Lecture notes for Math 61CM, Analysis, Version 2018

Lenya Ryzhik*

November 28, 2018

Nothing found here is original except for a few mistakes and misprints here and there. These notes are simply a record of what I cover in class, to spare the students some of the necessity of taking the lecture notes. The readers should consult the original books for a better presentation and context. The material from the following books will be likely used: L. Simon ”An Introduction to Multivariable Mathematics”, V. Zorich ”Mathematical Analysis I”, and maybe W. Rudin ”Principle of Mathematical Analysis”.

1 The real numbers

In this section, we will introduce the set of real numbers $\mathbb{R}$ via a collection of several groups of axioms: we will postulate that the real numbers (1) form what is known as a field, (2) they are ordered, and (3) satisfy the completeness axiom. The latter may be less intuitive at the first sight but, as we will soon see, is absolutely crucial for the subject known as ”mathematical analysis”. The intuition behind the axioms is that they should formalize what we are know as ”real numbers” from ”past experience”.

1.1 Axioms of addition and multiplication

First, we need to take care of the arithmetic of real numbers. We will assume that the operations of addition and multiplication are defined for real numbers, and satisfy the following properties:

(A1) The addition operation is associative:

$$ (a + b) + c = a + (b + c) \text{ for all } a, b, c \in \mathbb{R}, $$

(A2) There exists a zero element denoted by 0, so that

$$ x + 0 = 0 + x = x \text{ for all } x \in \mathbb{R}. $$

(A3) For each $x \in \mathbb{R}$ there exists an element $(-x) \in \mathbb{R}$ such that

$$ x + (-x) = 0. $$

(A4) The addition operation is commutative:

$$ a + b = b + a \text{ for all } a, b \in \mathbb{R}. $$

Axioms (A1)-(A3) say that $\mathbb{R}$ is a group with respect to addition, and (A4) says that this group is commutative. We will go over the general notion of a group again in the linear algebra portion of the class.
The next group of axioms concerns multiplication. We will sometimes write the product of two numbers \(a, b \in \mathbb{R}\) as \(ab\) and sometimes as \(a \cdot b\), hoping this causes no confusion.

(M1) The multiplication operation is associative:

\[(ab)c = a(bc)\] for all \(a, b, c \in \mathbb{R}\).

(M2) There exists an identity element denoted by 1, so that \(1 \neq 0\), and

\[x \cdot 1 = 1 \cdot x = x\] for all \(x \in \mathbb{R}\).

(M3) For each \(x \in \mathbb{R}\) such that \(x \neq 0\), there exists an element \(x^{-1} \in \mathbb{R}\) such that

\[x \cdot x^{-1} = 1\].

(M4) The multiplication operation is commutative:

\[ab = ba\] for all \(a, b \in \mathbb{R}\).

These axioms mean that the set \(\mathbb{R} \setminus \{0\}\) is a commutative group with respect to multiplication. The last algebraic axiom, when taken together with the previous axioms, says that \(\mathbb{R}\) is a field, and is simply the distributive law with respect to addition and multiplication:

\[(F1)\quad a(b + c) = ab + ac\] for all \(a, b, c \in \mathbb{R}\).

The familiar arithmetic properties of real numbers follow from these axioms relatively easily. For example, to show that there is only one zero in the set of real numbers, let us assume that \(0_1\) and \(0_2\) are both zeros and write:

\[0_1 = 0_1 + 0_2 = 0_2 + 0_1 = 0_2.\]

Make sure that you understand which axioms were used in each identity above! To show that each element \(x \in \mathbb{R}\) has a unique negative element, let us assume that for some \(x \in \mathbb{R}\) there exist \(x_1\) and \(x_2\) such that

\[0 = x + x_1, \quad 0 = x + x_2.\]

Then we can write:

\[x_1 = x_1 + 0 = x_1 + (x + x_2) = (x_1 + x) + x_2 = 0 + x_2 = x_2.\]

Again, make sure that you understand which axioms were used in each identity above. The next exercise lists some of the other elementary arithmetic properties of real numbers.

**Exercise 1.1** Show using the above axioms that the following algebraic properties hold:

1. Given any \(a, b \in \mathbb{R}\), equation

   \[a + x = b\]

   has a unique solution \(x \in \mathbb{R}\).

2. For each \(x \in \mathbb{R}\) there exists a unique \(y\) such that \(xy = 1\).

3. For each \(a \neq 0\) and each \(b \in \mathbb{R}\) there exists a unique solution to the equation \(a \cdot x = b\).

4. For any \(x \in \mathbb{R}\) we have \(x \cdot 0 = 0\).

5. If \(xy = 0\) then either \(x = 0\) or \(y = 0\).

6. For any \(x \in \mathbb{R}\) we have \(-x = (-1) \cdot x\).

7. For any \(x \in \mathbb{R}\) we have \((-1) \cdot (-x) = x\).

8. For any \(x \in \mathbb{R}\) we have \((-x) \cdot (-x) = x \cdot x\).
1.2 The order axioms

The next set of axioms has little to do with arithmetic but rather ordering: it says that the real numbers form an ordered set. That is, there is a relation \( \leq \) between the real numbers such that

1. (O1) For each \( x \in \mathbb{R} \) we have \( x \leq x \).

2. (O2) If \( x \leq y \) and \( y \leq x \) then \( x = y \).

3. (O3) If \( x \leq y \) and \( y \leq z \) then \( x \leq z \).

These axioms mean that \( \mathbb{R} \) is a (partially) ordered set. We postulate it is totally ordered:

4. (O4) For each \( x, y \in \mathbb{R} \) either \( x \leq y \) or \( y \leq x \).

To connect the axioms of addition, multiplication to the order axioms, we add the following two axioms.

1. (OA) If \( x \leq y \) and \( z \in \mathbb{R} \) then \( x + z \leq y + z \).

2. (OM) If \( 0 \leq x \) and \( 0 \leq y \) then \( 0 \leq xy \).

If \( x \leq y \) and \( x \neq y \) then we use the notation \( x < y \). The relations \( \geq \) and \( > \) are defined in the obvious way.

Note that we have \( 0 < 1 \). Indeed, the multiplication axioms imply that \( 0 \neq 1 \). If we have \( 1 < 0 \), then adding \( -1 \) to both sides gives \( 0 < -1 \), and (OM) implies that \( 0 < (-1)(-1) = 1 \), which is a contradiction to the assumption \( 1 < 0 \). We used claim (8) of Exercise 1.1 in the last step.

The next exercise shows that ordering is actually a non-trivial structure – not every field can be ordered.

**Exercise 1.2** Show that not every set satisfying the addition and multiplication axioms can be ordered. Hint: consider the set \( \{0, 1\} \) with \( 0 + 0 = 0, 0 + 1 = 1, 1 + 1 = 0 \) and \( 0 \cdot 0 = 1 \cdot 0 = 0, 1 \cdot 1 = 1 \). Show that it satisfies the addition and multiplication axioms but can not be ordered.

1.3 The completeness axiom

As you can readily check, the axioms of addition, multiplication and order do not yet narrow down the intuition we have for the set of real numbers that should, at the very least, include what we know as rational numbers and irrational numbers. Indeed, these axioms are satisfied by the set of rational numbers alone and hence are not yet sufficient. As a side note, strictly speaking, to define the rational numbers we need to know what integers are but let us disregard this for the moment. What is clear is that we need another axiom that would not be satisfied by what we think of as "rational numbers". This is achieved by the completeness axiom.

**The completeness axiom.** If \( X \) and \( Y \) are non-empty sets of real numbers such that for any \( x \in X \) and \( y \in Y \) we have \( x \leq y \), then there exists \( c \in \mathbb{R} \) such that \( x \leq c \leq y \) for all \( x \in X \) and \( y \in Y \).

In a sense, this axiom says that \( \mathbb{R} \) has no holes. On other hand, the set \( \mathbb{Q} \) of rational numbers does have holes: if we take

\[
X = \{ r \in \mathbb{Q} : r^2 < 2 \}, \quad Y = \{ r \in \mathbb{Q} : r^2 > 2 \text{ and } r > 0 \},
\]

then there is no \( c \in \mathbb{Q} \) such that for any \( x \in X \) and \( y \in Y \) we have \( x \leq c \leq y \) – this is because \( \sqrt{2} \) is irrational. No rational number squared can equal to 2 and if a rational \( c \) separating the sets \( X \) and \( Y \) were to exist, it is not hard to see that it would have to satisfy \( c^2 = 2 \) – we will come back to this a little later. Thus, the completeness axiom really distinguishes the reals from the set of rational numbers. To see how this axiom can be used, let us give some definitions.

**Definition 1.3** (1) A set \( S \subseteq \mathbb{R} \) is bounded from above if there exists a number \( K \in \mathbb{R} \) such that any \( x \in S \) satisfies \( x \leq K \). Such \( K \) is called an upper bound for \( S \).

(2) A set \( S \subseteq \mathbb{R} \) is bounded from below if there exists a number \( K \in \mathbb{R} \) such that any \( x \in S \)
satisfies $K \leq x$. Such $K$ is called a lower bound for $S$.

(3) A set $S \subseteq \mathbb{R}$ is bounded if it is bounded both from above and from below.

**Exercise 1.4** Let us define $|x|$ as $|x| = x$ if $0 \leq x$ and $|x| = -x$ if $0 \leq -x$. Show that a set $S$ is bounded if and only if there exists $M \in \mathbb{R}$ such that $|x| \leq M$ for all $x \in S$.

**Definition 1.5** We say that $x \in S$ is the maximum of a set $S$ if for any $y \in S$ we have $y \leq x$. Similarly, we say that $x \in S$ is the minimum of a set $S$ if for any $y \in S$ we have $x \leq y$.

Note that the maximum and minimum of a set belong to the set, by definition. However, not every set has a maximum or a minimum.

**Exercise 1.6** Show that the set $S = \{x : 0 < x < 1\}$ has neither a minimum nor a maximum.

**Definition 1.7** (1) A number $K$ is the least upper bound for a set $S \subseteq \mathbb{R}$, denoted as $K = \sup S$, if $K$ is an upper bound for $S$ and any upper bound $M$ for $S$ satisfies $K \leq M$.

(2) A number $K$ is the greatest lower bound for a set $S \subseteq \mathbb{R}$, denoted as $K = \inf S$, if $K$ is a lower bound for $S$ and any lower bound $M$ for $S$ satisfies $M \leq K$.

In other words, $\sup S$ is the minimum of the set of the upper bounds for $S$, and $\inf S$ is the maximum of the set of the lower bounds for $S$. As we have seen, not every set has a maximum or a minimum. The completeness axiom for the real numbers has the following consequence showing that the set of upper bounds for any set has a minimum.

**Lemma 1.8** (The least upper bound principle). Every non-empty set of real numbers that is bounded from above has a unique least upper bound.

**Proof.** Uniqueness of the least upper bound follows from the fact that (if it exists) it is the minimum of the set of all upper bounds, and every set can have have at most one minimum. Thus, we only need to show that a least upper bound exists. Let $Y$ be the set of all upper bounds for a non-empty set $X$. Since $X$ is bounded, the set $Y$ is not empty, and by the definition of an upper bound, we know that for any $x \in X$ and $y \in Y$ we have $x \leq y$. The completeness axiom then implies that there exists $c \in \mathbb{R}$ such that for any $x \in X$ and $y \in Y$ we have $x \leq c \leq y$. The first inequality implies that $c$ is an upper bound for $X$ and the second implies that it is the least upper bound. □

**Exercise 1.9** Show that the conclusion of Lemma 1.8 does not hold for the set $\mathbb{Q}$ of rational numbers. That is, exhibit a set $S$ of rational numbers such that there is no smallest rational upper bound for $S$.

One may wonder if one can construct a version of real numbers explicitly and verify that it satisfies the axioms. A standard way is to start with the rational numbers and use what is known as the Dedekind cuts. We will not do this here for lack of time, and simply note that an interested reader should have no difficulty finding it in a variety of books or on the Internet.

### 1.4 The natural numbers and the principle of mathematical induction

We understand commonly that the natural numbers are $1, 1+1, 1+1+1, \ldots$ In order to make this more formal, we give the following definition.

**Definition 1.10** A set $X \subseteq \mathbb{R}$ is inductive if for each $x \in X$, the set $X$ also contains $x + 1$.

**Definition 1.11** The set $\mathbb{N}$ of natural numbers is the intersection of all inductive sets that contain $1$. 
Exercise 1.12 Show that $\mathbb{N}$ is an inductive set.

Lemma 1.13 The sum of two natural numbers is a natural number.

Proof. Let $S$ be the set of all natural numbers $n$ such that for any $m \in \mathbb{N}$ we have $n + m \in \mathbb{N}$. Then $S$ is not empty because $1 \in S$ (this, in turn, is because $\mathbb{N}$ is an inductive set). Moreover, if $n \in S$ then $n + 1$ is also in $S$ because for any $m \in \mathbb{N}$ we have

$$(n + 1) + m = n + (m + 1) \in \mathbb{N},$$

because $n \in S$ and $m + 1 \in \mathbb{N}$. Thus, $S$ is an inductive set, and, as it is a subset of $\mathbb{N}$, it has to be equal to $\mathbb{N}$. $\square$

Exercise 1.14 Show that the product of two natural numbers is a natural number.

Exercise 1.15 Show that if $y \in \mathbb{N}$ and $y \neq 1$ then $y - 1 \in \mathbb{N}$. Hint: let $E$ be the set of all real numbers of the form $n - 1$ where $n \in \mathbb{N}$ and $n > 1$. Show that $E = \mathbb{N}$.

Lemma 1.16 If $n \in \mathbb{N}$ then there are no natural numbers $x$ such that $n < x < n + 1$.

Proof. We will show that the set $S$ of all natural numbers for which this assertion holds contains 1 and is an inductive set. To show that $1 \in S$, let

$$M = \{x \in \mathbb{N} : x = 1 \text{ or } x \geq 2\}.$$ Then, obviously $1 \in M$ and if $x \in M$ then either $x = 1$ so that $x + 1 = 2 \in M$ or $x \geq 2$ and then $x + 1 \geq 2$, thus $x + 1 \in M$ also in that case. Thus, $M$ is a subset of $\mathbb{N}$ that is an inductive set containing 1, hence $M = \mathbb{N}$. This means that the assertion of the lemma holds for 1 and $1 \in S$. Next, assume that $x \in S$—we will show that $x + 1 \in S$. Assume that there is $y \in \mathbb{N}$ such that $n + 1 < y < n + 2$. Then $y > 1$, thus $y - 1$ is in $\mathbb{N}$ and $n < y - 1 < n + 1$, meaning that $n \notin S$, which is a contradiction. $\square$

The set $\mathbb{Z}$ of integers consists of numbers $x \in \mathbb{R}$ such that either $x = 0$, or $x \in \mathbb{N}$, or $-x \in \mathbb{N}$. Now, we can define the rational numbers as usual: they have the form $\frac{m}{n}$ with $m, n \in \mathbb{Z}$, $n \neq 0$. The irrational numbers are those that are not rational. In order to show that irrational numbers exist, we make the following observation.

Exercise 1.17 (1) Show that if $s > 0$ and $s^2 < 2$ then there exists $\delta > 0$ so that $(s + \delta)^2 < 2$. Hint: you should be able find an explicit $\delta > 0$ in terms of $s$.
(2) Show that if $s > 0$ and $s^2 > 2$ then there exists $\delta > 0$ so that $(s - \delta)^2 < 2$. Hint: again, pick an explicit $\delta > 0$ in terms of $s$.
(3) Let $X = \{x > 0 : x^2 < 2\}$. Show that $X$ is bounded from above and set $\bar{s} = \sup S$. Show that $\bar{s}^2 = 2$ and that $\bar{s}$ is irrational.

1.5 The Archimedes principle

The Archimedes principle is a key to the approximation theory, and says the following.

Lemma 1.18 (The Archimedes principle) For any fixed $h > 0$ and any $x \in \mathbb{R}$ there exists an integer $k \in \mathbb{Z}$ so that

$$(k - 1)h \leq x < kh.$$
Proof. We first note that \( \mathbb{N} \) is not bounded from above. Indeed, if it is bounded, set \( s = \sup \mathbb{N} \). If \( s \notin \mathbb{N} \), then, by the definition of supremum, there must be \( n \in \mathbb{N} \) such that \( n > s - 1 \), but then \( s < n + 1 \) and thus \( s \) is not an upper bound for \( \mathbb{N} \). However, if \( s \in \mathbb{N} \) then \( s + 1 \notin \mathbb{N} \) meaning that \( s \) is again not an upper bound for \( \mathbb{N} \). This shows that \( \mathbb{N} \) is not bounded from above, and thus \( \mathbb{Z} \) is not bounded from above or from below. A similar argument shows that if \( S \) is a set of integers bounded from above, it must contain a maximal element, and if it is a set of integers bounded from below, it must contain a minimal element. Now, turning to the proof of the Archimedes principle, let \( S = \{ n \in \mathbb{Z} : \frac{x}{h} < n \} \). This set is bounded from below and thus contains a minimal element \( k \). Then we have

\[
k - 1 \leq \frac{x}{h} < k,
\]

and the conclusion of Lemma 1.18 follows. □

Here are some important corollaries of the Archimedes principle.

**Corollary 1.19**

(i) For any \( \varepsilon > 0 \) there exists \( n \in \mathbb{N} \) such that \( 0 < 1/n < \varepsilon \).

(ii) For any \( a, b \in \mathbb{R} \) such that \( a < b \) there exists \( r \in \mathbb{Q} \) such that \( a < r < b \).

**Exercise 1.20** Prove the assertions of Corollary 1.19.

2 The nested intervals lemma and the Heine-Borel theorem

In this section, we will introduce the first ideas related to the notion of compactness.

2.1 The nested interval lemma

**Definition 2.1** An infinite sequence of sets \( X_1, X_2, \ldots, X_n, \ldots \) is nested if for each \( n \in \mathbb{N} \) we have \( X_{n+1} \subseteq X_n \).

A closed interval \([a, b]\) is simply the set \([a, b] = \{ x : a \leq x \leq b \}\), and an open interval is the set \((a, b) = \{ x : a < x < b \}\).

**Lemma 2.2** (The nested intervals lemma). Let \( I_1 \supseteq I_2 \supseteq \cdots \supseteq I_n \supseteq \cdots \) be a nested sequence of closed intervals. Then there exists \( c \in \mathbb{R} \) that belongs to all \( I_k \), \( k \in \mathbb{N} \). If, in addition, for any \( \varepsilon > 0 \) there exists \( k \) such that \( |I_k| < \varepsilon \), then there is exactly one point \( c \) common to all intervals.

**Proof.** Since \( I_k \) is a nested sequence of intervals, we have \( a_k \leq a_m \leq b_m \leq b_k \) for all \( m \geq k \). Thus, the sets \( A = \{ a_n : n \in \mathbb{N} \} \) and \( B = \{ b_n : n \in \mathbb{N} \} \) satisfy the assumptions of the completeness axiom. Hence, there exists \( c \in \mathbb{R} \) such that \( a_m \leq c \leq b_k \) for all \( m, k \in \mathbb{N} \), and, in particular, we have \( a_n \leq c \leq b_n \) for all \( n \in \mathbb{N} \), thus \( c \in I_n \) for all \( n \in \mathbb{N} \).

If there are two points \( c_1 \neq c_2 \) that belong to all \( I_k \), say, with \( c_1 < c_2 \), then we have

\[
a_n \leq c_1 < c_2 \leq b_n \text{ for all } n,
\]

hence \( |b_n - a_n| > c_2 - c_1 \), and the length of all intervals \( I_n \) is larger than \( c_2 - c_1 \). □

**Exercise 2.3** Where did we use the fact that the intervals \( I_n \) are closed in the above proof? Give an example of a nested sequence of open intervals \( I_n = (a_n, b_n) \) that has an empty intersection.
2.2 The Heine-Borel lemma

Definition 2.4 A collection $\mathcal{F}$ of sets is said to cover a set $Y$ if for every element $y \in Y$ there exists a set $X \in \mathcal{F}$ such that $y \in X$.

A collection of sets $\mathcal{G}$ is a subcollection of $\mathcal{F}$ if every set $X \in \mathcal{G}$ is also in the collection $\mathcal{F}$.

Lemma 2.5 (The Heine-Borel lemma) Every collection $\mathcal{F}$ of open intervals covering a closed interval $[a, b]$ contains a finite subcollection that also covers $[a, b]$.

