that $f : [a, b] \to \mathbb{R}$ is continuous and that $F : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$, differentiable on (a, b) and satisfies $F'(x) = f(x)$ for all $x \in (a, b)$. Then:*

$$\int_a^b f(t)dt = F(b) - F(a).$$

Proof. Define the function $G : [a, b] \to \mathbb{R}$ by:

$$G(x) = \int_a^x f(t)dt.$$

By the previous Theorem we have that G is continuous on $[a, b]$, differentiable on (a, b) and $G'(x) = F'(x)$ for $x \in (a, b)$. By the Mean Value Theorem we have that:

$$F(x) = G(x) + C,$$

for some constant $C \in \mathbb{R}$ (See Exercise A.6). We calculate:

$$\begin{aligned} F(b) - F(a) &= G(b) - G(a) \\ &= \int_a^b f(t)dt - \int_a^a f(t)dt \\ &= \int_a^b f(t)dt. \end{aligned}$$

□

This justifies the usual calculations that one does when computing integrals: it is enough to find an anti-derivative and then evaluate it at the endpoints of the integral.

Exercise 5.4. Define $f : [-1, 1] \to \mathbb{R}$ by:

$$f(x) := \begin{cases} 0 & x \leq 0, \\ 1 & x > 0. \end{cases}$$

a) Show that f is integrable, and compute:

$$F(x) = \int_{-1}^x f(t)dt$$

b) Show that $F : [-1, 1] \to \mathbb{R}$ is *not* differentiable at $x = 0$.

Exercise 5.5 (*)**.** Suppose $f : [a, b] \to \mathbb{R}$ is integrable and that $F : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$, differentiable on (a, b) and satisfies $F'(x) = f(x)$ for all $x \in (a, b)$.

a) Let $\mathcal{P} = (x_0, \ldots, x_k)$ be a partition. Show that for each subinterval $[x_j, x_{j+1}]$ with $j = 0, \ldots, k-1$ there exists $t_j \in [x_j, x_{j+1}]$ such that:

$$F(x_{j+1}) - F(x_j) = f(t_j)\Delta x_j.$$

b) Show that for the tagged partition $(\mathcal{P}, \tau)$ with $\tau = (t_0, \ldots, t_{k-1})$ we have:

$$R(f, \mathcal{P}, \tau) = F(b) - F(a).$$

c) Deduce that the assumption in Corollary 2.14 that f is continuous may be weakened to integrability.

2.2.1 Change of variables formula

We can use the Fundamental Theorem of Calculus to prove a useful result, sometimes called the change of variables formula or substitution formula.

Theorem 2.15. *Let $I_1 := [\alpha, \beta], I_2 := [\gamma, \delta]$ be finite intervals. Suppose that $\phi : I_1 \to (\gamma, \delta)$ is a differentiable function with ϕ' continuous on (α, β). Suppose that $f : I_2 \to \mathbb{R}$ is continuous. Then if $[a, b] \subset (\alpha, \beta)$:*

$$\int_{\phi(a)}^{\phi(b)} f(t)dt = \int_a^b f(\phi(t))\phi'(t)dt.$$

Proof. Since $f(t)$ is continuous, by Theorem 2.13 there exists $F : [\gamma, \delta] \to \mathbb{R}$ with $F'(x) = f(x)$ for all $x \in (\gamma, \delta)$. By the chain rule, we have that the function $F \circ \phi : [\alpha, \beta] \to \mathbb{R}$ is differentiable on (α, β), with:

$$(F \circ \phi)'(x) = F'(\phi(x))\phi'(x), \qquad \text{for all } x \in (\alpha, \beta).$$

We now apply Corollary 2.14 to find:

$$\begin{aligned}\int_a^b f(\phi(t))\phi'(t)dt &= \int_a^b (F \circ \phi)'(t)dt \\ &= (F \circ \phi)(b) - (F \circ \phi)(a) = F(\phi(b)) - F(\phi(a)) \\ &= \int_{\phi(a)}^{\phi(b)} f(t)dt.\end{aligned}$$

□

The slight subtlety regarding the domains of definition in the statement of the theorem above is due to the fact that we need to ensure that ϕ' is continuous on the closed interval $[a, b]$ in order to conclude that the product $f(\phi(t))\phi'(t)$ is integrable on this interval, and to allow us to invoke the Fundamental Theorem of Calculus in the form that we have proved it.

Exercise 5.6. a) By making the change of variables given by $\phi(t) = \frac{1}{t}$, show that:

$$\int_a^b \frac{1}{1+t^2}dt = \int_{1/b}^{1/a} \frac{1}{1+t^2}dt,$$

for $0 < a < b$.

b) Setting $a = -1, b = 1$ in the above formula, we obtain:

$$\int_{-1}^{1} \frac{1}{1+x^2} dx = -\int_{-1}^{1} \frac{1}{1+x^2} dx.$$

What has gone wrong?

Exercise 5.7 (*)**.** Let $I_1 := [\alpha, \beta], I_2 := [\gamma, \delta]$ be finite intervals. Suppose that $\phi : I_1 \to I_2$ is continuous on I_1, and differentiable on (α, β) with ϕ' integrable on I_1. Suppose that $f : I_2 \to \mathbb{R}$ is continuous. Show that:

$$\int_{\phi(\alpha)}^{\phi(\beta)} f(t) dt = \int_{\alpha}^{\beta} f(\phi(t))\phi'(t) dt.$$

2.3 Interchanging limits

It is often the case that we wish to interchange various limits with integrals. For example, suppose that we have a sequence of integrable functions $f_j : [a, b] \to \mathbb{R}$ such that $f_j \to f$ in some appropriate sense. When can we say that:

$$\lim_{j\to\infty} \int_a^b f_j(x) dx = \int_a^b f(x) dx?$$

We have to be cautious in situations like this. Integration is a limiting operation, and experience teaches us that interchanging different limiting operations can lead to difficulties.

2.3.1 Pointwise convergence does not commute with integration

For $j \in \mathbb{N}$, consider the function $f_j : [0, 1] \to \mathbb{R}$ given by:

$$f_j(x) = \begin{cases} 2^{2j}x & 0 \leq x < 2^{-j}, \\ 2^j \left(2 - 2^j x\right) & 2^{-j} \leq x < 2^{-(j-1)} \\ 0 & x \geq 2^{-(j-1)}, \end{cases}$$

For each j, the graph of this function is an isosceles triangle, with base of length $2^{-(j-1)}$ and height 2^j. Therefore, we have:

$$\int_0^1 f_j(x) dx = 1.$$

However, we have that for all $x \in [0, 1]$:

$$\lim_{j\to\infty} f_j(x) = 0.$$

This is an example of a sequence of integrable functions for which $f_j(x) \to f(x)$ for all $x \in [a, b]$, but:

$$\lim_{j\to\infty} \int_a^b f_j(x) dx \neq \int_a^b f(x) dx.$$

This tells us that in general, pointwise convergence does *not* commute with integration.

Exercise 5.8. By modifying the example of Section 2.3.1, show that there exists a family $(f_j)_{j=0}^{\infty}$ of integrable functions $f_j : [0,1] \to \mathbb{R}$ such that:

$$f_j(x) \to 0 \qquad \text{for all } x \in [0,1],$$

while:

$$\int_0^1 f_j(x)dx \to \infty.$$

2.3.2 Uniform convergence does commute with integration

It may or may not surprise you to discover that the situation is much better if instead of pointwise convergence we instead consider *uniform* convergence. From our previous work we know that this would already be enough to rule out the bad behaviour inherent in our previous example.

Theorem 2.16. *Suppose that $(f_i)_{i=0}^{\infty}$ is a sequence of integrable functions $f : [a,b] \to \mathbb{R}$ such that $f_i \to f$ uniformly on $[a,b]$. Then f is integrable, and*

$$\lim_{i\to\infty} \int_a^b f_i(x)dx = \int_a^b f(x)dx.$$

(*) Proof. 1. Fix $\epsilon > 0$. By the uniform convergence of $(f_i)_{i=0}^{\infty}$ to f, there exists $N \in \mathbb{N}$ such that for any $i \geq N$ we have:

$$|f_i(x) - f(x)| < \frac{\epsilon}{2(b-a)},$$

for all $x \in [a,b]$. Henceforth assume $i \geq N$. Now, if $I \subset [a,b]$ is any subinterval, we have that:

$$\begin{aligned}\sup_{x\in I} f(x) &= \sup_{x\in I} (f(x) - f_i(x) + f_i(x)) \\ &\leq \sup_{x\in I} (f(x) - f_i(x)) + \sup_{x\in I} f_i(x) \\ &\leq \sup_{x\in I} f_i(x) + \frac{\epsilon}{2(b-a)}.\end{aligned}$$

From here, we conclude that for any partition $\mathcal{P}$:

$$\begin{aligned}U(f,\mathcal{P}) &= \sum_{j=0}^{k-1} \sup_{x\in[x_j,x_{j+1}]} f(x)\Delta x_j \\ &\leq \sum_{j=0}^{k-1} \left[\sup_{x\in[x_j,x_{j+1}]} f_i(x) + \frac{\epsilon}{2(b-a)} \right] \Delta x_j \\ &= U(f_i,\mathcal{P}) + \frac{\epsilon}{2(b-a)} \sum_{j=0}^{k-1} (x_{j+1} - x_j) \\ &= U(f_i,\mathcal{P}) + \frac{\epsilon}{2}\end{aligned}$$

Taking the infimum over $\mathcal{P}$ and using that f_i is integrable, we conclude:

$$\overline{\int_a^b} f(x)dx \leq \int_a^b f_i(x)dx + \frac{\epsilon}{2} < \int_a^b f_i(x)dx + \epsilon.$$

2. Similarly, if $I \subset [a,b]$ is any interval, we have that:

$$\begin{aligned}\inf_{x\in I} f(x) &= \inf_{x\in I} (f(x) - f_i(x) + f_i(x)) \\ &\geq \inf_{x\in I} (f(x) - f_i(x)) + \inf_{x\in I} f_i(x) \\ &\geq \inf_{x\in I} f_i(x) - \frac{\epsilon}{2(b-a)}.\end{aligned}$$

so that:

$$L(f,\mathcal{P}) \geq L(f_i,\mathcal{P}) - \frac{\epsilon}{2}.$$

and taking the supremum, we have:

$$\underline{\int_a^b} f(x)dx \geq \int_a^b f_i(x)dx - \frac{\epsilon}{2} > \int_a^b f_i(x)dx - \epsilon.$$

3. Putting this together, we have:

$$\int_a^b f_i(x)dx - \epsilon < \underline{\int_a^b} f(x)dx \leq \overline{\int_a^b} f(x)dx < \int_a^b f_i(x)dx + \epsilon$$

from which we first conclude:

$$0 \leq \overline{\int_a^b} f(x)dx - \underline{\int_a^b} f(x)dx < 2\epsilon.$$

Since ϵ is arbitrary, we must have:

$$\overline{\int_a^b} f(x)dx = \underline{\int_a^b} f(x)dx = \int_a^b f(x)dx$$

and f is integrable. Moreover, we have:

$$\left| \int_a^b f_i(x)dx - \int_a^b f(x)dx \right| < \epsilon,$$

so indeed:

$$\lim_{i\to\infty} \int_a^b f_i(x)dx = \int_a^b f(x)dx. \qquad \square$$

In Theorem 1.3, we established that if $(f_i)_{i=0}^{\infty}$ is a sequence of uniformly continuous maps $f_i : (a,b) \to \mathbb{R}$ such that $f_i \to f$ uniformly, then $f : (a,b) \to \mathbb{R}$ is uniformly continuous. We are now in a position to extend this to the situation where the f_i's are *differentiable* functions, provided the derivatives also converge uniformly.

Theorem 2.17. *Suppose $(f_i)_{i=0}^{\infty}$ is a sequence of uniformly continuous, differentiable maps $f_i : (a,b) \to \mathbb{R}$ such that each $f_i' : (a,b) \to \mathbb{R}$ is uniformly continuous. Suppose further that:*

$$f_i \to f, \qquad f_i' \to g$$

uniformly as $i \to \infty$. Then f is differentiable on (a,b) and $f' = g$.

Proof. Let us pick $c \in (a,b)$. By Corollary 2.14, we may write:

$$f_i(x) - f_i(c) = \int_c^x f_i'(t)dt.$$

Now, since f_i converges uniformly (hence pointwise), as $i \to \infty$ we have:

$$f_i(x) - f_i(c) \to f(x) - f(c).$$

Further, since the convergence of the f_i' is uniform we know that g is continuous, and by Theorem 2.16 we have:

$$\lim_{i\to\infty} \int_c^x f_i'(x)dx = \int_c^x g(t)dt,$$

so that:

$$f(x) = f(c) + \int_c^x g(t)dt.$$

Applying Theorem 2.13, we have that f is differentiable on (a,b) with $f' = g$. □

It is clear that we can iterate this result to show that if $f_i^{(m)} \to g_m$ for $m = 0, \ldots, k$, then $f := g_0$ will be $k-$times differentiable and $f^{(m)} = g_m$. Combining these results with the Weierstrass $M-$test we have:

Corollary 2.18. *Suppose that $(a_i)_{i=0}^{\infty}$ is a sequence of real numbers $a_i \in \mathbb{R}$, and suppose there exists $r > 0$ such that the sum*

$$\sum_{i=0}^{\infty} |a_i|\, r^i$$

converges. Then the power series:

$$f(x) = \sum_{i=0}^{\infty} a_i x^i$$

converges uniformly to a smooth function on the interval $(-r, r)$, and we can differentiate under the sum.

Proof. We already know from Corollary 1.8 that the sum converges uniformly. We need to show that $f(x)$ is smooth. Fix $k \in \mathbb{N}$ and consider the partial sums

$$S_j(x) = \sum_{i=0}^{j} a_i x^i.$$

for $j \geq k$. These are smooth on $(-r, r)$ with:

$$S_j^{(m)}(x) = \sum_{i=k}^{j} \frac{i!}{(i-m)!} a_i x^{i-m},$$

for $m \leq k$. Let us pick ρ with $0 < \rho < r$. If $x \in [-\rho, \rho]$, then we have:

$$\left| \frac{i!}{(i-m)!} a_i x^{i-m} \right| \leq \frac{i!}{(i-m)!} |a_i| \rho^{i-m} \leq \rho^{-m} \left[\left(\frac{\rho}{r} \right)^i \frac{i!}{(i-m)!} \right] |a_i| r^i = \lambda_i |a_i| r^i,$$

where

$$\lambda_i = \rho^{-m} \left(\frac{\rho}{r} \right)^i \frac{i!}{(i-m)!} \to 0$$

as $i \to \infty$, since $\rho/r < 1$, and $i!/(i-m)!$ is a polynomial in i, so eventually is beaten by the exponential that it's multiplied by. By the comparison test, we have that

$$\sum_{i=0}^{\infty} \lambda_i |a_i| r^i$$

converges, so by the Weierstrass $M-$test, we deduce that $S_j^{(m)}$ converges uniformly on $(-\rho, \rho)$, thus by the previous result, we have that $f(x)$ is $k-$times differentiable on $(-\rho, \rho)$, and moreover:

$$f^{(m)}(x) = \sum_{i=0}^{\infty} \frac{i!}{(i-m)!} a_i x^{i-m}$$

for $m \leq k$ and $x \in [-\rho, \rho]$. Since ρ<r and k were arbitrary, the result follows. □

2.3.3 Improper integrals

It is often the case that the integrals one wishes to consider do not fall into the class allowed by our definition. Either the integrand, or the range of integration may be unbounded. Typically the best approach to handle these problems is to consider limits of problems that we do understand. We will give here a few examples of the situations that can occur, and urge caution in working with them.

Example 2.5. Consider the integral:

$$\int_0^1 \frac{1}{\sqrt{x}} dx.$$

Since $x \mapsto x^{-\frac{1}{2}}$ is unbounded on the interval $[0, 1]$, it is not integrable on this interval. However, for any $\epsilon > 0$, it is integrable on $[\epsilon, 1]$. We are perfectly justified in calculating:

$$\int_\epsilon^1 \frac{1}{\sqrt{x}} dx = 2\left(1 - \sqrt{\epsilon}\right).$$

We therefore will define the *improper integral*:

$$\int_0^1 \frac{1}{\sqrt{x}}dx := \lim_{\epsilon\to 0}\int_\epsilon^1 \frac{1}{\sqrt{x}}dx = 2.$$

Of course, this sort of limit may or may not exist in general. For example:

$$\int_\epsilon^1 \frac{1}{x}dx = -\log\epsilon \to \infty \qquad \text{as } \epsilon \to 0.$$

In this case, we would say that the integral

$$\int_0^1 \frac{1}{x}dx$$

does not exist, or fails to converge.

Example 2.6. We can also consider integrating a function which is continuous and diverges at a single point. For example:

$$\int_{-1}^1 \frac{1}{\sqrt{|x|}}dx.$$

In this case we split up the integral into two regions as:

$$\int_{-1}^1 \frac{1}{\sqrt{|x|}}dx = \int_{-1}^0 \frac{1}{\sqrt{|x|}}dx + \int_0^1 \frac{1}{\sqrt{|x|}}dx,$$

and we demand that *both* the improper integrals on the right hand side exist in the sense of Example 2.5. In this case, that is so by the previous part and we have:

$$\int_{-1}^1 \frac{1}{\sqrt{|x|}}dx = 4.$$

One must be careful here. This is a stronger condition than:

$$\lim_{\epsilon\to 0}\left(\int_{-1}^{-\epsilon} \frac{1}{\sqrt{|x|}}dx + \int_\epsilon^1 \frac{1}{\sqrt{|x|}}dx\right)$$

existing. For example, it is true that:

$$\lim_{\epsilon\to 0}\left(\int_{-1}^{-\epsilon} \frac{1}{x}dx + \int_\epsilon^1 \frac{1}{x}dx\right) = \lim_{\epsilon\to 0} 0 = 0, \tag{2.2}$$

however since the limits:

$$\lim_{\epsilon\to 0}\left(\int_{-1}^{-\epsilon} \frac{1}{x}dx\right), \qquad \lim_{\epsilon\to 0}\left(\int_\epsilon^1 \frac{1}{x}dx\right),$$

do not exist, we say that the integral:

$$\int_{-1}^1 \frac{1}{x}dx \tag{2.3}$$

does not converge. The expression (2.2) is sometimes called the Cauchy Principal Value of the integral (2.3). Such quantities often arise naturally in complex analysis.

Example 2.7. One also wishes to consider integrating over an infinite range, that is one wishes to give meaning to expressions such as:

$$\int_0^\infty f(x)dx.$$

This again is not covered by our definition of integration. We define the improper integral to be:

$$\int_a^\infty f(x)dx := \lim_{R\to\infty}\int_a^R f(x)dx,$$

if the limit exists. For example:

$$\int_1^\infty \frac{1}{x^2}dx = \lim_{R\to\infty}\int_a^R \frac{1}{x^2}dx = \lim_{R\to\infty}\left[1-\frac{1}{R}\right] = 1.$$

For a doubly infinite range, we define:

$$\int_{-\infty}^\infty f(x)dx = \int_{-\infty}^0 f(x)dx + \int_0^\infty f(x)dx,$$

provided both improper integrals on the right hand side exist. Again, note that in general the existence of the limit:

$$\lim_{r\to\infty}\int_{-R}^R f(x)dx$$

is a weaker condition than the condition that both $\int_{-\infty}^0 f(x)dx$, $\int_0^\infty f(x)dx$ exist.

In all cases of improper integrals one has to be very cautious in applying the results that we have established for the standard integral, as these may or may not be true depending on the circumstances. The best approach is to explicitly reduce to the standard integral, and then attempt to take a limit.

Exercise 6.1. Suppose that $f : [0,\infty) \to [0,\infty)$ is a non-negative, continuous, monotone decreasing function, and let $N, M \in \mathbb{Z}$.

a) Show that:

$$\sum_{n=N}^M f(n) \geq \int_N^{M+1} f(x)dx.$$

b) Show that:

$$\sum_{n=N+1}^M f(n) \leq \int_N^M f(x)dx.$$

c) Establish the inequality:

$$\int_N^{M+1} f(x)dx \leq \sum_{n=N}^M f(n) \leq f(N) + \int_N^M f(x)dx$$

d) Show that
$$\sum_{n=N}^{\infty} f(n)$$
converges if and only if the improper integral:
$$\int_N^{\infty} f(x)dx$$
exists.

e) Show that the harmonic series:
$$\sum_{n=1}^{\infty} \frac{1}{n}$$
diverges.

Exercise 6.2. Let $n \in \mathbb{N}$ with $n \geq 2$. By choosing a suitable partition to approximate the integral:
$$\int_1^n \log x \, dx$$
by upper and lower sums, show that:
$$\sum_{k=1}^{n-1} \log k \leq n \log n - n + 1 \leq \sum_{k=2}^{n} \log k.$$
Deduce that:
$$n^n e^{-n+1} \leq n! \leq (n+1)^{n+1} e^{-n}.$$

2.4 Examples and applications

2.4.1 Complex and vector valued integrals

In some circumstances we wish to integrate a function which maps some interval $[a, b]$ to $\mathbb{C}$ or alternatively to $\mathbb{R}^m$. We can incorporate this into our existing set-up quite simply. Suppose $f : [a, b] \to \mathbb{C}$. Then we can write:
$$f(t) = u(x) + iv(x)$$
We say that f is integrable if both of the functions $u, v : [a, b] \to \mathbb{R}$ are integrable, and we write:
$$\int_a^b f(t)dt = \int_a^b u(t)dt + i \int_a^b v(t)dt.$$

In a similar way, if $f : [a, b] \to \mathbb{R}^m$, then we can write:
$$f(t) = e_1 f^1(t) + e_2 f^2(t) + \ldots + e_m f^m(t)$$
where $\{e_i\}_{i=1}^m$ is the standard basis for $\mathbb{R}^m$ and $f_i : [a, b] \to \mathbb{R}$ are real valued functions for $i = 1, \ldots, m$. We say that f is integrable if each f_i is integrable, and we define:
$$\int_a^b f(t)dt = e_1 \int_a^b f^1(t)dt + e_2 \int_a^b f^2(t)dt + \ldots + e_m \int_a^b f^m(t)dt$$

Exercise 6.3. Suppose $f : [a, b] \to \mathbb{R}^m$ is integrable.

a) Show that if $x, y \in [a, b]$:

$$\big| \, ||f(x)|| - ||f(y)|| \, \big| \leq \sum_{l=1}^{m} \left| f^l(x) - f^l(y) \right|,$$

where $f(t) = e_1 f^1(t) + e_2 f^2(t) + \cdots + e_m f^m(t)$.

b) Deduce that if I is any subinterval of $[a, b]$:

$$\sup_{x \in I} ||f(x)|| - \inf_{y \in I} ||f(y)|| \leq \sum_{l=1}^{m} \left(\sup_{x \in I} f^l(x) - \inf_{y \in I} f^l(y) \right).$$

c) Prove that $||f|| : [a, b] \to \mathbb{R}$ is integrable.

Exercise 6.4. Let $f : [a, b] \to \mathbb{R}^m$ be an integrable function, and write

$$f(t) = e_1 f^1(t) + e_2 f^2(t) + \cdots + e_m f^m(t)$$

for $f^i : [a, b] \to \mathbb{R}$.

a) For a tagged partition $(\mathcal{P}, \tau)$, show that:

$$R(f, \mathcal{P}, \tau) = e_1 R(f^1, \mathcal{P}, \tau) + e_2 R(f^2, \mathcal{P}, \tau) + \ldots + e_m R(f^m, \mathcal{P}, \tau)$$

b) Show that if $((\mathcal{P}_l, \tau_l))_{l=0}^{\infty}$ is a sequence of tagged partitions of $[a, b]$ with mesh $\mathcal{P}_l \to 0$, then

$$R(f, \mathcal{P}_l, \tau_l) \to \int_a^b f(x)dx,$$

as $l \to \infty$.

c) Show that for a tagged partition $(\mathcal{P}, \tau)$:

$$||R(f, \mathcal{P}, \tau)|| \leq R(||f|| \, , \mathcal{P}, \tau).$$

d) Deduce that:

$$\left|\left| \int_a^b f(t)dt \right|\right| \leq \int_a^b ||f(t)|| \, dt.$$

Note that an identical argument shows that if $f : [a, b] \to \mathbb{C}$ then:

$$\left| \int_a^b f(t)dt \right| \leq \int_a^b |f(t)| \, dt.$$

Exercise 6.5 (*).

Suppose $f, g, h : [a, b] \to \mathbb{R}^m$ are integrable. We define:

$$(f, g) := \int_a^b \langle f(t), g(t) \rangle \, dt, \qquad ||f||_{L^2} = \sqrt{(f, f)}.$$

a) Show that:

$$(f, g) = (g, f), \qquad (f + h, g) = (f, g) + (h, g), \qquad (\lambda f, g) = \lambda(f, g),$$

where $\lambda \in \mathbb{R}$.

b) For $t \in \mathbb{R}$, show that:

$$||f + tg||_{L^2}^2 = ||f||_{L^2}^2 + 2t(f, g) + t^2 \, ||g||_{L^2}^2 \geq 0. \tag{2.4}$$

c) By thinking of (2.4) as a quadratic in t and considering its possible roots, deduce the *Cauchy-Schwartz* inequality:

$$|(f, g)| \leq ||f||_{L^2} \, ||g||_{L^2} \, .$$

When does equality hold?

d) Deduce that:

$$||f + g||_{L^2} \leq ||f||_{L^2} + ||g||_{L^2} \, .$$

2.4.2 (*)The spaces $C^k([a, b])$

Recall that in Section 1.2.3, we defined the space $C^0([a, b])$ to consist of all continuous functions $f : [a, b] \to \mathbb{R}$, equipped with the norm function:

$$||f||_{C^0} = \max_{x \in [a,b]} |f(x)| \, .$$

We showed that this satisfied certain conditions analogous to those of the norm in Euclidean space. We also defined Cauchy sequences with respect to this norm and showed that they converge uniformly to a continuous function. We're now in a position to upgrade this definition to consider functions which have a certain amount of differentiability.

Definition 2.3. $C^k([a, b])$ is the set of all $k-$times differentiable functions $f : [a, b] \to \mathbb{R}$ such that f is continuous on $[a, b]$ and $f^{(m)} : (a, b) \to \mathbb{R}$ is uniformly continuous for $m = 1, \ldots, k$. We define a function $||\cdot||_{C^k}$ by:

$$||f||_{C^k} = \sum_{m=0}^{k} \sup_{x \in (a,b)} \left| f^{(m)}(x) \right| ,$$

for all $f \in C^k([a, b])$. We call $||\cdot||_{C^k}$ the *norm* on $C^k([a, b])$.

It is not immediately obvious, but for any $f \in C^k([a, b])$, we have that $||f||_{C^k} < \infty$. This follows from the following Lemma:

Lemma 2.19. *Suppose $f : (a, b) \to \mathbb{R}$ is uniformly continuous. Then f is bounded.*

Proof. Since f is uniformly continuous, there exists $\delta > 0$ such that if $x, y \in (a, b)$ with $|x - y| < \delta$, then $|f(x) - f(y)| < 1$. Let

$$y_j = a + \frac{j}{N}(b - a), \qquad j = 1, 2, \ldots, N$$

where N is sufficiently large that $\frac{(b-a)}{N} < \delta$. Set $M = \max\{f(y_1), f(y_2), \ldots, f(y_N)\}$. For any $x \in (a, b)$ there exists some j such that $|x - y_j| < \delta$, so we conclude:

$$|f(x)| = |f(x) - f(y_j) + f(y_j)| \leq |f(x) - f(y_j)| + |f(y_j)| < 1 + M,$$

so f is bounded on (a, b). □

Lemma 2.20. *The the norm on $C^k([a, b])$ has the following properties:*

i) For all $f \in C^k([a, b])$, we have $||f||_{C^k} \geq 0$, with $||f||_{C^0} = 0$ if and only if $f \equiv 0$.

ii) For all $f \in C^k([a, b])$, if $a \in \mathbb{R}$, then $||af||_{C^0} = |a| \, ||f||_{C^k}$

iii) The norm satisfies the triangle inequality:

$$||f + g||_{C^k} \leq ||f||_{C^0} + ||g||_{C^k}, \qquad \forall \; f, g, \in C^k. \tag{2.5}$$

Proof. i) It is clear that since each term in the sum defining $||f||_{C^k}$ is the supremum of a set of non-negative numbers, it must be non-negative, thus $||f||_{C^k}$ is non-negative. If $||f||_{C^k} = 0$, then in particular:

$$\sup_{x \in [a,b]} |f(x)| = 0,$$

so that $0 \leq |f(x)| \leq 0$ and $f(x) = 0$ for all $x \in [a, b]$.

ii) If $a \in \mathbb{R}$, then:

$$||af||_{C^k} = \sum_{m=0}^{k} \sup_{x \in (a,b)} \left| a f^{(m)}(x) \right| = \sum_{m=0}^{k} \sup_{x \in (a,b)} |a| \left| f^{(m)}(x) \right| = |a| \, ||f||_{C^k}.$$

iii) We estimate:

$$\begin{aligned}
||f + g||_{C^k} &= \sum_{m=0}^{k} \sup_{x \in (a,b)} \left| f^{(m)}(x) + g^{(m)}(x) \right| \leq \sum_{m=0}^{k} \sup_{x \in (a,b)} \left(\left| f^{(m)}(x) \right| + \left| g^{(m)}(x) \right| \right) \\
&\leq \sum_{m=0}^{k} \left(\sup_{x \in (a,b)} \left| f^{(m)}(x) \right| + \sup_{x \in (a,b)} \left| g^{(m)}(x) \right| \right) \\
&= \sum_{m=0}^{k} \sup_{x \in (a,b)} \left| f^{(m)}(x) \right| + \sum_{m=0}^{k} \sup_{x \in (a,b)} \left| g^{(m)}(x) \right| = ||f||_{C^k} + ||g||_{C^k}.
\end{aligned}$$

Here we have used the triangle inequality for $|\cdot|$, as well as the fact that $\sup(a + b) \leq \sup a + \sup b$. □

The observant reader will note that the proof here is almost identical to that of the properties of the norm on $C^0([a,b])$.

Now, we can give a definition of convergence in $C^k([a,b])$ as follows.

Definition 2.4. A sequence $(f_j)_{j=0}^{\infty}$ of functions $f_j \in C^k([a,b])$ converges to $f \in C^k([a,b])$ if the following holds: for any $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that if $j \geq N$, then

$$||f_j - f||_{C^k} < \epsilon.$$

A sequence $(f_j)_{j=0}^{\infty}$ of functions $f_j \in C^k([a,b])$ is *Cauchy* if the following holds: for any $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that if $i, j \geq N$, then

$$||f_i - f_j||_{C^k} < \epsilon.$$

Notice that since $||\cdot||_{C^k}$ is a sum of positive terms, we have in particular that:

$$\left|\left|f^{(m)}\right|\right|_{C^0} < ||f||_{C^k}$$

for all $m \leq k$, so that if $f_j \to f$ in C^k, then in particular $f_j^{(m)} \to f^{(m)}$ uniformly for each m. Similarly, if $(f_j)_{j=0}^{\infty}$ is a Cauchy sequence in $C^k([a,b])$, then $\left(f_j^{(m)}\right)_{j=0}^{\infty}$ is a Cauchy sequence[3] in $C^0([a,b])$ for each $m \leq k$.

Theorem 2.21. *A sequence $(f_j)_{j=0}^{\infty}$ of functions $f_j \in C^k([a,b])$ converges in $C^k([a,b])$ if and only if it is a Cauchy sequence.*

Proof. From the observations above, we see that $(f_j)_{j=0}^{\infty}$ converges in $C^k([a,b])$ if and only if $\left(f_j^{(m)}\right)_{j=0}^{\infty}$ converges in $C^0([a,b])$ for all $m \leq k$. By Theorem 1.6 we know that $\left(f_j^{(m)}\right)_{j=0}^{\infty}$ converges in $C^0([a,b])$ if and only if it is a Cauchy sequence in $C^0([a,b])$. Finally, we know that $\left(f_j^{(m)}\right)_{j=0}^{\infty}$ is a Cauchy sequence. $C^0([a,b])$ for all $m \leq k$ if and only if $(f_j)_{j=0}^{\infty}$ is a Cauchy sequence in $C^k([a,b])$. □

Thus the space $C^k([a,b])$ has the property of *completeness* that we introduced in Section 1.2.3 and so is another example of a Banach space.

[3]Strictly speaking, we should use the result of Exercise 3.6 to argue that a uniformly continuous function on (a,b) extends to a continuous function on $[a,b]$.

Chapter 3

Differentiation

3.1 The derivative of a map $f : \mathbb{R}^n \to \mathbb{R}^m$

So far, when differentiating functions, we've restricted ourselves to the situation where the function depends only on one variable. This covers lots of situations that we're interested in, but of course we often wish to consider functions of more than one variable. In this chapter we will see how the idea of differentiation can be extended to functions which map (subsets of) $\mathbb{R}^n$ to $\mathbb{R}^m$. The basic idea will be that the derivative of a function at a point p should be the *best linear approximation* to the function at p.

3.1.1 Definition and examples

Before we think about how to define a derivative of a map in higher dimensions, let's first note some of the potential challenges. In one dimension, we say that f is differentiable at p if the limit:

$$\lim_{x \to p} \frac{f(x) - f(p)}{x - p}$$

exists. If $x, p \in \mathbb{R}^n$ and $f(x), f(p) \in \mathbb{R}^m$ then we obviously have a problem: we don't even know how to make sense of 'dividing by $x - p$', and it's not clear what sort of object we should end up with.

To try and find a way through this impasse, let's just remind ourselves how the derivative is introduced in one dimension. By approximating with successive *chords*, we consider the *tangent* to the graph of f at p (see Figure 3.1). Let us think a little about how the tangent is characterised. Any (non-vertical) straight line passing through $(p, f(p))$ will be the graph of the affine function:

$$A_\Lambda : x \mapsto \Lambda(x - p) + f(p)$$

for some $\Lambda \in \mathbb{R}$. Let's consider the difference between f and such an affine function:

$$f(x) - A_\Lambda(x) = f(x) - f(p) - \Lambda(x - p).$$

In general, from the continuity of f we know that:

$$\lim_{x \to p} [f(x) - A_\Lambda(x)] = 0 \qquad (3.1)$$

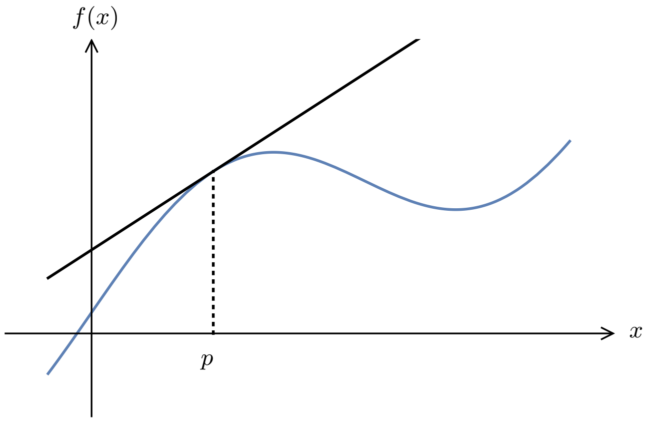

Figure 3.1 The tangent to f at p.

for any $\Lambda \in \mathbb{R}$. If f is differentiable, however, there is a particular choice of Λ that allows us to make a stronger statement. If f is differentiable, there exists a unique $\Lambda \in \mathbb{R}$ such that:

$$\lim_{x \to p} \frac{|f(x) - A_\Lambda(x)|}{|x - p|} = 0.$$

This is a stronger statement than (3.1) because it tells us that $f(x) - A_\Lambda(x)$ is going to zero *faster* than $|x - p|$ as $x \to p$. We can make this hand waving discussion a little more precise with:

Lemma 3.1. *The function $f : (a, b) \to \mathbb{R}$ is differentiable at $p \in (a, b)$ if and only if there exists $\Lambda \in \mathbb{R}$ such that if $A_\Lambda : x \mapsto \Lambda(x - p) + f(p)$ then*

$$\lim_{x \to p} \frac{|f(x) - A_\Lambda(x)|}{|x - p|} = 0.$$

Proof. We can re-write

$$\frac{|f(x) - f(p) - \Lambda(x - p)|}{|x - p|} = \left| \frac{f(x) - f(p)}{x - p} - \Lambda \right|,$$

so that:

$$\lim_{x \to p} \frac{|f(x) - A_\Lambda(x)|}{|x - p|} = 0, \qquad \Longleftrightarrow \qquad \lim_{x \to p} \frac{f(x) - f(p)}{x - p} = \Lambda,$$

which is the usual definition of differentiability. □

Now let's try thinking about a function $f : \mathbb{R}^n \to \mathbb{R}^m$. What is the appropriate generalisation of a 'straight line through $f(p)$'? It seems reasonable to propose that the correct generalisation is the affine function:

$$\begin{array}{rcccl} A_\Lambda & : & \mathbb{R}^n & \to & \mathbb{R}^m \\ & & x & \mapsto & \Lambda(x - p) + f(p), \end{array}$$

where $\Lambda \in L(\mathbb{R}^n; \mathbb{R}^m)$ is a linear map $\mathbb{R}^n \to \mathbb{R}^m$. This guides the following definition:

Definition 3.1. Suppose $\Omega \subset \mathbb{R}^n$ is open. The function $f : \Omega \to \mathbb{R}^m$ is differentiable at $p \in \Omega$ if and only if there exists $\Lambda \in L(\mathbb{R}^n; \mathbb{R}^m)$ such that if $A_\Lambda : x \mapsto \Lambda(x-p) + f(p)$ then

$$\lim_{x \to p} \frac{||f(x) - A_\Lambda(x)||}{||x-p||} = 0.$$

If this is the case, we write:

$$Df(p) := \Lambda,$$

and call $Df(p)$ the *differential* of the map f at the point p.

If we choose the canonical bases for $\mathbb{R}^n, \mathbb{R}^m$ then we can express the differential as an $m \times n$ matrix[1], which is called the *Jacobian.* For the purposes of this course, we won't make a big deal of the difference between a linear map and its matrix representation with respect to the canonical basis, so will use the words differential and Jacobian essentially indistinguishably.

It is often useful to have the following (totally equivalent) characterisation of differentiability: $f : \Omega \to \mathbb{R}$ is differentiable at $p \in \Omega$ if and only if there exists $\Lambda \in L(\mathbb{R}^n; \mathbb{R}^m)$ such that:

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

Note also that a differentiable function is necessarily continuous. From basic properties of limits, we know that:

$$\lim_{h \to 0} \frac{||f(p+h) - f(p) - \Lambda h||}{||h||} = 0 \qquad \Longrightarrow \qquad \lim_{h \to 0} ||f(p+h) - f(p) - \Lambda h|| = 0$$

which since linear functions are continuous implies:

$$0 = \lim_{h \to 0} [f(p+h) - f(p) - \Lambda h] = \lim_{h \to 0} f(h+p) - f(p).$$

Example 3.1. By Lemma 3.1 any function $f : (a,b) \to \mathbb{R}$ which is differentiable at p satisfies the conditions of 3.1 with $Df(p) = f'(p)$. Notice that a 1×1 matrix is simply a real number.

Example 3.2. Let $B \in L(\mathbb{R}^n; \mathbb{R}^m)$ be a linear map from $\mathbb{R}^n \to \mathbb{R}^m$. Then the map:

$$\begin{array}{rccc} f_B & : & \mathbb{R}^n & \to & \mathbb{R}^m \\ & & x & \mapsto & Bx, \end{array}$$

is differentiable at each $p \in \mathbb{R}^n$, and $Df(p) = B$. To see this, note that if $\Lambda = B$, then $A_\Lambda(x) = B(x-p) + Bp = Bx$, so that:

$$f_B(x) - A_\Lambda(x) = 0.$$

[1] The convention is that an $m \times n$ matrix has m *rows* and n *columns*

Example 3.3. The function:

$$\begin{array}{rcccl} x & : & \mathbb{R}^n & \to & \mathbb{R} \\ & & x & \mapsto & ||x||^2 , \end{array}$$

is differentiable at each $p \in \mathbb{R}^n$ and $Df(p)$ is the linear map:

$$\begin{array}{rcccl} Df(p) & : & \mathbb{R}^n & \to & \mathbb{R} \\ & & h & \mapsto & 2\langle p, h\rangle , \end{array}$$

To see this, let us expand:

$$f(p+h) = ||p+h||^2 = ||p||^2 + 2\langle p, h\rangle + ||h||^2 ,$$

so that:

$$\frac{||f(p+h) - f(p) - 2\langle p, h\rangle||}{||h||} = ||h|| \to 0, \qquad \text{as } h \to 0.$$

As a matrix, we have that $Df(p) = 2p^t$, so the Jacobian is a row vector for this map.

Example 3.4. Suppose $f^i : (a,b) \to \mathbb{R}$ are differentiable at $p \in (a,b)$ for $i = 1, \ldots, m$. Then the map:

$$\begin{array}{rcccl} f & : & (a,b) & \to & \mathbb{R}^m \\ & & x & \mapsto & \begin{pmatrix} f^1(x) \\ \vdots \\ f^m(x) \end{pmatrix} , \end{array}$$

is differentiable at p, and the differential $Df(p) : \mathbb{R} \to \mathbb{R}^m$ has the matrix representation:

$$Df(p) = \begin{pmatrix} {f^1}'(p) \\ \vdots \\ {f^m}'(p) \end{pmatrix} .$$

To see this, we note that:

$$f(p+h) - f(p) - \begin{pmatrix} {f^1}'(p) \\ \vdots \\ {f^m}'(p) \end{pmatrix} h = \begin{pmatrix} f^1(p+h) - f^1(p) - {f^1}'(p)h \\ \vdots \\ f^m(p+h) - f^m(p) - {f^m}'(p)h \end{pmatrix}$$

so that, using Exercise 1.1 f) ii):

$$\frac{||f(p+h) - f(p) - Df(p)[h]||}{||h||} \leq \sqrt{m} \max_{j=1,\ldots,m} \frac{\left| f^j(p+h) - f^j(p) - {f^j}'(p)h \right|}{|h|} \to 0,$$

since each f^j is differentiable at p. In this case the Jacobian is a *column* vector. Notice here that the expression $Df(p)[h]$ means the vector in $\mathbb{R}^m$ that results from applying the linear map $Df(p)$ to $h \in \mathbb{R}^n$.

Implicitly in the discussion above, we've assumed that $Df(p)t$, if it exists, must be unique. Of course, this is something that we need to prove.

Theorem 3.2. *The differential, if it exists, is unique.*

Proof. Suppose $\Omega \subset \mathbb{R}^n$ is open, $f : \Omega \to \mathbb{R}^m$, $p \in \Omega$ and that Λ and Λ' satisfy:

$$\lim_{x \to p} \frac{||f(x) - A_\Lambda(x)||}{||x-p||} = \lim_{x \to p} \frac{||f(x) - A_{\Lambda'}(x)||}{||x-p||} = 0.$$

Let e be any unit vector, then for any $\alpha > 0$ we have:

$$\frac{\Lambda(\alpha e)}{\alpha} = \Lambda e.$$

Now, let $(\alpha_j)_{j=0}^\infty$ be a sequence with $\alpha_j > 0$ and $\alpha_j \to 0$. We can calculate:

$$\begin{aligned}
\big|\big|\Lambda e - \Lambda' e\big|\big| &= \left|\left|\frac{\Lambda(\alpha_j e)}{\alpha_j} - \frac{\Lambda'(\alpha_j e)}{\alpha_j}\right|\right| \\
&= \lim_{j\to\infty} \frac{||\Lambda(\alpha_j e) - \Lambda'(\alpha_j e)||}{||\alpha_j e||} \\
&= \lim_{j\to\infty} \frac{||-f(p+\alpha_j e) + f(p) + \Lambda(\alpha_j e) + f(p+\alpha_j e) - f(p) - \Lambda'(\alpha_j e)||}{||\alpha_j e||} \\
&\le \lim_{j\to\infty} \frac{||f(p+\alpha_j e) - f(p) - \Lambda(\alpha_j e)||}{||\alpha_j e||} + \lim_{j\to\infty} \frac{||f(p+\alpha_j e) - f(p) - \Lambda'(\alpha_j e)||}{||\alpha_j e||} \\
&= 0,
\end{aligned}$$

since $\alpha_j e \to 0$ as $j \to \infty$. Thus for any unit vector e we have $\Lambda e = \Lambda' e$, which implies that (as linear maps) $\Lambda = \Lambda'$. □

We are thus fully justified in referring to *the* differential. We will see a much more straightforward means of calculating the differential shortly.

Exercise 7.1. Suppose $f : \mathbb{R}^n \to \mathbb{R}^n$ is given by:

$$f(x) = x.$$

Show that f is differentiable at each $p \in \mathbb{R}^n$ and:

$$Df(p) = \iota,$$

where $\iota : \mathbb{R}^n \to \mathbb{R}^n$ is the identity.

Exercise 7.2. Show that the map $f : \mathbb{R}^2 \to \mathbb{R}$ given by:

$$f : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto x^2 + y^2,$$

is differentiable at all points $p = (\xi, \eta)^t \in \mathbb{R}^2$ with Jacobian:

$$Df(p) = (2\xi \quad 2\eta)$$

Exercise 7.3. One might hope that the differential can be calculated by finding:

$$\lim_{x \to p} \frac{f(x) - f(p)}{||x - p||}.$$

By considering the example of Exercise 7.1 or otherwise, show that this limit may not always exist, even if f is differentiable at p.

Exercise 7.4. Suppose that $\Omega \subset \mathbb{R}^n$ is open, and $f, g : \Omega \to \mathbb{R}^m$ are differentiable at $p \in \Omega$. Show that $h = f + g$ is differentiable at p and

$$Dh(p) = Df(p) + Dg(p)$$

3.1.2 Chain rule

In the case of one-dimensional differentiability, an important result is the chain rule, which tells us how to differentiate the composition of two functions. If $f, g : \mathbb{R} \to \mathbb{R}$, with g differentiable at p and f is differentiable at $g(p)$, then $f \circ g$ is differentiable at p with:

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

Now, suppose that $g : \mathbb{R}^n \to \mathbb{R}^m$ and $f : \mathbb{R}^m \to \mathbb{R}^l$ with g differentiable at p and f is differentiable at $g(p)$. Let $h = f \circ g$. We know that $Dg(p) : \mathbb{R}^n \to \mathbb{R}^m$ and $Df(g(p)) : \mathbb{R}^m \to \mathbb{R}^l$ are linear maps, so it certainly makes sense to consider $Df(g(p)) \circ Dg(p)$, where here $\circ$ is composition of linear maps (i.e. matrix multiplication). This will be a linear map from $\mathbb{R}^n$ to $\mathbb{R}^l$, which is the right kind of object to be $Dh(p)$. In fact, it is the case that $h = f \circ g$ is differentiable at p with:

$$Dh(p) = Df(g(p)) \circ Dg(p)$$

Theorem 3.3. *Suppose $\Omega \subset \mathbb{R}^n$, $\Omega' \subset \mathbb{R}^m$ are open, and that $g : \Omega \to \Omega'$, $f : \Omega' \to \mathbb{R}^l$ are differentiable at p, $g(p)$ respectively. Then $h = f \circ g$ is differentiable at p with:*

$$Dh(p) = Df(g(p)) \circ Dg(p).$$

(*) Proof. Let $g(p) = q$, $\lambda = Dg(p)$, $\mu = Df(q)$. We define:

$$\begin{aligned}
\phi(x) &= g(x) - g(p) - \lambda(x - p), \\
\psi(y) &= f(y) - f(q) - \mu(y - q), \\
\tau(x) &= f(g(x)) - f(g(p)) - \mu\left[\lambda(x - p)\right].
\end{aligned}$$

We know that:

$$0 = \lim_{x \to p} \frac{\phi(x)}{||x - p||}, \tag{3.2}$$

$$0 = \lim_{y \to q} \frac{\psi(y)}{||y - q||}, \tag{3.3}$$

and we would like to show that:

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

Now,

$$\begin{aligned}
\tau(x) &= f(g(x)) - f(g(p)) - \mu\left[\lambda(x-p)\right] \\
&= f(g(x)) - f(g(p)) - \mu\left[g(x) - g(p) - \phi(x)\right] \\
&= f(g(x)) - f(g(p)) - \mu\left[g(x) - g(p)\right] + \mu(\phi(x)) \\
&= \psi(g(x)) + \mu(\phi(x)).
\end{aligned}$$

Since μ is linear, and hence continuous by Example 6, we have that:

$$\lim_{x \to p} \frac{\mu(\phi(x))}{||x-p||} = \lim_{x \to p} \mu\left(\frac{\phi(x)}{||x-p||}\right) = \mu\left(\lim_{x \to p} \frac{\phi(x)}{||x-p||}\right) = 0.$$

Now, fix $\epsilon > 0$. It follows from (3.3) that there exists $\delta > 0$ such that for $y \in \Omega'$ with $0 < ||y-q|| < \delta$ we have:

$$\frac{||\psi(y)||}{||y-q||} < \epsilon \qquad \implies ||\psi(y)|| < \epsilon \, ||y-q||$$

On the other hand, since g is continuous, there exists δ_1 such that if $x \in \Omega$ with $0 \leq ||x-p|| < \delta_1$ then

$$||g(x) - g(p)|| = ||g(x) - q|| < \delta.$$

Thus if $x \in \Omega$ with $0 \leq ||x-p|| < \delta_1$, we have:

$$\begin{aligned}
||\psi(g(x))|| &< \epsilon \, ||g(x) - q|| \\
&= \epsilon \, ||\phi(x) + \lambda(x-p)|| \\
&\leq \epsilon \, ||\phi(x)|| + \epsilon M \, ||x-p|| \,,
\end{aligned}$$

where we have used the fact that we derived in Example 6 for any linear map $\lambda : \mathbb{R}^n \to \mathbb{R}^m$ there exists M such that $||\lambda x|| \leq M \, ||x||$ for all $x \in \mathbb{R}^n$. Dividing through by $||x-p||$ and taking the limit, we deduce that:

$$\lim_{x \to p} \frac{||\psi(g(x))||}{||x-p||} \leq \epsilon M.$$

Since $\epsilon > 0$ was arbitrary, we conclude:

$$\lim_{x \to p} \frac{||\psi(g(x))||}{||x-p||} = 0,$$

and we're done. □

Exercise 7.5. Suppose $\Omega, \Omega' \subset \mathbb{R}^n$ are open, $g : \Omega \to \Omega'$ and $f : \Omega' \to \Omega$ are functions such that g is differentiable at $p \in \Omega$ and f is differentiable at $g(p) \in \Omega'$ and moreover:

$$\begin{aligned} f \circ g(x) &= x, \qquad \forall\, x \in \Omega. \\ g \circ f(x) &= x, \qquad \forall\, x \in \Omega'. \end{aligned}$$

Show that:

$$Df(g(p)) = (Dg(p))^{-1}.$$

Exercise 7.6 (*)**.** a) Show that the map $P : \mathbb{R}^2 \to \mathbb{R}$ given by:

$$P : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto xy$$

is differentiable at each point $p = \begin{pmatrix} \xi \\ \eta \end{pmatrix} \in \mathbb{R}^2$, with Jacobian:

$$DP(p) = \begin{pmatrix} \eta & \xi \end{pmatrix}.$$

b) Suppose that $f, g : \mathbb{R}^n \to \mathbb{R}$ are differentiable at $q \in \mathbb{R}^n$. Show that the map $Q : \mathbb{R}^n \to \mathbb{R}^2$ given by:

$$Q : z \mapsto \begin{pmatrix} f(z) \\ g(z) \end{pmatrix}$$

is differentiable at q and:

$$DQ(q) = \begin{pmatrix} Df(q) \\ Dg(q) \end{pmatrix}$$

c) Show that $F : \mathbb{R}^n \to \mathbb{R}$ given by $F(z) = f(z)g(z)$ for all $z \in \mathbb{R}^n$ is differentiable at q, and:

$$DF(q) = g(q)Df(q) + f(q)Dg(q)$$

3.2 Directional derivatives

3.2.1 Definition and examples

The differential of a map tells us something about the best affine approximation to f. This is a rather abstract notion. For a function $f : (a, b) \to \mathbb{R}$, we are familiar with the idea of $f'(p)$ telling us something about the rate of change of $f(x)$ as we vary x near $p \in (a, b)$. We can connect the differential to this sort of concept with the *directional derivative*. Let us suppose that we are given a function $f : \mathbb{R}^3 \to \mathbb{R}$, which is supposed to represent the temperature of some three dimensional body which is not changing in time. Suppose we start at the origin $0 \in \mathbb{R}^3$ and travel along the curve $t : vt$ for some $v \in \mathbb{R}^3$, that is we move along a straight line with velocity v passing through the origin at time 0. We can record the temperature of our surroundings as a function of time, $\theta(t)$ and we will find $\theta(t) = f(vt)$. Suppose we ask what the rate of change of temperature is at $t = 0$. This will of course be $\theta'(0)$. Now, we notice that we can write:

$$\theta = f \circ V$$

where V is the linear map $V : \mathbb{R} \to \mathbb{R}^3$ given by $V(t) = vt$. Now, we can use the chain rule to calculate $\theta'(0) = D\theta(0)$ and we find:

$$\theta'(0) = D\theta(0) = Df(0) \circ DV(0).$$

Now, since V is a linear map, we have $DV(0) = V$ and we conclude:

$$\theta'(0) = Df(0)[v].$$

This gives us a nice interpretation of the differential $Df(0)$. When we apply $Df(0)$ to a vector v, we find the rate of change of f at 0 as we travel along a line with velocity v. More generally, we can consider travelling along the line given by $V(t) = p + tv$ for some $p \in \mathbb{R}^3$. Then at $t = 0$, we are passing through the point $p \in \mathbb{R}^3$. Setting $\theta(t) = f(p + tv)$, We call the quantity:

$$\theta'(0) = D\theta(p) = Df(p)[v]$$

the directional derivative of f at x in the direction v. Sometimes the notation

$$\frac{\partial f}{\partial v}(p) := \lim_{t\to 0} \frac{1}{t}\left[f(p + vt) - f(p)\right] = Df(p)[v]$$

is used for the directional derivative.