Proof. Let $\mathcal{F}$ be a collection of open intervals covering a closed interval $[a, b]$ and let $S$ be the set of points $y$ in $[a, b]$ with the following property: the interval $[a, y]$ can be covered by finitely many intervals from $\mathcal{F}$. As the point $a$ itself is covered by some interval $(\alpha, \beta) \in \mathcal{F}$, with $\alpha < a < \beta$, we know that $S$ is not empty: it contains $a$, as well as all points $z \in [a, b]$ such that $a \leq z < \beta$. Moreover, the set $S$ is bounded from above: any element $y \in S$ satisfies $y \leq b$. It is easy to see that if $y \in S$ then any $x$ such that $a \leq x \leq y$ is also in $S$. Let $\bar{s} = \operatorname{sup} S$, and note that $\bar{s} \in [a, b]$. We claim that $\bar{s} \in S$ and $\bar{s} = b$. To see that $\bar{s} \in S$, assume that $\bar{s} \notin S$, and observe that the point $\bar{s}$ is covered by some open interval $(\alpha', \beta') \in \mathcal{F}$, with $\alpha' < \bar{s} < \beta'$. Since $\bar{s}$ is the least upper bound for $S$, and $\bar{s} \notin S$, there must be a point $x \in S$ such that $\alpha' < x \leq \bar{s}$. Since $x \in S$, we can cover the interval $[a, b]$ by finitely many intervals from $\mathcal{F}$. Adding the interval $(\alpha', \beta') \in \mathcal{F}$ to this collection gives us a finite collection covering the interval $[a, \bar{s}]$, contradicting the assumption that $\bar{s} \notin S$. Therefore, we have $\bar{s} \in S$. However, this also shows that unless $\bar{s} = b$, the points in the interval $[\bar{s}, \min(\beta', b)]$ are also in $S$, thus $\bar{s}$ is not an upper bound for $S$, which is a contradiction. Thus, we have $b = \bar{s}$. As we have shown that $\bar{s} \in S$, we know that $b \in S$, and we are done, by the definition of the set $S$. $\square$

Exercise 2.6 Where did we use the fact that the intervals in the covering collection $\mathcal{F}$ are open? Where did we use the fact that the interval $[a, b]$ is closed? Construct a collection $\mathcal{F}$ of closed intervals $[a, b]$ that covers the closed interval $[0, 1]$ but no finite sub-collection of $\mathcal{F}$ covers $[0, 1]$. Next, construct a collection $\mathcal{G}$ of open intervals $(a, b)$ that covers the open interval $(0, 1)$ but no finite sub-collection of $\mathcal{G}$ covers $(0, 1)$.

Exercise 2.7 Devise an alternative proof of the Heine-Borel lemma as follows. Assume that there is no finite sub-collection of open intervals from $\mathcal{F}$ that covers $I_0 = [a, b]$. Then either there is no finite sub-collection that covers $[a, a_1]$ or there is no finite sub-collection that covers $[a_1, b]$, where $a_1 = (a + b)/2$. Choose the interval $I_2$ out of these two that can not be covered, split it into two sub-intervals, choose a sub-interval $I_3$ that can not be covered by a finite sub-collection and repeat this procedure. This will give you a sequence of nested closed intervals $I_1 \supseteq I_2 \supseteq I_3 \supseteq \cdots \supseteq I_n \supseteq \cdots$. The Nested Intervals Lemma implies that there is a point $c$ that belongs to all of these intervals. As $c \in [a, b]$, it is covered by some open interval $I$ from the collection $\mathcal{F}$. Use this to arrive at a contradiction to the fact that none of $I_k$ can be covered by finitely many intervals from $\mathcal{F}$.

2.3 Limit points and the Bolzano-Weierstrass lemma

Let us recall that a neighborhood of a point $x$ is an open interval $(a, b)$ that contains $x$.

Definition 2.8 A point $z \in \mathbb{R}$ is a limit point of a set $X$ if any neighborhood of $z$ contains infinitely many elements of $X$.

Exercise 2.9 Show that $z$ is a limit point of a set $X$ if and only if any neighborhood of $z$ contains at least one element of $X$ different from $z$ itself.
As an example, part (i) of the Corollary 1.19 of the Archimedes principle shows that \( z = 0 \) is a limit point of the set \( \{1/n : n \in \mathbb{N}\} \).

**Exercise 2.10** Use the Archimedes principle to show that every \( x \in \mathbb{R} \) is a limit point of the set \( \mathbb{Q} \) of rational numbers.

The following lemma is a cornerstone of many things to come.

**Lemma 2.11** (The Bolzano-Weierstrass lemma) Every bounded set of real numbers has at least one limit point.

**Proof.** Let \( X \) be an infinite bounded subset of \( \mathbb{R} \). As \( X \) is bounded, there exists \( M > 0 \) so that \( X \) is contained in the interval \( I = [-M, M] \). Assume that no point in \( I \) is a limit point of \( X \). Then for each \( x \in I \) we can find an open interval \( I_x = (a_x, b_x) \) that contains \( x \) such that there are only finitely many points of \( X \) inside \( I_x \). The open intervals \( I_x \) form an open cover of the closed interval \( I \), and the Heine-Borel lemma implies that there exists a sub-cover of \( I \) by a finite sub-collection of intervals \( I_{x_k} = (a_{x_k}, b_{x_k}) \), \( k = 1, \ldots, N \) with some \( N \in \mathbb{N} \). Each interval \( I_{x_k} \) contains only finitely many elements of \( X \), hence there are only finitely many elements of \( X \) in the union of all these intervals. As no elements of \( X \) can lie outside of this union, there are only finitely many elements in \( X \) which is a contradiction. □

Another property we will use often is the following.

**Lemma 2.12** Let \( S \) be a bounded from above set that does not have a maximum. Then \( \sup S \) is a limit point of \( S \).

**Exercise 2.13** Prove Lemma 2.12.

3 Limits

3.1 Definition of the limit of a sequence

We first define a limit point of a sequence. One may informally think of them as points where the sequence "bunches up". Here is a way to formalize this.

**Definition 3.1** We say that \( z \in \mathbb{R} \) is a limit point of a sequence \( a_n \) if for any \( \varepsilon > 0 \) there exist infinitely many \( k \in \mathbb{N} \) such that \( |a_k - z| < \varepsilon \).

Note a subtle difference between a limit point of a sequence and a limit point of the set of its values. For instance, for a constant sequence \( 1, 1, 1, 1, \ldots \), that is, \( a_n = 1 \) for all \( n \in \mathbb{N} \), the set of its values is \( \{1\} \) -- it consists of one point and has no limit points. However, \( x = 1 \) is a limit point of this sequence. This small difference will be of no "serious" importance for us but one should keep it in mind.

**Exercise 3.2** Show that \( z \in \mathbb{R} \) is a limit point of a sequence \( a_n \) if and only if any open interval \( (c, d) \) that contains \( z \), also contains infinitely many elements of the sequence \( a_n \): there exist infinitely many \( n \) so that \( a_n \in (c, d) \).

A sequence may have more than one limit point. For instance, the sequence \( 1, -1, 1, -1, 1, \ldots \), that is, \( a_n = (-1)^{n+1} \) has exactly two limit points \( z = 1 \) and \( z = -1 \).

**Exercise 3.3** (1) Construct a sequence \( a_n \) that has no limit points. (2) Construct a sequence \( b_k \) that has infinitely many limit points. (3) Construct a sequence \( c_k \) such that any point \( z \in \mathbb{R} \) is a limit point of \( b_k \).
Next, we introduce the notion of the limit of a sequence.

**Definition 3.4** A point \( z \in \mathbb{R} \) is the limit of a sequence \( a_n \) if for any open interval \((c, d)\) that contains \( z \) there exists \( N \) so that all \( a_k \) with \( k \geq N \) lie inside \((c, d)\). We write this as \( a_n \to z \) as \( n \to +\infty \), or as \( \lim_{n \to \infty} a_n = z \).

The following exercise gives a slightly more "practical" definition of the limit.

**Exercise 3.5** Show that a point \( z \) is the limit of a sequence \( a_n \) if and only if for any \( \varepsilon > 0 \) there exists \( N(\varepsilon) \) so that \( |a_k - z| < \varepsilon \) for all \( k \geq N(\varepsilon) \).

A small advantage of Definition 3.4 over the one in Exercise 3.5, and the reason to introduce it first, is that it will be easy to generalize to spaces other than \( \mathbb{R} \) where the notion of a distance may be not defined but the notion of a neighborhood may be. However, the statement in Exercise 3.5 is much more "practical" and we will mostly use it rather than Definition 3.4 directly.

**Important:** unless otherwise specified, we will usually refer to Exercise 3.5 as the definition of the limit of a sequence. For the first reading, the reader should really think of Exercise 3.5 as the definition of the limit of a sequence.

The next important exercise gives another informal way to think of the limit of a sequence: this is the only point where the sequence "bunches up".

**Exercise 3.6** Show that a point \( z \) is the limit of a bounded sequence \( a_n \) if and only if \( z \) is the only limit point of \( a_n \).

Here are some examples of limits.

**Exercise 3.7** Show that

(a) \( \lim_{n \to \infty} \frac{1}{n} = 0 \)

(b) \( \lim_{n \to \infty} \frac{\sin n}{n} = 0 \)

(c) \( \lim_{n \to \infty} \frac{1}{q^n} = 0 \) for any \( q > 1 \).

Hints: for part (a) – use part (i) in Corollary 1.19; for part (c) – write \( q = 1 + \delta \) with \( \delta > 0 \) and use induction to show that \( q^n \geq 1 + n\delta \). Then apply part (a).

### 3.2 Basic properties of converging sequences

**Definition 3.8** (1) A sequence \( a_n \) is bounded if there exists \( M \in \mathbb{R} \) so that \( |a_k| \leq M \) for all \( k \in \mathbb{N} \).

(2) A sequence \( a_n \) is bounded from below if there exists \( M \in \mathbb{R} \) so that \( a_k \geq M \) for all \( k \in \mathbb{N} \).

(3) A sequence \( a_n \) is bounded from above if there exists \( M \in \mathbb{R} \) so that \( a_k \leq M \) for all \( k \in \mathbb{N} \).

(4) A sequence \( a_n \) is increasing if \( a_{k+1} \geq a_k \) for all \( k \in \mathbb{N} \).

(5) A sequence \( a_n \) is strictly increasing if \( a_{k+1} > a_k \) for all \( k \in \mathbb{N} \).

(6) A sequence \( a_n \) is decreasing if \( a_{k+1} \leq a_k \) for all \( k \in \mathbb{N} \).

(7) A sequence \( a_n \) is strictly decreasing if \( a_{k+1} < a_k \) for all \( k \in \mathbb{N} \).

(8) A sequence \( a_n \) is monotone if it is either decreasing or increasing.

(9) A sequence is converging if it has a limit.

**Theorem 3.9** A convergent sequence is bounded.
Proof. Let $a_n$ be a convergent sequence, with $\lim_{n \to \infty} a_n = \ell$. Take $\varepsilon = 1$ in the definition of the limit of a sequence (again, in Exercise 3.5) – this implies existence of $N$ so that $|a_k - \ell| < 1$ for all $k \geq N$. It follows that all $a_k$ with $k \geq N$ satisfy $\ell - 1 \leq a_k \leq \ell + 1$. There are only finitely many elements $a_j$ with $j < N$. Thus, if we set 

$$m = \min(a_1, \ldots, a_{N-1}, \ell - 1), \ M = \max(a_1, \ldots, a_{N-1}, \ell + 1),$$

then all $a_k$ satisfy $m \leq a_k \leq M$, hence the sequence $a_n$ is bounded. □

An immediate consequence is that any unbounded sequence can not have a limit.

Exercise 3.10 Not all bounded sequences converge: give an example of a bounded sequence that has no limit.

The next theorem gives a condition for convergence that is actually very useful in many applications as monotone sequences arise very often.

Theorem 3.11 If $a_n$ is monotone and bounded, then it converges. Moreover, if $S = \{a_1, a_2, \ldots, a_k, \ldots\}$, then (1) If $a_n$ is increasing and bounded from above then $\lim_{n \to \infty} a_n = \sup S$.

(2) If $a_n$ is decreasing and bounded from below, the $\lim_{n \to \infty} a_n = \inf S$.

Proof. We will only prove (1). Let us assume that $a_n$ is increasing and bounded from above. Then the set $S$ is bounded from above, thus $\ell = \sup S$ exists. Let $\varepsilon > 0$, then, as $\ell - \varepsilon$ is not an upper bound for $S$ (this follows from the definition of $\sup S$), there exists $N(\varepsilon)$ so that $a_{N(\varepsilon)} > \ell - \varepsilon$.

As the sequence $a_n$ is monotonically increasing, it follows that for all $k \geq N(\varepsilon)$ we have $a_k \geq \ell - \varepsilon$ as well. However, as $\ell$ is an upper bound for the sequence $a_n$, we also know that $a_k \leq \ell$. Thus, we have $|a_k - \ell| < \varepsilon$ for all $k \geq N(\varepsilon)$ and we are done. □

Theorem 3.12 Let $a_n$ and $b_n$ be convergent sequences with $\lim_{n \to \infty} a_n = A$ and $\lim_{n \to \infty} b_n = B$.

(1) The sequence $a_n + b_n$ also converges and $\lim_{n \to \infty} (a_n + b_n) = A + B$.

(2) The sequence $a_nb_n$ also converges and $\lim_{n \to \infty} (a_nb_n) = AB$.

(3) If, in addition, $B \neq 0$, then the sequence $a_n/b_n$ also converges and $\lim_{n \to \infty} (a_n/b_n) = A/B$.

Proof. We will only prove (2). Let us write

$$a_nb_n - AB = a_nb_n - a_nB + a_nB - AB = a_n(b_n - B) + (a_n - A)B.$$

The triangle inequality implies that

$$|a_nb_n - AB| \leq |a_n||b_n - B| + |a_n - A||B|.$$

Since $a_n$ is a convergent sequence, there exists $M$ so that $|a_n| \leq M$ for all $n \in \mathbb{N}$, which leads to

$$|a_nb_n - AB| \leq M|b_n - B| + |a_n - A||B|.$$

Now, given $\varepsilon > 0$ find $N_1$ so that $|a_n - A| \leq \varepsilon/(2|B|)$ for all $n \geq N_1$, and $N_2$ so that $|a_n - A| \leq \varepsilon/(2M)$ for all $n \geq N_1$. Then, for all $n \geq N = \max(N_1, N_2)$, we have

$$|a_nb_n - AB| \leq M|b_n - B| + |a_n - A||B| \leq M\frac{\varepsilon}{2M} + \frac{\varepsilon}{2|B|}|B| = \varepsilon.$$

Hence, the sequence $a_nb_n$ converges to $AB$. □
Exercise 3.13 (1) Prove assertion (1) in Theorem 3.12.
(2) Prove assertion (3) in Theorem 3.12. Hint: it is easier to first show that the sequence $1/b_n$ converges to $1/B$.
(3) Let $a_n = 1/n$, find a sequence $b_n$ such that the limit of $a_n/b_n$ exists, and a sequence $d_n$ such that the limit of $a_n/b_n$ does not exist.
(4) Assume that $\lim_{n \to \infty} a_n = A > 0$, and $\lim_{n \to \infty} b_n = 0$. Show that the sequence $a_n/b_n$ does not converge.

Theorem 3.14 Assume that $a_n$ and $b_n$ are two convergent sequences with

$$\lim_{n \to \infty} a_n = A \text{ and } \lim_{n \to \infty} b_n = B.$$ 

If $A < B$ then there exists $N$ so that $a_n < b_n$ for all $n \geq N$.

Exercise 3.15 Prove this theorem.

3.3 The number $e$

Theorem 3.16 The sequence $x_n = \left(1 + \frac{1}{n}\right)^n$ converges. Its limit is denoted as $e$:

$$e = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n \quad (3.1)$$

Proof. We will look instead at the sequence

$$y_n = \left(1 + \frac{1}{n}\right)^{n+1}$$

and show that it is decreasing. For that, we need the inequality

$$(1 + \alpha)^n \geq 1 + n\alpha \quad (3.2)$$

that holds for all $\alpha \geq 0$ and all $n \in \mathbb{N}$. Recall that we have seen this inequality in Exercise 3.7. Now, we write

$$\frac{y_{n-1}}{y_n} = \frac{\left(1 + \frac{1}{n-1}\right)^n}{\left(1 + \frac{1}{n}\right)^{n+1}} = \frac{n^{n-1}}{(n-1)^n} \cdot \frac{n^{n+1}}{(n+1)^n} = \frac{n^2}{(n^2-1)^n} \cdot \frac{n}{n+1} = \frac{n^2}{n^2-1} \cdot \frac{n}{n+1}.$$ 

Now, we use (3.2) with $\alpha = 1/(n^2 - 1)$, to get

$$\frac{y_{n-1}}{y_n} \geq \left(1 + \frac{n}{n^2-1}\right) \cdot \frac{n}{n+1} \geq \left(1 + \frac{1}{n}\right) \cdot \frac{n}{n+1} = 1.$$ 

Thus, the sequence $y_n$ is a positive decreasing sequence, and Theorem 3.11 implies that $\lim_{n \to \infty} y_n$ exists. Thus, the sequence $x_n$ can be written as a product of two converging sequences:

$$x_n = \left(1 + \frac{1}{n}\right)^n = \left(1 + \frac{1}{n}\right)^{n+1} \frac{1}{1 + 1/n},$$

hence $x_n$ is itself converging. \(\square\)
Let us explain informally why (3.1) agrees with what we may know from a standard calculus course. It follows from (3.1) that we can write
\[ e = \left( 1 + \frac{1}{n} \right)^n + \alpha_n, \]  
(3.3)
with a sequence \( \alpha_n \) that goes to zero as \( n \to +\infty \). Taking the logarithm with respect to base \( e \), that we will denote simply by log, gives
\[ 1 = n \log \left( 1 + \frac{1}{n} \right) + \beta_n, \]  
(3.4)
with
\[ \beta_n = \log \left[ 1 + \alpha_n \left( 1 + \frac{1}{n} \right)^{-n} \right]. \]

**Exercise 3.17** Show that \( \lim_{n \to \infty} \beta_n = 0 \). Hint: use the fact that \( \alpha_n \to 0 \).

Using the result of Exercise 3.17 in (3.4) gives
\[ \log \left( 1 + \frac{1}{n} \right) = \frac{1}{n} + \frac{\beta_n}{n}. \]  
(3.5)

**Exercise 3.18** Recalling the calculus definition of the derivative (that we officially do not know yet), show that (3.5) implies that if the derivative of \( f(x) = \log x \) exists at \( x = 1 \), then \( f'(x) = 1 \). Show that \( e \) is the unique number \( a \) such that
\[ \frac{d}{dx} (\log_a x) \bigg|_{x=1} = 1. \]

### 3.4 The Cauchy criterion

**Definition 3.19** We say that \( a_n \) is a Cauchy sequence if for any \( \varepsilon > 0 \) there exists \( m \in \mathbb{N} \) so that \( |x_n - x_m| < \varepsilon \) for all \( n, m \geq N \).

**Theorem 3.20** A sequence \( a_n \) converges if and only if \( a_n \) is a Cauchy sequence.

**Proof.** One direction is easy. Assume that \( a_n \) converges and \( A = \lim a_n \). Given \( \varepsilon > 0 \) we can find \( N \in \mathbb{N} \) so that \( |a_n - A| < \varepsilon/2 \) for all \( n \geq N \). Then, for all \( n, m \geq N \) we have
\[ |a_n - a_m| \leq |a_n - A| + |A - a_m| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon, \]
hence \( a_n \) is a Cauchy sequence.

Next, assume that \( a_n \) is a Cauchy sequence. First, we claim that \( a_n \) is bounded. Indeed, taking \( \varepsilon = 1 \), we can find \( N \in \mathbb{N} \) such that \( |a_n - a_m| < 1 \) for all \( n, m \geq N \). In particular, we have \( a_N - 1 < x_m < a_N + 1 \) for all \( m \geq N \). In addition, there are only finitely many elements of the sequence \( a_n \) with \( n < N \), so the set \( \{a_1, a_2, \ldots, a_{N-1}\} \) is a bounded set. Hence, \( \{a_n\} \) is a union of two bounded sets, hence it is also bounded, and the sequence \( a_n \) is bounded as well. Thus, for each \( n \in \mathbb{N} \) we can define
\[ x_n = \inf_{k \geq n} a_k, \quad y_n = \sup_{k \geq n} a_k. \]

It is clear from the definition that \( x_n \leq x_{n+1} \leq y_{n+1} \leq y_n \) for all \( n \in \mathbb{N} \), so that the sequence \( x_n \) is increasing and the sequence \( y_n \) is decreasing. By the nested intervals theorem there exists a point \( A \) common to all intervals \( [x_n, y_n] \):
\[ x_n \leq A \leq y_n \text{ for all } n \in \mathbb{N}. \]
In addition, we have
\[ x_n \leq a_n \leq y_n \] for all \( n \in \mathbb{N} \),
and it follows that
\[ |A - a_n| \leq |x_n - y_n|. \tag{3.6} \]
However, given any \( \varepsilon > 0 \) we can find \( N \) so that for all \( m, N \geq N \) we have
\[ |a_n - a_m| < \frac{\varepsilon}{10}, \]
and in particular,
\[ |a_n - a_N| < \frac{\varepsilon}{10}. \]
Now, it follows from the definition of \( x_n \) and \( y_n \) that for all \( n \geq N \) we have
\[ |x_n - a_N| \leq \frac{\varepsilon}{10}, \quad |y_n - a_N| \leq \frac{\varepsilon}{10}, \]
hence
\[ |x_n - y_n| \leq \frac{2\varepsilon}{10} < \varepsilon. \]
We conclude from this and (3.6) that \( |A - a_n| < \varepsilon \) for all \( n \geq N \), thus \( a_n \) converges to \( A \) as \( n \to +\infty \). \( \square \)

**Exercise 3.21** Use the Cauchy criterion to show that the sequence
\[ a_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \]
does not converge. Hint: show that \( a_{2n} - a_n \geq 1/2 \) for all \( n \).

**Exercise 3.22** (i) Show that the claim of Theorem 3.20 does not hold for the set \( \mathbb{Q} \) of rational numbers. In other words, show that a Cauchy sequence of rational numbers does not necessarily converge to a rational number. This is another crucial difference between \( \mathbb{Q} \) and \( \mathbb{R} \). (ii) However, show, without using the results about real numbers, that if a sequence \( a_k \) of rational numbers is Cauchy and a subsequence \( a_{n_k} \) converges to a rational number \( \ell \in \mathbb{Q} \), then the sequence \( a_n \) converges to \( \ell \). In other words, a Cauchy sequence that has a converging subsequence must converge.