Now, if we take $\{e_1, e_2, e_3\}$ to be the canonical basis vectors for $\mathbb{R}^3$, then we can write $v = v^1 e_1 + v^2 e_2 + v^3 e_3$ for $v^i \in \mathbb{R}$. Doing this, and recalling that $Df(p)$ is a linear map, we have:

$$\begin{aligned} \frac{\partial f}{\partial v}(p) &= Df(p)\left[v^1 e_1 + v^2 e_2 + v^3 e_3\right] && (3.4) \\ &= v^1 Df(0)\left[e_1\right] + v^2 Df(0)\left[e_2\right] + v^3 Df(0)\left[e_3\right] \\ &= v^1 D_1 f(p) + v^2 D_2 f(p) + v^3 D_3 f(p). && (3.5) \end{aligned}$$

In other words, we can find any directional derivative at p, provided we know the three numbers:

$$D_i f(p) = \frac{\partial f}{\partial e_i}(p), \qquad i = 1, 2, 3.$$

called the *partial derivatives* of f at p. These we can define by:

$$D_i f(p) := \lim_{t\to 0} \frac{f(p + te_i) - f(p)}{t}.$$

If we write:

$$f : \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto f(x, y, z),$$

then this simplifies to:

$$D_1 f(x, y, z) = \lim_{t\to 0} \frac{f(x + t, y, z) - f(x, y, z)}{t} =: \frac{\partial f}{\partial x}(x, y, z),$$

where we've introduced yet more notation. The expression $\frac{\partial f}{\partial x}$ you should think of as meaning 'differentiate f with respect to x, while treating y, z as constants. Returning to (3.5), we see that for any $v = (v^1, v^2, v^3)^t$, we have:

$$Df(p)[v] = \begin{pmatrix} D_1 f(p) & D_2 f(p) & D_3 f(p) \end{pmatrix} \begin{pmatrix} v^1 \\ v^2 \\ v^3 \end{pmatrix},$$

so that the Jacobian of f at p is given by:

$$Df(p) = \begin{pmatrix} D_1 f(p) & D_2 f(p) & D_3 f(p) \end{pmatrix}.$$

To introduce even more notation, we sometimes write:

$$\nabla f(p) = \begin{pmatrix} D_1 f(p) \\ D_2 f(p) \\ D_3 f(p) \end{pmatrix},$$

which is called the *gradient* of f at p, and with this notation:

$$Df(p) = \nabla f(p)^t.$$

We can extend all of these notions to more general range and domains, which leads us to the definitions:

Definition 3.2. Suppose $\Omega \subset \mathbb{R}^n$ is open and $f : \Omega \to \mathbb{R}^m$ is differentiable at $p \in \Omega$. For any vector $v \in \mathbb{R}^n$, the directional derivative of f at p in the direction v is given by:

$$\frac{\partial f}{\partial v}(p) := \lim_{t \to 0} \frac{f(p + tv) - f(p)}{t} = Df(p)[v]$$

The partial derivatives of f at p are given by:

$$D_i f(p) := \frac{\partial f}{\partial e_i}(p) = \lim_{t \to 0} \frac{f(p + te_i) - f(p)}{t}, \quad i = 1, \ldots, n.$$

Notice that $f(x)$ is now a vector in $\mathbb{R}^m$, so expressions like $\lim_{t\to 0} \frac{f(p+tv)-f(p)}{t}$ have to be understood as limits in $\mathbb{R}^m$, so that $\frac{\partial f}{\partial v}(p)$ will be a $m-$dimensional column vector. If we write:

$$f(x) = \begin{pmatrix} f^1(x) \\ \vdots \\ f^m(x) \end{pmatrix},$$

then:

$$D_i f(p) = \begin{pmatrix} D_i f^1(p) \\ \vdots \\ D_i f^m(p) \end{pmatrix}.$$

Theorem 3.4. *Suppose $\Omega \subset \mathbb{R}^n$ is open and $f : \Omega \to \mathbb{R}^m$ is differentiable at $p \in \Omega$. Then the Jacobian of f at p is given by:*

$$Df(p) = \begin{pmatrix} D_1 f^1(p) & \dots & D_n f^1(p) \\ \vdots & \ddots & \vdots \\ D_1 f^m(p) & \dots & D_n f^m(p) \end{pmatrix}.$$

Proof. Let $\{e_i\}$ be the canonical basis for $\mathbb{R}^n$. For any $v \in \mathbb{R}^n$, we write $v = \sum_{i=1}^n v^i e_i$. Then by the linearity of $Df(p)$ we have:

$$\begin{aligned} Df(p)[v] = Df(p)\left[\sum_{i=1}^n v^i e_i\right] &= \sum_{i=1}^n v^i Df(p)\left[e_i\right] = \sum_{i=1}^n v^i D_i f(p). \\ &= \begin{pmatrix} \sum_{i=1}^n v^i D_i f^1(p) \\ \vdots \\ \sum_{i=1}^n v^i D_i f^m(p) \end{pmatrix} \\ &= \begin{pmatrix} D_1 f^1(p) & \dots & D_n f^1(p) \\ \vdots & \ddots & \vdots \\ D_1 f^m(p) & \dots & D_n f^m(p) \end{pmatrix} \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix} \end{aligned}$$ □

This allows us to restate the chain rule in terms of the partial derivatives of the functions.

Corollary 3.5. *Suppose $\Omega \subset \mathbb{R}^n$, $\Omega' \subset \mathbb{R}^m$ are open, and that $g : \Omega \to \Omega'$, $f : \Omega' \to \mathbb{R}^l$ are differentiable at p, $g(p)$ respectively. Then $h = f \circ g$ is differentiable at p with Jacobian:*

$$Dh(p) = \begin{pmatrix} D_1 f^1(g(p)) & \dots & D_m f^1(g(p)) \\ \vdots & \ddots & \vdots \\ D_1 f^l(g(p)) & \dots & D_m f^l(g(p)) \end{pmatrix} \begin{pmatrix} D_1 g^1(p) & \dots & D_n g^1(p) \\ \vdots & \ddots & \vdots \\ D_1 g^m(p) & \dots & D_n g^m(p) \end{pmatrix}$$

In the one dimensional case, we often use the derivative to search for turning points, i.e. maxima and minima, since a differentiable function will have vanishing derivative at a local maximum or minimum. A similar result holds in the higher dimensional case:

Lemma 3.6. *Let $\Omega \subset \mathbb{R}^n$ be open and $f : \Omega \to \mathbb{R}$ be differentiable at each point in Ω. Suppose that f has a local maximum at $p \in \Omega$. Then:*

$$Df(p) = 0.$$

Similarly if p is a local minimum.

Proof. Pick $v \in \mathbb{R}^n$. Since Ω is open, there exists $\epsilon > 0$ such that $p + tv \in \Omega$ for $t \in (-\epsilon, \epsilon)$. Consider the function

$$\begin{aligned} g_v : (-\epsilon, \epsilon) &\to \mathbb{R} \\ t &\mapsto f(p + tv). \end{aligned}$$

Since f has a local maximum at p, g_v has a local maximum at 0 and moreover, g_v is differentiable by the chain rule, so we deduce:

$$0 = g_v'(0) = Df(p)[v].$$

Since v was arbitrary, we have that $Df(p) = 0$. A similar argument deals with the case where p is a minimum. □

3.2.2 Relation between partial derivatives and differentiability

We have seen above that for a function $f : \mathbb{R}^n \to \mathbb{R}$ which is differentiable at some point p, the limits:

$$D_i f(p) := \lim_{t \to 0} \frac{f(p + te_i) - f(p)}{t} \tag{3.6}$$

exist for $i = 1, \ldots n$, and moreover these limits completely determine the differential of f at p. One might hope, based on this, that in order for f to be differentiable at p it is enough to know that the partial derivatives (i.e. the limits in (3.6)) of f at p all exist. Unfortunately, this is not the case. To see this, let us consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by:

$$f : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{cases} 0 & x = y = 0 \\ \frac{xy}{\sqrt{x^2+y^2}} & \text{otherwise} \end{cases}$$

First note that this function is continuous at the origin. Since $|xy| \leq \frac{1}{2}\left(x^2 + y^2\right)$, we have that for $p = (x, y) \neq (0, 0)$:

$$|f(p)| \leq \frac{1}{2}\sqrt{x^2 + y^2},$$

so that:

$$\lim_{p \to 0} f(p) = 0.$$

Now consider the partial derivatives. We have that:

$$D_1 f(0) = \lim_{t \to 0} \frac{1}{t}\left[f(te_1) - f(0)\right] = \lim_{t \to 0} \frac{0 - 0}{t} = 0$$

since $f(te_1) = 0$ for all t. Similarly, we also have:

$$D_2 f(0) = \lim_{t \to 0} \frac{1}{t}\left[f(te_2) - f(0)\right] = \lim_{t \to 0} \frac{0 - 0}{t} = 0$$

Thus, *if* f is differentiable, then it must be that $Df = 0$, so all directional derivatives at 0 exist and are equal to zero. However, let $h = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. We have for $t > 0$:

$$\frac{f(th) - f(0)}{t} = \frac{t^2}{t\sqrt{2t^2}} = \frac{1}{\sqrt{2}},$$

which contradicts the differentiability of f at the origin. Thus *even though* the partial derivatives exist for this function, the function is not differentiable. Away from the origin, the function is a composition of smooth functions so is differentiable. We can calculate the partial derivatives at a point $p = (x, y) \neq (0, 0)$ and we find:

$$D_1 f(p) = \frac{y}{\sqrt{x^2+y^2}} - \frac{x^2 y}{(x^2+y^2)^{\frac{3}{2}}} = \frac{y^3}{(x^2+y^2)^{\frac{3}{2}}},$$

and by symmetry:

$$D_2 f(p) = \frac{x^3}{(x^2+y^2)^{\frac{3}{2}}}.$$

I claim that the function $g : \mathbb{R}^2 \setminus \{0\}$ given by

$$g : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \frac{x^3}{(x^2+y^2)^{\frac{3}{2}}}$$

has no limit as $p = (x, y) \to (0, 0)$. To see this, let $p = (r\cos\theta, r\sin\theta)$ for some $r \in (0, \infty)$, $\theta \in [0, 2\pi)$. Then:

$$g(p) = \cos^3\theta,$$

so there can be no limit as $r \to 0$, since g approaches a different value depending on which angle we approach from.

As it happens, the fact that the partial derivatives are not continuous in a neighbourhood of the origin is the only barrier to differentiability there.

Theorem 3.7. *Let $\Omega \subset \mathbb{R}^n$ be open and $f : \Omega \to \mathbb{R}$. Suppose the partial derivatives:*

$$D_i f(x) := \lim_{t\to 0} \frac{f(x + te_i) - f(x)}{t}$$

exist for all $x \in \Omega$, and moreover suppose that the maps:

$$x \mapsto D_i f(x)$$

are continuous at $p \in \Omega$ for all $i = 1, \ldots, n$. Then f is differentiable at p.

Proof(*). Since Ω is open, there exists $r > 0$ such that $B_r(p) \subset \Omega$. Suppose $h \in B_r(0)$ has components h^i, so that $h = \sum_{i=1}^n h^i e_i$. We consider:

$$\begin{aligned} f(p+h) - f(p) &= f\left(p + \sum_{i=1}^{n} h^i e_i\right) - f(p) \\ &= f\left(p + \sum_{i=1}^{n} h^i e_i\right) - f\left(p + \sum_{i=1}^{n-1} h^i e_i\right) \\ &\quad + f\left(p + \sum_{i=1}^{n-1} h^i e_i\right) - f\left(p + \sum_{i=1}^{n-2} h^i e_i\right) \\ &\quad + \ldots \\ &\quad + f(p + h^1 e_1) - f(p). \end{aligned}$$

Let's consider a typical line in our expansion:

$$f\left(p+\sum_{i=1}^{k} h^i e_i\right) - f\left(p+\sum_{i=1}^{k-1} h^i e_i\right) = f(q+h^k e_k) - f(q),$$

where $k \in \{1,\ldots,n\}$ and $q = p + \sum_{i=1}^{k-1} h^i e_i$. Now, applying the mean value theorem to the function $g(t) = f(q+te_k)$, which is differentiable by assumption, there exists[2] $s \in \left[-\left|h^k\right|, \left|h^k\right|\right)$ such that:

$$f(q+h^k e_k) - f(q) = h^k D_k f(q+se_k) = h^k D_k f(p+c_k),$$

where $c_k = \sum_{i=1}^{k-1} h^i e_i + se_k$. Now, note that since $|s| \leq \left|h^k\right|$, we have:

$$||c_k|| \leq ||h|| \,.$$

Putting this together, we conclude that there exists $c_1,\ldots,c_n \in \mathbb{R}^n$ with $||c_k|| \leq ||h||$ such that:

$$f(p+h) - f(p) = \sum_{i=1}^{n} h^k D_k f(p+c_k).$$

From here we can estimate using the Cauchy-Schwartz identity:

$$\begin{aligned}\left|f(p+h) - f(p) - \sum_{i=1}^{n} h^k D_k f(p)\right| &\leq \sum_{i=1}^{n} h^k \left|D_k f(p+c_k) - D_k f(p)\right| \\ &\leq ||h|| \left(\sum_{i=1}^{n} \left|D_k f(p+c_k) - D_k f(p)\right|^2\right)^{\frac{1}{2}},\end{aligned}$$

so that:

$$\frac{\left|f(p+h) - f(p) - \sum_{i=1}^{n} h^k D_k f(p)\right|}{||h||} \leq \left(\sum_{i=1}^{n} \left|D_k f(p+c_k) - D_k f(p)\right|^2\right)^{\frac{1}{2}}.$$

Now, fix $\epsilon > 0$. Since $x \mapsto D_k f(x)$ is continuous at p, for each $k = 1,\ldots,n$, there exists δ_k such that if $||c|| < \delta_k$ we have:

$$\left|D_k f(p+c) - D_k f(p)\right| < \frac{\epsilon}{\sqrt{n}}.$$

Suppose $||h|| < \min\{\delta_1,\ldots,\delta_n\} =: \delta$. Then as $||c_k|| \leq ||h||$, we deduce:

$$\frac{\left|f(p+h) - f(p) - \sum_{i=1}^{n} h^k D_k f(p)\right|}{||h||} < \left(\sum_{i=1}^{n} \frac{\epsilon^2}{n}\right)^{\frac{1}{2}} = \epsilon,$$

and we conclude that f is differentiable at p, with differential:

$$Df(p)\,[h] = \sum_{i=1}^{n} D_k f(p) h^k. \qquad \square$$

---

[2] One has to consider separately the cases $h^k > 0$, $h^k < 0$ and $h^k = 0$.

Exercise 8.1. Show that $f : \mathbb{R}^2 \to \mathbb{R}$ is everywhere differentiable, and find the differential when:

a) $f : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto x^2 + y^2 - x - xy$

b) $f : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \dfrac{1}{\sqrt{1 + x^2 + y^2}}$

c) $f : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto x^5 y^2$

Exercise 8.2. Find the minimum of the function $f : \mathbb{R}^3 \to \mathbb{R}$ given by:

$$f : \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto x^4(y^2 + x^2) + z^2 - 4z$$

3.3 Higher derivatives

Suppose that $\Omega \subset \mathbb{R}^n$ is open, and $f : \Omega \to \mathbb{R}^m$ is differentiable at each point $p \in \Omega$. We can consider the map:

$$\begin{array}{rcccl} Df & : & \Omega & \to & L(\mathbb{R}^n; \mathbb{R}^m) \\ & & p & \mapsto & Df(p). \end{array}$$

With respect to the canonical basis, we can of course think of this as a map from Ω to the set of $(m \times n)$-matrices. Now, we can identify the set of $(m \times n)$-matrices with $\mathbb{R}^{mn}$, so that $Df : \Omega \to \mathbb{R}^{mn} = \mathbb{R}^M$. We can, of course, then consider whether this map Df is continuous, or differentiable at a point p. If Df is differentiable at p, the differential is a linear map $DDf(p) \in L(\mathbb{R}^n; \mathbb{R}^{nm})$ which takes an $n-$vector to an $(m \times n)$ matrix. We can write down the components of this map. We know from previous results that the partial derivative maps

$$\begin{array}{rcccl} D_i f^j & : & \Omega & \to & \mathbb{R} \\ & & x & \mapsto & D_i f^j(x). \end{array}$$

exist for all $x \in \Omega$ and if Df is differentiable at p, then the second partial derivatives:

$$D_k D_i f^j(p) := \lim_{t \to 0} \frac{D_i f^j(p + t e_k) - D_i f^j(p)}{t}$$

will exist. Then we can write:

$$\left(DDf(p)\,[h]\right)^j{}_i = \sum_{k=1}^{n} D_k D_i f^j(p) h^k$$

where $\left(DDf(p)\,[h]\right)^j{}_i$ corresponds to the component in the i^{th} column and j^{th} row of the $m \times n$ matrix $DDf(p)[h]$.

Obviously, we can generalise this to consider a function which is $k-$times differentiable. In practice, the condition of $k-$times differentiable at a point p can be difficult to establish, one instead asks that all of the $k-$th partial derivatives exist and are continuous in a neighbourhood of p. This is a slightly stronger condition, which implies $k-$times differentiability at p by Theorem 3.7.

Example 3.5. Consider the map $f : \mathbb{R}^2 \to \mathbb{R}$ given by:

$$f : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto x^3 + y^3 + 5x^2 y.$$

This is differentiable at each point $p = (x, y)^t \in \mathbb{R}^2$, and the partial derivatives are:

$$D_1 f(p) = 3x^2 + 10xy, \qquad D_2 f(p) = 3y^2 + 5x^2.$$

To find the second partial derivatives, we consider the maps:

$$D_1 f : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto 3x^2 + 10xy,$$

and

$$D_2 f : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto 3y^2 + 5x^2,$$

and differentiate them. The second partial derivatives are thus:

$$\begin{aligned} D_1 D_1 f(p) &= 6x + 10y \\ D_2 D_1 f(p) &= 10x \\ D_1 D_2 f(p) &= 10x \\ D_2 D_2 f(p) &= 6y \end{aligned}$$

Notice that:

$$D_2 D_1 f(p) = D_1 D_2 f(p).$$

This is not in fact a coincidence.

3.3.1 Symmetry of mixed partial derivatives

We will state a result here, but we postpone the proof until later on. This is not the optimal result in this direction, but it is perfectly adequate for most purposes.

Theorem 3.8 (Schwartz' Theorem)**.** *Suppose $\Omega \subset \mathbb{R}^n$ is open and $f : \Omega \to \mathbb{R}$ is differentiable at each $p \in \Omega$. Suppose further that for some $i, j \in \{1, \ldots, n\}$ the second partial derivatives:*

$$D_i D_j f(p), \qquad D_j D_i f(p),$$

exist and are continuous at all $p \in \Omega$. Then:

$$D_i D_j f(p) = D_j D_i f(p).$$

If $f : \Omega \to \mathbb{R}$, the matrix of second partial derivatives at the point p:

$$\text{Hess}\, f(p) := [D_i D_j f(p)]_{i,j=1,\dots,n}$$

is called the Hessian. Assuming the hypotheses on the second partial derivatives hold, Schwartz' Theorem states that the Hessian is a symmetric matrix.

Exercise 8.3. Suppose A is a symmetric $(n \times n)$ matrix. Consider the function:

$$\begin{aligned} f \ : \ \mathbb{R}^n &\to \mathbb{R} \\ x &\mapsto x^t A x. \end{aligned}$$

a) Show that f is differentiable at all points $p \in \mathbb{R}^n$, with:

$$Df(p) = 2p^t A$$

b) Find:

$$\text{Hess}\, f(p).$$

Exercise 8.4 (*)**.** Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by:

$$f : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{cases} \dfrac{xy^3 - x^3 y}{x^2 + y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0). \end{cases}$$

a) Show that:

$$D_1 f : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{cases} \dfrac{y^3 - 3x^2 y}{x^2 + y^2} - \dfrac{2x\left(xy^3 - x^3 y\right)}{\left(x^2 + y^2\right)^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0). \end{cases}$$

and

$$D_2 f : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{cases} \dfrac{3y^2 x - x^3}{x^2 + y^2} - \dfrac{2y\left(xy^3 - x^3 y\right)}{\left(x^2 + y^2\right)^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0), \end{cases}$$

and show that these functions are both continuous at $(0,0)$.

b) Show that:

$$\lim_{t \to 0} \frac{1}{t} \left(D_1 f(te_2) - D_1 f(0)\right) = 1$$

and

$$\lim_{t \to 0} \frac{1}{t} \left(D_2 f(te_1) - D_2 f(0)\right) = -1$$

c) Conclude that both $D_2 D_1 f(0)$ and $D_1 D_2 f(0)$ exist, but that:

$$D_2 D_1 f(0) \neq D_1 D_2 f(0)$$

3.3.2 Taylor's theorem

A powerful result concerning differentiable functions of one variable is Taylor's theorem, which permits us to approximate a function in the neighbourhood of a point p by a polynomial, with an error term that goes to zero at a controlled rate as we approach p. In order to state Taylor's theorem for higher dimensions, it's useful to introduce some new notation.

When dealing with partial derivatives of high orders, the notation can get rather messy. To mitigate this, it's convenient to introduce *multi-indices*. We define a multi-index α to be an element of $(\mathbb{N})^n$, i.e. a $n-$vector of non-negative integers $\alpha = (\alpha_1, \ldots, \alpha_n)$. We define $|\alpha| = \alpha_1 + \ldots + \alpha_n$ and

$$D^\alpha f := (D_1)^{\alpha_1} (D_2)^{\alpha_2} \cdots (D_n)^{\alpha_n} f,$$

It's convenient to also introduce, for a vector $h = (h^1, \ldots, h^n)$:

$$h^\alpha := \left(h^1\right)^{\alpha_1} \left(h^2\right)^{\alpha_2} \cdots (h^n)^{\alpha_n}$$

as well as the multi-index factorial $\alpha! = \alpha_1!\alpha_2! \cdots \alpha_n!$,

Theorem 3.9. *Suppose that $p \in \mathbb{R}^n$ and $f : B_r(p) \to \mathbb{R}$ is $k-$times continuously differentiable at all points $q \in B_r(p)$, and suppose $||h|| < r$. Then we have that*

$$f(p+h) = \sum_{|\alpha| \leq k-1} \frac{h^\alpha}{\alpha!} D^\alpha f(p) + R_k(p,h).$$

The sum is taken over all multi-indices $\alpha = (\alpha_1, \ldots, \alpha_n)$ with $|\alpha| \leq k-1$ and the remainder term is given by:

$$R_k(p,h) = \sum_{|\alpha|=k} \frac{h^\alpha}{\alpha!} D^\alpha f(x).$$

for some x with $0 < ||x-p|| < ||h||$.

Proof (*). The result follows from the one-dimensional Taylor's theorem. First, we note that there exists $\epsilon > 0$ such that $||h|| < \frac{r}{1+\epsilon}$. Let us define the function:

$$\begin{aligned} g \ : \ (-1-\epsilon, 1+\epsilon) &\to \mathbb{R} \\ t &\mapsto f(p+th). \end{aligned}$$

By the chain rule, this function is k-times differentiable on the interval $(-1-\epsilon, 1+\epsilon)$, and $[0,1] \subset (-1-\epsilon, 1+\epsilon)$, so by Taylor's theorem we have:

$$g(1) = g(0) + g'(0) + \frac{g''(0)}{2!} + \ldots + \frac{g^{(k-1)}(0)}{(k-1)!} + R_k,$$

where:

$$R_k = \frac{g^{(k)}(\xi)}{k!},$$

for some $\xi \in (0,1)$. We will be done if we can show that for $j = 0, 1, \dots k$ we have:

$$g^{(j)}(t) = j! \sum_{|\alpha|=j} \frac{h^\alpha}{\alpha!} D^\alpha f(p+th). \tag{3.7}$$

This is certainly true for $j = 0$. Suppose it's true for some j. Then we have:

$$\begin{aligned} g^{(j+1)}(t) &= \sum_{l=1}^{n} h^l D_l \left[j! \sum_{|\alpha|=j} \frac{h^\alpha}{\alpha!} D^\alpha f \right] (p+th) \\ &= j! \sum_{l=1}^{n} \sum_{|\alpha|=j} \frac{h^\alpha h^l}{\alpha!} D_l D^\alpha f(p+th) \end{aligned}$$

Clearly, the right hand side is a sum of terms proportional to $h^\beta D^\beta f(p+th)$ where $|\beta| = j+1$. Suppose $\beta = (\beta_1, \dots, \beta_n)$, then the coefficient of the term proportional to $h^\beta D^\beta f(p+th)$ is:

$$\begin{aligned} \frac{j!}{(\beta_1-1)!\beta_2!\cdots\beta_n!} + \frac{j!}{\beta_1!(\beta_2-1)!\cdots\beta_n!} + \dots + \frac{j!}{\beta_1!\beta_2!\cdots(\beta_n-1)!} \\ = \frac{(j+1)!}{\beta_1!\beta_2!\cdots\beta_n!} = \frac{(j+1)!}{\beta!}, \end{aligned}$$

by a result from combinatorics (see Exercise 8.5). Thus we have:

$$\begin{aligned} g^{(j+1)}(t) &= j! \sum_{l=1}^{n} \sum_{|\alpha|=j} \frac{h^\alpha h^l}{\alpha!} D_l D^\alpha f(p+th) \\ &= (j+1)! \sum_{|\beta|=j+1} \frac{h^\beta}{\beta!} D^\beta f(p+th) \end{aligned}$$

By induction we conclude that (3.7) holds for all $j = 0, \dots, k$ and the result follows. □

Exercise 8.5 (*). a) By induction on $m \in \mathbb{N}$, show that if $x_1, \dots, x_m \in \mathbb{R}$ and $j \in \mathbb{N}$ then:

$$(x_1 + \dots + x_m)^j = \sum_{\substack{n_1,\dots,n_m \in \mathbb{N} \\ n_1+n_2+\dots+n_m=j}} \begin{pmatrix} j \\ n_1, n_2, \dots, n_m \end{pmatrix} x_1^{n_1} \cdots x_m^{n_m},$$

where the multinomial coefficient is defined by:

$$\begin{pmatrix} j \\ n_1, n_2, \dots, n_m \end{pmatrix} = \frac{j!}{n_1! n_2! \cdots n_m!}.$$

b) By writing $(x_1 + \dots + x_m)^j = (x_1 + \dots + x_m)(x_1 + \dots + x_m)^{j-1}$ and expanding, show that for $j \geq 1$:

$$\begin{aligned} \begin{pmatrix} j \\ n_1, n_2, \dots, n_m \end{pmatrix} = \frac{(j-1)!}{(n_1-1)!n_2!\cdots n_m!} + \frac{(j-1)!}{n_1!(n_2-1)!\cdots n_m!} + \dots \\ \dots + \frac{(j-1)!}{n_1!n_2!\cdots(n_m-1)!} \end{aligned}$$

holds for any $n_1, \dots, n_m \in \mathbb{N}$, with $n_1 + n_2 + \dots + n_m = j$

3.4 Inverse function theorem

Suppose $f : \mathbb{R} \to \mathbb{R}$ is continuously differentiable in an interval around $p \in \mathbb{R}$, and moreover that $f'(p) \neq 0$, say $f'(p) > 0$. Then there is an open interval I with $p \in I$ such that $f'(x) > 0$ for all $x \in I$. This (by the mean value theorem) implies that f is strictly monotone increasing on I and hence $f : I \to f(I)$ is bijective. In particular, there exists an inverse function $f^{-1} : f(I) \to I$. With a little work, one can establish that f^{-1} is differentiable, and moreover, by an application of the chain rule:

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

A very important result tells us that this statement generalises to higher dimensions. It goes by the name of the inverse function theorem:

Theorem 3.10 (Inverse Function Theorem)**.** *Suppose that $f : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable in some open set containing p, and suppose moreover that $Df(p)$ is invertible. Then there exist open sets $U, V \subset \mathbb{R}^n$ with $p \in U$, $f(p) \in V$ such that $f : U \to V$ is a bijection, and $f^{-1} : V \to U$ is continuously differentiable for all $y \in V$ with:*

$$Df^{-1}(y) = \left[Df(f^{-1}(y))\right]^{-1}.$$

Remark. Notice that since the Jacobian $Df(p)$ is an $n \times n$ matrix, the statement that it is invertible is equivalent to the statement that $\det Df(p) \neq 0$.

Example 3.6. Consider the map $f : \mathbb{R}^2 \to \mathbb{R}^2$:

$$f : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x + y + 5xy \\ y - x^2 \end{pmatrix}$$

This is differentiable at all points $p = (\xi, \eta)^t \in \mathbb{R}^2$, with:

$$Df(p) = \begin{pmatrix} 1 + 5\eta & 1 + 5\xi \\ -2\xi & 1 \end{pmatrix}.$$

Since:

$$Df(0) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

is invertible, we have that f is invertible in some neighbourhood of the origin, and moreover:

$$Df^{-1}(0) = [Df(0)]^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$$

While it is possible to obtain an explicit formula for the inverse function[3] it is very unwieldy and it is difficult to directly compute $Df^{-1}(0)$.

---

[3] you may like to try this with a computer algebra package

Exercise 8.6. Let $\Omega = \{(x,y)^t \in \mathbb{R}^2 : x > 0\}$. Consider the function $f : \Omega \to \mathbb{R}^2$ given by:

$$f : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \sin y \\ x \cos y \end{pmatrix}.$$

a) Show that f is differentiable at all $p = (\xi, \eta)^t \in \Omega$, with:

$$Df(p) = \begin{pmatrix} \sin \eta & \xi \cos \eta \\ \cos \eta & -\xi \sin \eta \end{pmatrix}.$$

b) Show that $Df(p)$ is invertible for all $p \in \Omega$.

c) Show that $f : \Omega \to \mathbb{R}^2$ is not injective. Deduce that the restriction to open sets U, V in the inverse function theorem is necessary.

Exercise 8.7. a) Suppose $f : \mathbb{R} \to \mathbb{R}$ is continuously differentiable in a neighbourhood of the origin, and $f'(0) = 0$. Give an example to show that f may nevertheless be bijective.

b) Suppose $f : \mathbb{R}^n \to \mathbb{R}^n$ is bijective, differentiable at the origin, and $\det Df(0) = 0$. Show that f^{-1} is not differentiable at $f(0)$.

3.4.1 (*) Proof of the inverse function theorem

We will prove the inverse function theorem in parts, with several Lemmas. The proof is long, as I've included all the details here. It can certainly be made more concise. There are broadly three parts to the proof:

1. We first establish that if we restrict the domain to a sufficiently small ball about p, then f is injective (Lemma 3.12).

2. Next we show that the range of f includes a small ball around $f(p)$ (Lemma 3.13). This is enough to establish the existence of open sets U, V with $p \in U$, $f(p) \in V$ such that $f : U \to V$ is a bijection (Corollary 3.14).

3. Finally, we establish that $f^{-1} : V \to U$ is differentiable (Lemma 3.15).

Before we start, we first prove a result about functions with bounded derivatives. We show that such a function is necessarily Lipschitz (see Exercise 2.6).

Lemma 3.11. *Let $B_r(p) \subset \mathbb{R}^n$ be a ball and let $f : B_r(p) \to \mathbb{R}^n$ be continuously differentiable. Suppose there exists M such that $\left|D_j f^i(x)\right| < M$ for all $x \in B_r(p)$. Then:*

$$||f(x) - f(y)|| \leq nM \, ||x - y||$$

for all $x, y \in B_r(p)$.

Proof. For each j, consider the function $g^j : [0,1] \to \mathbb{R}$ given by:

$$g^j(t) = f^j\left(x + t(y-x)\right)$$

Applying the mean value theorem, there exists $\tau \in (0,1)$ such that:

$$g^j(1) - g^j(0) = (g^j)'(\tau).$$

By the chain rule, this implies:

$$f^j(y) - f^j(x) = \sum_{i=1}^n (y-x)^i D_i f^j\left(x + \tau(y-x)\right).$$

We estimate with the Cauchy-Schwartz inequality:

$$\begin{aligned} \left|f^j(y) - f^j(x)\right| &\leq ||y-x|| \sqrt{\sum_{i=1}^n \left[D_i f^j\left(x+\tau(y-x)\right)\right]^2} \\ &\leq \sqrt{n} M \, ||x-y|| \end{aligned}$$

Now since $||x|| \leq \sqrt{n} \max \left|x^j\right|$, we deduce:

$$||f(y) - f(x)|| \leq n \max_{j=1,\ldots,n} \left|f^j(y) - f^j(x)\right| \leq nM \, ||x-y||$$

□

Next we show that, provided we restrict our attention to a sufficiently small neighbourhood of p, f is injective (i.e. one-to-one).

Lemma 3.12. *Suppose that $f : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable in some open set containing p, and suppose moreover that $Df(p)$ is invertible. Then for δ sufficiently small, $f : B_\delta(p) \to \mathbb{R}^n$ is injective.*

Proof. First, let λ be the linear transformation $\lambda = Df(p)$. By assumption λ is non-singular, and clearly $f : B_\delta(p) \to \mathbb{R}^n$ is injective if and only if $g = \lambda^{-1} \circ f : B_\delta(p) \to \mathbb{R}^n$ is injective. We calculate:

$$Dg(p) = D(\lambda^{-1}(f(p)) \circ Df(p) = \lambda^{-1} \circ Df(p) = \iota,$$

where $\iota : \mathbb{R}^n \to \mathbb{R}^n$ is the identity transformation.

Since g is continuously differentiable on an open set containing p, there exists δ such that:

$$\left|D_i g^j(x) - D_i g^j(p)\right| < \frac{1}{2n},$$

for all $x \in B_\delta(p)$. Since $Dg(p) = \iota$, this implies that if we set $h(x) = g(x) - x$, we have:

$$\left|D_i h^j(x)\right| < \frac{1}{2n}.$$

Applying Lemma 3.11 to h, we find:

$$||g(x_1) - x_1 - (g(x_2) - x_2)|| \leq \frac{1}{2} ||x_1 - x_1|| ,$$

for all $x_1, x_2 \in B_\delta(p)$. The reverse triangle inequality gives us:

$$||x_1 - x_2|| - ||g(x_1) - g(x_2)|| \leq ||x_1 - x_2 - (g(x_1) - g(x_2))|| \leq \frac{1}{2} ||x_1 - x_1|| ,$$

which on rearranging gives us:

$$||x_1 - x_2|| \leq 2 ||g(x_1) - g(x_2)|| . \tag{3.8}$$

Now suppose that $g(x_1) = g(x_2)$. We conclude that $||x_1 - x_2|| = 0$, so that $x_1 = x_2$ and g is indeed injective on $B_\delta(p)$. □

This tells us that the map $f : B_\delta(p) \to f(B_\delta(p))$ is one-to-one and onto (by definition of $f(B_\delta(p))$), hence it is invertible. The next step will be to show that $f(B_\delta(p))$ contains a small open ball around $f(p)$.

Lemma 3.13. *Under the same hypotheses as the previous Lemma, there exists $\kappa > 0$, $\delta > 0$ such that for all $y \in B_\kappa(f(p))$ there exists a unique $x \in B_\delta(p)$ with $f(x) = y$.*

Proof. We first note that the map $x \mapsto \det Df(x)$ is continuous in a neighbourhood of p, so we can find $\delta_1 > 0$ such that $\det Df(x) \neq 0$ for $x \in B_{\delta_1}(p)$. Next we apply Lemma 3.12 to deduce the existence of $\delta_2 > 0$ such that $f : B_{2\delta_2}(p) \to \mathbb{R}^n$ is injective. We set $\delta = \min\{\delta_1, \delta_2\}$.

Next, we have that $\partial B_\delta(p) = \{x \in \mathbb{R}^n : ||x - p|| = \delta\}$ is compact, so by Exercise 2.4 we have that $f(\partial B_\delta(p))$ is compact. Since the norm is a continuous function, the map:

$$\begin{array}{rccc} h & : & f(\partial B_\delta(p)) & \to & \mathbb{R} \\ & & y & \mapsto & ||y - f(p)|| , \end{array}$$

is continuous, and so by Theorem I.9, we know that $h(y)$ achieves its minimum for some $y_0 \in f(\partial B_\delta(p))$. By the injectivity of $f : B_{2\delta}(p) \to \mathbb{R}^n$, we have that $y_0 \neq f(p)$, and so $||f(x) - f(p)|| \geq ||y_0 - f(p)|| = d$ for all $x \in \partial B_\delta(p)$. Let us set $\kappa = d/2$. By construction, we have that for all $x \in \partial B_\delta(p)$, $y \in B_\kappa(f(p))$:

$$||y - f(p)|| < ||y - f(x)|| . \tag{3.9}$$

Now we will show that for any $y \in B_\kappa(f(p))$ there is a unique $x \in B_\delta(p)$ with $f(x) = y$. To see this, fix $y \in B_\kappa(f(p))$ and define the function:

$$\begin{array}{rccc} g & : & \overline{B_\delta(p)} & \to & \mathbb{R} \\ & & x & \mapsto & ||y - f(x)||^2 , \end{array}$$

This function is continuous, and hence achieves a minimum at $x_0 \in \overline{B_\delta(p)}$. Now, if $x_0 \in \partial B_\delta(p)$, then by (3.9) we have $g(p) < g(x_0)$, a contradiction, so it must be that $x_0 \in B_\delta(p)$. By the chain rule and Example 3.3, we have that g is differentiable and:

$$Dg(x)[h] = 2 \langle y - f(x), Df[h] \rangle = 2 \sum_{i,j=1}^{n} (y^i - f^i(x)) D_j f^i(x) h_j .$$

By Lemma 3.6, we know that $Dg(x_0) = 0$, so we conclude:

$$0 = 2 \sum_{i,j=1}^{n} (y^i - f^i(x_0)) D_j f^i(x_0),$$

for all $j = 1, \ldots, n$. However, for $x \in B_\delta(p)$, we know that $\det Df(x) \neq 0$, so that the matrix $Df(x_0)$ is invertible, thus we conclude that:

$$y^i - f^i(x_0) = 0$$

for all $i = 1, \ldots, n$, so that $f(x_0) = y$. Since $f : B_\delta(p) \to \mathbb{R}^n$ is injective by construction, x_0 is unique. □

Let's sum up the content of these two Lemmas as follows:

Corollary 3.14. *Suppose that $f : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable in some open set containing p, and suppose moreover that $Df(p)$ is invertible. Then there exist open sets $U, V \subset \mathbb{R}^n$ with $p \in U$, $f(p) \in V$ such that $f : U \to V$ is a bijection.*

Proof. With the notation of Lemma 3.13, we set $V = B_\kappa(f(p))$ and $U = f^{-1}(V)$. By Lemma 3.13, we have that $f : U \to V$ is a bijection. V is certainly open. By Exercise 2.4, U is also open. □

To complete the proof of the inverse function theorem, it remains to establish that $f^{-1} : V \to U$ is differentiable.

Lemma 3.15. *Under the assumptions of Corollary 3.14, the inverse function $f^{-1} : V \to U$ is continuous differentiable on V with:*

$$Df^{-1}(y) = \left[Df(f^{-1}(y))\right]^{-1},$$

for all $y \in V$.

Proof. First, note that by (3.8) we have for any $x_1, x_2 \in U$:

$$||x_1 - x_2|| \leq 2 \left|\left|Df(p)^{-1}[f(x_1) - f(x_2)]\right|\right| \leq 2M \left|\left|f(x_1) - f(x_2)\right|\right|,$$

for some $M > 0$ by Example 6. Rewriting this, we conclude that there exists $M > 0$ such that:

$$\left|\left|f^{-1}(y_1) - f^{-1}(y_2)\right|\right| \leq 2M \left|\left|y_1 - y_2\right|\right|, \tag{3.10}$$

for any $y_1, y_2 \in V$, which implies f^{-1} is (uniformly) continuous on V.

Now, suppose $y \in V$ and $x \in U$ with $y = f(x)$. Let $\mu = Df(x)$. We will show that f^{-1} is differentiable at y with $Df^{-1} = \mu^{-1}$. To do this, we recall that since f is differentiable at x, for $x_1 \in U$ we have:

$$f(x_1) = f(x) + \mu(x_1 - x) + \varphi(x_1 - x) \tag{3.11}$$

where the remainder function ϕ satisfies:

$$\lim_{x_1 \to x} \frac{\varphi(x_1 - x)}{||x_1 - x||} = 0. \tag{3.12}$$

We can rewrite (3.11) as:

$$\mu^{-1}\left(f(x_1) - f(x)\right) = x_1 - x + \mu^{-1}\varphi(x_1 - x).$$

Since $f : U \to V$ is bijective, for any $y_1 \in V$ there exists $x_1 \in U$ with $f(x_1) = y_1$, and thus:

$$\mu^{-1}(y_1 - y) = f^{-1}(y_1) - f^{-1}(y) + \mu^{-1}\varphi\left(f^{-1}(y_1) - f^{-1}(y)\right).$$

which rearranges to give:

$$f^{-1}(y_1) = f^{-1}(y) + \mu^{-1}(y_1 - y) - \mu^{-1}\varphi\left(f^{-1}(y_1) - f^{-1}(y)\right).$$

To establish the differentiability of f^{-1} at y, it suffices to show:

$$\lim_{y_1 \to y} \frac{\mu^{-1}\varphi\left(f^{-1}(y_1) - f^{-1}(y)\right)}{||y_1 - y||} = 0.$$

Note that by the linearity of μ^{-1}, and Example 6 we have:

$$\lim_{y_1 \to y} \frac{\mu^{-1}\varphi\left(f^{-1}(y_1) - f^{-1}(y)\right)}{||y_1 - y||} = \mu^{-1}\left(\lim_{y_1 \to y} \frac{\varphi\left(f^{-1}(y_1) - f^{-1}(y)\right)}{||y_1 - y||}\right)$$

so it's enough to show that:

$$\lim_{y_1 \to y} \frac{\varphi\left(f^{-1}(y_1) - f^{-1}(y)\right)}{||y_1 - y||} = 0.$$

By the injectivity of f^{-1} we can write:

$$\frac{\varphi\left(f^{-1}(y_1) - f^{-1}(y)\right)}{||y_1 - y||} = \frac{\varphi\left(f^{-1}(y_1) - f^{-1}(y)\right)}{||f^{-1}(y_1) - f^{-1}(y)||} \frac{\left|\left|f^{-1}(y_1) - f^{-1}(y)\right|\right|}{||y_1 - y||}.$$

Fix $\epsilon > 0$. By (3.12) there exists δ such that if $0 < ||x_1 - x|| < \delta$, we have:

$$\frac{||\varphi(x_1 - x)||}{||x_1 - x||} < \frac{\epsilon}{1 + 2M}.$$

with M as in (3.10). On the other hand, since f^{-1} is continuous, there exists δ' such that if $0 < ||y_1 - y|| < \delta'$ then:

$$||x_1 - x|| = \left|\left|f^{-1}(y_1) - f^{-1}(y)\right|\right| < \delta.$$

We also know from (3.10) that:

$$\frac{\left|\left|f^{-1}(y_1) - f^{-1}(y)\right|\right|}{||y_1 - y||} \le 2M.$$

Putting this together, if $0 < ||y_1 - y|| < \delta'$, then:

$$\frac{\left|\left|\varphi\left(f^{-1}(y_1) - f^{-1}(y)\right)\right|\right|}{||y_1 - y||} < \frac{2M\epsilon}{1 + 2M} < \epsilon,$$

thus

$$\lim_{y_1 \to y} \frac{\varphi\left(f^{-1}(y_1) - f^{-1}(y)\right)}{||y_1 - y||} = 0,$$

and we're done. □

3.4.2 (*) Implicit function theorem

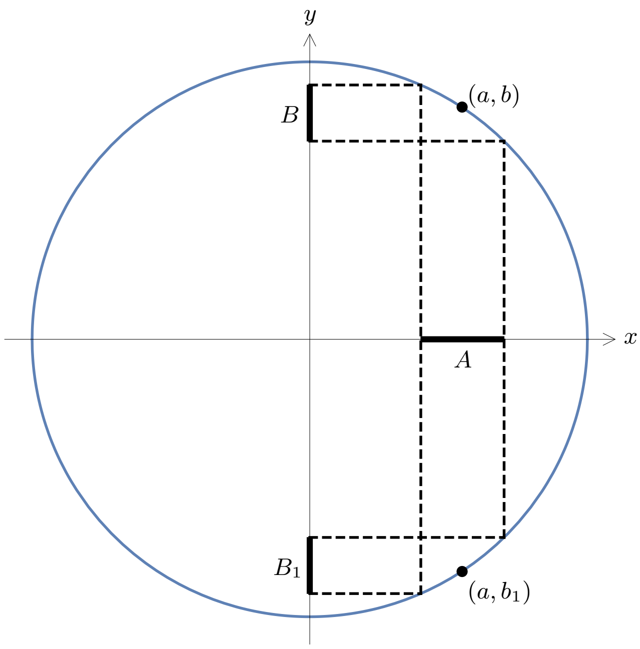

Figure 3.2 The set $x^2 + y^2 - 1 = 0$, and the intervals A, B, B_1.

A useful corollary of the inverse function theorem is the implicit function theorem. One can think of the inverse function theorem as providing a sufficient condition to be able to solve a system of equations for some variables in terms of others. To motivate this, let us consider the simple case of the equation:

$$x^2 + y^2 - 1 = 0.$$

We can think of this as the equation:

$$f(x, y) = 0,$$

where the function $f : \mathbb{R}^2 \to \mathbb{R}^2$ is given by:

$$f(x, y) = x^2 + y^2 - 1.$$

Now, suppose (a, b) satisfies $F(a, b) = 0$, and $a \neq 1, -1$. Then there is an open interval A containing a and an open interval B containing b with the property that for each $x \in A$ there is a unique $y \in B$ such that $F(x, y) = 0$. This permits us to define a function $g : A \to B$ by $g(x) = y$, so that $F(x, g(x)) = 0$. We can think of this as 'locally solving for y as a function of x'. If $b > 0$ then $g(x) = \sqrt{1 - x^2}$. For the problem at hand, there is in fact another number b_1 such that $F(a, b_1) = 0$. Associated to this point there is an open interval B_1 containing b_1 and a function $g_1 : A \to B_1$ such that $F(x, g_1(x)) = 0$. (If $b > 0$, then $b_1 < 0$ and $g_1(x) = -\sqrt{1 - x^2}$). Both g, g_1 are differentiable. See Figure 3.2.

The implicit function theorem tells us gives us a sufficient condition for the same to be true in a more general setting.

Theorem 3.16 (Implicit Function Theorem)**.** *Suppose $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ is continuously differentiable in an open set containing $p = (a, b)^t \in \mathbb{R}^n \times \mathbb{R}^m$. Suppose also that $f(p) = 0$. Let M be the $m \times m$ matrix:*

$$\left(D_{n+j} f^i(p)\right), \qquad 1 \leq i, j \leq m.$$

If $\det M \neq 0$ there is an open set $A \subset \mathbb{R}^n$ containing a, and an open set $B \subset \mathbb{R}^m$ containing b with the following property: for each $x \in A$ there is a unique $g(x) \in B$ such that

$$f(x, g(x)) = 0.$$

The function g is differentiable.

Proof. We define a new function $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m$ by $F(x, y) = (x, f(x, y))$. The Jacobian of F at $p = (a, b)$ is given by:

$$DF(p) = \left(\begin{array}{c|c} I & 0 \\ \hline N & M \end{array}\right)$$

Here I is the $n \times n$ identity matrix, M is as above and N is the $m \times n$ matrix with components:

$$\left(D_j f^i(p)\right), \qquad 1 \leq i \leq m, \quad 1 \leq j \leq n.$$

We have that $\det DF(p) \neq 0$, so we can apply the inverse function theorem to deduce the existence of open sets $U, V \subset \mathbb{R}^n \times \mathbb{R}^m$ with $(a, b) \in U$, $(a, 0) \in V$ such that $F : U \to V$ has a continuously differentiable inverse $h : V \to U$. By shrinking U if necessary, we can assume that $U = A \times B$ for some open $A \subset \mathbb{R}^n$, $B \subset \mathbb{R}^m$. The function h must have the form $h(x, y) = (x, k(x, y))$ for some continuously differentiable function k (since F has this form). Let $\pi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ be the projection map $\pi(x, y) = y$. Then $f = \pi \circ F$. Now, by the associativity of composition of functions, we deduce:

$$\begin{aligned} f(x, k(x, y)) &= f \circ h(x, y) = (\pi \circ F) \circ h(x, y) \\ &= \pi \circ (F \circ h)(x, y) = \pi(x, y) = y. \end{aligned}$$

Thus $f(x, k(x, 0)) = 0$, so we can take $g(x) = k(x, 0)$. □

Chapter 4

Integration in higher dimensions

4.1 Integration along a curve

4.1.1 Integration of a function along a curve

In many situations, we want to integrate over more interesting shapes than a simple interval. Consider for example the kind of integrals that appear in complex analysis: we want to integrate over some contour in the complex plane. We can in fact do this with the machinery that we've already established, but we first need to make a bit more precise what we mean by a curve.

Definition 4.1. A C^1-curve in $\mathbb{R}^n$ is a differentiable map $\gamma : [a, b] \to \mathbb{R}^n$ such that $\gamma' : (a, b) \to \mathbb{R}^n$ given by $\gamma'(p) = D_p\gamma$ is uniformly continuous and non-zero. If $\gamma(a) = \gamma(b)$, we say that the curve is closed.

Example 4.1. The map $\gamma : [0, 2\pi] \to \mathbb{R}^2$ given by:

$$\gamma(\theta) = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$$

is a closed C^1-curve.

We can integrate a function defined in the neighbourhood of a C^1-curve as follows. Suppose $\gamma : [a, b] \to \mathbb{R}^n$ is a C^1-curve, that $\Omega \subset \mathbb{R}^n$ with $\gamma([a, b]) \subset \Omega$ and that $f : \Omega \to \mathbb{R}$ is a continuous function. We define the integral of f along γ to be:

$$\int_\gamma f d\ell := \int_a^b f(\gamma(t)) \left|\left|\gamma'(t)\right|\right| dt.$$

Notice that the assumption that f is continuous and γ' is uniformly continuous is stronger than necessary: it is sufficient to assume that $(f \circ \gamma) \times ||\gamma'|| : [a, b] \to \mathbb{R}$ is integrable.

Example 4.2. Take γ to be the curve in Example 4.1 and suppose $f : \mathbb{R}^2 \setminus \{0\} \to \mathbb{R}$ is given by:

$$f(x, y) = \frac{x^2}{x^2 + y^2}$$

Then we calculate

$$f(\gamma(\theta)) = f(\cos\theta, \sin\theta) = \cos^2\theta,$$

and:

$$\gamma'(t) = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$$

so that $||\gamma'(t)|| = 1$ and:

$$\int_\gamma f d\ell = \int_0^{2\pi} cos^2\theta d\theta = \pi.$$

Suppose $\gamma : [a, b] \to \mathbb{R}^n$ is a C^1-curve, that $\Omega \subset \mathbb{R}^n$ with $\gamma([a, b]) \subset \Omega$ and that $f : \Omega \to \mathbb{R}$ is a continuous function. Suppose also that $s : [c, d] \to [a, b]$ satisfies $s(c) = a$, $s(d) = b$ and that $s' : (c, d) \to \mathbb{R}$ is uniformly continuous and positive. Then, by the mean value theorem, s is strictly monotone increasing and continuous, and hence maps $[c, d] \to [a, b]$ bijectively. If we define $\tilde{\gamma} = \gamma \circ s : [c, d] \to \mathbb{R}^n$, then $\tilde{\gamma}$ is a C^1-curve. Moreover,

$$\gamma([a, b]) = \tilde{\gamma}([c, d])$$

so that the curves γ, $\tilde{\gamma}$ coincide as subsets of $\mathbb{R}^n$. We say $\tilde{\gamma}$ is a C^1-*reparameterisation* of γ. By the chain rule, we have for any $t \in (c, d)$ that $\tilde{\gamma}'(t) = s'(t)\gamma'(s(t))$, so that:

$$||\tilde{\gamma}'(t)|| = s'(t) \, ||\gamma'(s(t))|| \, ,$$

using that $s'(t) > 0$. Now, we have:

$$\begin{aligned}
\int_{\tilde{\gamma}} f d\ell &= \int_c^d f(\tilde{\gamma}(t)) \, ||\tilde{\gamma}'(t)|| \, dt \\
&= \int_c^d f(\gamma(s(t))) \, ||\gamma'(s(t))|| \, s'(t) dt \\
&= \int_a^b f(\gamma(s)) \, ||\gamma'(s)|| \, ds = \int_\gamma f d\ell,
\end{aligned}$$

by making a change of variables. Thus, the integral along a C^1-curve depends on the curve, rather than the parameterisation. We define the length of a C^1-curve to be:

$$L(\gamma) := \int_\gamma 1 d\ell.$$

By the argument above, this definition of length doesn't depend on the parameterisation of the curve.

Exercise 9.1. Let $\gamma : [0, 2\pi] \to \mathbb{R}^2$ be given by:

$$\gamma(\theta) = \begin{pmatrix} \theta - \sin\theta \\ 1 - \cos\theta \end{pmatrix}.$$

a) Sketch γ.

b) Find

$$L(\gamma) = \int_\gamma 1 d\ell.$$

This curve is called the *cycloid*, and it is the path traced out by a point on a unit circle which rolls without slipping along the $x-$axis.

4.1.2 Integration of a vector field along a curve

In many physical and mathematical applications, we want to integrate *vectorial* quantities along a curve. For instance, the classical Ampère's law of electromagnetism relates the flux of a stationary current through a surface bounded by a curve to an integral of the magnetic field (a vector) around that curve. We can incorporate this idea into the machinery that we've already developed. Suppose that $\Omega \subset \mathbb{R}^n$ is an open set and that $\gamma : [a, b] \to \Omega$ is a C^1-curve. A *vector field*[1] defined on Ω is simply a map $V : \Omega \to \mathbb{R}^n$, which attaches a vector in $\mathbb{R}^n$ to each point of Ω.

If V is continuous on Ω, we define the integral of V along γ to be:

$$\int_\gamma \langle V, d\boldsymbol{\ell} \rangle = \int_a^b \big\langle V(\gamma(t)), \gamma'(t) \big\rangle \, dt,$$

Notice that since V and γ' are both vectors in $\mathbb{R}^n$, the inner product makes sense at each point $t \in [a, b]$.

As in the case of integration of a function along γ, we would like to show that this quantity doesn't depend on how we parameterise the curve. Let us again suppose that $s : [a, b] \to [c, d]$ satisfies $s(c) = a$, $s(d) = b$ and that $s' : (c, d) \to \mathbb{R}$ is uniformly continuous and positive. Defining $\tilde{\gamma} = \gamma \circ s : [c, d] \to \mathbb{R}^n$, we compute:

$$\begin{aligned}
\int_{\tilde{\gamma}} \langle V, d\boldsymbol{\ell} \rangle &= \int_c^d \big\langle V(\tilde{\gamma}(t)), \tilde{\gamma}'(t) \big\rangle \, dt \\
&= \int_c^d \big\langle V(\gamma(s(t))), \gamma'(s(t)) s'(t) \big\rangle \, dt \\
&= \int_c^d \big\langle V(\gamma(s(t))), \gamma'(s(t)) \big\rangle \, s'(t) dt \\
&= \int_a^b \big\langle V(\gamma(t)), \gamma'(t) \big\rangle \, dt = \int_\gamma \langle V, d\boldsymbol{\ell} \rangle .
\end{aligned}$$

so that the value of the integral of V along γ again depends on the curve, rather than the particular parameterisation that we choose.

Again, the condition that γ is C^1 and V is continuous is stronger than we really require to define the integral of V along γ. It's enough that $\langle V \circ \gamma, \gamma' \rangle : [a, b] \to \mathbb{R}$ is integrable.

[1]This definition is more than adequate for our purposes, but in more exciting situations where Ω is not an open subset of $\mathbb{R}^n$, but a more general manifold the definition of a vector field becomes more subtle. The Manifolds course next year will deal with these very interesting ideas.

Example 4.3. Let us take $\gamma : [0, 2\pi] \to \mathbb{R}^2$ to be given by:

$$\gamma(\theta) = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix},$$

so that γ is the unit circle traversed anticlockwise. We will take as our vector field $V : \mathbb{R}^2 \to \mathbb{R}^2$

$$V : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} -y \\ x \end{pmatrix}$$

Now, we calculate:

$$\gamma'(\theta) = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix},$$

and

$$V(\gamma(t)) = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix},$$

so that $\langle V(\gamma(t)), \gamma'(\theta)\rangle = 1$ and:

$$\int_\gamma \langle V, d\boldsymbol{\ell}\rangle = \int_0^{2\pi} 1 dt = 2\pi.$$

4.2 Integration over an area: the Darboux integral on $\mathbb{R}^2$

In this section we are going to consider how to define the integral of a function $f : A \to \mathbb{R}$, where $A \subset \mathbb{R}^2$. The restriction to $\mathbb{R}^2$ is for clarity and to keep the notation sensible. The same basic idea easily extends to integrating over functions of more than two variables. To begin with we will take A to be a rectangular domain: $A = [a_1, b_1] \times [a_2, b_2]$. This is the natural generalisation to two dimensions of the integral over an interval. We will again partition the domain of integration into smaller pieces and approximate the integral above and below by finite sums. This time, rather than representing the area of a set of rectangles, the sums will represent the volume of a set of cuboids, which approximate from above and below the volume below the graph of f. See Figures 4.1, 4.2 for a graphical depiction of this.