### 3.5 The Bolzano-Weierstrass theorem, \( \limsup \) and \( \liminf \)

**Definition 3.23** If \( x_n \) is a sequence, and \( n_k \in \mathbb{N} \) is an increasing sequence: \( n_1 < n_2 < n_3 \cdots < n_k < \ldots \), then the sequence \( y_k = x_{n_k} \) is called a subsequence of \( x_n \).

**Theorem 3.24** Every bounded sequence of real number contains a convergent subsequence.

**Proof.** Let \( E \) be the set of values of a bounded sequence \( x_n \). If \( E \) is finite, then there exists \( a \in \mathbb{R} \) so that \( x_n = a \) for infinitely many \( n \). It is then an easy exercise to see that there exists a subsequence \( x_{n_k} \) of \( x_n \) such that \( x_{n_k} = a \) for all \( k \), hence \( \lim_{k \to \infty} x_{n_k} = a \) and we are done. If the set \( E \) is infinite, then, as \( x_n \) is a bounded sequence, \( E \) is an infinite bounded set, hence by the Bolzano-Weierstrass Lemma 2.11, it has a limit point \( A \). Then one can choose \( n_1 \) so that \( |x_{n_1} - A| < 1 \). Next, note that, since \( A \) is a limit point of \( x_n \), the interval \((A - 1/2, A + 1/2)\) contains infinitely many elements of the sequence, hence one can choose \( n_2 > n_1 \) so that \( |x_{n_2} - A| < 1/2 \), and so on, at each step choosing \( n_{k+1} > n_k \) so that \( |x_{n_{k+1}} - A| < 1/(k + 1) \). As \( \lim_{k \to \infty} 1/(k + 1) = 0 \), the sequence \( x_{n_k} \) converges to \( A \), and we are done. \( \square \)
Definition 3.25 If there exists a subsequence \( x_{n_k} \) such that \( \ell = \lim_{k \to \infty} x_{n_k} \), then we say \( \ell \) is a limit of \( x_n \) along a subsequence.

We now define \( \limsup \) and \( \liminf \) of a sequence \( x_n \). As we have done in the proof of the Cauchy criterion for convergence, let us set
\[
a_n = \inf_{k \geq n} x_k, \quad b_n = \sup_{k \geq n} x_k.
\]
As we have observed in that proof, if \( x_n \) is bounded from below, then \( a_n \) are well-defined, and the sequence \( a_n \) is increasing, while if \( x_n \) is bounded from above, then \( b_n \) are well-defined, and the sequence \( b_n \) is decreasing. We denote the corresponding limits by
\[
\limsup_{n \to \infty} x_n = \lim_{n \to \infty} \left( \sup_{k \geq n} x_k \right), \quad \liminf_{n \to \infty} x_n = \lim_{n \to \infty} \left( \inf_{k \geq n} x_k \right).
\]

Proposition 3.26 Let \( x_n \) be a bounded sequence, then \( \liminf_{n \to \infty} x_n \) and \( \limsup_{n \to \infty} x_n \) are the largest and the smallest limits of \( x_n \) along a subsequence.


4 Continuity and limits of a function

4.1 Limit of a function at a point

Let \( f(x) \) be a function \( f: [c, d] \to \mathbb{R} \) and \( a \in [c, d] \).

Definition 4.1 We say that
\[
\lim_{x \to a} f(x) = A,
\]
in the sense of Cauchy if for every \( \varepsilon > 0 \) there exists \( \delta > 0 \) so that for any \( x \neq a \) such that \( |x - a| < \delta \) and \( x \in [c, d] \) we have \( |f(x) - A| < \varepsilon \).

Note that we exclude \( x = a \) in the above definition. The reason is that we do not want the value of \( \lim_{x \to a} f(x) \) to depend on the value of \( f(a) \). For instance, if \( f(x) = 1 \) for \( x \neq 0 \) and \( f(0) = 0 \), we would like to have \( \lim_{x \to 0} f(x) = 1 \).

An alternative definition is as follows.

Definition 4.2 We say that \( A \) is a sequential limit of \( f(x) \) as \( x \to a \) if for any sequence \( x_n \to a \), with \( x_n \neq a \) and \( x_n \in [c, d] \), we have
\[
\lim_{n \to \infty} f(x_n) = A.
\]

Theorem 4.3 The two definitions of the limit of \( f(x) \) at a point \( a \) are equivalent.

Proof. Let us first assume that
\[
\lim_{x \to a} f(x) = A,
\]
in the sense of Cauchy, and let \( x_n \) be a sequence such that \( x_n \to a \). Given \( \varepsilon > 0 \) there exists \( \delta > 0 \) so that for any \( x \neq a \) such that \( |x - a| < \delta \) we have \( |f(x) - A| < \varepsilon \). As \( x_n \to a \), given this \( \delta > 0 \), we can find \( N \) so that \( |x_n - a| < \delta \) for all \( n \geq N \), and then \( |f(x_n) - A| < \varepsilon \) for all \( n \geq N \), by the above, hence \( f(x_n) \to A \).
Next, assume that \( f(x) \) converges in the sequential sense to \( A \) as \( x \to a \) but that \( A \) is not the limit of \( f(x) \) in the Cauchy sense as \( x \to a \). This means that there exists \( \varepsilon_0 > 0 \) such that for any \( \delta > 0 \) there exists \( x \neq a \) such that \( |x - a| < \delta \) and \( |f(x) - A| > \varepsilon_0 \). Let us take \( \delta_n = 1/n \) – this will generate the corresponding sequence \( x_n \). It is easy to see that \( x_n \to a \) but \( f(x_n) \) does not converge to \( A \), which is a contradiction. □

**Exercise 4.4** Verify that the usual arithmetic and inequality properties hold for the limit of a function at a point.

### 4.2 Continuous functions and their basic properties

**Definition 4.5** (1) A function \( f(x) : [c,d] \to \mathbb{R} \) is continuous at a point \( a \in [c,d] \) if

\[
\lim_{x \to a} f(x) = f(a).
\]

(2) A function \( f : [c,d] \to \mathbb{R} \) is continuous on the interval \([c,d]\) if it is continuous at every point of \([c,d]\).

We summarize some basic properties of continuous functions in the following proposition.

**Proposition 4.6** (i) Let \( f(x) \) defined on an interval \([c,d]\) be continuous at a point \( z \in [c,d] \) and \( f(z) \neq 0 \), then there is a neighborhood \( U \) of \( z \) such that \( f(x) \) has the same sign as \( f(z) \) for all \( x \in U \) such that \( x \in [c,d] \).

(ii) If the functions \( f \) and \( g \) defined on \([c,d]\) are continuous at \( z \in [c,d] \), then so are the functions \( f + g \) and \( fg \). Also, the function \( f/g \) is continuous at \( z \) provided that \( g(z) \neq 0 \).

(iii) Let \( f \) be a continuous function defined on an interval \([c_1,d_1]\) and \( g \) be a continuous function defined on an interval \([c_2,d_2]\), such that \( g(x) \in [c_1,d_1] \) for all \( x \in [c_2,d_2] \), so that the composition \((f \circ g)(x) = f(g(x))\) is defined for all \( x \in [c_2,d_2] \). Show that if \( f \) is continuous on \([c_1,d_1]\) and \( g \) is continuous on \([c_2,d_2]\), then \( f \circ g \) is continuous on \([c_2,d_2]\). be two functions such that the composition

**Exercise 4.7** Prove Proposition 4.6, pay special attention to (i) and (iii).

The next theorem, and Corollary 4.9 show that a continuous function "can not skip values".

**Theorem 4.8** (Intermediate Value Theorem) Let \( f \) be a continuous function on a closed interval \([a,b]\) such that \( f(a) \) and \( f(b) \) have different signs, that is, \( f(a)f(b) \leq 0 \). Then there exists \( c \in [a,b] \) such that \( f(c) = 0 \).

**Proof.** If either \( f(a) = 0 \) or \( f(b) = 0 \), we are done, so let us assume without loss of generality that \( f(a) < 0 \) and \( f(b) > 0 \). Let us denote \( I_1 = [a,b] \) and divide \( I_1 \) in half. If the value of \( f \) at the mid-point is zero, we are done, otherwise, one of the two closed intervals has the same property: the signs of \( f \) at the two end-points are different. Let us call that interval \( I_2 = [a_1,b_1] \), divide it at its half-point, and continue. The process terminates when the value of \( f \) at one of the mid-points will be zero, which means we are done – we have found a point on \([a,b]\) where \( f \) vanishes. Otherwise, we get an infinite sequence of nested intervals \( I_k = [a_k,b_k] \) such that \( f(a_k) \) and \( f(b_k) \) have different signs, and \( I_{k+1} \subset I_k \). We also know that \(|I_k| = |a-b|/2^{k-1}\) goes to zero as \( k \to +\infty \). The Nested Intervals Lemma implies that there is a unique point \( c \in [a,b] \) in the intersection of all \( I_k \). Let \( a'_k \) be the endpoint of \( I_k \) such that \( f(a'_k) < 0 \) and \( b'_k \) be the endpoint of \( I_k \) such that \( f(b'_k) > 0 \). Note
that \( a_k^* \to c \) and \( b_k^* \to c \) as \( k \to +\infty \). As \( c \in [a, b] \), the function \( f \) is continuous at \( c \). Continuity of \( f \) at \( c \) implies that
\[
    f(c) = \lim_{k \to \infty} f(a_k^*), \quad f(c) = \lim_{k \to \infty} f(b_k^*).
\]
The first equality above implies that \( f(x) \leq 0 \) and the second implies that \( f(c) \geq 0 \). It follows that \( f(c) = 0 \). \( \square \)

**Corollary 4.9** Let \( f \) be a continuous function on a closed interval \([a, b]\) such that \( f(a) < A \) and \( f(b) > A \), or the other way around. Then there exists \( c \in [a, b] \) such that \( f(c) = A \).

**Exercise 4.10** Prove Corollary 4.9. Hint: apply the Intermediate Value Theorem to the function \( g(x) = f(x) - A \).

**Theorem 4.11** A continuous function on a closed interval \([a, b]\) is bounded and attains both its maximum and its minimum on \([a, b]\).

**Proof.** Since \( f \) is continuous on \([a, b]\), for every point \( x \in [a, b] \) there exists an open interval \( I_x \) containing \( x \) such that
\[
    f(x) - 1 \leq f(y) \leq f(x) + 1 \text{ for all } y \in I_x \text{ such that } y \in [a, b].
\]
The open intervals \( I_x \) form a cover of the closed interval \([a, b]\), hence the Heine-Borel lemma implies that there exists a finite sub-collection \( I_{x_1}, \ldots, I_{x_N} \) that also covers \([a, b]\). Let
\[
    M = 1 + \max(f(x_1), \ldots, f(x_N)),
\]
and
\[
    m = -1 + \min(f(x_1), \ldots, f(x_N)).
\]
As the intervals \( I_{x_1}, \ldots, I_{x_N} \) cover \([a, b]\), we know that every point \( z \in [a, b] \) belongs to some \( I_{x_k} \). It follows that
\[
    m - 1 \leq f(z) \leq M + 1,
\]
for all \( z \in [a, b] \), hence the function \( f \) is bounded on \([a, b]\).

To see that the maximum and minimum are attained, let
\[
    M = \sup_{x \in [a, b]} f(x) \quad (4.1)
\]
and assume that \( f(x) \neq M \) for all \( x \in [a, b] \). Then the function \( g(x) = 1/(M - f(x)) \) is continuous on \([a, b]\), thus, by what we have just proved, it is bounded: there exists \( K \) so that
\[
    g(x) \leq K \text{ for all } x \in [a, b].
\]
But then we have
\[
    M - f(x) \geq \frac{1}{K}, \text{ for all } x \in [a, b],
\]
so that
\[
    f(x) \leq M - \frac{1}{K}, \text{ for all } x \in [a, b],
\]
which contradicts the definition of \( M \) in (4.1). Thus, there has to exist \( c \in [a, b] \) such that \( f(c) = M \), hence \( f \) attains its maximum on \([a, b]\). The proof that \( f \) attains its minimum on \([a, b]\) is almost verbatim the same. \( \square \)
Exercise 4.12 Prove that a continuous function on a closed interval \([a, b]\) is bounded and attains its minimum on \([a, b]\).

Exercise 4.13 Give an alternative proof of Theorem 4.11 with the following outline that relies on the sequential definition of continuity: assume that \(M = \sup_{x \in [a, b]} f(x)\) but \(f(x) \neq M\) for any \(x \in [a, b]\).

Exercise 4.14 (i) Give an example of a continuous function on the open interval \((0, 1)\) that is unbounded. (ii) Give an example of a continuous function on the open interval \((0, 1)\) that is bounded but does not attain its maximum or minimum on \((0, 1)\). (iii) Give an example of a discontinuous function on the closed interval \([0, 1]\) that is unbounded. (iv) Give an example a discontinuous function on the closed interval \([0, 1]\) that is bounded but does not attain its maximum or minimum on \([0, 1]\).

4.3 Uniform continuity

Definition 4.15 A function \(f\) is uniformly continuous on a set \(E\) if for every \(\varepsilon > 0\) there exists \(\delta > 0\) so that \(|f(x) - f(y)| < \varepsilon\) for all \(x, y \in E\) such that \(|x - y| < \delta\).

Exercise 4.16 Show that the function \(f(x) = \sin(1/x)\) is continuous but not uniformly continuous on the open interval \((0, 1)\). Hint: consider the points \(x_k = 1/(2\pi k + \pi/2)\) and \(y_k = 1/(2\pi k - \pi/2)\) with \(k \in \mathbb{N}\).

Exercise 4.17 (i) Let a function \(f\) be continuous on the open interval \((a, b)\) and assume that \(f\) is bounded on \((a, b)\): there exists \(M\) so that \(|f(x)| \leq M\) for all \(x \in (a, b)\) and also that \(f\) takes each value \(y \in [-M, M]\) at most finitely many times. Show that then \(\lim_{x \to a} f(x)\) exists.

(ii) Let a function \(f\) be uniformly continuous on the open interval \((a, b)\). Show that then \(\lim_{x \to a} f(x)\) exists.

Proposition 4.18 A uniformly continuous function defined on a bounded set \(E\) is bounded.

Proof. Assume that \(f\) is unbounded on \((a, b)\). Then there exists a sequence of points \(x_k \in (a, b)\) such that

\[ |f(x_{k+1})| \geq |f(x_k)| + k. \]

It follows, in particular, that

\[ |f(x_m) - f(x_k)| > k \text{ for any } k \text{ and } m > k. \] (4.2)

Uniform continuity of \(f\) implies that there exists \(\delta_0 > 0\) so that \(|f(x) - f(y)| < 1\) for all \(x, y \in E\) such that \(|x - y| < \delta_0\). The sequence \(x_k\) is bounded, hence it has a convergent subsequence \(x_{n_k}\). Therefore, the sequence \(x_{n_k}\) is Cauchy. In particular, there exists \(N\) such that \(|x_{n_k} - x_{n_m}| < \delta_0\) for all \(k, m \geq N\). But then \(|f(x_{n_k}) - f(x_{n_m})| < 1\) for all \(k, m \geq N\), which contradicts (4.2). Thus, the function \(f\) has to be bounded on \(E\). □

Theorem 4.19 A function that is continuous on a closed interval \([a, b]\) is uniformly continuous on that interval.
Proof. Let $f$ be continuous on $[a, b]$. Then, given $\varepsilon > 0$ for each $x \in [a, b]$ there exists $\delta_x > 0$ such that $|f(x) - f(y)| < \varepsilon/10$ for all $y$ such that $|y - x| < \delta_x$. Let us consider the family of smaller open intervals $I_x = (x - \delta_x/2, x + \delta_x/2)$. As they cover the closed interval $[a, b]$, it follows from the Heine-Borel lemma that there exists a finite sub-collection $I_{x_1}, \ldots, I_{x_N}$ of such intervals that also covers $[a, b]$. Let us set $\delta = \min(\delta_{x_1}/2, \ldots, \delta_{x_N}/2)$. We will show that for any $x, y \in [a, b]$ such that $|x - y| < \delta$ we have $|f(x) - f(y)| < \varepsilon - \delta$ this will prove the uniform continuity of $f$. Given such $x$ and $y$, since the intervals $I_{x_1}, \ldots, I_{x_N}$ cover $[a, b]$, there exists $k \in \{1, \ldots, N\}$ such that $x \in I_{x_k}$, and $|x_k - x| < \delta_{x_k}/2$. Then we also have

$$|y - x_k| \leq |y - x| + |x - x_k| < \delta + \frac{\delta_{x_k}}{2} \leq \frac{\delta_{x_k}}{2} + \frac{\delta_{x_k}}{2} = \delta_{x_k}.$$ 

Hence, both $x$ and $y$ belong to the interval $(x_k - \delta_{x_k}, x_k + \delta_{x_k})$. It follows that

$$|f(x) - f(x_k)| < \frac{\varepsilon}{10}, \quad |f(y) - f(x_k)| < \frac{\varepsilon}{10}.$$ 

The triangle inequality now implies that

$$|f(x) - f(y)| \leq |f(x) - f(x_k)| + |f(y) - f(x_k)| < \frac{\varepsilon}{10} + \frac{\varepsilon}{10} < \varepsilon.$$ 

Thus, we have shown that for any $\varepsilon > 0$ we can find $\delta > 0$ so that for any $x, y \in [a, b]$ such that $|x - y| < \delta$ we have $|f(x) - f(y)| < \varepsilon$, which means that the function $f$ is uniformly continuous on $[a, b]$. $\square$

5 Open, closed and compact sets in $\mathbb{R}^n$

5.1 Open and closed sets

Let us recall that the distance between two points $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$ is

$$d(x, y) = [x_1 - y_1]^2 + \cdots + (x_n - y_n)^2]^{1/2},$$

also often denoted as

$$||x - y|| = [x_1 - y_1]^2 + \cdots + (x_n - y_n)^2]^{1/2}.$$ 

We will use the notation $B(x, r)$ for a ball centered at a point $x \in \mathbb{R}^n$ of radius $r$:

$$B(x, r) = \{y \in \mathbb{R}^n : d(x, y) < r\},$$

and the closed ball as

$$\overline{B}(x, r) = \{y \in \mathbb{R}^n : d(x, y) \leq r\}.$$ 

Definition 5.1 A set $U \subset \mathbb{R}^n$ is open if for every $x \in U$ there is an open ball $B(x, r)$ that is contained in $U$.

Definition 5.2 A point $x \in \mathbb{R}^n$ is a limit point of a set $G$ if for any open ball $B(x, r)$ with $r > 0$ there exists a point $y \in B(x, r)$ such that $y \in G$ and $y \neq x$.

Definition 5.3 A set $C \subset \mathbb{R}^n$ is closed if all limit points of $C$ are contained in $C$.

Definition 5.4 The complement of a set $S \subset \mathbb{R}^n$ is the set $S^c = \mathbb{R}^n \setminus S$. 

18
**Exercise 5.5** (i) Let $F$ be a collection of sets and let $U = \bigcup_{S \in F} S$. Show that $U^c = \bigcap_{S \in F} S^c$. 
(ii) Let $F$ be a collection of sets and let $G = \bigcap_{S \in F} S$. Show that $G^c = \bigcup_{S \in F} S^c$.

**Theorem 5.6** A set $C \subset \mathbb{R}^n$ is closed if and only if its complement $U = \mathbb{R}^n \setminus C$ is open.

**Proof.** Let $C$ be a closed set and $x \in U = \mathbb{R}^n \setminus C$. As $C$ is closed, we know that $x$ is not a limit point of $C$. Thus, there exists a ball $B(x, r)$ such that there is no point from $C$, except, possibly, $x$, in $B(x, r)$. As $x \in U = \mathbb{R}^n \setminus C$, we know that $x \notin C$, hence the ball $B(x, r)$ contains no points from $C$, hence $B(x, r) \subset U$. This shows that $U$ is an open set.

Next, assume a set $C$ is such that $U = \mathbb{R}^n \setminus C$ be an open set, and let $x$ be a limit point of $C$. We will show that $x \in C$. Indeed, if $x \notin C$, then $x \in U$ and, as $U$ is open, there exists a ball $B(x, r)$ that is contained in $U$. Hence, no points in $B(x, r)$ are in $C$. This contradicts the assumption that $x$ is a limit point of $C$. Thus, $C$ contains all its limit points and is closed. □

**Theorem 5.7** (i) The union of any collection of open sets is itself an open set.
(ii) The intersection of any collection of closed sets is a closed set.
(iii) The intersection $\bigcap_{i=1}^N U_i$ of a finite collection of open sets $U_1, \ldots, U_N$ is an open set.
(iv) The union $\bigcup_{i=1}^N F_i$ of a finite collection of closed sets $F_1, \ldots, F_N$ is a closed set.

**Proof.** Let us prove (i). Let $U_\alpha$, $\alpha \in A$ (here, $A$ is just some set of indices, finite or infinite) be a collection of open sets and let $x \in U = \bigcup_{\alpha \in A} U_\alpha$. Then there exists some $U_\alpha$ with $\alpha \in A$ such that $x \in U_\alpha$. As $U_\alpha$ is an open set, there exists $r > 0$ such that the ball $B(x, r)$ is contained in $U_\alpha$. But then $B(x, r) \subseteq U$, hence $U$ is an open set. Note that (i) is equivalent to (ii) because of Theorem 5.6 and Exercise 5.5, hence (ii) is also proven.