4.2.1 Partitions of a square

Suppose that $A = [a_1, b_1] \times [a_2, b_2]$. A partition $\mathcal{P}$ of A is a pair $(\mathcal{P}_1, \mathcal{P}_2)$, where $\mathcal{P}_1 = (x_0, \ldots, x_k)$ is a partition of $[a_1, b_1]$ and $\mathcal{P}_2 = (y_0, \ldots, y_l)$ is a partition of $[a_2, b_2]$. The partition $\mathcal{P}$ divides A into $k \times l$ subdomains:

$$\pi_{ij} = [x_i, x_{i+1}] \times [y_j, y_{j+1}], \qquad i = 0, \ldots, k-1, \quad j = 0, \ldots, l-1.$$

We define $|\pi_{ij}| := (x_{i+1} - x_i) \times (y_{j+1} - y_j)$ to be the area of the square π_{ij}. We also define

$$\text{mesh } \mathcal{P} = \max\{\text{mesh } \mathcal{P}_1, \text{mesh } \mathcal{P}_2\}.$$

We say that $\mathcal{Q} = (\mathcal{Q}_1, \mathcal{Q}_2)$ refines $\mathcal{P}$ if $\mathcal{P}_1 \preceq \mathcal{Q}_1$ *and* $\mathcal{P}_2 \preceq \mathcal{Q}_2$. In this case, each subdomain of $\mathcal{P}$ is split into a finite number of subdomains of $\mathcal{Q}$. An example of a

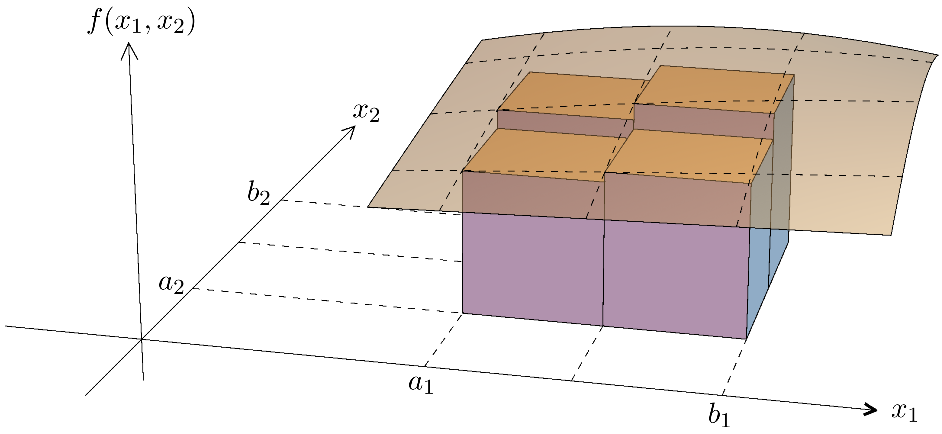

Figure 4.1 Approximating the volume beneath a graph over the plane

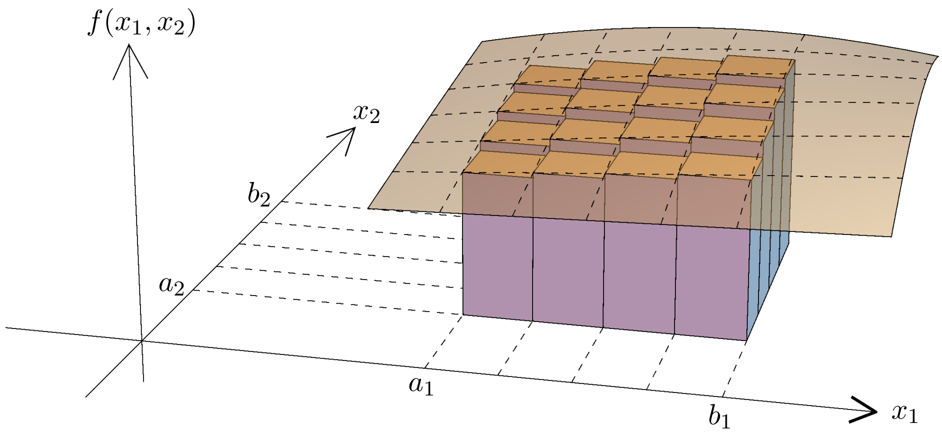

Figure 4.2 Approximating with a finer partition

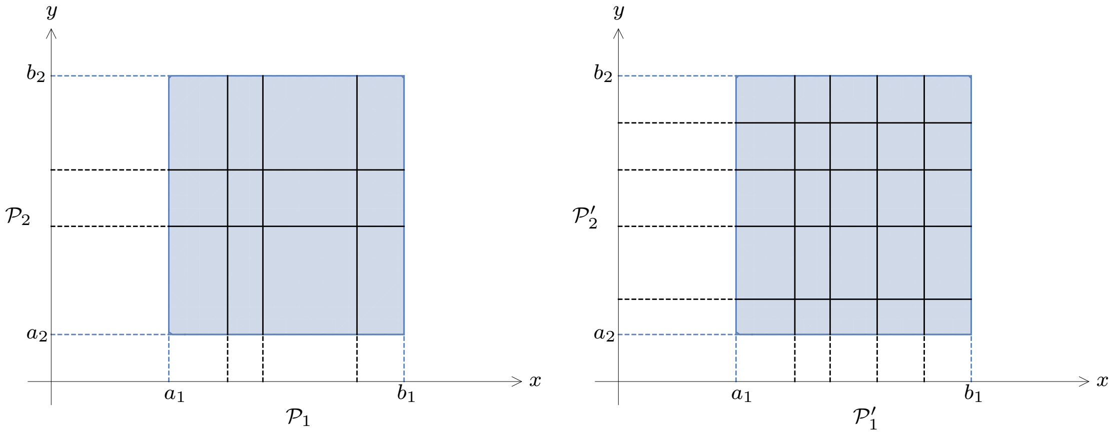

Figure 4.3 Two partitions $\mathcal{P} = (\mathcal{P}_1, \mathcal{P}_2)$ and $\mathcal{P}' = (\mathcal{P}'_1, \mathcal{P}'_2)$ of the domain A, with $\mathcal{P} \preceq \mathcal{P}'$.

partition and a refinement of that partition are shown in Figure 4.3. If $\mathcal{P} = (\mathcal{P}_1, \mathcal{P}_2)$ and $\mathcal{P}' = (\mathcal{P}'_1, \mathcal{P}'_2)$ are any two partitions, we define their common refinement to be $\mathcal{R} = (\mathcal{R}_1, \mathcal{R}_2)$, where $\mathcal{R}_1$ is the common refinement of $\mathcal{P}_1, \mathcal{P}'_1$ and $\mathcal{R}_2$ is the common refinement of $\mathcal{P}_2, \mathcal{P}'_2$.

4.2.2 Upper and Lower Darboux Sums

We now consider a bounded function $f : A \to \mathbb{R}$. For each $i = 0, \ldots, k-1$ and $j = 0, \ldots, l-1$, we define:

$$m_{ij} := \inf_{x \in \pi_{ij}} f(x),$$
$$M_{ij} := \sup_{x \in \pi_{ij}} f(x),$$

which are respectively the greatest lower and least upper bounds for f in the subdomain π_{ij}. As we've assumed f to be bounded these are well defined. We then define the lower and upper sums of f associated to the partition $\mathcal{P}$ is the obvious way:

$$L(f, \mathcal{P}) := \sum_{i=0}^{k-1} \sum_{j=0}^{l-1} m_{ij} \, |\pi_{ij}|$$
$$U(f, \mathcal{P}) := \sum_{i=0}^{k-1} \sum_{j=0}^{l-1} m_{ij} \, |\pi_{ij}| \, .$$

Abusing notation slightly, we'll sometimes write $\pi \in \mathcal{P}$ to mean π is a subdomain of $\mathcal{P}$. The allows us to write the sums more compactly as:

$$L(f, \mathcal{P}) = \sum_{\pi \in \mathcal{P}} \inf_{x \in \pi} f(x) \, |\pi|$$
$$U(f, \mathcal{P}) = \sum_{\pi \in \mathcal{P}} \sup_{x \in \pi} f(x) \, |\pi| \, .$$

where by the sum over $\pi \in \mathcal{P}$ we mean summing over all distinct subdomains of $\mathcal{P}$. As in the one-dimensional case, since $m_{ij} \leq M_{ij}$, we certainly have:

$$L(f, \mathcal{P}) \leq U(f, \mathcal{P}).$$

In fact, almost all of the results that we proved for upper and lower sums in the one-dimensional case carry over to the higher dimensional setting. In particular, we have:

Theorem 4.1. *Let $A = [a_1, b_1] \times [a_2, b_2]$ and $f : A \to \mathbb{R}$ be bounded. Suppose that $\mathcal{P}$, $\mathcal{Q}$ are partitions of A with $\mathcal{Q}$ a refinement of $\mathcal{P}$. Then:*

$$L(f, \mathcal{P}) \leq L(f, \mathcal{Q}) \leq U(f, \mathcal{Q}) \leq U(f, \mathcal{P}).$$

Proof. Write $\mathcal{P} = (\mathcal{P}_1, \mathcal{P}_2)$ and $\mathcal{Q} = (\mathcal{Q}_1, \mathcal{Q}_2)$. Suppose first that $\mathcal{P}_1 = (x_0, \ldots, x_k)$, $\mathcal{Q}_1 = (x_0, \ldots, x_l, c, x_{l+1}, \ldots x_k)$, and $\mathcal{P}_2 = \mathcal{Q}_2 = (y_0, \ldots, y_m)$, so that $\mathcal{P}$ and $\mathcal{Q}$ differ only in that a single subinterval $[x_l, x_{l+1}]$ in the $x-$direction has been split into $[x_l, c]$ and $[c, x_{l+1}]$, with the partition in the $y-$direction left alone. From basic properties of the infimum, we have:

$$\inf_{x \in [x_l, c] \times [y_j, y_{j+1}]} f(x) \geq \inf_{x \in [x_l, x_{l+1}] \times [y_j, y_{j+1}]} f(x),$$
$$\inf_{x \in [c, x_{l+1}] \times [y_j, y_{j+1}]} f(x) \geq \inf_{x \in [x_l, x_{l+1}] \times [y_j, y_{j+1}]} f(x),$$

for any $y = 0, \ldots, m-1$. Now, if we take the difference of the lower sums associated to $\mathcal{P}$ and $\mathcal{Q}$, then the only terms which will not cancel will be those involving the subinterval $[x_l, x_{l+1}]$, see Figure 4.4:

$$\begin{aligned}
L(f, \mathcal{Q}) - L(f, \mathcal{P}) &= \sum_{j=0}^{m-1} (c - x_l)\,(y_{j+1} - y_j) \inf_{x \in [x_l, c] \times [y_j, y_{j+1}]} f(x) \\
&\quad + \sum_{j=0}^{m-1} (x_{l+1} - c)\,(y_{j+1} - y_j) \inf_{x \in [c, x_{l+1}] \times [y_j, y_{j+1}]} f(x) \\
&\quad - \sum_{j=0}^{m-1} (x_{l+1} - x_l)\,(y_{j+1} - y_j) \inf_{x \in [x_l, x_{l+1}] \times [y_j, y_{j+1}]} f(x) \\
&\geq \sum_{j=0}^{m-1} \left[(c - x_l) + (x_{l+1} - c) - (x_{l+1} - x_l)\right] \\
&\qquad \times (y_{j+1} - y_j) \times \inf_{x \in [x_l, x_{l+1}] \times [y_j, y_{j+1}]} f(x), \\
&= 0.
\end{aligned}$$

Thus we have:

$$L(f, \mathcal{P}) \leq L(f, \mathcal{Q}).$$

Now, note that by Exercise 9.2 *a*), we have $U(f, \mathcal{P}) = -L(-f, \mathcal{P})$, so that:

$$U(f, \mathcal{Q}) = -L(-f, \mathcal{Q}) \leq -L(-f, \mathcal{P}) = U(f, \mathcal{P}),$$

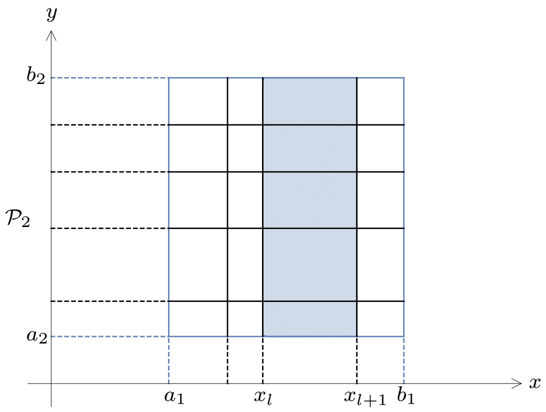

Figure 4.4 The partitions $\mathcal{P}$, $\mathcal{Q}$ from the proof of Theorem 4.1. The subdomains giving rise to terms in the upper and lower sums which do not cancel are shaded.

so that we have established

$$L(f,\mathcal{P}) \le L(f,\mathcal{Q}) \le U(f,\mathcal{Q}) \le U(f,\mathcal{P}).$$

in the special case where $\mathcal{Q}$ is obtained from $\mathcal{P}$ by adding one point in the $x-$partition. An identical argument establishes the result in the case where $\mathcal{P}_1 = \mathcal{Q}_1$ and $\mathcal{P}_2$, $\mathcal{Q}_2$ differ only by one point. For the general result, we observe that if $\mathcal{P} \preceq \mathcal{Q}$ then we can obtain $\mathcal{Q}$ by inserting a finite number of points into $\mathcal{P}$, and we can apply our special case result at each stage to obtain the general result. □

Corollary 4.2. *Let $A = [a_1, b_1] \times [a_2, b_2]$ and let $f : A \to \mathbb{R}$ be bounded. Suppose $\mathcal{P}, \mathcal{P}'$ are any two partitions of A. Then:*

$$L(f,\mathcal{P}) \le U(f,\mathcal{P}').$$

Proof. Let $\mathcal{Q}$ be the common refinement of $\mathcal{P}, \mathcal{P}'$. Then by Theorem 4.1 we have:

$$L(f,\mathcal{P}) \le L(f,\mathcal{Q}) \le U(f,\mathcal{Q}) \le U(f,\mathcal{P}'),$$

which gives the result. □

Exercise 9.2. Let $A = [a_1, b_1] \times [a_2, b_2]$ be a square domain in $\mathbb{R}^2$ and let $\mathcal{P}$ be any partition of A. Suppose $f, g : A \to \mathbb{R}$ are bounded functions, and that $\lambda \ge 0$. Show that:

a) $U(-f,\mathcal{P}) = -L(f,\mathcal{P})$,

b) $L(-f,\mathcal{P}) = -U(f,\mathcal{P})$,

c) $U(\lambda f,\mathcal{P}) = \lambda U(f,\mathcal{P})$,

d) $L(\lambda f,\mathcal{P}) = \lambda L(f,\mathcal{P})$,

e) $U(f+g,\mathcal{P}) \le U(f,\mathcal{P}) + U(g,\mathcal{P})$,

f) $L(f+g,\mathcal{P}) \ge L(f,\mathcal{P}) + L(g,\mathcal{P})$.

4.2.3 The Darboux integral

We are now ready to define the integral:

Definition 4.2. The *lower* and *upper Darboux integrals* are defined as follows:

$$\underline{\int_A} f(x)dx := \sup_{\mathcal{P}} L(f,\mathcal{P}) \qquad\qquad \overline{\int_A} f(x)dx := \inf_{\mathcal{P}} U(f,\mathcal{P})$$

where the supremum and infimum are taken over all partitions of $[a,b]$. By Corollary 4.2 we have that:

$$\underline{\int_A} f(x)dx \leq \overline{\int_A} f(x)dx.$$

We say that a bounded function $f:[a,b]\to\mathbb{R}$ is *Darboux integrable*, or simply *integrable*, if it is bounded, and the upper and lower Darboux integrals are equal. In this case, we write:

$$\int_A f(x)dx := \underline{\int_A} f(x)dx = \overline{\int_A} f(x)dx.$$

Notice that the lower and upper Darboux integrals are always well defined for a bounded function f.

As a notational convention, x appears in the integral as a 'dummy variable', much as in the one-dimensional case. Sometimes to emphasise that this is a two dimensional integral, it is written:

$$\int_A f(x)d^2x, \quad \text{or} \quad \int_A f(x)dx^1dx^2.$$

The results that we established in Chapter 2 concerning integrable functions defined on the interval carry over wholesale to the two dimensional setting. We'll state them again, but only repeat the proofs where some modification is necessary. You should look back at the proofs from Chapter 2 and check that you understand how they extend to the two dimensional case.

Theorem 4.3 (Linearity of the integral)**.** *Let $A=[a_1,b_1]\times[a_2,b_2]$. Suppose that $f:A\to\mathbb{R}$ and $g:A\to\mathbb{R}$ are both integrable, and that $\lambda\in\mathbb{R}$. Then $f+\lambda g$ is integrable and we have:*

$$\int_A (f(x)+\lambda g(x))\,dx = \int_A f(x)dx + \lambda\int_A g(x)dx.$$

Proof. Repeat proof of Theorem 2.4 □

Lemma 4.4. *Let $A=[a_1,b_1]\times[a_2,b_2]$. A function $f:A\to\mathbb{R}$ is integrable if and only if for any $\epsilon>0$ there exists a partition $\mathcal{P}$ such that:*

$$U(f,\mathcal{P}) - L(f,\mathcal{P}) < \epsilon. \tag{4.1}$$

Proof. Repeat proof of Lemma 2.5 □

Theorem 4.5 (Additivity of the integral)**.** *Let $A = [a_1, b_1] \times [a_2, b_2]$. Suppose $c \in (a_1, b_1)$ and set $A_L = [a_1, c] \times [a_2, b_2], A_R = [c, b_1] \times [a_2, b_2]$. Then $f : A \to \mathbb{R}$ is integrable if and only if $f|_{A_L}$ and $f|_{A_R}$ are integrable, and:*

$$\int_A f(x)dx = \int_{A_L} f(x)dx + \int_{A_R} f(x)dx.$$

Proof. Repeat proof of Theorem 2.6 □

Theorem 4.6. *Let $A = [a_1, b_1] \times [a_2, b_2]$. Suppose $f : A \to \mathbb{R}$ and $g : A \to \mathbb{R}$ are both integrable. Then so is:*

i) $|f|$,

ii) f^2,

iii) fg.

Proof. Repeat proof of Theorem 2.8 □

Theorem 4.7. *Let $A = [a_1, b_1] \times [a_2, b_2]$. Suppose that $f : A \to \mathbb{R}$ is continuous on A. Then f is integrable.*

Proof. This time we need to slightly modify the proof. Let $|A| := (b_1 - a_1) \times (b_2 - a_2)$. Fix $\epsilon > 0$. Since f is continuous on A, it is uniformly continuous. Thus there exists $\delta > 0$ such that for all $x, y \in A$ with $||x - y|| < \delta$ we have $|f(x) - f(y)| < \frac{1}{2|A|}\epsilon$. Pick a partition $\mathcal{P} = (\mathcal{P}_1, \mathcal{P}_2)$ with $\mathcal{P}_1 = (x_0, \dots, x_k)$ and $\mathcal{P}_2 = (y_0, \dots, y_l)$ such that:

$$\sqrt{(x_{i+1} - x_i)^2 + (y_{j+1} - y_j)^2} < \delta$$

for all $i = 0, \dots k - 1, j = 0, \dots, l - 1$. We can do this by (for example) setting:

$$x_i = a_1 + \frac{b_1 - a_1}{k} i, \qquad i = 0, \dots, k,$$
$$y_j = a_2 + \frac{b_2 - a_2}{l} j, \qquad j = 0, \dots, l,$$

where k, l are chosen sufficiently large that

$$(b_1 - a_1)\sqrt{2} < k\delta, \qquad (b_2 - a_2)\sqrt{2} < l\delta.$$

Now, by construction, for $x, y \in \pi_{ij}$ we have:

$$|f(x) - f(y)| < \frac{1}{2|A|}\epsilon,$$

which implies that:

$$\sup_{x \in \pi_{ij}} f(x) - \inf_{x \in \pi_{ij}} f(x) \le \frac{1}{2|A|}\epsilon < \frac{1}{|A|}\epsilon.$$

Now, we can estimate:

$$\begin{aligned}
U(f,\mathcal{P}) - L(f,\mathcal{P}) &= \sum_{i=0}^{k-1}\sum_{j=0}^{l-1}\left[\sup_{x\in\pi_{ij}} f(x) - \inf_{x\in\pi_{ij}} f(x)\right]|\pi_{ij}| \\
&< \sum_{i=0}^{k-1}\sum_{j=0}^{l-1}\frac{\epsilon}{|A|}|\pi_{ij}| \\
&= \frac{\epsilon}{|A|}\sum_{i=0}^{k-1}\sum_{j=0}^{l-1}(x_{i+1}-x_i)\times(y_{j+1}-y_j) \\
&= \frac{\epsilon}{|A|}\sum_{i=0}^{k-1}(x_{i+1}-x_i)\times\sum_{j=0}^{l-1}(y_{j+1}-y_j) \\
&= \epsilon,
\end{aligned}$$

as we have a pair of telescoping sums. Thus for any $\epsilon > 0$, we have found a partition $\mathcal{P}$ such that:

$$U(f,\mathcal{P}) - L(f,\mathcal{P}) < \epsilon,$$

which implies f is integrable. □

Again, notice that it's enough that A is uniformly continuous on $(a_1, b_1) \times (a_2, b_2)$ for this argument to work. Finally, we will state without proof the following result concerning Riemann sums. A tagged partition is a partition $\mathcal{P}$ together with a choice of a point in each subdomain $\tau = \{t_\pi | \pi \in \mathcal{P}\}$. The Riemann sum corresponding to the tagged partition $(\mathcal{P}, \tau)$ is:

$$R(f,\mathcal{P},\tau) = \sum_{\pi\in\mathcal{P}} f(t_\pi)\,|\pi|\,.$$

Theorem 4.8. *Let $A = [a_1, b_1] \times [a_2, b_2]$ and let $f : [a, b] \to \mathbb{R}$ be an integrable function. Suppose $((\mathcal{P}_l, \tau_l))_{l=0}^{\infty}$ is a sequence of tagged partitions with mesh $\mathcal{P}_l \to 0$. Then we have:*

$$R(f,\mathcal{P}_l,\tau_l) \to \int_A f(x)dx.$$

Exercise 9.3. Let $A = [-1, 1] \times [-1, 1]$. Show that the function $f : A \to \mathbb{R}$ given by:

$$f(x,y) = \begin{cases} 1 & x \notin \mathbb{Q}, \text{ and } y = 0, \\ 0 & \text{otherwise.} \end{cases}$$

is integrable, and:

$$\int_A f(z)dz = 0.$$

Exercise 9.4. Let $A = [a_1, b_1] \times [a_2, b_2]$ and set $|A| = (b_1 - a_1) \times (b_2 - a_2)$. Suppose $f : A \to \mathbb{R}$, $g : A \to \mathbb{R}$ are integrable.

a) Show that:

$$|A| \inf_{x\in A} f(x) \leq \int_A f(x)dx \leq |A| \sup_{x\in A} f(x).$$

b) Establish the estimate:

$$\left|\int_A f(x)dx\right| \leq |A| \sup_{x\in A} |f(x)|$$

c) Show that if $0 \leq f(x)$ for all $x \in A$ then:

$$0 \leq \int_A f(x)dx$$

d) Show that if $f(x) \leq g(x)$ for all $x \in A$ then:

$$\int_A f(x)dx \leq \int_A g(x)dx$$

e) Prove that:

$$\left|\int_A f(x)dx\right| \leq \int_A |f(x)|\, dx$$

Exercise 9.5. Let $A = [a_1, b_1] \times [a_2, b_2]$. Suppose that $f : A \to \mathbb{R}$ is continuous. Show that the map $F : [a_2, b_2] \to \mathbb{R}$ given by:

$$F(y) = \int_{a_1}^{b_1} f(x, y)dx$$

is continuous.

4.2.4 Relation to repeated integrals

When it comes to evaluating an integral over a square domain in practice, it is useful to connect the integral to repeated one-dimensional integrals. Roughly speaking, once can imagine first sending the mesh of the partition in the $x-$direction to zero, and then the mesh of the partition in the $y-$direction. More concretely, if $A = [a_1, b_1] \times [a_2, b_2]$ and $f : A \to \mathbb{R}$ is continuous, for each $y \in [a_2, b_2]$ we can consider the function:

$$\begin{array}{rccc} g_y & : & [a_1, b_1] & \to & \mathbb{R} \\ & & x & \mapsto & f(x, y). \end{array}$$

For each y this will be a continuous function, so we can define:

$$G_y = \int_{a_1}^{b_1} g_y(x)dx = \int_{a_1}^{b_1} f(x, y)dx$$

By Exercise 9.5 we have that the map $G : y \mapsto G_y$ is continuous, and so it makes sense to compute:

$$I_{12} := \int_{a_2}^{b_2} G(y)dy = \int_{a_2}^{b_2} \left(\int_{a_1}^{b_1} f(x, y)dx\right) dy.$$

Similarly, we could first integrate over y to find:

$$I_{21} := \int_{a_1}^{b_1} \left(\int_{a_2}^{b_2} f(x,y)dy \right) dx$$

One might expect that these double integrals are related to

$$I = \int_A f(z)dz,$$

and indeed we shall show:

$$I_{12} = I_{21} = I.$$

For a function $f : A \to \mathbb{R}$ which is merely integrable (rather than continuous) we have the problem that the integral over x may not be defined at all values of y (and vice versa). For example, the function in Exercise 9.3 is integrable over $[-1,1] \times [-1,1]$, however:

$$x \mapsto f(x,0)$$

is not integrable, so the expression:

$$\int_{-1}^{1} \left(\int_{-1}^{1} f(x,y)dx \right) dy$$

does not make sense. We can fix this by replacing the inner integral with an upper (or lower) integral.

Theorem 4.9 (Fubini's theorem)**.** *Let $A = [a_1, b_1] \times [a_2, b_2]$ and suppose $f : A \to \mathbb{R}$ is integrable. Then the two maps $\mathcal{U}, \mathcal{L} : [a_2, b_2] \to \mathbb{R}$ given by:*

$$\mathcal{U} : y \mapsto \overline{\int_{a_1}^{b_1}} f(x,y)dx,$$
$$\mathcal{L} : y \mapsto \underline{\int_{a_1}^{b_1}} f(x,y)dx,$$

are both integrable, and:

$$\int_A f(z)dz = \int_{a_2}^{b_2} \left(\underline{\int_{a_1}^{b_1}} f(x,y)dx \right) dy = \int_{a_2}^{b_2} \left(\overline{\int_{a_1}^{b_1}} f(x,y)dx \right) dy.$$

Proof (*). Suppose $\mathcal{P} = (\mathcal{P}_1, \mathcal{P}_2)$ be a partition of A with $\mathcal{P}_1 = (x_0, \dots, x_k)$ and $\mathcal{P}_2 = (y_0, \dots, y_l)$.