To prove (iii), let the sets $U_1, \ldots, U_N$ be open, and take $x \in U = \bigcap_{i=1}^N U_i$. As all $U_i$ are open sets, there exist $r_1, \ldots, r_N > 0$ such that $B(x, r_k) \in U_k$ for all $k \in \{1, \ldots, N\}$. Let $r = \min(r_1, \ldots, r_N)$, then $B(x, r) \in U$, hence $U$ is an open set. Note that (iii) is equivalent to (iv) because of Theorem 5.6 and Exercise 5.5, hence (iv) is also proven. □

Here is a bunch of related definitions.

**Definition 5.8** (i) A point $x$ is an interior point of a set $U \subset \mathbb{R}^n$ if there is a ball $B(x, r)$ that is contained in $U$.
(ii) A point $x$ is a boundary point of a set $U \subset \mathbb{R}^n$ if for any $r > 0$ the ball $B(x, r)$ contains both points in $U$ and not in $U$. The boundary of $U$ is denoted as $\partial U$.
(iii) The closure $\bar{U}$ of a set $U$ is the union of $U$ and the set of its limit points.

**Exercise 5.9** Show that $\bar{U}$ is the union of $U$ and the boundary $\partial U$.

**Exercise 5.10** Show that the closure $\bar{S}$ of any set $S$ is a closed set.

**Exercise 5.11** Show that a set $F$ is closed if and only $F = \overline{\overline{F}}$. 

19
5.2 Compact sets in $\mathbb{R}^n$

**Definition 5.12** A set $K \subset \mathbb{R}^n$ is compact if from every covering of $K$ by sets that are open in $\mathbb{R}^n$, one can extract a finite sub-covering of $K$.

The Heine-Borel lemma implies that a closed interval $[a, b]$ is a compact set in $\mathbb{R}$. Here is a generalization of this result to $\mathbb{R}^n$. A closed $n$-dimensional interval is a set of the form

$$\bar{I}_{ab} = \{x \in \mathbb{R}^n : a_i \leq x_i \leq b_i, \quad i = 1, \ldots, n\},$$

with some fixed numbers $a_i < b_i$. We may also define the open $n$-dimensional interval as

$$I'_{ab} = \{x \in \mathbb{R}^n : a_i < x_i < b_i, \quad i = 1, \ldots, n\},$$

**Exercise 5.13** A closed $n$-dimensional interval is a closed set in $\mathbb{R}^n$.

**Exercise 5.14** An open $n$-dimensional interval is an open set in $\mathbb{R}^n$.

**Proposition 5.15** A closed $n$-dimensional interval is a compact set in $\mathbb{R}^n$.

**Proof.** Let us fix $a_i < b_i$, $i = 1, \ldots, N$, and consider the corresponding closed $n$-dimensional interval $\bar{I}_{ab}$. Assume that there is a covering $\mathcal{G}$ of $\bar{I}_{ab}$ by open sets such that no finite sub-collection of $\mathcal{G}$ covers $\bar{I}_{ab}$. Let us bisect each interval $a_i \leq x_i \leq b_i$ in half – this partitions $\bar{I}_{ab}$ into $2^n$ closed $n$-dimensional sub-intervals. Note that least one of these $n$-dimensional sub-intervals cannot be covered by a finite sub-collection of $\mathcal{G}$. Let us call this $n$-dimensional sub-interval $\bar{I}^{(1)}$. Continuing this process, we obtain a nested sequence of closed $n$-dimensional intervals $\bar{I}^{(k)}$ such that $\bar{I}^{(k+1)}$ is contained in $\bar{I}^{(k)}$, and none of which admit a finite sub-covering. Each $\bar{I}^{(k)}$ has the form

$$\bar{I}^{(k)} = \{x \in \mathbb{R}^n : a_{i}^{(k)} \leq x_i \leq b_{i}^{(k)}, \quad i = 1, \ldots, n\},$$

with some $a_{i}^{(k)} < b_{i}^{(k)}$, $i \in \{1, \ldots, n\}$. By construction, for each $i \in \{1, \ldots, n\}$, the intervals $[a_{i}^{(k)}, b_{i}^{(k)}]$ form a nested collection of closed intervals whose length goes to zero as $k \to +\infty$. Hence, by the Nested Intervals Lemma, for each $i$ there exists a unique $c_i \in \mathbb{R}$ such that $c_i$ belongs to all intervals $[a_{i}^{(k)}, b_{i}^{(k)}]$, for all $k \geq 1$. Then, the point $c = (c_1, \ldots, c_n) \in \bar{I}_{ab}$ belongs to all $n$-dimensional intervals $\bar{I}^{(k)}$, for all $k \geq 1$. Since $c \in \bar{I}_{ab}$, there exists an open set $U \in \mathcal{G}$ so that $c \in U$. As $U$ is an open set, there exists $r > 0$ so that the ball $B(c, r)$ is contained in $U$. However, by construction, then there exists $N \in \mathbb{N}$ so that all $\bar{I}^{(n)}$ with $n \geq N$ are contained in $U$.

**Exercise 5.16** Prove this last assertion. Hint: first, show that $a_{i}^{(k)} \to c_i$ and $b_{i}^{(k)} \to c_i$ as $k \to +\infty$ for each $i \in \{1, \ldots, n\}$.

This contradicts the fact that none of $\bar{I}^{(k)}$ can be covered by finitely many sets from $\mathcal{G}$. Hence, a finite sub-cover of $\bar{I}_{ab}$ by sets from $\mathcal{G}$ has to exist, thus the set $\bar{I}_{ab}$ is compact. $\square$

**Proposition 5.17** Any compact set $K$ in $\mathbb{R}^n$ is closed.

**Proof.** Let $K$ be a compact set and $z$ be not in $K$. For each $x \in K$, we take $r_x = \|x - z\|/3$ and consider the open ball $B(x, r_x)$. Note that $z \not\in B(x, r_x)$ for any $x \in K$ and, moreover, the balls $B(x, r_x)$ and $B(z, r_x)$ do not intersect for any $x \in K$. On the other hand, the open balls $B(x, r_x)$ form a cover $\mathcal{G}$ of $K$ by open sets. As the set $K$ is compact, there exists a finite sub-cover of $K$ by the sets in $\mathcal{G}$. That is, there exist $x_1, \ldots, x_N \in K$ such that $K$ is contained in the union $\bigcup_{i=1}^{N} B(x_i, r_{x_i})$. 
Let $R = \min(r_1, \ldots, r_N)$, then, by construction, the ball $B(z, R)$ does not intersect any of the balls $B(x_k, r_{x_k})$, $k = 1, \ldots, N$. As these balls cover $K$, it follows that the ball $B(z, R)$ contains no points from $K$, hence $z$ is not a limit point of $K$. Thus, $K$ is a closed set. □

It is not true that all closed sets are compact in $\mathbb{R}^n$. For example, the whole space $\mathbb{R}^n$ is a closed set that is not compact, as is the closed half-space \( \{x = (x_1, \ldots, x_n) \in \mathbb{R}^n : x_1 \geq 0\} \), or the complement of the open unit ball \( \{x \in \mathbb{R}^n : |x| \geq 1\} \). The next proposition shows that a closed set that is contained in some compact set is itself compact.

**Proposition 5.18** Let a set $K \subset \mathbb{R}^n$ be compact and $K_1$ be a subset of $K$ that is a closed set. Then $K_1$ is a compact set.

**Proof.** Let $K_1 \subset \mathbb{R}^n$ be a closed set that is a subset of a compact set $K \subset \mathbb{R}^n$, and let $\mathcal{G}_1$ be a collection of open sets that covers $K_1$. Since $K_1$ is a closed set, the set $U_1 = K_1^c = \mathbb{R}^n \setminus K_1$ is open. Let us add this set to the collection $\mathcal{G}_1$ – the resulting collection $\mathcal{G}$ covers not just $K_1$ but also the set $K$. Since $K$ is compact, there is a finite sub-collection $O_1, \ldots, O_N$ of sets in $\mathcal{G}$ that covers $K$. If none of the sets $O_k$ is $U_1$, then this is a sub-collection not only of $\mathcal{G}$ but also of the original collection $\mathcal{G}_1$. Moreover, as this sub-collection covers $K_1$, it also covers $K$ – thus we have found a finite sub-collection of sets in $\mathcal{G}_1$ that covers $K_1$. On the other hand, if one of the sets $O_k$, with $k = 1, \ldots, N$, say, $O_N$ is $U_1$, then we can remove it and consider the collection $O_1, \ldots, O_{N-1}$. Note that, since the collection $O_1, \ldots, O_N$ covers $K$, it also covers $K_1$. Therefore, for any $x \in K_1$ we can find $k = 1, \ldots, N$ such that $x \in O_k$. However, since $U_1$ does not intersect $K_1$ and we assume that $O_N = U_1$, it is impossible that $k = N$ for any $x \in K_1$. Therefore, the collection $O_1, \ldots, O_{N-1}$ covers $K_1$. However, all sets $O_k$, $k = 1, \ldots, N-1$ are actually not only in $\mathcal{G}$ but also in the original collection $\mathcal{G}_1$. Thus, we have found a finite sub-collection of $\mathcal{G}_1$ that covers $K_1$. Since $\mathcal{G}_1$ is an arbitrary collection of open sets that covers $K_1$, it follows that $K_1$ is a compact set. □

**Definition 5.19** A set $S \subset \mathbb{R}^n$ is bounded if there exists $R > 0$ so that $S$ is contained in the ball $B(0, R)$.

In other words, a set $S$ is bounded if there exists $R > 0$ so that for any $x \in S$ we have $|x| < R$.

**Proposition 5.20** If a set $K \subset \mathbb{R}^n$ is compact then $K$ is bounded.

**Proof.** Let $K$ be a compact set, and consider the cover of $K$ by open balls of the form $B(x, 1)$, for all $x \in K$. As the set $K$ is compact, there exists a finite sub-cover of $K$ by such open balls, that is, a finite collection of open balls $B(x_1, 1), B(x_2, 1), \ldots, B(x_N, 1)$ with $x_k \in K$, that covers $K$. Let us set

$$R = 2 + \max_{1 \leq i \leq N} |x_i|.$$ 

As the balls $B(x_k, 1)$, $k = 1, \ldots, N$, cover $K$, for any $x \in K$ we can find $k$ such that $x \in B(x_k, 1)$. Then we have, by the triangle inequality,

$$|x| \leq |x - x_k| + |x_k| \leq 1 + \max_{1 \leq i \leq N} |x_i| < R,$$

hence $K$ is a bounded set. □

**Theorem 5.21** A set $K \subset \mathbb{R}^n$ is compact if and only if it is closed and bounded.

**Proof.** Propositions 5.17 and 5.20 show that if $K \subset \mathbb{R}^n$ is a compact set then it is closed and bounded.

Let us now assume that a set $K \subset \mathbb{R}^n$ is closed and bounded. Since $K$ is bounded, it is contained in some closed $n$-dimensional interval $I$. Proposition 5.15 shows that $I$ is a compact set. Thus, $K$ is a closed (by assumption) subset of a compact set. Now, Proposition 5.18 implies that $K$ is a compact set and we are done. □
6 Continuous mappings in \(\mathbb{R}^n\)

6.1 Convergence of sequences in \(\mathbb{R}^m\)

Let \(x^{(n)} \in \mathbb{R}^m\) be a sequence with values in \(\mathbb{R}^m\). In other words, \(x^{(n)} = (x_1^{(n)}, x_2^{(n)}, \ldots, x_m^{(n)})\) is an \(m\)-tuple of real-valued sequences \(x_1^{(n)}, x_2^{(n)}, \ldots, x_m^{(n)}\).

**Definition 6.1** Let \(\ell \in \mathbb{R}^m\). We say that \(\lim_{n \to \infty} x^{(n)} = \ell\) if for every \(\varepsilon > 0\) there exists \(N\) so that for all \(n \geq N\) we have \(\|x^{(n)} - \ell\| < \varepsilon\).

Note that this definition is equivalent to the following: \(\lim_{n \to \infty} x^{(n)} = \ell\) if \(\lim_{n \to \infty} \|x^n - \ell\| = 0\), with the limit understood simply in the sense of convergence of the sequence \(\|x_n - \ell\|\) of real numbers to zero.

**Exercise 6.2** Show that a sequence \(\{x^{(1)}, x^{(2)}, \ldots, x^{(n)}, \ldots\} \in \mathbb{R}^m\) converges to \(\ell = (\ell_1, \ldots, \ell_m) \in \mathbb{R}^m\) if and only if

\[
\lim_{n \to \infty} x_k^{(n)} = \ell_k, \text{ for all } 1 \leq k \leq m.
\]

In other words, convergence of sequences with values in \(\mathbb{R}^m\) is equivalent to the convergence of all \(m\) components of the sequence. The key to this exercise is the following pair of inequalities that holds for all \(x, y \in \mathbb{R}^m\):

\[
\begin{align*}
|x_i - y_i| &\leq \|x - y\| \text{ for all } 1 \leq i \leq m, \\
\|x - y\| &\leq \sqrt{m} \max_{1 \leq i \leq m} |x_i - y_i|.
\end{align*}
\]

(6.1)

The first inequality in (6.1) is obvious from the definition of \(\|x - y\|\) and the second also follows immediately from this definition:

\[
\|x - y\|^2 = (x_1 - y_1)^2 + \cdots + (x_m - y_m)^2 \leq m \max_{1 \leq i \leq n} |x_i - y_i|^2.
\]

Taking the square root in both sides gives the second inequality in (6.1).

**Definition 6.3** Let \(x^{(n)}\) be a sequence in \(\mathbb{R}^m\). We say that \(x^{(n)}\) is Cauchy if for every \(\varepsilon > 0\) there exists \(N \in \mathbb{N}\) so that for all \(j, k \geq N\) we have \(\|x^{(j)} - x^{(k)}\| < \varepsilon\).

**Proposition 6.4** A sequence \(\{x_n\} \in \mathbb{R}^m\) is Cauchy if and only if it is convergent.

**Exercise 6.5** Prove this proposition, using inequalities (6.1) and the fact that sequences of real numbers are Cauchy if and only if they are convergent.

6.2 Continuity of maps from \(\mathbb{R}^n\) to \(\mathbb{R}^m\)

We will now consider maps \(f : \mathbb{R}^n \to \mathbb{R}^m\) and generalize the notions of continuity and convergence that we have defined for real-valued functions on \(\mathbb{R}^n\) and sequences of real numbers. Such map is simply a collection of \(m\) real-valued functions \(f_1(x), \ldots, f_m(x)\) that each maps \(\mathbb{R}^n\) to \(\mathbb{R}\). For example,

\[
f(x_1, x_2) = (x_1^2 + x_2^2, x_2)
\]
is a map from $\mathbb{R}^2$ to $\mathbb{R}^2$ with $f_1(x_1, x_2) = x_1^2 + x_2^2$ and $f_2(x_1, x_2) = x_2$, while
\[
g(x_1, x_2, x_3) = (x_1 + x_2, \cos(x_1 + x_3), e^{x_1 + x_2 + x_3}, \sin(x_3))
\]
is a map from $\mathbb{R}^3$ to $\mathbb{R}^4$ with
\[
g_1(x_1, x_2, x_3) = x_1 + x_2, \quad g_2(x_1, x_2, x_3) = \cos(x_1 + x_3), \quad g_3(x_1, x_2, x_3) = e^{x_1 + x_2 + x_3}, \quad g_4(x_1, x_2, x_3) = \sin(x_3).
\]
We define the limit of a mapping $f : \mathbb{R}^n \to \mathbb{R}^m$ similarly to what we did for real-valued functions.

**Definition 6.6** Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a mapping, $a \in \mathbb{R}^n$, and $\ell \in \mathbb{R}^m$. We say that $\lim_{x \to a} f(x) = \ell$ if for every $\varepsilon > 0$ there exists $\delta > 0$ so that for all $x \in B(a, \delta)$ such that $x \neq a$ we have $f(x) \in B(\ell, \varepsilon)$.

Note that we use a slightly different notation than in $\mathbb{R}$: instead of writing $\|x - a\| < \delta$ we write $x \in B(a, \delta)$, and instead of saying $\|f(x) - \ell\| < \delta$ we say $f(x) \in B(\ell, \varepsilon)$ but the meaning is literally the same. The reason for this choice is to emphasize the geometry of what is going on.

**Definition 6.7** We say that a mapping $f : U \to \mathbb{R}^m$ is continuous at $x \in U \subseteq \mathbb{R}^n$ if
\[
\lim_{x' \to x} f(x') = f(x).
\]

This definition is, of course, exactly the same as in $\mathbb{R}$. The same proofs as in $\mathbb{R}$ show that if a map $f : U \to \mathbb{R}^m$ is continuous at $x \in U$ and $\lambda \in \mathbb{R}$ then $\lambda f$ is continuous at $x$, and that if two maps $f : U \to \mathbb{R}^m$ and $g : U \to \mathbb{R}^m$ are both continuous at $x \in U$ then so is $f + g$. The same is true for composition of continuous maps.

**Proposition 6.8** Let $f : E \to \mathbb{R}^m$ be a map defined on an open set $E \subseteq \mathbb{R}^n$, and $g : U \to \mathbb{R}^k$ be a map defined on an open set $U \subseteq \mathbb{R}^m$. Assume that $f$ is continuous at $x \in E$, that $y = f(x)$ is in $U$, and that $g$ is continuous at $y = f(x)$. Then the composition $g \circ f$ is defined in a ball around $x$ and is continuous at $x$.

**Proof.** Given $\varepsilon > 0$, since $g$ is continuous at $y$ we can find $r > 0$ so that $\|g(y') - g(y)\| < \varepsilon$ for all $y' \in B(y, r)$. Moreover, as $f$ is continuous at $x$, we can find $\delta > 0$ so that $\|f(x') - y\| < r$ for all $x' \in B(x, \delta)$. It follows that $\|g(f(x')) - g(f(x))\| < \varepsilon$ for all $x' \in B(x, r)$ and we are done. \(\square\)

### 6.3 Successive limits vs. the limit

One should not think that the limit of a function of several variables can be computed by taking successively the limits in each variable. The standard example when this fails is
\[
f(x_1, x_2) = \begin{cases} 
x_1x_2 & \text{if } (x_1, x_2) \neq (0, 0), \\
x_1^2 + x_2^2 & \text{if } (x_1, x_2) = (0, 0). 
\end{cases}
\]
Then we have
\[
\lim_{x_2 \to 0} f(x_1, x_2) = 0 \text{ for any } x_1 \neq 0,
\]
and
\[
\lim_{x_1 \to 0} f(x_1, x_2) = 0 \text{ for any } x_2 \neq 0,
\]
so that
\[
\lim_{x_2 \to 0} \left( \lim_{x_1 \to 0} f(x_1, x_2) \right) = 0,
\]
\[
\lim_{x_1 \to 0} \left( \lim_{x_2 \to 0} f(x_1, x_2) \right) = 0,
\]
23
but \( f(x, x) = 1/2 \), so that the limit \( \lim_{(x_1, x_2) \to (0, 0)} f(x_1, x_2) \) does not exist.

Here are some other variations on this theme. Consider

\[
f(x_1, x_2) = \begin{cases} 
  \frac{x_1^2 - x_2^2}{x_1^2 + x_2^2}, & \text{if } (x_1, x_2) \neq (0, 0), \\
  0, & \text{if } (x_1, x_2) \neq (0, 0).
\end{cases}
\]

Then each of the successive limits exists but they are different:

\[
\lim_{x_2 \to 0} f(x_1, x_2) = \frac{x_1^2}{x_2^2} = 1 \text{ for any } x_1 \neq 0,
\]
and

\[
\lim_{x_1 \to 0} f(x_1, x_2) = -\frac{x_2^2}{x_2^2} = -1 \text{ for any } x_2 \neq 0,
\]

so that

\[
\lim_{x_2 \to 0} \left( \lim_{x_1 \to 0} f(x_1, x_2) \right) = -1,
\]
and

\[
\lim_{x_1 \to 0} \left( \lim_{x_2 \to 0} f(x_1, x_2) \right) = 1.
\]

It is easy to see that \( \lim_{(x_1, x_2) \to (0, 0)} f(x_1, x_2) \) does not exist.

The next interesting example is

\[
f(x_1, x_2) = \begin{cases} 
  \frac{x_1^2 x_2}{x_1^2 + x_2^2}, & \text{if } (x_1, x_2) \neq (0, 0), \\
  0, & \text{if } (x_1, x_2) \neq (0, 0).
\end{cases}
\]

Then the limit of \( f(x_1, x_2) \) is zero along any ray \( x_1 = \lambda t, x_2 = \mu t \), in the sense that for any \( \lambda, \mu \in \mathbb{R} \) fixed, with at least one of \( \lambda, \mu \) not equal to zero, we have

\[
\lim_{t \to 0} f(\lambda t, \mu t) = \lim_{t \to 0} \frac{\lambda^2 \mu t^3}{\lambda^4 t^4 + \mu^2 t^2} = 0.
\]

However, if we take \( x_1 = t, x_2 = t^2 \), then \( f(t, t^2) = 1/2 \), hence, again, \( \lim_{(x_1, x_2) \to (0, 0)} f(x_1, x_2) \) does not exist.