1. We first estimate:

$$\begin{aligned} L(f, \mathcal{P}) &= \sum_{\pi \in \mathcal{P}} \inf_{z \in \pi} f(z) \, |\pi| \\ &= \sum_{j=0}^{l-1} \left(\sum_{i=0}^{k-1} \inf_{z \in \pi_{ij}} f(z)(x_{i+1} - x_i) \right) (y_{j+1} - y_j) \end{aligned}$$

Note that if $\pi_{ij} = [x_i, x_{i+1}] \times [y_j, y_{j+1}]$ then:

$$\inf_{z\in\pi_{ij}} f(z) \leq \inf_{x\in[x_{i+1},x_i]} f(x,y), \qquad \text{for all } y \in [y_{j+1}, y_j].$$

So that:

$$\begin{aligned}\sum_{i=0}^{k-1} \inf_{z\in\pi_{ij}} f(z)(x_{i+1}-x_i) &\leq \sum_{i=0}^{k-1} \inf_{x\in[x_{i+1},x_i]} f(x,y)(x_{i+1}-x_i)\\ &= L(g_y, \mathcal{P}_1)\\ &\leq \underline{\int_{a_1}^{b_1}} f(x,y)dx = \mathcal{L}(y).\end{aligned}$$

for all $y \in [y_{j+1}, y_j]$. Taking the infimum over $y \in [y_{j+1}, y_j]$, we have:

$$\sum_{i=0}^{k-1} \inf_{z\in\pi_{ij}} f(z)(x_{i+1}-x_i) \leq \inf_{y\in[y_{j+1},y_j]} \mathcal{L}(y).$$

We deduce that:

$$L(f,\mathcal{P}) \leq \sum_{j=0}^{l-1} \inf_{y\in[y_{j+1},y_j]} \mathcal{L}(y)(y_{j+1}-y_j) = L(\mathcal{L}, \mathcal{P}_2).$$

2. Next we estimate:

$$\begin{aligned}U(f,\mathcal{P}) &= \sum_{\pi\in\mathcal{P}} \sup_{z\in\pi} f(z)\,|\pi|\\ &= \sum_{j=0}^{l-1}\left(\sum_{i=0}^{k-1} \sup_{z\in\pi_{ij}} f(z)(x_{i+1}-x_i)\right)(y_{j+1}-y_j)\end{aligned}$$

Note that if $\pi_{ij} = [x_i, x_{i+1}] \times [y_j, y_{j+1}]$ then:

$$\sup_{z\in\pi_{ij}} f(z) \geq \sup_{x\in[x_{i+1},x_i]} f(x,y), \qquad \text{for all } y \in [y_{j+1}, y_j].$$

So that:

$$\begin{aligned}\sum_{i=0}^{k-1} \sup_{z\in\pi_{ij}} f(z)(x_{i+1}-x_i) &\geq \sum_{i=0}^{k-1} \sup_{x\in[x_{i+1},x_i]} f(x,y)(x_{i+1}-x_i)\\ &= U(g_y, \mathcal{P}_1)\\ &\geq \overline{\int_{a_1}^{b_1}} f(x,y)dx = \mathcal{U}(y).\end{aligned}$$

for all $y \in [y_{j+1}, y_j]$. Taking the supremum over $y \in [y_{j+1}, y_j]$, we have:

$$\sum_{i=0}^{k-1} \sup_{z \in \pi_{ij}} f(z)(x_{i+1} - x_i) \geq \sup_{y \in [y_{j+1}, y_j]} \mathcal{U}(y).$$

We deduce that:

$$U(f, \mathcal{P}) \geq \sum_{j=0}^{l-1} \sup_{y \in [y_{j+1}, y_j]} \mathcal{U}(y)(y_{j+1} - y_j) = U(\mathcal{U}, \mathcal{P}_2).$$

3. Recalling that $\mathcal{L}(y) \leq \mathcal{U}(y)$ for all $y \in [a_2, b_2]$ we have:

$$L(f, \mathcal{P}) \leq L(\mathcal{L}, \mathcal{P}_2) \leq U(\mathcal{L}, \mathcal{P}_2) \leq U(\mathcal{U}, \mathcal{P}_2) \leq U(f, \mathcal{P}).$$

Since f is integrable, we have:

$$\inf_{\mathcal{P}} U(f, \mathcal{P}) = \sup_{\mathcal{P}} L(f, \mathcal{P}) = \int_A f(z)dz.$$

We deduce that:

$$\inf_{\mathcal{P}_2} U(\mathcal{L}, \mathcal{P}_2) = \sup_{\mathcal{P}_2} L(\mathcal{L}, \mathcal{P}_2) = \int_A f(z)dz,$$

so that $\mathcal{L}$ is integrable and:

$$\int_A f(z)dz = \int_{a_2}^{b_2} \left(\underline{\int_{a_1}^{b_1}} f(x,y)dx \right) dy.$$

The equivalent statement for $\mathcal{U}$ follows from the inequalitites:

$$L(f, \mathcal{P}) \leq L(\mathcal{L}, \mathcal{P}_2) \leq L(\mathcal{U}, \mathcal{P}_2) \leq U(\mathcal{U}, \mathcal{P}_2) \leq U(f, \mathcal{P}).$$ □

Obviously, there was nothing special about our decision to integrate first over the $x-$coordinate and then the y. We could equally have established that:

$$\int_A f(z)dz = \int_{a_1}^{b_1} \left(\underline{\int_{a_2}^{b_2}} f(x,y)dy \right) dx = \int_{a_1}^{b_1} \left(\overline{\int_{a_2}^{b_2}} f(x,y)dy \right) dx.$$

In the case where f is continuous, we have the immediate corollary:

Corollary 4.10. *Suppose $f : A \to \mathbb{R}$ is continuous. Then:*

$$\int_A f(z)dz = \int_{a_1}^{b_1} \left(\int_{a_2}^{b_2} f(x,y)dy \right) dx = \int_{a_2}^{b_2} \left(\int_{a_1}^{b_1} f(x,y)dx \right) dy.$$

Proof. Since f is continuous, so is the map $g_y : x \mapsto f(x, y)$ for each $y \in [a_2, b_2]$, thus g_y is integrable and we have:

$$\underline{\int_{a_1}^{b_1}} f(x,y)dx = \int_{a_1}^{b_1} f(x,y)dx,$$

and equivalently for the integrals keeping x fixed. □

4.2.5 Non-square domains

Thus far we have restricted our attention to functions define on a rectangle $A = [a_1, b_1] \times [a_2, b_2]$. Of course, we often wish to integrate over more exotic shapes (circles, ellipses, stars, etc.). Recall the characteristic function of a set $C \subset \mathbb{R}^2$ is the function $\chi_C : \mathbb{R}^2 \to \mathbb{R}$ given by:

$$\chi_C(x) = \begin{cases} 1 & x \in C \\ 0 & x \notin C, \end{cases}$$

If $C \subset A$ for some closed rectangle and $f : A \to \mathbb{R}$ is bounded, then:

$$\int_C f(x)dx = \int_A \chi_C(x)f(x)dx,$$

provided $\chi_C(x)f(x)$ is integrable. By Theorem 4.6 this will certainly be the case if χ_C and f are separately integrable.

Definition 4.3. A bounded set C is called *Jordan measurable* if χ_C is integrable. For such a set we can define the *Jordan measure* to be:

$$m(C) = \int_C 1dx = \int_A \chi_C(x)dx,$$

where A is any rectangle enclosing C.

The Jordan measure is a notion of area for a reasonably large class of sets. Not every set is Jordan measurable. For example, if we take

$$C = \left\{(x, y)^t \middle| -1 < x, y < 1, \quad x, y \in \mathbb{Q}\right\}$$

then by a similar argument to Example 2.3 we can show that χ_C is not integrable. In fact, in the next section we shall see that the criterion that a set be Jordan measurable is that the boundary be 'small' in some appropriate sense.

A word of caution. Although the nomenclature 'Jordan measure' is reasonably established, one should be careful because the word "measure" has a technical meaning which you will be introduced to in the measure theory course, however the Jordan measure is not a measure in this sense.

4.2.6 (*) Lebesgue's criterion for integrability

Sets of measure zero

A domain π is a closed rectangle if it is of the form $\pi = [a_1, b_1] \times [a_2, b_2]$ and if so, then $|\pi| = (b_1 - a_1) \times (b_2 - a_2)$. A set $A \subset \mathbb{R}^2$ is said to have *measure zero* if the following property holds: given $\epsilon > 0$, we can find a countable collection of rectangular domains $\{\pi_i\}_{i=1}^{\infty}$ such that:

$$A \subset \bigcup_{i=1}^{\infty} \pi_i,$$

and:

$$\sum_{i=1}^{\infty} |\pi_i| < \epsilon.$$

A set has measure zero if it can be completely contained in a set of arbitrarily small area.

Example 4.4. Any countable set $A \subset \mathbb{R}^2$ has measure zero. To see this, we write $A = \{x_1, x_2, \ldots\}$ with $x_i = (x_i^1, x_i^2)^t \in \mathbb{R}^2$. Let us then define:

$$\pi_i = [x_i^1 - \epsilon 2^{-\frac{i}{2}-2}, x_i^1 + \epsilon 2^{-\frac{i}{2}-2}] \times [x_i^2 - \epsilon 2^{-\frac{i}{2}-1}, x_i^2 + \epsilon 2^{-\frac{i}{2}-1}]$$

We have that:

$$|\pi_i| = \epsilon 2^{-i-1}$$

so that:

$$\sum_{i=1}^{\infty} |\pi_i| = \frac{\epsilon}{2} < \epsilon.$$

In fact, we can make a more general statement

Theorem 4.11. *Suppose that $A_1, A_2, \ldots$ is a countable collection of sets of measure zero. Then*

$$A = \bigcup_{j=1}^{\infty} A_j$$

has measure zero.

Proof. Fix $\epsilon > 0$. Since each A_j has measure zero, we can find a countable collection $\mathcal{U}_j = \{\pi_{j,i}\}_{i=1}^{\infty}$ of closed rectangles such that:

$$A_j \subset \bigcup_{i=1}^{\infty} \pi_{j,i},$$

and

$$\sum_{i=1}^{\infty} |\pi_{j,i}| < \epsilon 2^{-j-1}.$$

Now, the set:

$$\mathcal{U} = \bigcup_{j=1}^{\infty} \mathcal{U}_j$$

is a countable union of countable sets, so it is countable. We can therefore write $\mathcal{U} = \{\Pi_k\}_{k=1}^{\infty}$. We certainly have that:

$$A \subset \bigcup_{k=1}^{\infty} \Pi_k,$$

and moreover:

$$\sum_{k=1}^{\infty} |\Pi_k| = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} |\pi_{j,i}| = \sum_{j=1}^{\infty} \epsilon 2^{-j-1} = \frac{\epsilon}{2} < \epsilon.$$ □

Theorem 4.12. *Suppose $\gamma : [a, b] \to \mathbb{R}^2$ is a C^1-curve. Then $\gamma([a, b])$ has measure zero.*

Proof. We write:

$$\gamma(s) = \begin{pmatrix} x(s) \\ y(s) \end{pmatrix}$$

for $x, y : [a, b] \to \mathbb{R}$. First, note that since $\gamma'(a, b) \to \mathbb{R}^2$ is uniformly continuous, so are $x', y' : (a, b) \to \mathbb{R}$ and so x', y' extend continuously to $[a, b]$ by Exercise 3.6. By the extreme value theorem, we conclude that both x' and y' are bounded, so that there exists K with:

$$\left|x'(s)\right| \leq K, \qquad \left|y'(s)\right| \leq K.$$

By Exercise 2.6 we deduce that if $s_1, s_2 \in [a, b]$, then:

$$|x(s_1) - x(s_2)| \leq K |s_1 - s_2|, \qquad |y(s_1) - y(s_2)| \leq K |s_1 - s_2|$$

Fix $\epsilon > 0$. Let us define for $n \in \mathbb{N}^*$:

$$s_i = a + (b - a)\frac{i}{n}, \qquad i = 0, \ldots, n.$$

We define the rectangles:

$$\pi_i = [x(s_i) - k, x(s_i) + k] \times [y(s_i) - k, y(s_i) + k],$$

where:

$$k = K\frac{(b - a)}{n}.$$

Now, for any $s \in [a, b]$, there is an s_i with $|s - s_i| < \frac{(b-a)}{n}$. If this is the case, we know that:

$$|x(s) - x(s_i)| \leq K\frac{(b - a)}{n}, \qquad |y(s) - y(s_i)| \leq K\frac{(b - a)}{n},$$

which implies that $\gamma(s) \in \pi_i$. Thus we have:

$$\gamma([a, b]) \subset \bigcup_{i=0}^{n} \pi_i.$$

On the other hand,

$$|\pi_i| = 4K^2\frac{(b - a)^2}{n^2}$$

so that:

$$\sum_{i=0}^{n} |\pi_i| = 4K^2\frac{(b - a)^2}{n^2}(n + 1).$$

By taking n sufficiently large, we can ensure that the right hand side of this expression is as small as we like, in particular less than ϵ, and so we have exhibited a collection of rectangles which contains $\gamma([a, b])$ and has total area less than ϵ. □

Lebesgue's criterion

We shall state without proof a very useful characterisation of the integrable functions.

Theorem 4.13 (Lebesgue's criterion for integrability)**.** *Let $A = [a_1, b_1] \times [a_2, b_2]$ and let $f : A \to \mathbb{R}$ be a bounded function. Then f is integrable if, and only if, the set:*

$$\big\{p \in A \big| f \textit{ is } \text{not } \textit{continuous at } p\big\}$$

has measure zero.

Those students interested in seeing the proof of this result can find it in the book "Calculus on Manifolds" by Spivak, Theorem 3-8.

A very useful consequence of this result is the following:

Corollary 4.14. *A set $B \subset A = [a_1, b_1] \times [a_2, b_2]$ is Jordan measurable if, and only if, ∂B has measure zero.*

Proof. We need to show that $\chi_B : A \to \mathbb{R}$ is integrable. By Lebesgue's criterion, this is true if and only if the set of points at which χ_B is discontinuous has measure zero. Thus what we need to show is that χ_B is not continuous at p if and only if $p \in \partial B$.

First, suppose that $p \in B^\circ$. Then there exists $\delta > 0$ such that $B_\delta(p) \in B$, and thus $\chi_B(q) = 1$ for any q with $||q - p|| < \delta$. For such q, we have:

$$|\chi_B(q) - \chi_B(p)| = 0,$$

so χ_B is continuous at p.

Next, suppose that $p \in \overline{B}^c$. Since the complement of a closed set is open, there exists $\delta > 0$ such that $B_\delta(p) \subset \overline{B}^c$. Thus if $|q - p| < \delta$ we have $\chi_B(p) = \chi_B(q) = 0$ and again χ_B is continuous at p. Thus χ_B is continuous on $\overline{B}^c$ and B°. It remains to show that χ_B is discontinuous on $\partial B = \overline{B} \setminus B^\circ$.

Finally, suppose that $p \in \partial B$. Then since $p \notin B^\circ$, for any $\delta > 0$ we have that $B_\delta(p) \cap B^c \neq \emptyset$. In particular, for any $i \in \mathbb{N}$ we can find q_i with $q_i \notin B$ and $||q_i - p|| < 2^{-i}$. Thus $q_i \to p$ as $n \to \infty$. We have:

$$\chi_B(q_i) = 0 \to 0,$$

as $n \to \infty$. On the other hand, since $p \in \overline{B}$, there exists a sequence $(q_i)_{i=0}^{\infty}$ with $q'_i \in B$ and $q'_i \to p$. For this sequence we have:

$$\chi_B(q'_i) = 1 \to 1,$$

which implies that χ_B is not continuous at p. □

With this result, we are in a position to show that many 'reasonable' domains are Jordan integrable. For example any domain whose boundary can be (piecewise) parameterised by a C^1-curve will necessarily be Jordan integrable.

4.2.7 (*) Generalisation to $\mathbb{R}^n$

It is clear that the same basic approach we've described above can be generalised to integrate over a hypercube $A = [a_1, b_1] \times \cdots \times [a_n, b_n] \subset \mathbb{R}^n$. The only real challenge is in handling the notation, which gets very cumbersome very quickly. We define a partition of A to be a $n-$tuple of partitions $\mathcal{P} = (\mathcal{P}^1, \ldots, \mathcal{P}^n)$, where $\mathcal{P}^i = (x^i_0, \ldots, x^i_{k_i})$ is a partition of the interval $[a_i, b_i]$. The subdomains of $\mathcal{P}$ are the hypercubes of the form:

$$\pi_{i_1,\ldots,i_n} = [x^1_{i_1}, x^1_{i_1+1}] \times [x^2_{i_2}, x^2_{i_2+1}] \times \cdots \times [x^n_{i_n}, x^n_{i_n+1}],$$

where $0 \leq i_j \leq k_i - 1$. We define

$$|\pi_{i_1,\ldots,i_n}| = (x^1_{i_1+1} - x^1_{i_1})(x^2_{i_2+1} - x^2_{i_2}) \cdots (x^n_{i_n+1} - x^n_{i_n})$$

For $f : A \to \mathbb{R}$ a bounded function we define the lower and upper sums with respect to $\mathcal{P}$ in the obvious way:

$$L(f, \mathcal{P}) = \sum_{\pi \in \mathcal{P}} \inf_{x \in \pi} f(x) \, |\pi| \, ,$$
$$U(f, \mathcal{P}) = \sum_{\pi \in \mathcal{P}} \sup_{x \in \pi} f(x) \, |\pi| \, ,$$

where the sums are understood to be taken over all subdomains $\pi_{i_1,\ldots,i_n}$ of $\mathcal{P}$. From here, we can proceed as in Section 4.2.3, defining integrable functions and establishing the obvious generalisations of the results we found for the integral in two dimensions.

4.3 Green's Theorem in the plane

When we considered the integral in one dimension, one of the crucial results we obtained was the fundamental theorem of calculus, in particular Corollary 2.14. This states that if $f : [a, b] \to \mathbb{R}$ is continuous and $F : [a, b] \to \mathbb{R}$ satisfies $F'(x) = f(x)$ for all all $x \in (a, b)$ then:

$$\int_a^b f(t)dt = F(b) - F(a).$$

This result is the simplest of a family of results which relate integrals over a region $A \subset \mathbb{R}^n$ with quantities evaluated on the boundary ∂A. The general result is known as Stokes' Theorem, and is a deep and very important result in differential geometry. We shall establish a version for planar regions, known as Green's Theorem in the plane. This states that certain integrals over two-dimensional regions can be related to line integrals around the boundary of the region. We won't make any attempt to prove an optimal result.

4.3.1 Proof for simple domains

We say that a domain $A \subset \mathbb{R}^2$ is of *Type I* if there exist $a, b \in \mathbb{R}$ and continuous functions $\alpha, \beta : [a, b] \to \mathbb{R}$ such that:

$$A = \{(x, y)^t : a \leq x \leq b, \ \alpha(x) \leq y \leq \beta(x)\},$$

and moreover α and β are differentiable at all $x \in (a, b)$ with $\alpha', \beta' : (a, b) \to \mathbb{R}$ piecewise continuous.

Similarly, a domain $A \subset \mathbb{R}^2$ is of *Type II* if there exist $c, d \in \mathbb{R}$ and continuous functions $\lambda, \mu : [c, d] \to \mathbb{R}$ such that:

$$A = \{(x, y)^t : \lambda(y) \leq x \leq \mu(y), c \leq y \leq d\}.$$

and moreover λ and μ are differentiable at all $x \in (c, d)$ with $\lambda', \mu' : (c, d) \to \mathbb{R}$ piecewise continuous. See Figure 4.5 for examples of Type I and Type II domains.

A domain is *simple* if it is of both Type I and Type II, see Figure 4.6. We shall assert without proof the fact that in all three cases, A is Jordan measurable, so:

$$\int_A f(x)dx$$

makes sense whenever f is continuous on A

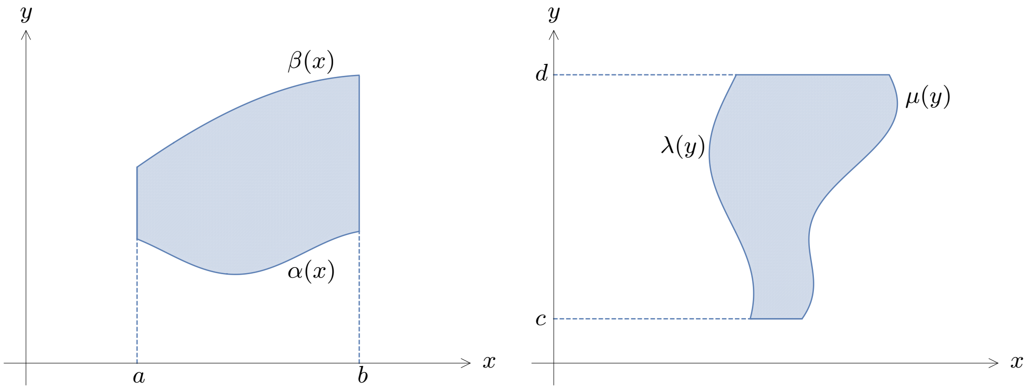

Figure 4.5 A type I (left) and type II (right) domain.

Lemma 4.15. *Suppose that $\Omega \subset \mathbb{R}^2$ is an open set, and that $A \subset \Omega$ is a type I domain. Suppose that $f : \Omega \to \mathbb{R}$ is differentiable on Ω, with continuous partial derivatives. Then:*

$$\int_A D_2 f(x)dx = \int_a^b f(t, \beta(t))dt - \int_a^b f(t, \alpha(t))dt.$$

Proof. We first apply Fubini's theorem to reduce the integral over A to a repeated integral:

$$\int_A D_2 f(x)dx = \int_a^b \left(\int_{\alpha(t)}^{\beta(t)} D_2 f(t, s)ds \right) dt. \tag{4.2}$$

Recalling that:

$$D_2 f(t, s) = \lim_{h \to 0} \frac{f(t, s + h), -f(t, s)}{h},$$

we note that the first integral may be written:

$$\int_{\alpha(t)}^{\beta(t)} D_2 f(t, s)ds = \int_{\alpha(t)}^{\beta(t)} F_t'(s)ds,$$

where $F_t : [\alpha(t), \beta(t)] \to \mathbb{R}$ is the map $s \mapsto f(t,s)$. Applying the fundamental theorem of calculus, we have:

$$\int_{\alpha(t)}^{\beta(t)} D_2 f(t,s) ds = f(t, \beta(t)) - f(t, \alpha(t)).$$

Inserting this into (4.2), we deduce

$$\int_A D_2 f(x) dx = \int_a^b \left(f(t, \beta(t)) - f(t, \alpha(t)) \right) dt,$$

and the result follows. □

Lemma 4.16. *Suppose that $\Omega \subset \mathbb{R}^2$ is an open set, and that $A \subset \Omega$ is a type II domain. Suppose that $g : \Omega \to \mathbb{R}$ is differentiable on Ω, with continuous partial derivatives. Then:*

$$\int_A D_1 g(x) dx = \int_c^d g(\mu(s), s) ds - \int_c^d g(\lambda(s), s) ds.$$

Proof. The proof proceeds almost exactly as in the previous Lemma. We first apply Fubini's theorem to reduce the integral over A to a repeated integral:

$$\int_A D_1 g(x) dx = \int_c^d \left(\int_{\lambda(s)}^{\mu(s)} D_1 g(t,s) dt \right) ds. \tag{4.3}$$

Recalling that:

$$D_1 g(t,s) = \lim_{h \to 0} \frac{g(t+h, s), -g(t,s)}{h},$$

we note that the first integral may be written:

$$\int_{\lambda(s)}^{\mu(s)} D_1 g(t,s) dt = \int_{\lambda(s)}^{\mu(s)} G_s'(t) dt,$$

where $G_s : [\lambda(s), \mu(s)] \to \mathbb{R}$ is the map $t \mapsto g(t,s)$. Applying the fundamental theorem of calculus, we have:

$$\int_{\lambda(s)}^{\mu(s)} D_1 g(t,s) dt = g(\mu(s), s) - g(\lambda(s), s).$$

Inserting this into (4.3), we deduce

$$\int_A D_1 g(x) dx = \int_c^d \left(g(\mu(s), s) - g(\lambda(s), s) \right) ds,$$

and the result follows. □

Combining these two results, we arrive at Green's theorem:

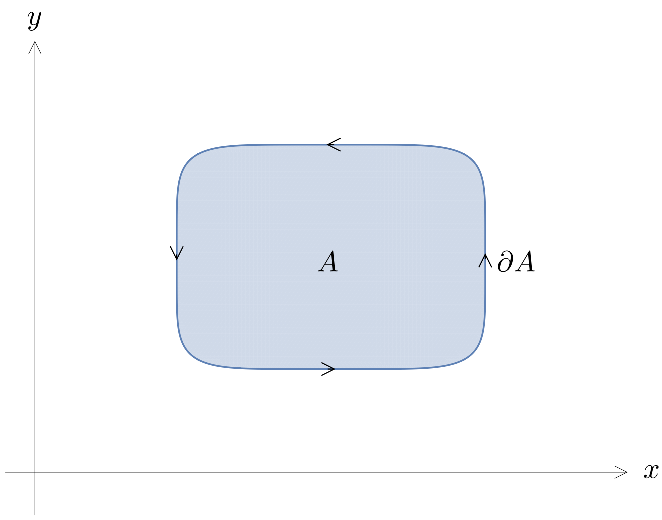

Figure 4.6 A simple domain, with the direction of the boundary indicated.

Theorem 4.17 (Green's Theorem in the plane)**.** *Suppose that* $\Omega \subset \mathbb{R}^2$ *is an open set, and that* $A \subset \Omega$ *is a simple domain. Suppose that* $V : \Omega \to \mathbb{R}^2$ *is differentiable at all points of* Ω*, with continuous partial derivatives. Then:*

$$\int_{\partial A} \langle V, d\boldsymbol{\ell} \rangle = \int_A \left[D_1 V^2(x) - D_2 V^1(x) \right] dx, \tag{4.4}$$

where we understand the integration over the boundary to be taken in the anti-clockwise sense.

Proof. 1. First, let us assume that $V^2 = 0$ on Ω. We make use of the fact that A is of Type I and apply Lemma 4.15 to the second term on the right of (4.4). We deduce:

$$\int_A -D_2 V^1(x) dx = \int_a^b V^1(t, \alpha(t)) dt - \int_a^b V^1(t, \beta(t)) dt.$$

The boundary ∂A of a Type I domain consists of at most four components (see Figure). We can parameterise these as:

$$\begin{aligned}
\gamma_1(t) &= (b, t)^t, & \alpha(b) \le t \le \beta(b) \\
\gamma_2(t) &= (-t, \beta(-t))^t & -b \le t \le -a \\
\gamma_3(t) &= (a, -t)^t, & -\beta(a) \le t \le -\beta(a) \\
\gamma_4(t) &= (t, \alpha(t))^t, & a \le t \le b.
\end{aligned}$$

so that we can compute:

$$\begin{aligned}
\gamma_1'(t) &= (0, 1)^t, & \alpha(b) \le t \le \beta(b) \\
\gamma_2'(t) &= (-1, -\beta'(-t))^t & -b \le t \le -a \\
\gamma_3'(t) &= (0, -1)^t, & -\beta(a) \le t \le -\beta(a) \\
\gamma_4'(t) &= (1, \alpha'(t))^t, & a \le t \le b.
\end{aligned}$$

Accordingly, we compute:

$$\int_{\gamma_1} \langle V, d\boldsymbol{\ell}\rangle = 0,$$

$$\int_{\gamma_2} \langle V, d\boldsymbol{\ell}\rangle = \int_{-b}^{-a} -V^1(-t,\beta(-t))dt = -\int_a^b V^1(t,\beta(t))dt,$$

$$\int_{\gamma_3} \langle V, d\boldsymbol{\ell}\rangle = 0,$$

$$\int_{\gamma_4} \langle V, d\boldsymbol{\ell}\rangle = \int_a^b V^1(t,\alpha(t))dt.$$

Putting this all together, we conclude that:

$$\int_{\partial A} \langle V, d\boldsymbol{\ell}\rangle = \int_A \left[D_1V^2(x) - D_2V^1(x)\right] dx,$$

when $V^2 = 0$.

2. Next, we consider the case when $V^1 = 0$ on Ω. We make use of the fact that A is of Type II and apply Lemma 4.16 to the first term on the right of (4.4). We deduce:

$$\int_A D_1V^2(x)dx = \int_c^d \left(V^2(\mu(s),s) - V^2(\lambda(s),s)\right) ds$$

The boundary ∂A of a Type II domain consists of at most four components (see Figure). We can parameterise these as:

$$\begin{aligned}
\gamma_1(t) &= (\mu(s),s)^t, & c \le s \le d\\
\gamma_2(t) &= (-s,d)^t & -\mu(d) \le s \le -\lambda(d)\\
\gamma_3(t) &= (\lambda(-s),-s)^t, & -d \le s \le -c\\
\gamma_4(t) &= (s,c)^t, & \lambda(c) \le s \le \mu(c).
\end{aligned}$$

so that we can compute:

$$\begin{aligned}
\gamma_1'(t) &= (\mu'(s),1)^t, & c \le s \le d\\
\gamma_2'(t) &= (-1,0)^t & -\mu(d) \le s \le -\lambda(d)\\
\gamma_3'(t) &= (-\lambda'(-s),-1)^t, & -d \le s \le -c\\
\gamma_4'(t) &= (1,0)^t, & \lambda(c) \le s \le \mu(c).
\end{aligned}$$

Accordingly, we compute:

$$\int_{\gamma_1} \langle V, d\boldsymbol{\ell}\rangle = \int_c^d V^2(\mu(s),s)ds,$$

$$\int_{\gamma_2} \langle V, d\boldsymbol{\ell}\rangle = 0,$$

$$\int_{\gamma_3} \langle V, d\boldsymbol{\ell}\rangle = \int_{-d}^{-c} V^2(\lambda(-s),-s)ds = -\int_c^d V^2(\lambda(s),s)ds$$

$$\int_{\gamma_4} \langle V, d\boldsymbol{\ell}\rangle = 0.$$

Putting this all together, we conclude that:

$$\int_{\partial A} \langle V, d\boldsymbol{\ell} \rangle = \int_A \left[D_1 V^2(x) - D_2 V^1(x) \right] dx,$$

when $V^1 = 0$.

3. Now, note that for any vector field V we can write $V = U + W$ for two vector fields U, W with $U^2 = W^1 = 0$ on Ω. Applying the results above together with the linearity of the integral, we deduce that:

$$\int_{\partial A} \langle V, d\boldsymbol{\ell} \rangle = \int_A \left[D_1 V^2(x) - D_2 V^1(x) \right] dx,$$

holds for any vector field V satisfying the conditions of the theorem. □

Note that while the condition of being simple is not by itself sufficient to ensure that ∂A is piecewise C^1, we nevertheless have that the integral:

$$\int_{\partial A} \langle V, d\boldsymbol{\ell} \rangle$$

always makes sense.

Example 4.5. Let's consider $A = \overline{B_1(0)}$, the unit disc centred at the origin, and we'll consider:

$$V : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} -y \\ x \end{pmatrix}.$$

The boundary ∂A is the unit circle traversed in an anti-clockwise direction. We've already computed the integral of this vector field along the unit circle in Example 4.3 and we found:

$$\int_{\partial A} \langle V, d\boldsymbol{\ell} \rangle = 2\pi.$$

On the other hand, we can compute:

$$D_1 V^2 - D_2 V^1 = 2,$$

so that:

$$\int_A 2dx = 2\pi.$$

Now, the disc is an example of a simple domain (check this) and so we deduce that the Jordan measure of the disc is $m(\overline{B_1(0)}) = \pi$.

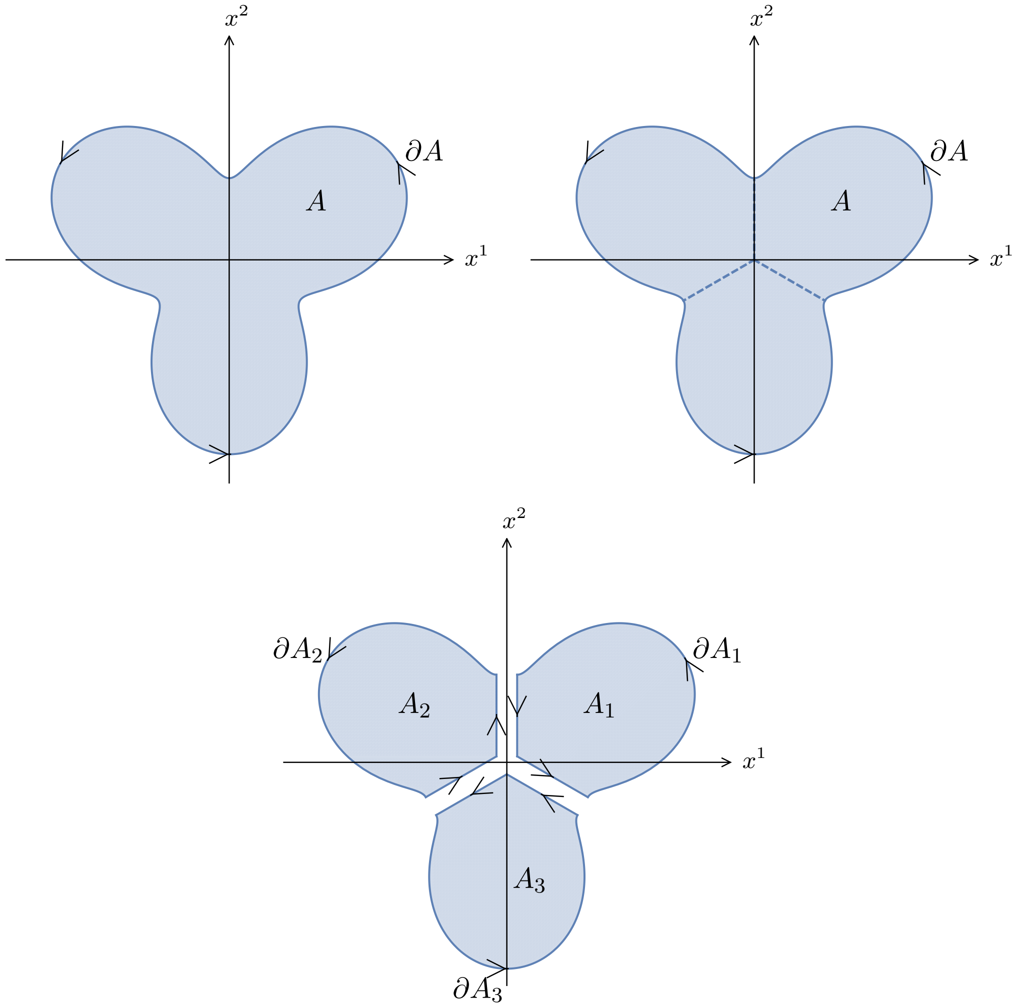

Figure 4.7 Breaking up a region into simple components.

4.3.2 Green's theorem in more general domains

The assumption that the domain is simple is rather a strong one. We can, however extend the results to more complicated domains by breaking them up into simple domains. We give here one example where this is possible, but the same basic idea can be applied in many other situations.

Consider the planar domain A shown in Figure 4.7. This is not of Type I or Type II. We can, however, split it up into three simple domains by making the cuts shown in the second diagram. The final diagram shows the three new domains A_1, A_2, A_3 slightly displaced to show the boundaries of these regions. If $f : A \to \mathbb{R}$ is any function which is integrable over A, then clearly:

$$\int_A f(x)dx = \int_{A_1} f(x)dx + \int_{A_1} f(x)dx + \int_{A_1} f(x)dx$$

Moreover, if V is any integrable vector field defined in a neighbourhood of A, then we have:

$$\int_{\partial A} \langle V, d\boldsymbol{\ell} \rangle = \int_{\partial A_1} \langle V, d\boldsymbol{\ell} \rangle + \int_{\partial A_2} \langle V, d\boldsymbol{\ell} \rangle + \int_{\partial A_3} \langle V, d\boldsymbol{\ell} \rangle$$

since the contributions from the cuts cancel, as the line segments are traversed twice in opposite directions. On each of the domains A_i, we can apply Green's theorem separately, so we deduce that Green's theorem holds for the domain A.

4.3.3 Proof of symmetry of mixed partial derivatives

When we discussed partial derivatives, we stated the following result without proving it:

Theorem 3.8 (Schwartz' Theorem). *Suppose $\Omega \subset \mathbb{R}^n$ is open and $f : \Omega \to \mathbb{R}$ is differentiable at each $p \in \Omega$. Suppose further that for some $i, j \in \{1, \ldots, n\}$ the second partial derivatives:*

$$D_i D_j f(p), \qquad D_j D_i f(p),$$

exist and are continuous at all $p \in \Omega$. Then:

$$D_i D_j f(p) = D_j D_i f(p).$$

We are now in a position to establish this result. We first need to establish the following Lemma

Lemma 4.18. *Let $\Omega \subset \mathbb{R}^2$ be open and let $f : \Omega \to \mathbb{R}$ be continuous. Suppose that:*

$$\int_{B_r(p)} f(x) dx = 0$$

for all $r > 0, p \in \Omega$ such that the disc $B_r(p) \subset \Omega$. Then $f = 0$ on Ω.

Proof. Suppose not, there there exists $p \in \Omega$ such that $f(p) = c \neq 0$. Without loss of generality we can assume $c > 0$ (otherwise apply the argument to $-f$). Since Ω is open, there exists $r > 0$ such that $B_r(p) \subset \Omega$. Since f is continuous, there exists $\delta < r$ such that if $||x - p|| < \delta$ then $|f(x) - f(p)| < \frac{c}{2}$. Thus, we have that for $x \in B_\delta(p)$ that $f(x) \geq \frac{c}{2} > 0$. We deduce that:

$$\int_{B_\delta(p)} f(x) dx \geq \pi \delta^2 \frac{c}{2} > 0,$$

contradicting the assumption that the integral of f over any disc vanishes. □

Now we are ready to prove the result.

Proof of Theorem 3.8. First note that it's enough to show the result for $n = 2$, since for higher n we can treat all of the variables except x^i, x^j as constant. Thus, suppose that $\Omega \subset \mathbb{R}^2$ is open and $f : \Omega \to \mathbb{R}$ is differentiable at each $p \in \Omega$. Suppose further that the second partial derivatives:

$$D_1 D_2 f(x), \qquad D_2 D_1 f(x),$$

exist and are continuous at all $x \in \Omega$. Let us fix $r > 0, p \in \Omega$ such that $B_r(p) \subset \Omega$. We compute using Green's theorem:

$$\int_{B_r(p)} \left(D_1 D_2 f(x) - D_2 D_1 f(x)\right) dx = \int_{\partial B_r(p)} \langle V, d\ell \rangle ,$$

where V is the vector field:

$$V(x) = \begin{pmatrix} D_1 f(x) \\ D_2 f(x) \end{pmatrix}.$$

Now, note that if $\gamma : [a, b] \to \mathbb{R}^2$ is an anti-clockwise parameterisation of $B_r(p)$, then:

$$\begin{aligned} \left\langle V(\gamma(t)), \gamma'(t) \right\rangle &= D_1 f(\gamma(t)) {\gamma^1}'(t) + D_2 f(\gamma(t)) {\gamma^2}'(t) \\ &= (f \circ \gamma)'(t), \end{aligned}$$

by the chain rule. Applying the fundamental theorem of calculus, we have:

$$\begin{aligned} \int_{\partial B_r(p)} \langle V, d\ell \rangle = \int_a^b \left\langle V(\gamma(t)), \gamma'(t) \right\rangle dt &= \int_a^b (f \circ \gamma)'(t) dt \\ &= f(\gamma(b)) - f(\gamma(a)) = 0, \end{aligned}$$

since $\partial B_r(p)$ is closed, so $\gamma(a) = \gamma(b)$. We conclude that:

$$\int_{B_r(p)} \left(D_1 D_2 f(x) - D_2 D_1 f(x)\right) dx = 0.$$

Since $B_r(p)$ was arbitrary, by the previous Lemma we deduce that:

$$D_1 D_2 f(x) - D_2 D_1 f(x) = 0$$

for all $x \in \Omega$, which is the result we require. □

4.3.4 Change of variables formula in two dimensions

From the fundamental theorem of calculus in one dimension we were able to establish the change of variables formula (Theorem 2.15). We will show how it's possible to use Green's theorem to establish a change of variables theorem in two dimensions. We shall make strong assumptions in order to establish this result. It is certainly the case that these assumptions may be weakened. In particular the restriction we make on the shape of the domain is much stronger than necessary.

Theorem 4.19. *Let $A = [a_1, b_2] \times [a_2, b_2]$ be a square domain in $\mathbb{R}^2$ and let $f : A \to \mathbb{R}$ be continuous. Suppose $\Omega \subset \mathbb{R}^2$ be open, and suppose $\phi : \Omega \to A$ is an injective, orientation preserving[2], twice continuously differentiable function. Finally suppose that $D \subset \Omega$ has the property that Green's theorem is valid on both D and $\phi(D)$. Then:*

$$\int_{\phi(D)} f(x) dx = \int_D f \circ \phi(x) \left| D\phi(x) \right| dx$$

Here $|D\phi|$ indicates the determinant of the Jacobian.

[2] A map is orientation preserving if $|D\phi| > 0$ everywhere

(*) Proof. We shall proceed by applying Green's theorem to relate the left hand side to a line integral around $\partial\phi(D)$ and then showing that this can be transformed into a line integral over ∂D. Finally another application of Green's theorem gives the result.

To set up some notation let us assume that $\gamma : [a, b] \to \mathbb{R}^2$ is the boundary of D, traversed in an anti-clockwise direction. The boundary of D will be $\tilde{\gamma} = \phi \circ \gamma : [a, b] \to \mathbb{R}^2$. The condition that ϕ is orientation preserving is required to ensure that $\tilde{\gamma}$ traverses ∂D in the correct sense.

1. First, we note that since f is continuous, there exists a continuous function $F : A \to \mathbb{R}$ such that $D_1F(x) = f(x)$ for all $x \in A$, which we can define by:

$$F(t, s) = \int_{a_1}^{t} f(t', s)dt'.$$

Now, applying Green's theorem on the domain ∂D to the vector field:

$$V(x) = \begin{pmatrix} 0 \\ F(x) \end{pmatrix},$$

we have:

$$\begin{aligned}\int_{\phi(D)} f(x)dx &= \int_{\phi(D)} D_1F(x)dx \\ &= \int_{\partial\phi(D)} \langle V, d\ell\rangle \\ &= \int_a^b F \circ \tilde{\gamma}(t)\, \tilde{\gamma}^{2\prime}(t)dt.\end{aligned}$$

2. Next, we can apply the chain rule to relate $\tilde{\gamma}^{2\prime}(t)$ to γ by:

$$\tilde{\gamma}^{2\prime}(t) = D_1\phi^2(\gamma(t))\gamma^{1\prime}(t) + D_2\phi^2(\gamma(t))\gamma^{2\prime}(t),$$

so that:

$$\begin{aligned}\int_a^b F \circ \tilde{\gamma}(t)\, \tilde{\gamma}^{2\prime}(t)dt &= \int_a^b F \circ \phi(\gamma(t)) \left[D_1\phi^2(\gamma(t))\gamma^{1\prime}(t) + D_2\phi^2(\gamma(t))\gamma^{2\prime}(t)\right] dt \\ &= \int_a^b \left\langle \begin{pmatrix} F \circ \phi\ D_1\phi^2 \\ F \circ \phi\ D_2\phi^2 \end{pmatrix} (\gamma(t)), \gamma'(t) \right\rangle dt \\ &= \int_{\partial D} \langle W, d\ell\rangle\,,\end{aligned}$$

where W is the vector field:

$$W(x) = \begin{pmatrix} F \circ \phi(x)\ D_1\phi^2(x) \\ F \circ \phi(x)\ D_2\phi^2(x) \end{pmatrix}.$$

3. Now we compute with the chain rule:

$$D_1 W^2(x) = \big(D_1F(\phi(x))D_1\phi^1(x) + D_2F(\phi(x))D_1\phi^2(x)\big)\, D_2\phi^2(x) + F(\phi(x))D_1D_2\phi(x),$$
$$D_2 W^1(x) = \big(D_1F(\phi(x))D_2\phi^1(x) + D_2F(\phi(x))D_2\phi^2(x)\big)\, D_1\phi^2(x) + F(\phi(x))D_2D_1\phi(x),$$

so that:

$$\begin{aligned}\int_{\partial D} \langle W, d\ell\rangle &= \int_D \big(D_1W^2(x) - D_2W^1(x)\big)\, dx \\ &= \int_D D_1F(\phi(x))\,\big[D_1\phi^1(x)D_2\phi^2(x) - D_2\phi^1(x)D_1\phi^2(x)\big]\, dx \\ &= \int_D f\circ\phi(x)\; |D\phi(x)|\, dx,\end{aligned}$$

where we use the fact that $D_1F = f$. Combining all of these identities, the result follows. □

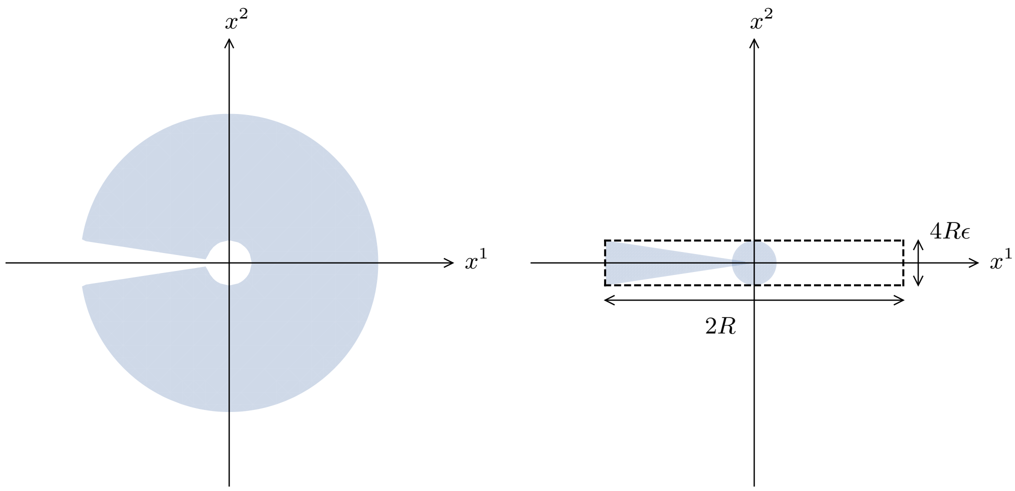

Figure 4.8 The regions $\phi(D)$, and $B_R(0) \setminus \phi(D)$.

Example 4.6. Fix $R > 0$. Let $A = [-2R, 2R] \times [-2R, 2R]$ and suppose that $f : A \to \mathbb{R}$ is continuous. Fix $\epsilon > 0$, and let us take $\Omega = (\epsilon R, 2R) \times (-\pi + \epsilon, \pi - \epsilon)$ The map:

$$\phi : \begin{pmatrix} r \\ \theta \end{pmatrix} \mapsto \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix}$$

is injective and smooth, and we can compute for $z = (r, \theta)^t$:

$$\begin{aligned}|D\phi(z)| &= D_1\phi^1(z)D_2\phi^2(z) - D_1\phi^2(z)D_2\phi^1(z) \\ &= \cos\theta \times r\cos\theta - \sin\theta \times (-r\cos\theta) \\ &= r.\end{aligned}$$

Let us take $D = [2\epsilon R, R] \times [-\pi + 2\epsilon, \pi - 2\epsilon]$. We can apply the above result to conclude (making use of Fubini's theorem):

$$\int_{\phi(D)} f(x)dx = \int_D f \circ \phi(z) \; |D\phi(z)| \, dz = \int_{2\epsilon R}^{R} \left(\int_{-\pi+2\epsilon}^{\pi-2\epsilon} f(r\cos\theta, r\sin\theta)d\theta \right) rdr.$$

Here we are justified in using Fubini since the function $(r, \theta) \mapsto f(r\cos\theta, r\sin\theta)$ is continuous (as it is a composition of continuous functions).

Now, let us consider the domain $B_R(0) \setminus \phi(D)$. We can estimate the area of this set by putting it inside a box of height $4R\epsilon$ and width $2R$, so that $|B_R(0) \setminus \phi(D)| \leq 8\epsilon R^2$. Since f is a continuous function on a bounded set, we have $|f(x)| \leq K$ for some K. We deduce:

$$\left| \int_{B_R(0)} f(x)dx - \int_{\phi(D)} f(x)dx \right| \leq 8\epsilon K R^2$$

Since ϵ was arbitrary, we deduce that:

$$\int_{B_R(0)} f(x)dx = \int_0^R \left(\int_{-\pi}^{\pi} f(r\cos\theta, r\sin\theta)d\theta \right) rdr.$$

which is the usual change of variables formula for polar coordinates in the plane. Notice that we had to introduce ϵ in order that the conditions of Theorem 4.19 are satisfied. This type of approach enables us to extend the change of variables formula in other cases where the hypotheses of Theorem 4.19 are not satisfied.

Exercise 9.6. a) Show that the improper integral:

$$\int_1^\infty e^{-u}du$$

converges.

b) By making use of the estimate:

$$\int_1^R e^{-u^2}du \leq \int_1^R e^{-u}du, \qquad \text{for } R > 1,$$

or otherwise, establish that the improper integral:

$$I := \int_{-\infty}^{\infty} e^{-u^2}du$$

converges.

c) Let $S_R = [-R, R] \times [-R, R]$. Show that for any function $f : \mathbb{R}^2 \to \mathbb{R}$ satisfying $f \geq 0$, the estimate:

$$\int_{B_R(0)} f(x)dx \leq \int_{S_R} f(x)dx \leq \int_{B_{\sqrt{2}R}(0)} f(x)dx$$

holds.

d) By taking $f : (u, v)^t \mapsto e^{-u^2-v^2}$, or otherwise, show that:

$$\pi\left(1 - e^{-R^2}\right) \leq \int_{-R}^{R}\left(\int_{-R}^{R} e^{-u^2-v^2}\, du\right) dv \leq \pi\left(1 - e^{-2R^2}\right)$$

e) Deduce that:

$$I = \sqrt{\pi}.$$

Appendix A

Differentiation in one dimension

A.1 Introduction

We'll start by heuristically motivating the idea of a tangent and derivative of a function. Suppose $f : (a, b) \to \mathbb{R}$ is some function which maps an interval (a, b) with $a < b$ to the real line $\mathbb{R}$. For the time being, let's assume that f is continuous. It's often useful to represent f graphically by sketching its graph (see Figure A.1)

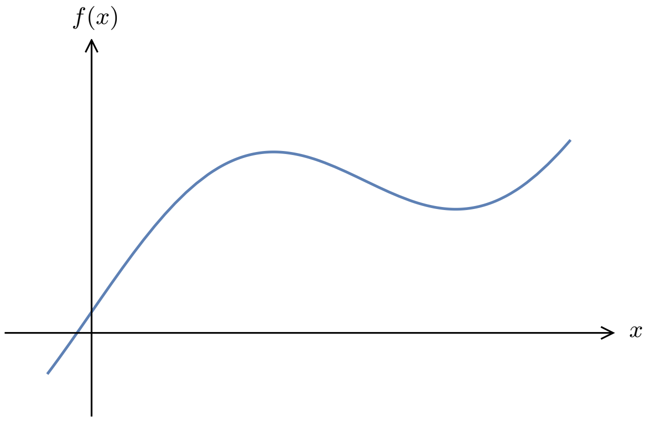

Figure A.1 The graph of a function f.

A very useful concept is the notion of the *tangent* to the graph at a point $p \in (a, b)$. We approximate this by constructing a straight line passing through $(p, f(p))$ and $(p + h, f(p + h))$. This line is the graph of the function:

$$x \mapsto \left(\frac{f(p + h) - f(p)}{h} \right) (x - p) + f(p).$$

Such a straight line is called a chord (see Figure A.2). As h becomes smaller, we might hope that the chords should get closer and closer together and n the limit where $h \to 0$ approach some line passing through p. We call this line the tangent to f at p (see Figure

A.3). The gradient of the tangent to f at p is an important quantity, called the *derivative* of f at p and denoted $f'(p)$.

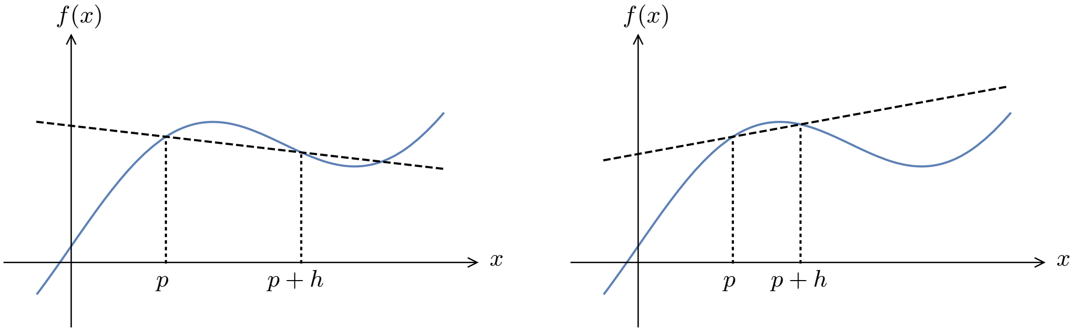

Figure A.2 Chords approximating the tangent to f at p for two values of h.

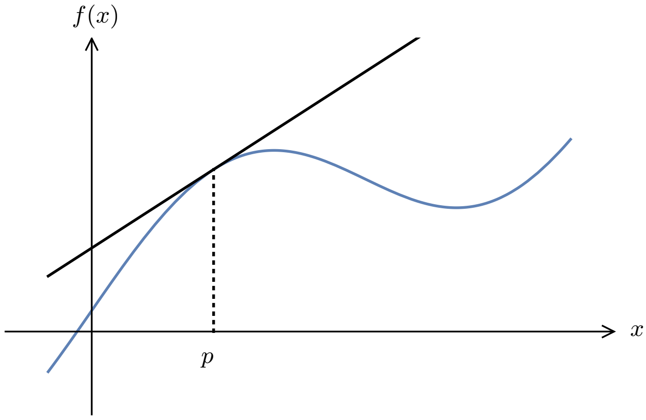

Figure A.3 The tangent to f at p.

This somewhat hand-waving motivation justifies the following definition of differentiability, which we shall unpack shortly:

Definition A.1. A function $f : (a, b) \to \mathbb{R}$ is *differentiable at* $p \in (a, b)$ if the limit

$$\lim_{x \to p} \frac{f(x) - f(p)}{x - p}$$

exists. If this is the case, we call the value of this limit the *derivative* of f at p and write:

$$f'(p) := \frac{f(x) - f(p)}{x - p}.$$

We can expand this definition by making use of the usual $\epsilon - \delta$ formulation for limits:

Definition A.2 (Alternative definition of differentiability)**.** A function $f : (a, b) \to \mathbb{R}$ is *differentiable at* $p \in (a, b)$ if there exists $\lambda \in \mathbb{R}$ such that for any $\epsilon > 0$ we can find $\delta > 0$ with:

$$\left| \frac{f(x) - f(p)}{x - p} - \lambda \right| < \epsilon$$

for all $x \in (a, b)$ with $0 < |x - p| < \delta$.

The notation:

$$\frac{df}{dx}(p) := f'(p)$$

is sometimes used for the derivative at p.

Exercise(∗). Show that Definitions A.1, A.2 are equivalent.

We will use either formulation of differentiability interchangeably. Notice that at the heart of the definition of differentiability is the assertion that a certain limit exists. In general this may or may not be the case as we shall see now in some examples.

Example A.1. Fix $\alpha, \beta \in \mathbb{R}$. Consider $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = \alpha x + \beta$, and pick $p \in \mathbb{R}$. We have:

$$f(x) - f(p) = \alpha(x - p)$$

so that

$$\frac{f(x) - f(p)}{x - p} = \alpha.$$

We conclude that f is differentiable at p for any $p \in \mathbb{R}$, and the derivative of f is given by $f'(p) = \alpha$.

Example A.2. Consider $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = |x|$, and take $p = 0$. We have:

$$\frac{f(x) - f(0)}{x} = \begin{cases} 1 & x > 0 \\ -1 & x < 0 \end{cases}$$

Thus for $\epsilon < 1$, there can exist no λ and $\delta > 0$ such that:

$$\left| \frac{f(x) - f(0)}{x} - \lambda \right| < \epsilon$$

for all x with $0 < |x| < \delta$. We conclude that $f(x) = |x|$ is not differentiable at $x = 0$.

This example shows that there exist continuous functions which are not differentiable. In the other direction, we have the following result:

Lemma A.1. *Suppose $f : (a, b) \to \mathbb{R}$ is differentiable at $p \in (a, b)$. Then f is continuous at p.*

Proof. Fix $\epsilon > 0$. Note that:

$$|f(x) - f(p)| = |x - p| \left| \frac{f(x) - f(p)}{x - p} - f'(p) + f'(p) \right|$$

Applying the triangle inequality, we have:

$$|f(x) - f(p)| \leq |x - p| \left| \frac{f(x) - f(p)}{x - p} - f'(p) \right| + |x - p| \left| f'(p) \right|$$

Now, since f is differentiable at p, we know that there exists δ' such that if $x \neq p$ satisfies $|x - p| < \delta'$ we have:

$$\left| \frac{f(x) - f(p)}{x - p} - f'(p) \right| < 1,$$

so that

$$|f(x) - f(p)| \leq |x - p| \left(1 + \left| f'(p) \right| \right).$$

Thus, if we take

$$|x - p| < \min \left\{ \frac{\epsilon}{1 + |f'(p)|}, \delta' \right\} =: \delta,$$

we have:

$$|f(x) - f(p)| < \epsilon.$$

This is precisely the statement that f is continuous at p. □

This Lemma tells us that differentiability is a strictly stronger condition than continuity. Not all functions that are continuous at p are differentiable at p, but all functions that are differentiable at p are continuous.

We can build differentiable functions by combining functions that we already know to be differentiable in various ways.