Finally, consider the function

\[
f(x_1, x_2) = \begin{cases} 
  x_1 + x_2 \sin \left( \frac{1}{x_1} \right), & \text{if } x_1 \neq 0, \\
  0, & \text{if } x_1 = 0.
\end{cases}
\]

Then we have \( |f(x_1, x_2)| \leq |x_1| + |x_2| \), which immediately implies that the limit \( \lim_{(x_1, x_2) \to (0, 0)} f(x_1, x_2) \) exists and equals to zero. However, the successive limit

\[
\lim_{x_2 \to 0} \left( \lim_{x_1 \to 0} f(x_1, x_2) \right)
\]

does not exist at all simply because the limit

\[
\lim_{x_1 \to 0} f(x_1, x_2)
\]

does not exist for any \( x_2 \neq 0 \).
6.4 Continuous functions on compact sets

**Definition 6.9** We say that a function \( f : E \rightarrow \mathbb{R}^m \) is uniformly continuous on a set \( E \subset \mathbb{R}^n \) if for every \( \varepsilon > 0 \) there exists \( \delta > 0 \) so that \( d(f(x_1), f(x_2)) < \varepsilon \) for all \( x_1, x_2 \in E \) such that \( d(x_1, x_2) < \delta \).

**Proposition 6.10** If a mapping \( f : K \rightarrow \mathbb{R}^m \) is continuous on a compact set \( K \subset \mathbb{R}^n \) then it is uniformly continuous on \( K \).

**Proposition 6.11** If a mapping \( f : K \rightarrow \mathbb{R}^m \) is continuous on a compact set \( K \subset \mathbb{R}^n \) then it is bounded on \( K \).

**Proposition 6.12** If a mapping \( f : K \rightarrow \mathbb{R}^m \) is continuous on a compact set \( K \subset \mathbb{R}^n \) then it attains its maximum and minimum on \( K \).

The proofs of all of the above propositions are exactly the same as for continuous real-valued functions on \( \mathbb{R} \) defined on a closed interval: these proofs used nothing but the Heine-Borel lemma for closed intervals, and that is exactly the definition of a compact set. Check this!

6.5 Connected sets

In order to generalize the intermediate value theorem, we will need the notion of a connected set.

**Definition 6.13** A continuous path is a continuous function \( \Gamma : [\alpha, \beta] \rightarrow \mathbb{R}^m \). We say that a set \( E \subset \mathbb{R}^m \) is pathwise connected if for every \( x, y \in E \) there exists a continuous path \( \Gamma : [a, b] \rightarrow \mathbb{R}^m \) such that \( \Gamma(\alpha) = x \), \( \Gamma(\beta) = y \) and \( \Gamma(t) \in E \) for all \( \alpha \leq t \leq \beta \). A domain \( U \subset \mathbb{R}^m \) is a pathwise connected set that is open.

Note that by re-defining \( \Gamma(t) \) as \( \tilde{\Gamma}(t) = \Gamma((1-t)a + tb) \) for \( 0 \leq t \leq 1 \), we can always assume that \( \alpha = 0 \) and \( \beta = 1 \) in the definition of the path.

An open ball \( B(a, r) \) and a closed ball \( \bar{B}(a, r) \) are both pathwise connected sets in \( \mathbb{R}^m \), for any \( a \in \mathbb{R}^m \). Indeed, given any two points \( x, y \in B(a, r) \) we can define the path \( s(t) = (1-t)x + ty \), with \( 0 \leq t \leq 1 \). Then \( s(0) = x \), \( s(1) = y \) and for each \( t \in [0, 1] \) we have

\[
\|s(t) - a\|^2 = (s_1(t) - a_1)^2 + \cdots + (s_m(t) - a_m)^2 \\
= ((1-t)x_1 + ty_1 - a_1)^2 + \cdots + ((1-t)x_m + ty_m - a_m)^2 \\
= ((1-t)(x_1 - a_1) + t(y_1 - a_1))^2 + \cdots + ((1-t)(x_m - a_m) + t(y_m - a_m))^2.
\]

We now use the inequality

\[
((1-t)a + tb)^2 \leq (1-t)a^2 + tb^2,
\]

that holds for all \( a, b \in \mathbb{R} \), and \( 0 \leq t \leq 1 \), to get

\[
\|s(t) - a\|^2 \leq (1-t)(x_1 - a_1)^2 + t(y_1 - a_1)^2 + \cdots + (1-t)(x_m - a_m)^2 + t(y_m - a_m)^2 \\
= (1-t)\|x - a\|^2 + t\|y - a\|^2 \leq (1-t)r^2 + tr^2 = r^2.
\]

It follows that \( s(t) \in B(a, r) \) for all \( t \), which proves that \( B(a, r) \) is a connected set. The proof for the closed ball \( \bar{B}(a, r) \) is identical.

**Exercise 6.14** Prove the inequality (6.2). Hint: use the fact that \( y = x^2 \) is a convex function for \( x \in \mathbb{R} \).
The following version of the intermediate value theorem holds for continuous functions defined on pathwise connected sets.

**Proposition 6.15** Let \( E \subset \mathbb{R}^n \) be a connected set, and \( f : E \to \mathbb{R} \) be a function that is continuous on \( E \). Assume that there exists \( x \in E \) and \( y \in E \) such that \( f(x) = A \) and \( f(x) = B \). Then, for any \( C \in \mathbb{R} \) that is between \( A \) and \( B \) there exists \( z \in E \) such that \( f(z) = C \).

**Proof.** Since \( E \) is a pathwise connected set and \( x, y \in E \), there exists a path \( \Gamma : [0, 1] \to E \) such that \( \Gamma(0) = x \) and \( \Gamma(1) = y \). The function \( g(t) = f(\Gamma(t)) \), defined for \( 0 \leq t \leq 1 \), is continuous, as a composition of two continuous functions, \( g : [0, 1] \to E \) and \( f : E \to \mathbb{R} \). In addition, we have \( g(0) = f(x) = A \) and \( g(1) = f(y) = B \). The intermediate value theorem for real valued functions on \([0, 1]\) then implies that there exists \( t_0 \in [0, 1] \) such that \( g(t_0) = C \). This means that \( f(z) = C \), with \( z = \Gamma(t_0) \). \( \square \)

7 Basic properties of metric spaces

We now take a detour to discuss some very basic properties of metric spaces, generalizing what we know about \( \mathbb{R} \) and \( \mathbb{R}^n \).

**7.1 Convergence in metric spaces**

If you go back to what we have really used about \( \|x\| \in \mathbb{R}^n \), we will discover that what we really needed in most situations was that the function \( d(x, y) = \|x - y\| \) was non-negative, equal to 0 only if \( x = y \), symmetric, so that \( d(x, y) = d(y, x) \), and satisfied the triangle inequality. We thus make the following definition.

**Definition 7.1** A metric space \((X, d)\) is a set \( X \) together with a map \( d : X \times X \to \mathbb{R} \) (called a distance function) such that

1. \( d(x, y) \geq 0 \) for all \( x, y \in X \), and \( d(x, y) = 0 \) if and only if \( x = y \).

2. \( d(x, y) = d(y, x) \) for all \( x, y \in X \),

3. (Triangle inequality) \( d(x, z) \leq d(x, y) + d(y, z) \) for all \( x, y, z \in X \).

Recall that a norm \( \|\cdot\| : V \to \mathbb{R} \) on a vector space \( V \) over \( \mathbb{R} \) (or \( \mathbb{C} \)) is a map that is

1. positive definite, i.e. \( \|x\| \geq 0 \) for all \( x \in V \) with equality if and only if \( x = 0 \),

2. absolutely homogeneous, i.e. \( \|\lambda x\| = |\lambda| \|x\| \) for \( \lambda \in \mathbb{R} \) (or \( \mathbb{C} \)) and \( x \in V \),

3. satisfies the triangle inequality, i.e. \( \|x + y\| \leq \|x\| + \|y\| \) for all \( x, y \in V \).

Then one easily checks that every normed vector space is a metric space with the induced metric \( d(x, y) = \|x - y\| \); for instance the triangle inequality for the metric follows from

\[
    d(x, z) = \|x - z\| = \|(x - y) + (y - z)\| \leq \|x - y\| + \|y - z\| = d(x, y) + d(y, z),
\]

where the inequality in the middle is the triangle inequality for norms.

There are many interesting metric spaces that are not normed vector spaces, for the simple reason that the distance function does not require that \( X \) is a vector space. For instance, for any set \( X \) we may define a metric on it setting \( d(x, y) = 0 \) if \( x = y \), \( d(x, y) = 1 \) if \( x \neq y \).
Another example of a metric space we will use often is the space $C(K)$ of continuous real-valued functions $f : K \rightarrow \mathbb{R}$, where $K$ is a compact subset of $\mathbb{R}^n$. The distance is defined as

$$d(f, g) = \sup_{x \in K} |f(x) - g(x)|.$$  

We will often simply look at $K = [a, b]$, a closed interval on $\mathbb{R}$.

Convergence in metric spaces is defined exactly as in $\mathbb{R}^n$, except we use the metric $d(x, y)$ rather than $\|x - y\|$.

**Definition 7.2** A sequence $x_n$ of points in a metric space $(X, d)$ converges to $x \in X$, denoted as $\lim x_n = x$, if for all $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $n \geq N$ we have $d(x_n, x) < \varepsilon$.

**Exercise 7.3** Show that the limit of a sequence $x_n$ is unique (if it exists).

**Exercise 7.4** Show that if $X$ is a normed space and $x_n \rightarrow x$ in $X$, then $\|x_n\| \rightarrow \|x\|$. Hint: show that $\|x_n\| - \|x\| \leq \|x - x_n\|$.

### 7.2 Continuous mappings between metric spaces

We now turn to continuity of functions $f : X \rightarrow Y$ where $(X, d_X)$, $(Y, d_Y)$ are metric spaces, generalizing what we have done for maps from $\mathbb{R}^n$ to $\mathbb{R}^m$.

**Definition 7.5** Suppose $(X, d_X)$, $(Y, d_Y)$ are metric spaces. A function $f : X \rightarrow Y$ is continuous at a point $a \in X$ if for all $\varepsilon > 0$ there exists $\delta > 0$ such for all $x \in X$ such that $d_X(x, a) < \delta$ we have $d_Y(f(x), f(a)) < \varepsilon$. A function is called continuous on $X$ if it is continuous at all $a \in X$.

As you recall, for functions on $\mathbb{R}$ we have shown that $f$ is continuous at $a \in \mathbb{R}$ if for any sequence $x_n \rightarrow a$ we have $f(x_n) \rightarrow f(a)$. In general metric spaces, we have the same property.

**Definition 7.6** Suppose $(X, d_X)$, $(Y, d_Y)$ are metric spaces. A function $f : X \rightarrow Y$ is sequentially continuous at the point $a \in X$ if for every sequence $x_n$ in $X$ which converges to $a$, we have $\lim_{n \rightarrow \infty} f(x_n) = f(a)$.

We then have

**Lemma 7.7** Suppose $(X, d_X)$, $(Y, d_Y)$ are metric spaces. A function $f : X \rightarrow Y$ is continuous at $a \in X$ if and only if it is sequentially continuous at $a \in X$.

**Proof.** The proof is verbatim as in $\mathbb{R}$, check this! □

Let us give some examples. Let $X = C[0, 1]$ and consider $F : X \rightarrow \mathbb{R}$ defined as $F(f) = f(3/4)$. A sequence of functions $f_n \in C[0, 1]$ converges to $f \in C[0, 1]$ if

$$\sup_{x \in [0, 1]} |f_n(x) - f(x)| \rightarrow 0 \text{ as } n \rightarrow +\infty.$$  

In other words, for any $\varepsilon > 0$ there exists $N$ so that

$$\sup_{x \in [0, 1]} |f_n(x) - f(x)| < \varepsilon \text{ for all } n \geq N.$$  

Rephrasing this again, this is equivalent to saying that for any $\varepsilon > 0$ there exists $N$ so that

$$|f_n(x) - f(x)| < \varepsilon \text{ for all } n \geq N \text{ and all } x \in [0, 1].$$
It follows, in particular, that for any \( \varepsilon > 0 \) there exists \( N \) so that
\[
|f_n(3/4) - f(3/4)| < \varepsilon \text{ for all } n \geq N,
\]
which is nothing but
\[
|F(f_n) - F(f)| < \varepsilon \text{ for all } n \geq N,
\]
which means that \( F(f_n) \) converges to \( F(f) \), thus \( F \) is continuous at all \( f \in C[0,1] \).

The Bolzano-Weierstrass theorem on \( \mathbb{R} \) stated that bounded sequences had convergent subsequences. In particular, if \( x_n \) is a sequence of points in a closed interval \( [a,b] \subset \mathbb{R} \), then it has a subsequence \( x_{n_k} \) converging to some \( x \in [a,b] \). This motivates the following generalization.

**Definition 7.8** A metric space \((X,d_X)\) is sequentially compact if every sequence \( x_n \) of points in \((X,d_X)\) has a convergent subsequence that converges to a point in \( X \).

Here the word ‘sequentially’ refers to the fact that there is in fact an equivalent (in the setting of metric spaces) definition, in terms of the Heine-Borel property, which is usually called compactness. However, sequential compactness, which is, again, equivalent in the setting of metric spaces, is equally convenient.

**Definition 7.9** A subset \( K \) of a metric space \((X,d_X)\) is sequentially compact if every sequence \( x_n \) of points in \( K \) has a convergent subsequence that converges to a point in \( K \).

Recall that in \( \mathbb{R}^n \) a set is compact if and only if it is closed and bounded. In general metric spaces only one implication is true.

**Proposition 7.10** Let \( K \) be a sequentially compact subset of a metric space \((X,d)\), then \( K \) is closed and bounded.

**Proof.** First, let us show that \( K \) is closed. Consider a limit point \( y \in X \) of the set \( K \). Then there is a sequence \( x_n \in K \) such that \( x_n \to y \). As \( K \) is sequentially compact, there is a subsequence \( x_{n_k} \) that converges to a point in \( K \). Since it is a subsequence of a convergent sequence, it must converge to the same limit, hence \( y \in K \) which shows that \( K \) is closed. To show that \( K \) is bounded, assume it is not. Then one can find a sequence \( x_k \in K \) such that \( d(x_n,0) > n \). As \( K \) is compact, \( x_n \) should have a convergent subsequence \( x_{n_k} \to y \in K \). However, we have, by the triangle inequality
\[
d(x_{n_k},0) - d(y,x_{n_k}) \leq d(y,0) \leq d(y,x_{n_k}) + d(x_{n_k},0).
\]
Hence, there exists \( N \) so that for all \( k \geq N \) we have
\[
d(y,0) \geq d(x_{n_k},0) - 1 \geq n_k - 1,
\]
which is a contradiction. Thus, \( K \) must be bounded. \( \square \)

The next proposition shows that the other direction fails in general.

**Proposition 7.11** The closed unit ball \( \bar{B}(0,1) \subset C[0,1] \) is a closed and bounded set that is not sequentially compact.

**Proof.** The closed unit ball in \( C[0,1] \) is the set
\[
\bar{B}(0,1) = \{ f \in C[0,1] : \sup_{x \in [0,1]} |f(x)| \leq 1 \},
\]
that is, the set of all $f \in C[0,1]$ such that $-1 \leq f(x) \leq 1$ for all $x \in [0,1]$. Consider the following sequence of functions

$$f_n(x) = \begin{cases} 
1, & 0 \leq x \leq 1/2 - 1/n, \\
2n(1/2 - x + 1/(2n)), & 1/2 - 1/n \leq x \leq 1/2 - 1/(2n), \\
0, & 1/2 - 1/(2n) \leq x \leq 1.
\end{cases}$$

Then all $f_n$ lie in $B(0,1) \subset C[0,1]$ but the sequence $f_n$ is not Cauchy. This is because for $m > 2n + 1$ we have $f_m(1/2 - 1/(2n)) = 1$ but $f_n(1/2 - 1/(2n)) = 0$, so that $\|f_n - f_m\| = 1$. Therefore, the sequence $f_n$ cannot have a convergent subsequence, and $B(0,1)$ is not a compact set.

**Theorem 7.12** Suppose $(X,d)$ is a sequentially compact metric space, and $f : X \to \mathbb{R}$ is continuous. Then $f$ is bounded, and it attains its maximum and minimum, i.e. there exist points $a, b \in X$ such that $f(a) = \sup\{f(x) : x \in X\}$, $f(b) = \inf\{f(x) : x \in X\}$.

**Proof.** The proof is verbatim as in $\mathbb{R}^n$ except in $\mathbb{R}^n$ we relied on the Heine-Borel definition of compactness, so let show how this is done using the sequential compactness. Let us show first that $f(x)$ is uniformly bounded from above: there exists $M > 0$ so that $f(x) \leq M$ for all $x \in X$. Indeed, suppose it is not, i.e. there is no upper bound for the sequence of functions $f_n$. Since $f_n(1/2 - 1/(2n)) = 1$ for all $n$, there exists $x_n \in X$ such that $f(x_n) > n$. By the sequential compactness of $X$, the sequence $x_n$ has a convergent subsequence, $x_{n_k}$. Let us say $x = \lim_{k \to \infty} x_{n_k} \in X$. Due to its continuity, $f$ is sequentially continuous at $x$, so

$$\lim_{k \to \infty} f(x_{n_k}) = f(x).$$

In particular, applying the definition of convergence with $\varepsilon = 1$, there exists $N$ such that $k \geq N$ implies $|f(x_{n_k}) - f(x)| < 1$. But then

$$|f(x_{n_k})| = |(f(x_{n_k}) - f(x)) + f(x)| \leq |f(x_{n_k}) - f(x)| + |f(x)| < |f(x)| + 1$$

for all $k \geq N$. Since $f(x_{n_k}) > n_k \geq k$ by the very choice of $x_{n_k}$ (note $n_k \geq k$ is true for any subsequence), this is a contradiction: choose any $k > \max(N, |f(x)| + 1)$, and then the two inequalities are contradictory. This completes the boundedness from above claim; a completely analogous argument shows boundedness from below.

Let us show that the maximum is actually attained, the case of the minimum is identical. The proof that we used in $\mathbb{R}$ works verbatim but let us give a slightly different proof. Now that we know that $f(x)$ is bounded from above, let $M = \sup_{x \in X} f(x)$. Then for all $n \in \mathbb{N}$, $M - \frac{1}{n}$ is not an upper bound for $f(x)$, so there exists $x_n \in X$ such that $f(x_n) > M - 1/n$. The sequence $x_n$ has a convergent subsequence, by the sequential compactness of $X$, and we let

$$a = \lim_{k \to \infty} x_{n_k} \in X.$$ 

We claim that $f(a) = M$. Indeed, due to its continuity, $f$ is sequentially continuous at $a$, so

$$\lim_{k \to \infty} f(x_{n_k}) = f(a).$$

On the other hand, since

$$M \geq f(x_{n_k}) > M - \frac{1}{n_k} \geq M - \frac{1}{k},$$

by the very choice of the $x_n$, we have by the sandwich theorem that

$$\lim_{k \to \infty} f(x_{n_k}) = M.$$ 

In combination with the just observed sequential continuity, this gives $f(a) = M$, as desired. □

A very similar proof gives
**Theorem 7.13** Suppose $(X,d_X)$, $(Y,d_Y)$ are metric spaces, $X$ is sequentially compact, and $f : X \rightarrow Y$ is continuous. Then $(f(X),d_Y)$ is sequentially compact.

**Proof.** Suppose that $y_n$ is a sequence in $f(X)$, that is, $y_n = f(x_n)$ for some $x_n$. We need to find a subsequence of $y_n$ which converges to a point in $f(X)$ in the metric $d_Y$.

Since $X$ is compact, $x_n$ has a convergent subsequence $x_{n_k}$. We let $x = \lim_{k \to \infty} x_{n_k} \in X$.

We claim that $y_{n_k}$ converges to $y = f(x) \in f(X)$ in the metric $d_Y$. Indeed, by the continuity of $f$, we know that $f$ is sequentially continuous at $x$, so

$$\lim_{k \to \infty} f(x_{n_k}) = f(x).$$

As $f(x_{n_k}) = y_{n_k}$, this proves the claim, and thus the theorem. □

8 Basic facts about series

8.1 Convergence of a series

Given real numbers $a_n$, we will use the following definition to make sense of an infinite sum

$$a_1 + a_2 + \cdots + a_n + \ldots,$$

that we will denote as

$$\sum_{k=1}^{\infty} a_k,$$

or sometimes simply as $\sum_k a_k$. We call $a_j$ the individual terms of the series. Note that there is no meaning yet assigned to the sum above, at the moment it is just a notation. The $n$-th partial sum of the series $\sum_{k=1}^{\infty} a_k$ is

$$S_n = a_1 + \cdots + a_n = \sum_{k=1}^{n} a_k.$$

**Definition 8.1** We say that a series $\sum_{k=1}^{\infty} a_k$ converges if the sequence $S_n$ of its partial sums converges, and call the limit of the sequence $S_n$ the sum of the series $\sum_{k=1}^{\infty} a_k$.

An immediate consequence of the definition is the Cauchy criterion for convergence that simply restates the Cauchy criterion for the sequence $S_n$ of the partial sums, using the fact that for all $m > n$ we have

$$S_m - S_n = \sum_{k=m+1}^{n} S_k.$$

**Theorem 8.2** (The Cauchy criterion for convergence) A series $\sum_{k=1}^{\infty} a_k$ is convergent if and only if for every $\varepsilon > 0$ there exists $N$ so that for all $m > n \geq N$ we have

$$|a_{n+1} + \cdots + a_m| < \varepsilon.$$

Here are two important immediate consequence of the Cauchy criterion for convergence of a series.

**Corollary 8.3** If only finitely many terms of a series are changed then the new series is convergent if and only if the original series is convergent.
The proof is a simple exercise: be careful about how you choose \( N \) given \( \varepsilon > 0 \) in the definition of the Cauchy property.

**Corollary 8.4** If a series \( \sum_{k=1}^{\infty} a_n \) converges then \( \lim_{n \to \infty} a_n = 0 \).

The proof is, again, an exercise in the application of the Cauchy property: take \( m = n + 1 \) and look at what \( S_m - S_n \) is.