Theorem A.2. *i) Suppose that $f, g : (a, b) \to \mathbb{R}$ are both differentiable at $p \in (a, b)$. Then*

$$\begin{array}{rcl} u : (a,b) & \to & \mathbb{R}, \\ x & \mapsto & f(x) + g(x) \end{array}$$

is differentiable at p with

$$u'(p) = f'(p) + g'(p),$$

ii) [Product rule] *Suppose that $f, g : (a, b) \to \mathbb{R}$ are both differentiable at $p \in (a, b)$. Then*

$$\begin{array}{rcl} v : (a,b) & \to & \mathbb{R}, \\ x & \mapsto & f(x)g(x) \end{array}$$

is differentiable at p with

$$v'(p) = f'(p)g(p) + f(p)g'(p).$$

iii) [Quotient rule] *Suppose that $f, g : (a,b) \to \mathbb{R}$ are both differentiable at $p \in (a,b)$ and that $g(p) \neq 0$. Then taking δ sufficiently small,*

$$\begin{array}{rcccl} w & : & (p-\delta, p+\delta) & \to & \mathbb{R}, \\ & & x & \mapsto & \frac{f(x)}{g(x)} \end{array}$$

is well defined. Moreover, w is differentiable at p and

$$w'(p) = \frac{f'(p)g(p) - f(p)g'(p)}{g(p)^2}$$

Proof. We will use quite heavily the results of Theorem I.7.

i) We write:

$$\frac{u(x) - u(p)}{x - p} = \frac{f(x) - f(p)}{x - p} + \frac{g(x) - g(p)}{x - p}.$$

Now, since f is differentiable, we know that the first term on the RHS has a limit, $f'(p)$ as $x \to p$. Similarly, the second term has a limit, $g'(p)$. Applying Theorem I.7 part i), we deduce the result.

ii) We write:

$$\begin{aligned} \frac{u(x) - u(p)}{x - p} &= \frac{f(x)g(x) - f(p)g(p)}{x - p} \\ &= \frac{[f(x) - f(p)][g(x) - g(p)] + [f(x) - f(p)]\, g(p) + [g(x) - g(p)]\, f(p)}{x - p} \\ &= \frac{f(x) - f(p)}{x - p}[g(x) - g(p)] + \frac{f(x) - f(p)}{x - p} g(p) + \frac{g(x) - g(p)}{x - p} f(p) \end{aligned} \tag{A.1}$$

Now, as $x \to p$, we have that

$$\lim_{x \to p} \frac{f(x) - f(p)}{x - p} = f'(p), \qquad \lim_{x \to p} [g(x) - g(p)] = 0,$$

so by Theorem I.7 part ii), the first term of (A.1) vanishes as $x \to p$. The other two terms are also easily dealt with in a similar fashion, and we find:

$$\lim_{x \to p} \frac{u(x) - u(p)}{x - p} = f'(p)g(p) + f(p)g'(p).$$

iii) First note that since g is differentiable at p, it is continuous at p, which implies that if $g(p) \neq 0$, then there exists $\delta > 0$ such that $g(x) \neq 0$ for $x \in (p - \delta, p + \delta)$. Moving on to differentiability, by making use of part ii), it's enough to consider the case where $f(x) \equiv 1$. Then we have, for $x \in (p - \delta, p + \delta)$:

$$\frac{w(x) - w(p)}{x - p} = \frac{\frac{1}{g(x)} - \frac{1}{g(p)}}{x - p} = -\frac{1}{g(x)g(p)} \frac{g(x) - g(p)}{x - p}$$

Now as $x \to p$ we have, making use of Theorem I.7 part iii):

$$\lim_{x\to p} -\frac{1}{g(x)g(p)} = \frac{-1}{g(p)^2}, \qquad \lim_{x\to p} \frac{g(x)-g(p)}{x-p} = g'(p).$$

By Theorem I.7 part ii) we conclude:

$$\lim_{x\to p} \frac{w(x)-w(p)}{x-p} = -\frac{g'(p)}{g(p)^2}.$$

□

Exercise(∗). Give a proof of Theorem A.2 parts $i), ii)$ directly using the $\epsilon - \delta$ formulation of differentiability (Definition A.2).

The next result tells us how to differentiate functions which are formed as the composition of two functions. The proof is slightly subtle, but in this form will be useful later when we come to the chain rule for functions of many variables.

Theorem A.3 (Chain Rule)**.** *Suppose that $g : (a, b) \to (c, d)$ and $f : (c, d) \to \mathbb{R}$. Suppose further that g is differentiable at p and f is differentiable at $g(p)$. Then*

$$\begin{array}{cccc} z & : & (a,b) & \to & \mathbb{R}, \\ & & x & \mapsto & f(g(x)) \end{array}$$

is differentiable at p and:

$$z'(p) = g'(p)f'(g(p))$$

Proof. Since f is differentiable at p, we know that:

$$\lim_{x\to p} \frac{f(x)-f(p)}{x-p} = f'(p).$$

We can package this piece of information up in a slightly different form by defining a new function $F : (c, d) \to \mathbb{R}$ by:

$$F(x) := \begin{cases} \frac{f(x)-f(p)}{x-p} - f'(p) & x \neq p \\ 0 & x = p. \end{cases}$$

The statement that f is differentiable at p is equivalent to the statement that F is continuous at p (check that you understand this). Notice that we can write:

$$f(x) = f(p) + \left[f'(p) + F(x)\right](x - p). \tag{A.2}$$

Similarly, letting $q = f(p)$, we define $G : (a, b) \to \mathbb{R}$ by:

$$G(y) := \begin{cases} \frac{g(y)-g(q)}{y-q} - f'(p) & y \neq q \\ 0 & y = q. \end{cases}$$

The statement that g is differentiable at $f(p)$ is equivalent to the statement that G is continuous at 0. We also have:

$$g(y) = g(q) + \left[g'(q) + G(y)\right](y - q). \tag{A.3}$$

Now, let's rewrite:

$$\begin{aligned} g(f(x)) - g(f(p)) &= g(q) + \left[g'(q) + G(f(x))\right](f(x) - q) - g(f(p)) \\ &= \left[g'(q) + G(f(x))\right](f(x) - q) \end{aligned}$$

Here we have used (A.3) to rewrite $g(y)$ with $y = f(x)$, and used $q = f(p)$. Next, we expand the $f(x)$ appearing in the second bracket using (A.2):

$$\begin{aligned} g(f(x)) - g(f(p)) &= \left[g'(q) + G(f(x))\right](f(x) - q) \\ &= \left[g'(q) + G(f(x))\right]\left[f'(p) + F(x)\right](x - p), \end{aligned}$$

again using $q = f(p)$ to get a cancellation. Thus we have:

$$\frac{g(f(x)) - g(f(p))}{x - p} = \left[g'(q) + G(f(x))\right]\left[f'(p) + F(x)\right].$$

Now, since F is continuous at p, we have $\lim_{x\to p} F(x) = 0$. Moreover, since f is continuous at p and G is continuous at $q = f(p)$, we have:

$$\lim_{x\to p} G(f(x)) = G\left(\lim_{x\to p} f(x)\right) = G(q) = 0.$$

Thus we have:

$$\lim_{x\to p} \frac{g(f(x)) - g(f(p))}{x - p} = f'(x)g'(f(x)),$$

the result we require. □

The results above allow us to establish that quite a few functions are differentiable. An important further result on differentiability concerns power series. We shan't prove this result at this stage, since the proof will follow as a corollary of some results we establish later in the course (see Corollary 2.18).

Theorem A.4. *Suppose f is given by the power series:*

$$f(x) = \sum_{k=0}^{\infty} a_k x^k,$$

with radius of convergence $R > 0$. Then f is differentiable at all points $x \in (-R, R)$ and moreover:

$$f'(x) = \sum_{k=1}^{\infty} k a_k x^{k-1}.$$

In other words, provided we are inside the radius of convergence, we can differentiate a power series term by term.

Exercise A.1. An alternative proposed proof of the chain rule runs as follows. We rewrite:

$$\frac{g(f(x)) - g(f(p))}{x-p} = \frac{g(f(x)) - g(f(p))}{f(x) - f(p)} \frac{f(x) - f(p)}{x-p},$$

and we claim:

$$\lim_{x \to p} \frac{g(f(x)) - g(f(p))}{f(x) - f(p)} = \lim_{y \to f(x)} \frac{g(y) - g(f(p))}{y - f(p)} = g'(f(p))$$

which together with:

$$\lim_{x \to p} \frac{f(x) - f(p)}{x - p} = f'(p)$$

gives the result. Why is this 'proof' not valid?

Exercise A.2. a) Suppose that $n \in \mathbb{N}$ and $f : \mathbb{R} \to \mathbb{R}$ is given by:

$$f(x) = x^n.$$

Show that f is differentiable at all points of $\mathbb{R}$ and $f'(x) = nx^{n-1}$.

b) Show that if $f : \mathbb{R} \to \mathbb{R}$ is any polynomial function of x, then f is differentiable at all points of $\mathbb{R}$.

Exercise A.3. Prove[1] that $f : \mathbb{R} \to \mathbb{R}$ is differentiable everywhere in $\mathbb{R}$, and find the derivative of:

a) $f(x) = \sin x$

b) $f(x) = \cos x$

c) $f(x) = \exp x$

d) $f(x) = e^{-x^2}$

Exercise A.4. Consider the function $f : \mathbb{R} \to \mathbb{R}$ given by:

$$f(x) := \begin{cases} x^2 \sin \frac{1}{x} & x \neq 0 \\ 0 & x = 0. \end{cases}$$

a) Show that for $x \neq 0$, f is differentiable at x with derivative:

$$f'(x) = -\cos\frac{1}{x} + 2x \sin\frac{1}{x}.$$

b) Show that f is differentiable at $x = 0$ with derivative:

$$f'(0) = 0.$$

c) Show that the function $x \mapsto f'(x)$ is not continuous at $x = 0$.

[1] You may assume

$$\sin x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1}, \quad \cos x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n}, \quad \exp x = \sum_{n=0}^{\infty} \frac{1}{n!} x^n.$$

A.1.1 Higher derivatives

Suppose $f : (a, b) \to \mathbb{R}$ is differentiable at each point $x \in (a, b)$. Then we can obviously consider the function:

$$\begin{array}{rcccl} f' & : & (a,b) & \to & \mathbb{R}, \\ & & x & \mapsto & f'(x). \end{array}$$

A priori, this function can be very badly behaved. From the example in Exercise A.4, we know that f' need not be continuous. If f' is continuous, we say that f is continuously differentiable. If f' is itself differentiable, we say that f is twice differentiable and we denote by f'' or $f^{(2)}$ the derivative of f' (the second derivative of f). Obviously we can inductively define a $k-$times differentiable function to be a function for which f' is $(k-1)-$times differentiable, and we denote the k^{th} derivative by $f^{(k)}$. If a function is $k-$times differentiable for all $k \in \mathbb{N}$, we say that f is *smooth* (or infinitely differentiable).

Exercise A.5. Show that if the power series:

$$f(x) = \sum_{k=0}^{\infty} a_k x^k,$$

has radius of convergence $R > 0$, then the power series:

$$g(x) = \sum_{k=1}^{\infty} k a_k x^{k-1}$$

also has radius of convergence R. Deduce that a power series is smooth inside its radius of convergence.

A.2 Rolle's theorem

Suppose we have a function which is continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b), and moreover, $f(a) = f(b)$. Intuitively, we expect that (unless f is constant), the graph of f will have to 'turn around' at some point between a and b (see Figure A.4). We'd also expect that this might mean the derivative of f should vanish somewhere in the interval (a,b). Rolle's theorem asserts that this is indeed the case.

Theorem A.5 (Rolle's theorem)**.** *Suppose $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and differentiable at each $x \in (a, b)$. Suppose further that $f(a) = f(b)$. Then there exists $c \in (a, b)$ such that $f'(c) = 0$.*

Proof. By the extreme value theorem (Theorem I.9) we know that f achieves a maximum and a minimum in $[a, b]$. Now, either f is constant (in which case the result holds trivially), or else one of the maxima or minima occurs at some $c \in (a, b)$. Suppose that $c \in (a, b)$ is a maximum, and consider for $x \neq c$:

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

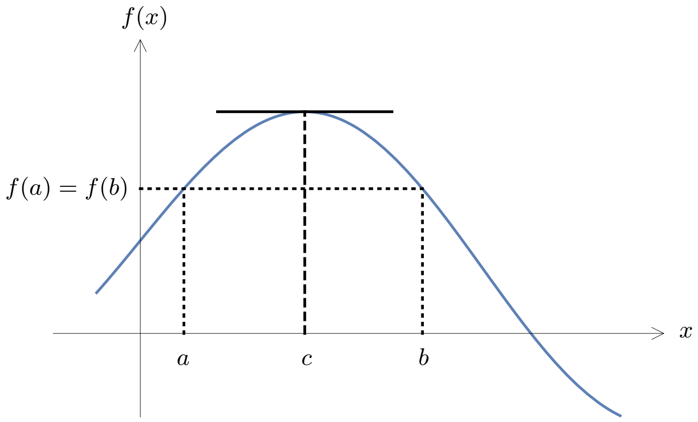

Figure A.4 The set-up for Rolle's theorem

Since f achieves its maximum value at c, we have $f(x) - f(c) \leq 0$. Thus we have $g(x) \geq 0$ for $x < c$ and $g(x) \leq 0$ for $x > c$. We also know that the limit $\lim_{x \to c} g(x)$ exists and is equal to $f'(c)$. We deduce that $f'(c) = 0$.

In the case where $c \in (a, b)$ is a minimum, we can apply the same argument to $-f(x)$, which has a maximum at c. □

A.3 Mean value theorem

The mean value theorem can be thought of as a 'tilted' version of Rolle's theorem. It asserts that a function which is continuous on $[a, b]$ and differentiable on (a, b) must somewhere have a slope equal to the 'average slope' over the interval, i.e. $\frac{f(b)-f(a)}{b-a}$ (see Figure A.5). Put in more human terms, this says:

> If you drive 50 miles, starting from London at 1pm and arriving in Brighton at 2pm, then at some time between 1pm and 2pm your car must have been travelling at 50mph.

This is true regardless of what style of driving one adopts.

Theorem A.6 (Mean value theorem)**.** *Suppose $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and differentiable at each $x \in (a, b)$. Then there exists $c \in (a, b)$ such that:*

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

Proof. We will use a clever trick to reduce this proof to an application of Rolle's theorem. Consider the function:

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

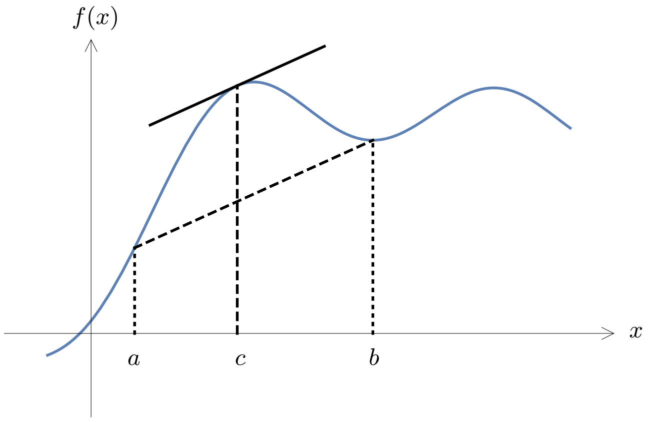

Figure A.5 The set-up for the mean value theorem

Clearly g is continuous on $[a, b]$ and differentiable on (a, b). Moreover, $g(a) = f(a) = g(b)$. Thus we can apply Rolle's theorem to g to deduce that there exists $c \in (a, b)$ such that $g'(c) = 0$. However,

$$0 = g'(c) = f'(c) - \frac{f(b) - f(a)}{b - a},$$

so we deduce that

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

□

Corollary A.7. *Suppose that $f : (a, b) \to \mathbb{R}$ satisfies $f'(x) = 0$ for $x \in (a, b)$. Then $f(x) = C$ for some $C \in \mathbb{R}$.*

Proof. Pick c, d with $a < c < d < b$. We can apply the mean value theorem on the interval (c, d) to deduce there exists $e \in (c, d)$ with

$$f'(e) = \frac{f(c) - f(d)}{c - d}.$$

However, we know that $f'(e) = 0$ and $c \neq d$, so we must have $f(c) = f(d)$. Since c, d were arbitrary points in (a, b) we deduce that f is constant, i.e. $f(x) = C$. □

Exercise A.6. Suppose that $f, g : (a, b) \to \mathbb{R}$ are differentiable on (a, b) and satisfy:

$$f'(x) = g'(x), \qquad \text{for all } x \in (a, b).$$

Show that:

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

Exercise A.7. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is differentiable and satisfies:

$$f'(x) = f(x), \quad \text{for all } x \in \mathbb{R} \qquad f(0) = 1. \tag{A.4}$$

a) Show that the function:

$$f(x) = \exp x$$

is a solution of (A.4).

b) Let $g(x) = \frac{f(x)}{\exp(x)}$. Show that:

$$g'(x) = 0, \quad \text{for all } x \in \mathbb{R} \qquad g(0) = 1.$$

Conclude that $f(x) = \exp x$ is the *unique* solution of (A.4).

c) By considering $f(x) = \frac{1}{\exp(-x)}$ or otherwise, show that $\exp(-x) = \frac{1}{\exp x}$.

Exercise A.8. Let $p_k : \mathbb{R} \to \mathbb{R}$ be given by:

$$p_k(x) = a_k x^k + a_{k-1}x^{k-1} + \ldots + a_0$$

for some $a_i \in \mathbb{R}$. Show that there is a unique $f : \mathbb{R} \to \mathbb{R}$ which is differentiable everywhere in $\mathbb{R}$ and satisfies:

$$f'(x) = p_k, \qquad f(0) = 0,$$

and that f is a polynomial of degree $k + 1$.

Exercise A.9. Suppose that $f, g : (a, b) \to \mathbb{R}$ are $k-$times differentiable on (a, b) and satisfy:

$$f^{(k)}(x) = g^{(k)}(x), \qquad \text{for all } x \in (a, b).$$

Show that:

$$f(x) = g(x) + p_{k-1}(x).$$

for some polynomial p_{k-1} of order $k - 1$.

A.4 Taylor's theorem

An important feature of $k-$times differentiable functions is that they can be locally approximated by a polynomial of degree $k - 1$. This important fact is encapsulated in *Taylor's theorem*:

Theorem A.8 (Taylor's Theorem). *Suppose that $f : (c, d) \to \mathbb{R}$ is $k-$times differentiable on (c, d) and suppose $[a, x] \subset (c, d)$. Then:*

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \ldots + \frac{f^{(k-1)}(a)}{(k-1)!}(x-a)^{k-1} + R_k(x),$$

where:

$$R_k(x) = \frac{f^{(k)}(\xi)}{k!}(x-a)^k$$

for some $\xi \in (a, x)$. The same result holds if $[x, a] \subset (c, d)$, the only modification being that now $\xi \in (x, a)$.

We shall proceed to prove this by first proving two simpler Lemmas, such that Taylor's theorem follows as a Corollary. This version of the proof is adapted from one that I first saw on Tim Gowers blog[2].

Lemma A.9 (Higher order Rolle's Theorem)**.** *Suppose that f satisfies the conditions of Theorem A.8, and that moreover,*

$$f^{(l)}(a) = 0, \quad \textit{for all } \ 0 \le l \le k-1, \qquad \textit{and } f(x) = 0.$$

Then there exists $\xi \in (a, x)$ such that $f^{(k)}(\xi) = 0$.

Proof. We proceed by induction on k. The case $k = 1$ is simply Rolle's theorem. Now suppose that the result holds for $k = K$, and that:

$$f^{(l)}(a) = 0, \quad \text{for all } \ 0 \le l \le K, \qquad \text{and } f(x) = 0.$$

Applying Rolle's theorem, we know that there exists $y \in (a, x)$ such that $f'(y) = 0$. Now let $g = f'$. The function g satisfies:

$$g^{(l)}(a) = 0, \quad \text{for all } \ 0 \le l \le K-1, \qquad \text{and } g(y) = 0.$$

By the induction hypothesis, there exists $\xi \in (a, y) \subset (a, x)$ such that $0 = g^{(K)}(\xi) = f^{(K+1)}(\xi)$, which establishes the result for $k = K + 1$. By induction we're done. □

Lemma A.10. *Suppose that f satisfies the conditions of Theorem A.8. There exists a polynomial of order k, P_k with the properties:*

i) $f^{(l)}(a) = P_k^{(l)}(a)$ *for all* $0 \le l \le k-1$.

ii) $f(x) = P_k(x)$.

Proof. First, note that for $m \in \mathbb{N}$, the function:

$$q_m : y \mapsto \frac{(y-a)^m}{m!}$$

satisfies:

$$q_m^{(l)}(a) = \begin{cases} 1 & m = l, \\ 0 & m \neq l. \end{cases}$$

Let us define Q_{k-1} to be the polynomial:

$$\begin{aligned} Q_{k-1}(y) &= \sum_{m=0}^{k-1} f^{(m)}(a) q_m(y) \\ &= f(a) + f'(a)(y-a) + \frac{f''(a)}{2!}(y-a)^2 + \ldots + \frac{f^{(k-1)}(a)}{(k-1)!}(y-a)^{k-1}. \end{aligned}$$

[2]https://gowers.wordpress.com/2014/02/11/taylors-theorem-with-the-lagrange-form-of-the-remainder/

This satisfies $f^{(l)}(a) = Q^{(l)}_{k-1}(a)$ for all $0 \leq l \leq k-1$. To construct P_k, we take:

$$P_k(y) = Q_{k-1}(y) + \lambda q_k(y) = Q_{k-1}(y) + \lambda \frac{(y-a)^k}{k!}.$$

This certainly satisfies $f^{(l)}(a) = P^{(l)}_k(a)$ for all $0 \leq l \leq k-1$. Setting $y = x$, we see that $f(x) = P_k(x)$ implies

$$\lambda = \frac{k!}{(x-a)^k}\left[f(x) - Q_{k-1}(x)\right].$$

□

Now, we are ready to prove Taylor's theorem, by combining these two results:

Proof of Taylor's Theorem. Consider the function $g = f - P_k$, where P_k is the polynomial constructed in Lemma A.10. This satisfies

$$g^{(l)}(a) = 0, \qquad \text{for all } \ 0 \leq l \leq k-1, \qquad \text{and } g(x) = 0,$$

so by Lemma A.9, we know there exists $\xi \in (a, x)$ such that $g^{(k)}(\xi) = 0$. Now, we can calculate $g^{(k)}(\xi)$, and we find:

$$\begin{aligned} 0 = g^{(k)}(\xi) &= f^{(k)}(\xi) - P^{(k)}_k(\xi) = f^{(k)}(\xi) - \lambda \\ &= f^{(k)}(\xi) - \frac{k!}{(x-a)^k}\left[f(x) - Q_{k-1}(x)\right]. \end{aligned}$$

Rearranging this, we deduce:

$$\begin{aligned} f(x) &= Q_{k-1}(x) + f^{(k)}(\xi)\frac{(x-a)^k}{k!}, \\ &= f(a) + f'(a)(y-a) + \frac{f''(a)}{2!}(y-a)^2 + \ldots + \frac{f^{(k-1)}(a)}{(k-1)!}(y-a)^{k-1} + f^{(k)}(\xi)\frac{(x-a)^k}{k!}, \end{aligned}$$

which is precisely the result we seek. □

As an example of the use of Taylor's theorem, let us consider the function:

$$f : y \mapsto \frac{1}{1+y},$$

which is smooth on the interval $(-1, \infty)$. A short exercise shows that

$$f^{(l)}(y) = (-1)^l \frac{l!}{(1+y)^{l+1}}.$$

Taylor's theorem tells us that for $x \in (0, 1)$:

$$f(x) = \sum_{l=0}^{k-1} (-x)^l + R_k(x), \tag{A.5}$$

where

$$R_k(x) = \frac{(-x)^k}{(1+\xi)^{k+1}}$$

for some $\xi \in (0, x)$. We have no control over which value ξ takes. We can however estimate:

$$\left| \frac{(-1)^k}{(1+\xi)^{k+1}} \right| \leq 1$$

for $\xi \in (0, x)$. We deduce $|R_k(x)| \leq x^k$. Rearranging (A.5), we have:

$$\left| \frac{1}{1+x} - \sum_{l=0}^{k-1} (-x)^l \right| \leq x^k$$

for[3] $x \in [0, 1)$ and all k. As $k \to \infty$, the right hand side tends to zero for $x \in [0, 1)$. We conclude that for x in this range:

$$\frac{1}{1+x} = \sum_{l=0}^{\infty} (-x)^l.$$

Notice that Taylor's theorem doesn't fail to hold for $x \geq 1$. However, our control over the remainder term becomes too weak for us to sum the series.

Taylor's theorem is very powerful, but is frequently misused, so lots of care must be taken in its application. Notice that Taylor's theorem makes a statement about a finite number of derivatives. One is often tempted to assume that a (smooth) function is equal to its Taylor series. The following example shows that this is not the case.

Consider the function:

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

First of all, we note that on any interval which avoids $x = 0$, this function is certainly continuous, indeed smooth. In fact, I claim that this function is smooth on the whole real line. To show this, we'll prove a slightly more general result which will imply the statement we require.

Lemma A.11. *Suppose p_k is a polynomial of degree k. Consider the function:*

$$f(x) = \begin{cases} 0 & x \leq 0, \\ p_k\left(\frac{1}{x}\right) e^{-\frac{1}{x}} & x > 0. \end{cases}$$

This function is differentiable on $\mathbb{R}$, with derivative:

$$f'(x) = \begin{cases} 0 & x \leq 0, \\ p_{k+2}\left(\frac{1}{x}\right) e^{-\frac{1}{x}} & x > 0. \end{cases}$$

where p_{k+2} is a polynomial of order $k+2$ given by:

$$p_{k+2}(t) = t^2 \left(p_k(t) - p_k'(t)\right)$$

[3] Note the result holds trivially with $x = 0$

Proof. First we note that as $x \to 0$, we have:

$$\lim_{x\to 0} p_k\left(\frac{1}{x}\right)e^{-\frac{1}{x}} = 0,$$

by Exercise A.10, so f is continuous. Next, observe that for any $p < 0$, the function is constant in a neighbourhood of $x = p$, so $f'(p) = 0$ if $p < 0$. For $p > 0$, we have that if x is close enough to $x = p$, then:

$$f(x) = p_k\left(\frac{1}{x}\right)e^{-\frac{1}{x}}.$$

Invoking theorems A.2, A.3 and Exercise A.2, we have that:

$$f'(p) = \left(\frac{1}{p^2}p_k\left(\frac{1}{p}\right) - \frac{1}{p^2}p_k'\left(\frac{1}{p}\right)\right)e^{-\frac{1}{p}} = p_{k+2}\left(\frac{1}{p}\right)e^{-\frac{1}{p}}.$$

It remains to show that f is differentiable at $x = 0$. We need to consider the limit:

$$\lim_{x\to 0}\frac{f(x)}{x}.$$

Now, we know that:

$$\frac{f(x)}{x} = \begin{cases} 0 & x \leq 0, \\ \frac{1}{x}p_k\left(\frac{1}{x}\right)e^{-\frac{1}{x}} & x > 0. \end{cases}$$

Again invoking Exercise A.10, we conclude that

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

This establishes the result we require. □

Corollary A.12. *The function:*

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

is smooth on $\mathbb{R}$*, and* $f^{(k)}(0) = 0$*, for all* $k \in \mathbb{N}$*.*

Now consider the Taylor series for this function. We have that every term in the Taylor series is identically zero. Therefore the Taylor's series converges everywhere in $\mathbb{R}$ (as it is simply the zero series). On the other hand, $e^{-\frac{1}{y}} \neq 0$ for $y > 0$. This means that even though the Taylor series converges everywhere, the function only equals its Taylor series at the point $x = 0$!

Exercise A.10. a) By using the power series or otherwise, show that for any $N \in \mathbb{N}^*$ and $x \geq 0$ that:

$$1 + \frac{x^N}{N!} \leq \exp x.$$

b) Deduce that for $y > 0$:

$$e^{-\frac{1}{y}} \leq \frac{1}{1 + \frac{1}{N!y^N}} \leq N!y^N.$$

c) Suppose p_k is a polynomial of degree k, and consider the function:

$$f(x) = \begin{cases} 0 & x \leq 0, \\ p_k\left(\frac{1}{x}\right) e^{-\frac{1}{x}} & x > 0. \end{cases}$$

Show that:

$$\lim_{y \to 0} f(y) = 0.$$

Exercise A.11 (*). Consider the function:

$$f(x) = \begin{cases} 0 & |x| \geq 1, \\ e^{-\frac{1}{1-x^2}} & |x| < 1. \end{cases}$$

Show that f is smooth and non-negative on $\mathbb{R}$.