The geometric series \( \sum_{k=1}^{\infty} r^k \) plays an incredibly important role in analysis, and especially in complex analysis that is outside the scope of this class.

**Exercise 8.5** (i) Show that the series \( \sum_{k=0}^{\infty} r^k \) converges if \( |r| < 1 \) and diverges if \( |r| > 1 \). Hint: recall that 

\[
\sum_{k=0}^{n} r^k = \frac{1 - r^{k+1}}{1 - r}.
\]

(ii) Show that the series \( \sum_{k=0}^{\infty} r^k \) diverges both when \( r = -1 \) and \( r = 1 \). Now, consider the function 

\[
S(r) = \sum_{k=1}^{\infty} r^k
\]
on the open interval \( r \in (-1, 1) \). Show that \( \lim_{r \to 1} S(r) \) does not exist but \( \lim_{r \to -1} S(r) \) exists. Why does it not contradict the divergence of the series \( \sum_{k=1}^{\infty} (-1)^k \)?

(iii) Show that the series \( \sum_{k=1}^{\infty} \frac{1}{k} \) diverges. Hint: note that we have 

\[
\frac{1}{n+1} + \cdots + \frac{1}{2n} \geq \frac{1}{2n} + \cdots + \frac{1}{2n} = \frac{n}{2n} = \frac{1}{2},
\]

and use the Cauchy criterion.

### 8.2 Absolute convergence of a series

**Definition 8.6** We say that a series \( \sum_{k=1}^{\infty} a_k \) converges absolutely if the series \( \sum_{k=1}^{\infty} |a_k| \) converges.

Note that since 

\[
|a_n + \cdots + a_m| \leq |a_n| + \cdots + |a_m|,
\]

the Cauchy criterion implies that if a series converges absolutely, then it converges. The converse is not true. As an example, consider the series 

\[
1 - 1 + \frac{1}{2} - \frac{1}{2} + \frac{1}{3} - \frac{1}{3} + \cdots
\]

that has partial sums equal either to zero or \( 1/n \). Thus, it converges. However, part (iii) of Exercise 8.5 shows that it does not converge absolutely.
Theorem 8.7 Let \( a_n \geq 0 \), then the series \( \sum_{k=1}^{\infty} a_n \) converges if and only if the sequence of its partial sums \( S_n = \sum_{k=1}^{n} a_n \) is bounded.

**Proof.** The sequence \( S_n \) is increasing because \( a_n \geq 0 \). Thus, \( S_n \) converges if and only if it is bounded. □

Theorem 8.8 (Comparison theorem) Assume that \( 0 \leq a_n \leq b_n \), then (i) if the series \( \sum_{k=1}^{\infty} b_n \) converges then so does the series \( \sum_{k=1}^{\infty} a_n \), and (ii) if the series \( \sum_{k=1}^{\infty} a_n \) diverges then so does the series \( \sum_{k=1}^{\infty} b_n \).

Exercise 8.9 Prove this theorem in two ways: (i) using the Cauchy criterion for convergence, and (ii) using Theorem 8.7.

Theorem 8.10 (The Weierstrass test) Assume that \( 0 \leq |a_n| \leq b_n \), then if the series \( \sum_{k=1}^{\infty} b_n \) converges then the series \( \sum_{k=1}^{\infty} a_n \) converges absolutely.

This is an immediate consequence of Theorem 8.8.

Criteria for series convergence

We now discuss some criteria for the convergence of a series, based on the Weierstrass and the convergence of the geometric series \( \sum r^n \) with \( |r| < 1 \).

Theorem 8.11 (The Cauchy test) Let \( \sum a_n \) be a series and set
\[
\alpha = \limsup_{n \to \infty} (|a_n|)^{1/n}.
\]

Then (i) if \( \alpha < 1 \) then the series \( \sum a_n \) converges absolutely, and (ii) if \( \alpha > 1 \) then the series diverges.

Note that if \( \alpha = 1 \) then the Cauchy criterion says nothing about the convergence. As an example of a divergent series with \( \alpha = 1 \), consider \( \sum_n (1/n) \), then
\[
\alpha = \limsup_{n \to \infty} \frac{1}{n^{1/n}} = 1,
\]
but the series diverges. On the other hand, the series \( \sum_n (1/n^2) \) has
\[
\alpha = \limsup_{n \to \infty} \frac{1}{n^{2/n}} = 1,
\]
but the series converges. This is because for \( n \geq 2 \) we have
\[
\frac{1}{n^2} \leq \frac{1}{n(n-1)} = \frac{1}{n-1} - \frac{1}{n}.
\]
and the series
\[ \sum_{n=1}^{\infty} \left( \frac{1}{n-1} - \frac{1}{n} \right) \]
converges because its partial sums are explicit:
\[ S_n = \left( 1 - \frac{1}{2} \right) + \left( \frac{1}{2} - \frac{1}{3} \right) + \left( \frac{1}{3} - \frac{1}{4} \right) + \cdots + \left( \frac{1}{n-1} - \frac{1}{n} \right) = 1 - \frac{1}{n}, \]
and \( S_n \to 1 \) as \( n \to +\infty \).

Proof of Theorem 8.11. Assume that \( \alpha < 1 \), then there exists \( N \) so that for all \( n > N \) we have
\[ |a_n|^{1/n} \leq \beta = \alpha + \frac{1 - \alpha}{2} = \frac{1 + \alpha}{2} < 1, \]
so that for all \( n > N \) we have \( |a_n| \leq \beta^n \). As \( 0 < \beta < 1 \), the series \( \sum \beta^n \) converges, hence the Weierstrass criterion implies that the series \( \sum a_n \) converges absolutely.

On the other hand, if \( \alpha > 1 \) then there exists \( N \) so that for all \( n > N \) we have
\[ |a_n|^{1/n} \geq \gamma = \alpha - \frac{\alpha - 1}{2} = \frac{1 + \alpha}{2} > 1, \]
hence \( |a_n| > \gamma^n \). As \( \gamma > 1 \), it follows that it is not true that \( a_n \to 0 \) as \( n \to +\infty \), hence the series \( \sum \beta^n \) diverges. □

Theorem 8.12 (The d’Alembert test) Suppose that the limit
\[ \alpha = \lim_{n \to +\infty} \left| \frac{a_{n+1}}{a_n} \right| \]
exists for a series \( \sum a_n \). Then (i) if \( \alpha < 1 \), the series converges absolutely, and (ii) if \( \alpha > 1 \), the series diverges.

Proof. (i) If \( \alpha < 1 \) then there exists \( N \) so that for all \( n \geq N \) we have
\[ |a_{n+1}| \leq |a_n|, \]
where \( r = (1 + \alpha)/2 < 1 \). Hence, for all \( n \geq N \) we have \( |a_n| \leq a_N r^{n-N} \). As \( r < 1 \), the series \( a_n \) converges absolutely.

(ii) If \( \alpha > 1 \), then there exists \( N \) so that for all \( n \geq N \) we have
\[ |a_{n+1}| \geq |a_n|, \]
where \( r = (1 + \alpha)/2 > 1 \). Hence, for all \( n > N \) we have \( |a_n| \geq |a_N| r^{n-N} \) with \( r > 1 \), hence it is impossible that \( a_n \to 0 \) as \( n \to +\infty \), because \( |a_n| \geq |a_N| \). □

The next proposition is both important, and its proof uses a very useful tool.

Theorem 8.13 The series \( \sum a_n (1/n^p) \) converges for all \( p > 1 \).

Proof. According to Theorem 8.7, it suffices to prove that the sequence of partial sums
\[ S_n = \sum_{k=1}^{n} \frac{1}{k^p} \]
is bounded. As $a_n \geq 0$, it is actually sufficient to show that the sequence $S_{2^n}$ is bounded. Note that $a_n = 1/n^p$ is a decreasing sequence. Hence we may write

$$S_{2^n} = a_1 + a_2 + a_3 + a_4 + \cdots + a_{2^n} \leq a_1 + a_2 + a_2 + a_4 + a_4 + a_4 + a_8 + \cdots + a_{2^n} = a_1 + 2a_2 + 4a_4 + 8a_8 + \cdots + 2^n a_{2^n}.$$  

Recalling that $a_n = 1/n^p$, we have

$$S_{2^n} \leq \sum_{k=1}^{n} 2^k \frac{1}{(2k)^p} = \sum_{k=1}^{n} \frac{1}{(2k)^p} \leq \sum_{k=1}^{n} \frac{1}{(2^p-1)^k} = \frac{1 - r^{n+1}}{1 - r} \leq \frac{1}{1 - r},$$

with $r = 1/2^p < 1$ — here we use the fact that $p > 1$. Therefore, the partial sums $S_{2^n}$ are bounded, and hence so are the partial sums $S_n$, and the series converges. □

### 8.3 Uniform convergence of functions and series

We now discuss the notion of uniform convergence of functions.

**Definition 8.14** A sequence of functions $f_n : E \to \mathbb{R}$ converges to a function $f : E \to \mathbb{R}$ uniformly on $E$ if for any $\varepsilon > 0$ there exists $N$ so that $|f(x) - f_n(x)| < \varepsilon$ for all $n \geq N$ and all $x \in E$. We sometimes use the notation $f_n \rightarrow f$ on $E$.

As an example of a uniformly convergent sequence, consider $f_n(x) \equiv 1/n$ on $\mathbb{R}$, or any other set $E \subset \mathbb{R}$. The sequence $f_n(x) = x/n$ converges uniformly on $[0, 1]$ to $f \equiv 0$ but is not uniformly convergent on $\mathbb{R}$, thus the notion of uniform convergence very much depends on the set $E$ on which we consider the convergence.

**Definition 8.15** A sequence of functions $f_n : E \to \mathbb{R}$ is uniformly Cauchy on $E$ if for any $\varepsilon > 0$ there exists $N$ so that $|f_m(x) - f_n(x)| < \varepsilon$ for all $n, m \geq N$ and all $x \in E$.

Here is a useful criterion for the uniform convergence.

**Theorem 8.16** A sequence of functions $f_n : E \to \mathbb{R}$ is uniformly convergent on $f : E \to \mathbb{R}$ if and only if $f_n$ is uniformly Cauchy on $E$.

**Proof.** First, let $f_n \rightarrow f$ on a set $E$. Then for any $\varepsilon > 0$ there exists $N$ so that $|f(x) - f_n(x)| < \varepsilon/10$ for all $n \geq n$ and all $x \in E$. The triangle inequality implies that for any $n, m \geq N$ and all $x \in E$ we have

$$|f_n(x) - f_m(x)| \leq |f_n(x) - f(x)| + |f(x) - f_m(x)| \leq \frac{\varepsilon}{10} + \frac{\varepsilon}{10} < \varepsilon,$$

hence $f_n$ is a uniformly Cauchy sequence on $E$.

Next, assume that $f_n$ is a uniformly Cauchy sequence on $E$. Then, for each $x \in E$ fixed, the sequence of numbers $f_n(x)$ is Cauchy, thus it converges to some limit that we denote by $f(x)$. As the sequence $f_n$ is uniformly Cauchy on $E$, given $\varepsilon > 0$ there exists $N$ so that for all $n, m \geq N$ and all $x \in E$ we have

$$|f_n(x) - f_m(x)| < \frac{\varepsilon}{10}.$$

Fixing $m > N$ and letting $n \to \infty$ above, we deduce that

$$|f(x) - f_m(x)| < \frac{\varepsilon}{10},$$

for all $m \geq n$ and all $x \in E$,

which means that $f_n \rightarrow f$ on $E$. □
Theorem 8.17 Let \( f_n : E \rightarrow \mathbb{R} \) be continuous functions on \( E \). Assume that \( f_n \) converge uniformly to \( f \) on \( E \), then \( f \) is continuous.

**Proof.** This is a homework problem, will be filled in later.

In the same way, we can talk about uniform convergence of a series of functions.

**Definition 8.18** Let \( a_n : E \rightarrow \mathbb{R} \) be functions from a set \( E \) to \( \mathbb{R} \). The series \( \sum_n a_n(x) \) converges uniformly on \( E \) if the corresponding sequence of partial sums \( S_n(x) = \sum_{k=1}^n a_k(x) \) converges uniformly on \( E \).

Restating Theorem 8.16 for the uniform convergence of a series gives the following.

**Theorem 8.19** A series \( \sum_n a_n(x) \) converges uniformly on \( E \) if and only if the corresponding sequence of partial sums \( S_n(x) = \sum_{k=1}^n a_k(x) \) is uniformly Cauchy on \( E \).

Restating Theorem 8.17 for the uniform convergence of series of continuous functions gives the next theorem.

**Theorem 8.20** Let \( a_n : E \rightarrow \mathbb{R} \) be continuous functions from a set \( E \) to \( \mathbb{R} \). If a series \( \sum_n a_n(x) \) converges uniformly on \( E \) to \( S(x) = \sum_{n=1}^{\infty} a_n(x) \), then the function \( S(x) \) is continuous.

Here is a very useful criterion for the uniform convergence of a series.

**Theorem 8.21** (The Weierstrass test) Let \( a_n : E \rightarrow \mathbb{R} \) be functions from a set \( E \) to \( \mathbb{R} \). If there exists a sequence of numbers \( b_n \geq 0 \) such that \( |a_n(x)| < b_n \), and the series \( \sum_n b_n \) converges, then the series \( \sum_n a_n(x) \) converges uniformly on \( E \). It also converges absolutely for each \( x \in E \). If, in addition, each function \( a_n(x) \) is continuous then the sum \( S(x) = \sum_{n=1}^{\infty} a_n(x) \) is a continuous function.

**Proof.** The absolute convergence follows simply from Theorem 8.10 (another Weierstrass test). For the uniform convergence, note that the partial sums

\[
S_n(x) = \sum_{k=1}^n a_k(x)
\]

satisfy

\[
|S_n(x) - S_m(x)| = \left| \sum_{k=m+1}^{n} a_k(x) \right| \leq \sum_{k=m+1}^{n} |a_k(x)| \leq \sum_{k=m+1}^{n} b_k,
\]

for all \( x \in E \). As the series \( \sum b_k \) is convergent, it is Cauchy, thus given any \( \varepsilon > 0 \) we can find \( N \) so that for all \( n, m \geq n \) we have

\[
\sum_{k=m+1}^{n} b_k < \varepsilon.
\]

It follows that then

\[
|S_n(x) - S_m(x)| < \varepsilon
\]

for all \( x \in E \). Thus, the series \( \sum a_n(x) \) is uniformly Cauchy, hence it is uniformly convergent. Finally, the continuity of \( S(x) \) follows from the uniform convergence of the series \( \sum_n a_n(x) \) on the set \( E \) that we have just proved, and Theorem 8.20. □
8.4 Convergence of power series

A series of the form

\[ a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + \cdots + a_n(x - x_0)^n + \ldots \]

is called a power series around \( x_0 \), or simply a power series. We are interested in two questions: first, if the series converges and second, if the sum is a continuous function. The is answered by the following theorem.

**Theorem 8.22** Let

\[ R^{-1} = \limsup_{n \to \infty} |a_n|^{1/n} \]

if the lim sup in the right side is different from zero, otherwise set \( R = +\infty \). Then the series

\[ \sum_{n=1}^{\infty} a_n(x - x_0)^n \]

converges for all \( x \) such that \( |x - x_0| < R \) and diverges for each \( x \) such that \( |x - x_0| > R \). Moreover, convergence is uniform on any interval \( |x - x_0| \leq r \) with \( 0 < r < R \), and the sum

\[ S(x) = \sum_{n=1}^{\infty} a_n(x - x_0)^n \]

is continuous on the open interval \( |x - x_0| < R \).

**Proof.** We will set \( x_0 = 0 \) without loss of generality. Note that the individual terms of the series are \( c_n(x) = a_n x^n \), and

\[ |c_n|^{1/n} = |x|(|a_n|)^{1/n}, \]

so that

\[ \limsup_{n \to \infty} |c_n|^{1/n} = |x| \limsup_{n \to \infty} |a_n|^{1/n} = \frac{|x|}{R}. \]

Now, the Cauchy criterion, Theorem 8.11 tells us that the series converges if \( |x| < R \) and diverges in \( |x| > R \), the first claim of the theorem. To show that the series converges uniformly on any interval \([−r, r]\) with \( r < R \), note that we can find \( N \) so that for all \( n > N \) we have

\[ |a_n|^{1/n} \leq \frac{1}{(R+r)/2}, \]

so that

\[ |a_n| \leq \left( \frac{2}{R+r} \right)^n. \]

Then, for all \( n \geq n \) and all \( x \in [−r, r] \) we have

\[ |a_n||x|^n \leq \left( \frac{2r}{R+r} \right)^n |x|^n \leq \left( \frac{2r}{R+r} \right)^n s^n, \quad s = \frac{2r}{R+r}. \]

As \( r < R \), we know that \( 0 < s < 1 \), so that the series \( \sum s^n \) converges. Now, the Weiestress test, Theorem 8.21, implies that the series converges uniformly and the sum is a continuous function on \([−r, r]\). As this is true for all \( 0 < r < R \), the sum is continuous on the whole open interval \((−R, R)\). □
Exercise 8.23  (i) Show that the power series
\[ 1 + x + x^2 + \cdots + x^n + \ldots \]
converges for \(|x| < 1\).
(ii) Show that the power series
\[ 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + \ldots \]
converges for all \(x \in \mathbb{R}\).
(iii) Show that the power series
\[ 1 + x + \frac{x^2}{2} + \cdots + \frac{x^n}{n} + \ldots \]
converges for \(|x| < 1\).

9 Differentiability in \(\mathbb{R}^n\)

9.1 Differentiability in \(\mathbb{R}\)

In this section, we deal with differentiability of functions defined on the real line, so \(x \in \mathbb{R}\) throughout.

Definition 9.1 Let \(f\) be a function from an open set \(U\) to \(\mathbb{R}\), and \(a \in U\) If the limit
\[ f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \]
exists then \(f'(a)\) is called the derivative of \(f\) at \(a\).
Equivalently, we \(f(x)\) is differentiable at \(a\) if
\[ \frac{f(x) - f(a)}{x - a} = f'(a) + \alpha(x), \quad (9.1) \]
and \(\alpha(x) \to 0\) as \(x \to a\). The notation is \(\alpha(x) = o(1)\) as \(x \to a\). We say that \(\alpha(x) = o(|x - a|^m)\) as \(x \to a\) if
\[ \lim_{x \to a} \frac{\alpha(x)}{|x - a|^m} = 0. \]
We say that \(\alpha(x) = O(|x - a|^m)\) as \(x \to a\) if
\[ \limsup_{x \to a} \frac{|\alpha(x)|}{|x - a|^m} < +\infty. \]
In other words, \(\alpha(x) = O(|x - a|^m)\) as \(x \to a\) if there exists \(r_0 > 0\) and \(C > 0\) so that
\[ \frac{|\alpha(x)|}{|x - a|^m} < C \text{ for all } x \text{ such that } |x - a| < r_0. \]
In other words, \(\alpha(x) = o(|x - a|^m)\) if \(\alpha(x) \to 0\) "faster" than \(|x - a|^m\) and \(\alpha(x) = O(|x - a|^m)\) if \(\alpha(x) \to 0\) "at least as fast" as \(|x - a|^m\).
Now, we can restate (9.1) as
\[ f(x) = f(a) + f'(a)(x - a) + o(|x - a|), \quad (9.2) \]
that is, \(f(x)\) is approximated, to the leading order by the linear function \(f(a) + f'(a)(x - a)\). We often write this as
\[ f(a + h) = f(a) + f'(a)h + o(|h|). \quad (9.3) \]
**Exercise 9.2** An important observation is that if $\alpha(h) = o(|h|)$ and $\beta(h)$ is bounded, then $(\alpha\beta)(h) = o(|h|)$ as well. Check this!

**Theorem 9.3** Let $U$ be an open set. If $f : U \to \mathbb{R}$ is differentiable at $x \in U$ then $f$ is continuous at $a$.

**Proof.** This follows immediately from (9.2). □

**Theorem 9.4** Let $U$ be an open set. If $f : U \to \mathbb{R}$ and $g : U \to \mathbb{R}$ are differentiable at $x \in U$ then $f + g$ and $fg$ are differentiable at $a$, with

$$(f + g)'(a) = f'(a) + g'(a), \quad (fg)'(a) = f(a)g'(a) + f'(a)g(a).$$

If, in addition, $g(a) \neq 0$, then $f/g$ is differentiable at $a$ and

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

**Proof.** We will only prove this for $fg$, and use the $o(h)$ notation, to get the reader used to it:

$$(fg)(a + h) - (fg)(a) = (f(a) + f'(a)h + o(|h|))(g(a) + g'(a)h + o(|h|)) - f(a)g(a)$$

$$= (f'(a)g(a) + f(a)g'(a))h + o(|h|).$$

For $1/g$ we note that for $\alpha \neq 0$ we have

$$\frac{1}{1 + \beta h + o(|h|)} = 1 - \beta h + o(|h|).$$

Warning: the $o(|h|)$ in the left and the right sides above are not the same functions. This identity means that if $\alpha(h)/h \to 0$ as $h \to 0$, then

$$\frac{1}{1 + \beta h + o(h)} - (1 - \beta h) = h\gamma(h),$$

where $\gamma(h) \to 0$ as $h \to 0$. Make sure you understand this! Now, we can write

$$\frac{1}{g(a + h)} - \frac{1}{g(a)} = \frac{1}{g(a) + g'(a)h + o(|h|)} - \frac{1}{g(a)} = \frac{1}{g(a)} \frac{1}{1 + (g'(a)/g(a))h + o(|h|)} - \frac{1}{g(a)}$$

$$= \frac{1}{g(a)} \left(1 - \frac{g'(a)}{g(a)}h + o(|h|)\right) - \frac{1}{g(a)} = -\frac{g'(a)}{g^2(a)}h + o(|h|),$$

so that

$$\left(\frac{1}{g}\right)'(a) = -\frac{g'(a)}{g^2(a)}.$$ 

**Theorem 9.5** Let $X$ and $Y$ be open subsets of $\mathbb{R}$. If $f : X \to \mathbb{R}$ is differentiable at a point $x \in X$, and $f(x) \in Y$, and $g : Y \to \mathbb{R}$ is differentiable at $y = f(x)$, then the composition $g \circ f$ is differentiable at $x$ and $(g \circ f)'(x) = g'(f(x))f'(x)$.

**Proof.** We have

$$f(x + h) = f(x) + f'(x)h + o(|h|) \text{ as } h \to 0,$$

and

$$g(y + t) = g(y) + g'(y)t + o(|t|) \text{ as } t \to 0.$$ 

(9.4)
Then we have
\[(g \circ f)(x + h) = g(f(x) + f'(x)h + o(|h|)) = g(y + t(h)), \quad (9.5)\]
where \(t(h) = f'(x)h + o(|h|)\). Note that \(t(h) \to 0\) as \(h \to 0\) and there exists \(r_\epsilon\) so that \(|t(h)| \leq 2|f'(x)|h\) for \(|h| < r_\epsilon\). If we write out the meaning of (9.4), it says that for any \(\epsilon > 0\) there exists \(\delta > 0\) so that if \(|t| < \delta\), then
\[
\frac{|g(y + t) - (g(y) + g'(y)t)|}{|t|} < \epsilon. \quad (9.6)
\]
Choosing \(|h|\) sufficiently small, we may ensure that \(|t(h)| < \delta\), so that
\[
\frac{|g(y + t(h)) - (g(y) + g'(y)t(h))|}{|t(h)|} < \epsilon, \quad (9.7)
\]
and then
\[
|g(y + t(h)) - (g(y) + g'(y)t(h))| < \epsilon|t(h)| < 2\epsilon|f'(x)|h. \quad (9.8)
\]
Going back to (9.5) it follows that
\[
|(g \circ f)(x + h) - g(y) - g'(y)t(h)| < 2\epsilon|f'(x)|h. \quad (9.9)
\]
Now, we recall that \(y = f(x)\) and \(t(h) = f'(x)h + o(|h|)\), which gives
\[
|(g \circ f)(x + h) - g(f(x)) - g'(f(x))f'(x)h - g'(f(x))o(|h|)| < 2\epsilon|f'(x)|h. \quad (9.10)
\]
It follows that
\[
(g \circ f)(x + h) - g(f(x)) - g'(f(x))f'(x)h = o(|h|), \quad (9.11)
\]
and we are done. \(\square\)

**Exercise 9.6** Go over the proof and make sure you understand at each step how the notation \(o(|h|)\) is used and why it makes sense and is all rigorous!

**Theorem 9.7** (Rolle’s theorem) If \(f : [a, b] \to \mathbb{R} \) is continuous, \(f(a) = f(b)\) and \(f\) is differentiable at each point \(x \in (a, b)\), then there is \(c \in (a, b)\) such that \(f'(c) = 0\).

**Proof.** If \(f(x) \equiv f(a)\) on \([a, b]\) then \(f'(x) = 0\) for all \(x \in (a, b)\). Let us assume that \(f\) is not equal identically to a constant on \([a, b]\). Then it either attains its maximum, or a minimum at some \(c \in (a, b)\). Then the ratios
\[
\frac{f(c + h) - f(c)}{h}
\]
have different signs for \(h > 0\) and \(h < 0\). It follows that \(f'(x) = 0\). \(\square\)

**Theorem 9.8** (Mean Value theorem) If \(f : [a, b] \to \mathbb{R} \) is continuous, and \(f\) is differentiable at each point \(x \in (a, b)\), then there is \(c \in (a, b)\) such that
\[
f'(c) = \frac{f(b) - f(a)}{b - a}.
\]
**Proof.** Apply Rolle’s theorem to the function
\[
g(x) = f(x) - \frac{f(b) - f(a)}{b - a}(x - a).
\]
Note that \(g(a) = g(b) = f(a)\). \(\square\)
9.2 The Taylor series representation

We now ask the question of when we can approximate a function to a higher order than linearly. So far, we have seen that if \( f'(x_0) \) exists then we can approximate \( f(x + h) \) by a linear function, up to an error of the order \( o(h) \). Here is a more general result.

**Theorem 9.9** Let \( f \) be a function differentiable up to order \( m + 1 \) on an interval \( (x_0 - r, x_0 + r) \). Then for any \( x \) such that \( |x - x_0| < r \), there exists \( c \) between \( x_0 \) and \( x \) such that

\[
f(x) = \sum_{n=0}^{m} \frac{f^{(n)}(x_0)}{n!} (x - x_0)^n + \frac{f^{(m+1)}(c)}{(m+1)!} (x - x_0)^{m+1}.
\]

**Proof.** Fix some \( x \) such that \( |x - x_0| < r \) and consider the function \( g(t) \) defined on \( |t - x_0| < r \) by

\[
g(t) = f(t) - \sum_{n=0}^{m} \frac{f^{(n)}(x_0)}{n!} (t - x_0)^n - M(t - x_0)^{m+1},
\]

with the constant \( M \) chosen so that \( g(x) = 0 \) (keep in mind that both \( x \) and \( x_0 \) are fixed, the independent variable is \( t \)):

\[
M = \frac{1}{(x - x_0)^{m+1}} \left( f(x) - \sum_{n=0}^{m} \frac{f^{(n)}(x_0)}{n!} (x - x_0)^n \right).
\]

As \( g(x_0) = 0 \) and \( g(x) = 0 \), we deduce that there exists \( c_1 \) between \( x_0 \) and \( x \) such that \( g'(c_1) = 0 \). However, we also \( g'(x_0) = 0 \) (check this!) – hence there exists \( c_2 \) between \( x_0 \) and \( c_1 \) (and still between \( x_0 \) and \( x \)) such that \( g''(c_2) = 0 \). But we also have \( g''(x_0) = 0 \) (check this!), so there exists \( c_3 \) between \( x_0 \) and \( c_2 \) (and still between \( x_0 \) and \( x \)) such that \( g'''(c_3) = 0 \) Note that

\[
g^{(k)}(x_0) = 0, \quad \text{for all } 0 \leq k \leq m,
\]

hence we can continue this argument until step \( m + 1 \), funding a point \( c_{m+1} \) between \( x_0 \) and \( x \) so that \( g^{(m+1)}(c_{m+1}) = 0 \). Note that for any \( |t - x_0| < r \) we have

\[
g^{(m+1)}(t) = f^{(m+1)}(t) - M(m+1)!,
\]

hence

\[
M = \frac{f^{(m+1)}(c)}{(m+1)!}.
\]

Using this in (9.12) says that

\[
f(x) = \sum_{n=0}^{m} \frac{f^{(n)}(x_0)}{n!} (x - x_0)^n + \frac{f^{(m+1)}(c_{m+1})}{(m+1)!} (x - x_0)^{m+1},
\]

and we are done. □

**Theorem 9.10** Let \( f(x) \) be differentiable to all orders in an interval \( |x - x_0| < r \) and assume that there exists a constant \( C > 0 \) so that

\[
\left| \frac{f^{(n)}(x)}{n!} \right| r^n \leq C, \quad \text{for all } n \in \mathbb{N} \text{ and all } x \text{ such that } |x - x_0| < r,
\]

then we have

\[
f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!} (x - x_0)^n, \quad \text{for all } x \text{ such that } |x - x_0| < r.
\]
Proof. By assumptions of this theorem and by Theorem 9.9, for any \( m \in \mathbb{N} \) and any \( x \in (x_0 - r, x_0 + r) \), we can find \( c_m(x) \) between \( x_0 \) and \( x \) so that
\[
f(x) - \sum_{n=0}^{m} \frac{f^{(n)}(x_0)}{n!} (x-x_0)^n = \frac{f^{(m+1)}(c_m(x))}{(m+1)!} (x-x_0)^{m+1}.
\]
Using the assumptions of the present theorem, we see that
\[
\left| f(x) - \sum_{n=0}^{m} \frac{f^{(n)}(x_0)}{n!} (x-x_0)^n \right| \leq \frac{|f^{(m+1)}(c_m(x))|}{(m+1)!} |x-x_0|^{m+1} = \frac{|f^{(m+1)}(c_m(x))|}{(m+1)!} r^{m+1} |x-x_0|^{m+1}
\]
and we are done. \( \square \)

9.3 Differentiability in \( \mathbb{R}^n \)

Recall that for a scalar function \( f : \mathbb{R} \to \mathbb{R} \) we say that a function \( f \) is differentiable at \( x \) if there exists a number \( a \in \mathbb{R} \) so that
\[
f(x+h) = f(x) + ah + o(|h|), \text{ as } h \to 0,
\]
and then we say that \( a = f'(x) \). For a scalar-valued function \( f : \mathbb{R}^n \to \mathbb{R} \) the definition is very similar. It is differentiable at a point \( x \in \mathbb{R}^n \) if there exists a linear function \( L : \mathbb{R}^n \to \mathbb{R} \) such that
\[
f(x+h) = f(x) + L(h) + o(\|h\|), \text{ as } h \to 0,
\]
in the sense that
\[
\lim_{h \to 0} \frac{\|f(x+h) - (f(x) + L(h))\|}{\|h\|} = 0.
\]
A linear function \( L : \mathbb{R}^n \to \mathbb{R} \) must have the form \( L(h) = (v \cdot h) \) with some \( v \in \mathbb{R}^n \). We call this vector the gradient of \( f \) at \( x \):
\[
v = \nabla f(x).
\]
In other words, a scalar-valued function \( f : \mathbb{R}^n \to \mathbb{R} \) is differentiable at a point \( x \in \mathbb{R}^n \) if
\[
f(x+h) = f(x) + (\nabla f(x) \cdot h) + o(\|h\|), \text{ as } h \to 0.
\]
This definition can be extended to mappings \( f : \mathbb{R}^n \to \mathbb{R}^m \) easily: a mapping \( f : \mathbb{R}^n \to \mathbb{R}^m \) is differentiable at a point \( x \in \mathbb{R}^n \) if there exists a linear map \( L : \mathbb{R}^n \to \mathbb{R}^m \) such that
\[
f(x+h) = f(x) + L(h) + o(\|h\|), \text{ as } h \to 0,
\]
in the sense that
\[
\lim_{h \to 0} \frac{\|f(x+h) - (f(x) + L(h))\|}{\|h\|} = 0.
\]
Recall that a linear map \( L : \mathbb{R}^n \to \mathbb{R} \) must have the form \( L(h) = Ah \) with some \( m \times n \) matrix \( A \). We call this matrix the derivative matrix of \( f \) (or the Jacobian matrix):
\[
A = Df(x).
\]

Lemma 9.11 If \( U \subset \mathbb{R}^n \) and \( f : U \to \mathbb{R}^m \) is differentiable at a point \( x_0 \in U \), then \( f \) is continuous at \( x_0 \).

Exercise 9.12 Adapt the proof of this fact for scalar valued functions of a real variable to prove Lemma 9.11.

Definition 9.13 For a scalar-valued function \( f(x) \), the gradient of \( f \) at \( x_0 \) is the vector \( Df(x_0) \) if it exists.
9.4 Directional derivatives

We now interpret the derivative matrix $Df(x)$ in more concrete terms. Let $U \subset \mathbb{R}^n$ be an open set, and $f$ a map from $U$ to $\mathbb{R}^m$. Then, for each $x \in U$ we can find a ball $B(x, r)$ that is contained in $U$. This means that for any $v \in \mathbb{R}^n$ the point $x + tv$ is in $U$ as long as $|t| < r/|v|$ is sufficiently small, so that the function

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

of a real variable $t \in \mathbb{R}$ is defined for $|t| < r/v$.

**Definition 9.14** Given a vector $v \in \mathbb{R}^n$, $v \neq 0$, the directional derivative $D_v f(x_0)$ is

$$D_v f(x_0) = \lim_{t \to 0} \frac{1}{t} (f(x_0 + tv) - f(x_0)),$$

whenever that limit exists.

Note that if $f$ maps $U$ to $\mathbb{R}^m$ then $D_v f(x_0)$ is a (row) vector in $\mathbb{R}^m$ as well.

**Exercise 9.15** Fix some $\lambda \in \mathbb{R}$ and $v \in \mathbb{R}^n$, and let $w = \lambda v$. Express $D_w f(x_0)$ in terms of $D_v f(x_0)$.

In the special case when $v = e_j$, the $j$-th vector of the standard basis, we denote $D_{e_j} f(x_0)$ as

$$D_{e_j} f(x_0) = \frac{\partial f(x_0)}{\partial x_j}.$$

Let us show that if the gradient matrix $D f(x_0)$ exists then directional derivatives can all be computed from it.

**Theorem 9.16** Let $U \subset \mathbb{R}^n$ be an open set and $f : U \to \mathbb{R}^m$ be a map that is differentiable at a point $x_0 \in U$. Then the directional derivative in a direction $v \in \mathbb{R}^n$, $v \neq 0$, has the form

$$D_v f(x_0) = (D f(x_0))v,$$

or, written component-wise, we have

$$D_v f_k(x) = \sum_{j=1}^n v_j \frac{\partial f_k(x_0)}{\partial x_j}.$$

**Proof.** First, as $U$ is an open set, there exists $r$ so that $x_0 + tv$ is in $U$ for $|t| < r$, and the function

$$g(t) = f(x_0 + tv)$$

is defined, as a mapping from $(-r, r)$ to $\mathbb{R}^m$. Also, as $f$ is differentiable at $x_0$, we have

$$f(x_0 + h) - f(x) = (D f(x_0))h + o(\|h\|), \quad \text{as } h \to 0,$$

with $h \in \mathbb{R}^n$. Note that if some function $\alpha(h) = o(\|h\|)$ as $h \to 0$, then $\alpha(tv) = o(t)$ for any $v \in \mathbb{R}^n$ fixed (check this!), hence using $h = tv$ above, we have

$$f(x_0 + tv) - f(x) = (D f(x_0))(tv) + o(t), \quad \text{as } t \to 0,$$

with $t \in (-r, r)$. This means exactly that

$$\lim_{t \to 0} \frac{f(x_0 + tv) - f(x_0)}{t} = (D f(x_0))v,$$
hence \( D_v f(x_0) = (Df(x_0))v \), as claimed. Going back to (9.13) and writing it separately for each coordinate \( f_k, k = 1, \ldots, m \), gives

\[
f_k(x_0 + tv) - f_k(x) = (Df_k(x_0))(tv) + o(t), \quad \text{as } t \to 0,
\]

(9.14)

However, \( Df_k \) is simply a row vector in \( \mathbb{R}^m \) and \( (Df_k)(v) \) is the dot-product \( \langle Df_k(x_0), v \rangle \). Taking \( v = e_j \) shows that

\[
(Df_k(x_0))_j = \frac{\partial f_k(x_0)}{\partial x_j}.
\]

It follows then that

\[
D_v f_k(x_0) = (Df_k(x_0))v = \sum_{j=1}^n v_j (Df_k(x_0))_j = \sum_{j=1}^n v_j \frac{\partial f_k(x_0)}{\partial x_j},
\]

and we are done. □

Note that existence of partial derivatives of \( f \) at \( x_0 \) does not imply that the derivative matrix \( Df(x_0) \) exists. To see that, we may borrow an example from Section 6.3. Consider a function \( f : \mathbb{R}^2 \to \mathbb{R} \) defined by

\[
f(x_1, x_2) = \begin{cases} x_1 + x_2 \sin \left( \frac{1}{x_1} \right), & \text{if } x_1 \neq 0, \\ 0, & \text{if } x_1 = 0. \end{cases}
\]

Then we have \( f(x_1, 0) = x_1 \) and \( f(0, x_2) = 0 \), which means that

\[
\frac{\partial f}{\partial x_1}(0, 0) = 1, \quad \frac{\partial f}{\partial x_2}(0, 0) = 0,
\]

so if \( Df(0, 0) \) exists then \( Df = (1, 0) \), but if we look at \( x_1 = x_2 = h \), then

\[
f(h, h) - f(0, 0) = h + h \sin \left( \frac{1}{h} \right),
\]

and the right side is not of the form \( \langle (1, 0), (h, h) \rangle + o(\|h\|) \).

Here is an extra condition that ensures that existence of partial derivatives implies existence of the derivative matrix.

**Theorem 9.17** Let \( f : U \to \mathbb{R}^n \), \( x_0 \in U \subset \mathbb{R}^n \) and assume that the partial derivatives \( D_k f(x) \) exists for all \( x \) in a ball \( B(x_0, \rho) \) with some \( \rho > 0 \) and are continuous at \( x_0 \). Then \( f \) is differentiable at \( x_0 \).

**Proof.** We will assume without loss of generality that \( m = 1 \), so that \( f \) is real-valued and the derivative ”matrix” is simply the gradient row-vector (if it exists). For general \( m > 1 \) one argue simply component by component. Take \( h = (h, \ldots, h_n) \) with \( |h| < \rho \) and write

\[
f(x_0 + h) - f(x_0) = (f(x_0 + h_1 e_1) - f(x_0)) + (f(x_0 + h_1 e_1 + h_2 e_2) - f(x_0 + h_1 e_1)) + \ldots + (f(x_0 + h_1 e_1 + \ldots + h_n e_n) - f(x_0 + h_1 e_1 + \ldots + h_{n-1} e_{n-1})).
\]

(9.15)

The intermediate value theorem for functions of one variable implies that there exist \( \xi_1 \in B(x_0, \|h\|), \xi_2 \in B(x_0, \|h\|), \ldots, \xi_n \in B(x_0, \|h\|) \), such that

\[
\begin{align*}
f(x_0 + h_1 e_1) - f(x_0) &= \frac{\partial f(\xi_1)}{\partial x_1} h_1, \\
f(x_0 + h_1 e_1 + h_2 e_2) - f(x_0 + h_1 e_1) &= \frac{\partial f(\xi_2)}{\partial x_2} h_2, \\
&\vdots \\
f(x_0 + h_1 e_1 + \ldots + h_n e_n) - f(x_0 + h_1 e_1 + \ldots + h_{n-1} e_{n-1}) &= \frac{\partial f(\xi_n)}{\partial x_n} h_n.
\end{align*}
\]

43
We now consider the chain rule for maps. Let

$$U$$

The chain rule

It follows from (9.22) and (9.19) that actually

$$\frac{\partial f(\xi_k)}{\partial x_k} \cdot h = o(1), \quad as \ h \to 0,$$

so that, as $$h_k o(1) = o(||h||)$$ for any $$1 \leq k \leq n$$ fixed (check this!), we have

$$f(x_0 + h) - f(x_0) = \frac{\partial f(x_0)}{\partial x_1} h_1 + \frac{\partial f(x_0)}{\partial x_2} h_2 + \cdots + \frac{\partial f(x_0)}{\partial x_n} h_n + o(||h||), \quad as \ h \to 0. \quad (9.17)$$

This means precisely that $$f$$ is differentiable at $$x_0$$ and

$$Df(x_0) = \left( \frac{\partial f(x_0)}{\partial x_1}, ..., \frac{\partial f(x_0)}{\partial x_n} \right),$$

and we are done. □

9.5 The chain rule

We now consider the chain rule for maps. Let $$U \subset \mathbb{R}^n$$ and $$V \subset \mathbb{R}^m$$ be two open sets, $$f$$ be a map from $$U$$ to $$V$$ and $$g$$ a map from $$V$$ to $$\mathbb{R}^k$$. Assume that $$f$$ is differentiable at $$x_0 \in U$$ and $$g$$ is differentiable at $$y_0 = f(x_0)$$. Then the composition $$g \circ f$$ is a map from $$U$$ to $$\mathbb{R}^k$$. We will show that $$(g \circ f)$$ is differentiable at $$x_0$$ and that the derivative matrix of $$(g \circ f)$$ at $$x_0$$ is

$$D (g \circ f)(x_0) = Dg(y_0) Df(x_0). \quad (9.18)$$

Note that $$Df(x_0)$$ is an $$m \times n$$ matrix, while $$Dg(y_0)$$ is a $$k \times m$$ matrix, so that the product $$Dg(y_0) Df(x_0)$$ makes sense and is a $$k \times n$$ matrix, as it should be for a derivative matrix of a map from $$\mathbb{R}^n$$ to $$\mathbb{R}^k$$. The proof is quite similar to that for real-valued functions. First, note that the continuity of $$f$$ and $$g$$ and the fact that $$V$$ is open imply that there exists $$r > 0$$ small so that $$f(x) \in V$$ for all $$x \in B(x_0, r)$$, so that $$g \circ f$$ is defined for all $$x \in B(x_0, r)$$. Next, we write

$$f(x_0 + h) = f(x_0) + Df(x_0) h + o(||h||), \quad as \ h \to 0, \quad (9.19)$$

with $$h \in \mathbb{R}^n$$, and

$$g(y_0 + t) = g(y_0) + Dg(y_0) t + o(||t||), \quad as \ t \to 0, \quad (9.20)$$

with $$t \in \mathbb{R}^m$$. Given $$h$$, let us take $$t = f(x_0 + h) - f(x_0)$$, then, as $$y = f(x_0)$$, (9.19), together with (9.20) say that

$$g(f(x_0) + h) = g(f(x_0)) + Dg(y_0) (f(x_0 + h) - f(x_0)) + r(h), \quad (9.21)$$

with a function $$r(h)$$ such that

$$\frac{r(h)}{||f(x_0 + h) - f(x_0)||} \to 0 \quad as \ h \to 0. \quad (9.22)$$

It follows from (9.22) and (9.19) that actually

$$\frac{r(h)}{||h||} \to 0 \quad as \ h \to 0, \quad (9.23)$$
that is, \( r(h) = o(h) \). Thus, (9.21) says that
\[
g(f(x_0) + h) = g(f(x_0)) + Dg(y_0)(f(x_0 + h) - f(x_0)) + o(h). \tag{9.24}
\]
Next, use (9.19) for the difference \( f(x_0 + h) - f(x_0) \) in the right side above:
\[
g(f(x_0) + h) = g(f(x_0)) + Dg(y_0)(Df(x_0)h + o(|h|)) + o(h). \tag{9.25}
\]

**Exercise 9.18** Go over the details on how we obtained (9.25), in particular, how (9.19) and (9.22) imply (9.23).

It remains to observe that
\[
Dg(y_0)o(\|h\|) = o(\|h\|),
\]

to conclude that
\[
g(f(x_0) + h) = g(f(x_0)) + Dg(y_0)Df(x_0)h + o(\|h\|), \tag{9.26}
\]
from which we conclude that \( g \circ f \) is differentiable at \( x_0 \) and
\[
D(g \circ f)(x_0) = Dg(y_0)Df(x_0), \tag{9.27}
\]
which is the chain rule.

An important special case is when \( g : \mathbb{R}^m \to \mathbb{R} \) is a real-valued function. Let us define
\[
p(x) = g(f(x)) = g(f_1(x), f_2(x), \ldots, f_m(x)).
\]
Then the chain rule says that
\[
Dp(x) = D(g \circ f) = Dg(f(x))Df(x),
\]
where \( Dg \) is a row-vector of length \( m \), and \( Df \) is an \( m \times n \) matrix, so the product \( Dp \) is a row vector of length \( n \), with the entries
\[
\frac{\partial p(x)}{\partial x_j} = \sum_{l=1}^{m} \frac{\partial g(f(x))}{\partial y_l} \frac{\partial f_l(x)}{\partial x_j}.
\]

\[
\frac{\partial^2 f}{\partial x_i \partial x_j},
\]
or as \( D_iD_j f \), and, more generally, \( k \)-th order partial derivatives of \( f \) as
\[
\frac{\partial^k f}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_k}},
\]
or as \( D_{i_1}D_{i_2} \ldots D_{i_k} f \), with some collection of indices \( 1 \leq i_k \leq n \).
Theorem 9.19 Let $U \subset \mathbb{R}^n$ be an open set, with $x_0 \in U$. Assume that a function $f$ is twice differentiable at $x_0$, and that the partial derivatives $D_iD_jf$ and $D_jD_if$ are continuous at $x_0$. Then $D_iD_jf(x_0) = D_jD_if(x_0)$.

Proof. Let $e_i$ and $e_j$ be the standard basis vectors, and for $h, t \in \mathbb{R}$ sufficiently small, consider the four points $x_0$, $x_0 + te_j$, $x_0 + he_i$, and $x_0 + te_j + he_i$. Then the “second order difference"

$$
\phi(h, t) = f(x_0 + he_i + te_j) + f(x_0) - f(x_0 + he_i) - f(x_0 + te_j)
$$

can be written in two ways: first, as

$$
\phi(h, t) = [f(x_0 + he_i + te_j) - f(x_0 + he_i)] - [f(x_0 + te_j) - f(x_0)].
$$

(9.29)

but also as

$$
\phi(h, t) = [f(x_0 + he_i + te_j) - f(x_0 + te_j)] - [f(x_0 + he_i) - f(x_0)].
$$

(9.30)

Looking at (9.30), let us define the function $g(x)$, with $h$ fixed, as

$$
g(x) = f(x + he_i) - f(x),
$$

(9.31)

so that (9.30) says that

$$
\phi(h, t) = g(x_0 + te_j) - g(x_0).
$$

(9.32)

The intermediate value theorem applied to (9.32) implies that there exists $t_1$ between 0 and $t$ so that

$$
\phi(h, t) = (D_jg)(x_0 + t_1e_j)t.
$$

(9.33)

Note that, differentiating (9.31) gives

$$
D_jg(x) = D_jf(x + he_i) - D_jf(x),
$$

(9.34)

Again, as $D_jf$ is differentiable, the intermediate value theorem applied to $D_jf$ implies that the difference in the right side of (9.34) can be written as

$$
D_jg(x) = hD_iD_jf(x + h_1e_i),
$$

(9.35)

with some $h_1$ between 0 and $h$. Using this in (9.33) with $x = x_0 + t_1e_j$ gives

$$
\phi(h, t) = ht(D_iD_jf)(x_0 + t_1e_j + h_1e_i).
$$

(9.36)

Next, let us use (9.29) rather than (9.30). Define the function $p(x)$, with $t$ fixed, as

$$
p(x) = f(x + te_j) - f(x),
$$

(9.37)

so that (9.29) says that

$$
\phi(h, t) = p(x_0 + he_i) - p(x_0).
$$

(9.38)

The intermediate value theorem implies that there exists $h_2$ between 0 and $h$ so that

$$
\phi(h, t) = (D_i p)(x_0 + h_2e_i)h.
$$

(9.39)

Note that, differentiating (9.37) gives

$$
D_i p(x) = D_i f(x + te_j) - D_i f(x).
$$

(9.40)
Again, as $D_i f$ is differentiable, the intermediate value theorem implies that the difference in the right side can be written as

$$D_i p(x) = tD_j D_i f(x + t_2 e_j), \quad (9.41)$$

with some $t_2$ between 0 and $t$. Using this in (9.39) with $x = x_0 + h_2 e_i$ gives

$$\phi(h, t) = ht(D_j D_i f)(x_0 + t_2 e_j + h_2 e_i). \quad (9.42)$$

Now, comparing (9.36) and (9.42) we deduce that

$$(D_i D_j f)(x_0 + t_1 e_j + h_1 e_i) = (D_j D_i f)(x_0 + t_2 e_j + h_2 e_i). \quad (9.43)$$

Note that $t_1, h_1, t_2, h_2$ in (9.43) all depend on $t$ and $h$ but tend to zero as $t, h \to 0$. Moreover, both $D_i D_j f$ and $D_j D_i f$ are continuous at $x_0$. Passing to the limit in (9.43) gives, therefore:

$$(D_i D_j f)(x_0) = (D_j D_i f)(x_0), \quad (9.44)$$

and we are done. □

As a corollary, we deduce the following: if $f$ is $k$ times differentiable at a point $x_0$, and all of its derivatives of order $k$ are continuous, then the partial derivative

$$D_{i_1} D_{i_2} \ldots D_{i_k} f$$

does not depend on the order in which these derivatives are taken.

### 9.7 Critical points in $\mathbb{R}$

Let us now investigate the critical points of a function $f : \mathbb{R} \to \mathbb{R}$, a real-valued function of a real variable. We say that $x_0$ is a local maximum of a function $f(x)$ if there exists $R_0 > 0$ so that we have $f(x) \leq f(x_0)$ for all $x \in (x_0 - R_0, x_0 + R_0)$. Similarly, $x_0$ is a local minimum of a function $f(x)$ if there exists $R_0 > 0$ so that we have $f(x) \geq f(x_0)$ for all $x \in (x_0 - R_0, x_0 + R_0)$. In other words, there exists an interval $I$ around $x_0$ so that $f$ attains its maximum (minimum) on $I$ at the point $x_0$. Recall that we have already proved that if $f$ attains it maximum or minimum over an open interval $I$ at a point $x_0 \in I$, then $f'(x_0) = 0$. In order to be able to say whether $x_0$ is a maximum or a minimum, we use the second derivative test. In order to make the presentation very easily adaptable to functions on $\mathbb{R}^n$, we begin with the following lemma.

**Lemma 9.20** Let $f$ be twice continuously differentiable on an interval $(a, b)$ and $x_0 \in (a, b)$. Assume that $f'(x_0) = f''(x_0) = 0$, then

$$f(x_0 + h) - f(x_0) = o(h^2).$$

**Proof.** The intermediate value theorem implies that, given $h$, we can find $\xi_1(h)$ that is between $x_0$ and $x_0 + h$ so that

$$f(x_0 + h) - f(x_0) = f'(\xi_1(h))h. \quad (9.45)$$

Next, the intermediate value theorem applied to the function $f'(x)$, implies that there exists $\xi_2(h)$ that is between $x_0$ and $\xi_1(h)$ so that

$$f'(\xi_1(h)) - f'(x_0) = f''(\xi_2(h))(\xi_1(h) - x_0). \quad (9.46)$$

As $f'(x_0) = 0$, this is simply

$$f'(\xi_2(h)) = f''(\xi_2(h))(\xi_1(h) - x_0). \quad (9.47)$$

47
Using this in (9.45) gives
\[ f(x_0 + h) - f(x_0) = f''(\xi_2(h))h(\xi_1(h) - x_0). \] (9.48)

As \(|\xi_1(h) - x_0| \leq h\), we deduce that
\[ |f(x_0 + h) - f(x_0)| \leq |f''(\xi_2(h))|h^2, \] (9.49)
so that
\[ \frac{|f(x_0 + h) - f(x_0)|}{h^2} \leq |f''(\xi_2(h))|. \] (9.50)

Now, as \(f''(x)\) is continuous at \(x_0\), and \(\xi_2(h) \to x_0\) as \(h \to 0\), we know that \(f''(\xi_2(h)) \to f''(x_0) = 0\) as \(h \to 0\). Thus, we have
\[ f(x_0 + h) - f(x_0) = o(h^2), \]
and we are done. □

**Theorem 9.21** Let \(f\) be twice continuously differentiable at a point \(x_0\). Assume that \(f'(x_0) = 0\), then (i) if \(f'(x_0) > 0\) then \(x_0\) is a local minimum of \(f\) and (ii) if \(f''(x_0) < 0\) then \(x_0\) is a local maximum of \(f\).

**Proof.** Let us assume that \(f''(x_0) > 0\). Consider the function
\[ g(x) = f(x) + \frac{1}{2}f''(x_0)(x - x_0)^2, \]
and set
\[ p(x) = f(x) - g(x). \]
Note that \(g'(x_0) = 0\) and \(g''(x_0) = f'(x_0)\), so that
\[ p(x_0) = 0, \quad p'(x_0) = 0, \quad p''(x_0) = 0. \]

It follows from Lemma 9.20 that \(p(x_0 + h) = o(h^2)\), so that
\[ f(x_0 + h) = g(x_0 + h) + o(h^2), \]
so that
\[ r(h) = f(x_0 + h) - f(x_0) - \frac{1}{2}f''(x_0)h^2 = o(h^2). \] (9.51)

Thus, we can choose \(R_0\) so that for all \(h \in (-R_0, +R_0)\) we have
\[ |r(h)| \leq \frac{1}{10}f''(x_0)h^2. \]

Using this in (9.51) gives
\[ \left| f(x_0 + h) - f(x_0) - \frac{1}{2}f''(x_0)h^2 \right| \leq \frac{1}{10}f''(x_0)h^2, \quad \text{for all } |h| < R_0, \]
and thus
\[ f(x_0 + h) - f(x_0) \geq \frac{1}{2}f''(x_0)h^2 - \frac{1}{10}f''(x_0)h^2 > 0, \quad \text{for all } |h| < R_0, \]
thus \(x_0\) is a local minimum of \(f\) if \(f''(x_0) > 0\). The case \(f'(x_0) < 0\) is essentially identical. □
9.8 A digression on the quadratic forms

A quadratic form is an expression of the form

$$F(\xi) = \sum_{i,j=1}^{n} q_{ij} \xi_i \xi_j.$$ 

Here, $\xi \in \mathbb{R}^n$ is a vector that we think of as an "independent variable", and the $n \times n$ symmetric matrix $Q$ with the entries $q_{ij}$ is fixed.

**Exercise 9.22** Why can we assume that $q_{ij} = q_{ji}$ so that the matrix $Q$ is symmetric?

This is a natural generalization of the quadratic functions in $\mathbb{R}$ that do not have a linear part: in that case, $F(\xi) = q \xi^2$, where $q \in \mathbb{R}$ is a number, and $\xi \in \mathbb{R}$ is an independent variable. Notice that, depending on the sign of the number $q$, we either have $F(\xi) \geq 0$ for all $\xi \in \mathbb{R}$ or $F(\xi) \leq 0$ for all $\xi \in \mathbb{R}$. In higher dimensions $n > 1$, it is also possible that a quadratic form may be positive for all $\xi \in \mathbb{R}^n$, for instance for

$$F(\xi) = a_1 \xi_1^2 + a_2 \xi_2^2 + \cdots + a_n \xi_n^2,$$

with all $a_j \geq 0$, $1 \leq j \leq n$. It may also happen that a quadratic form may be negative for all $\xi \in \mathbb{R}^n$, for instance for

$$F(\xi) = a_1 \xi_1^2 + a_2 \xi_2^2 + \cdots + a_n \xi_n^2,$$

with all $a_j \leq 0$, $1 \leq j \leq n$. It is also possible that a quadratic form may take both positive and negative values: take $n = 2$ and consider

$$F(\xi) = \xi_1^2 - \xi_2^2, \quad \xi = (\xi_1, \xi_2) \in \mathbb{R}^2.$$

Then, for instance $F(1,0) = 1 > 0$ and $F(0,1) = -1 < 0$.

In order to understand what is happening in general, let us write

$$F(\xi) = \sum_{i,j=1}^{n} q_{ij} \xi_i \xi_j = \sum_{j=1}^{n} \left( \sum_{i=1}^{n} q_{ij} \xi_j \right) \xi_i = \sum_{i=1}^{n} v_i \xi_i,$$

(9.52)

with the vector $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$ having the components

$$v_i = \sum_{j=1}^{n} q_{ij} \xi_j, \quad i = 1, \ldots, n.$$ 

Note that $v = Q\xi$ so that we can re-write (9.52) as

$$F(\xi) = (Q\xi \cdot \xi).$$

(9.53)

The matrix $Q$ is an $n \times n$ symmetric matrix. The spectral theorem tells us that it has $n$ eigenvalues $\lambda_1, \ldots, \lambda_n$ taken with multiplicities. We denote the corresponding eigenvectors $v_1, \ldots, v_n$:

$$Qv_k = \lambda_k v_k.$$ 

(9.54)

Note that if $\lambda_k \neq \lambda_j$ then the eigenvectors $v_k$ and $v_j$ are orthogonal. To see that, write

$$Qv_k = \lambda_k v_k \quad Qv_j = \lambda_j v_j,$$

(9.55)
and take the dot product of the first equation with \(v_j\) and of the second with \(v_k\). This gives
\[
(Qv_k \cdot v_j) = \lambda_k (v_k \cdot v_j) \\
(Qv_j \cdot v_k) = \lambda_j (v_j \cdot v_k).
\] (9.56)

Note that as the matrix \(Q\) is symmetric, we have \(Q^T = Q\), so that
\[
(Qv_k \cdot v_j) = (v_k \cdot (Q^Tv_j)) = (v_k \cdot Qv_j) = (Qv_j \cdot v_k).
\]

With this identity in hand, subtracting the second equation in (9.56) from the first gives
\[
(\lambda_k - \lambda_j)(v_k \cdot v_j) = 0.
\]

As \(\lambda_k \neq \lambda_j\), it follows that \((v_k \cdot v_j) = 0\).

**Exercise 9.23** Use the above to show that if \(Q\) is symmetric then there exists an orthonormal basis \(\{v_1, \ldots, v_n\}\) of \(\mathbb{R}^n\) so that each \(v_j\) is an eigenvector of \(Q\).

This basis is extremely useful for the quadratic form \(F(\xi) = (Q\xi \cdot \xi)\). Given a vector \(\xi\), let \(c_1(\xi), \ldots, c_n(\xi)\) be its coordinates in that basis:
\[
\xi = c_1(\xi)v_1 + \cdots + c_n(\xi)v_n.
\]

As the basis \(v_k\) is orthonormal, we have \((v_k \cdot v_j) = 0\) if \(k \neq j\) and \((v_k \cdot v_k) = 1\) for all \(1 \leq k \leq n\), so that
\[
|\xi|^2 = (\xi \cdot \xi) = |c_1(\xi)|^2 + \cdots + |c_n(\xi)|^2.
\] (9.57)

Furthermore, we may write \(Q\xi\) as
\[
Q\xi = c_1(\xi)Qv_1 + \cdots + c_n(\xi)Qv_n = c_1(\xi)\lambda_1 v_1 + \cdots + c_n(\xi)\lambda_n v_n,
\]
thus
\[
F(\xi) = (Q\xi \cdot \xi) = (c_1(\xi)\lambda_1 v_1 + \cdots + c_n(\xi)\lambda_n v_n) \cdot (c_1(\xi)v_1 + \cdots + c_n(\xi)v_n).\] (9.58)

As the basis \(v_k\) is orthonormal, we have \((v_k \cdot v_j) = 0\) if \(k \neq j\) and \((v_k \cdot v_k) = 1\) for all \(1 \leq k \leq n\). Using this in (9.58) gives
\[
F(\xi) = (Q\xi \cdot \xi) = \lambda_1|c_1(\xi)|^2 + \cdots + \lambda_n|c_n(\xi)|^2.
\] (9.59)

It follows that if we set
\[
m = \min_{1 \leq k \leq n} \lambda_k, \quad M = \max_{1 \leq k \leq n} \lambda_k,
\] (9.60)
then
\[
m(|c_1(\xi)|^2 + \cdots + |c_n(\xi)|^2) \leq F(\xi) = (Q\xi \cdot \xi) \leq M(|c_1(\xi)|^2 + \cdots + |c_n(\xi)|^2).
\] (9.61)

With the help of (9.57), this becomes
\[
m\|\xi\|^2 \leq F(\xi) = (Q\xi \cdot \xi) \leq M\|\xi\|^2,
\] (9.62)
with \(m\) and \(M\) as in (9.60). Let us summarize this as the following theorem.
Theorem 9.24 Let $Q$ be an $n \times n$ symmetric matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$ (taken with multiplicities), and set

$$m = \min_{1 \leq k \leq n} \lambda_k, \quad M = \max_{1 \leq k \leq n} \lambda_k,$$

(9.63)

then for any $\xi \in \mathbb{R}^n$ we have

$$m \|\xi\|^2 \leq (Q\xi \cdot \xi) \leq M \|\xi\|^2.$$

(9.64)

Moreover, there exists $\xi \in \mathbb{R}^n$, $\xi \neq 0$, so that $(Q\xi \cdot \xi) = m \|\xi\|^2$, and $\eta \in \mathbb{R}^n$, $\eta \neq 0$, so that $(Q\xi \cdot \xi) = M \|\xi\|^2$.

Definition 9.25 We say that a quadratic form $F(\xi)$ on $\mathbb{R}^n$ is positive definite if there exists $m > 0$ so that for any $\xi \in \mathbb{R}^n$ we have

$$m |\xi|^2 \leq F(\xi).$$

(9.65)

We say that a quadratic form $F(\xi)$ on $\mathbb{R}^n$ is negative definite if there exists $m < 0$ so that for any $\xi \in \mathbb{R}^n$ we have

$$m |\xi|^2 \geq F(\xi).$$

(9.66)

Corollary 9.26 (i) A quadratic form $F(\xi)$ defined on $\mathbb{R}^n$ is positive definite if all eigenvalues of the corresponding symmetric positive definite $n \times n$ matrix $Q$ are positive.

(ii) A quadratic form $F(\xi)$ defined on $\mathbb{R}^n$ is negative definite if all eigenvalues of the corresponding symmetric positive definite $n \times n$ matrix $Q$ are negative.

Another function one can associate to a symmetric $n \times n$ matrix $Q$ is a bilinear form $G(\xi, \eta)$ defined for a pair of vectors $\xi \in \mathbb{R}^n$ and $\eta \in \mathbb{R}^n$ as

$$G(\xi, \eta) = \sum_{i,j=1}^{n} q_{ij} \xi_i \eta_j = \sum_{i=1}^{n} \left( \sum_{j=1}^{n} q_{ij} \xi_j \right) \eta_i = (Q\xi \cdot \eta).$$

(9.67)

Since $Q$ is symmetric, we have the identity

$$(Q\xi \cdot \eta) = \frac{1}{4} \left[ (Q(\xi + \eta) \cdot (\xi + \eta)) - (Q(\xi - \eta) \cdot (\xi - \eta)) \right],$$

which allows us to write

$$G(\xi, \eta) = \frac{1}{4} [F(\xi + \eta) - F(\xi - \eta)],$$

hence the values of the bilinear form are determined by the values of the quadratic form. Note that the corresponding quadratic form is simply $F(\xi) = G(\xi, \xi)$.

Exercise 9.27 Show that the entries of the matrix $Q$ can be expressed in terms of the bilinear form as

$$q_{ij} = \frac{1}{4} [F(e_i + e_j) - F(e_i - e_j)].$$

9.9 The second order Taylor polynomial in $\mathbb{R}^n$

Let us now consider the analog of the Taylor series in $\mathbb{R}^n$. Recall that for an $(m + 1)$-differentiable function $f(x)$ of a single variable we have the Taylor formula

$$f(x_0 + h) = P_m(h) + \frac{f^{(m+1)}(\xi)}{(m+1)!} h^{m+1},$$

where $P_m(h)$ is the $m$th degree polynomial that interpolates $f(x)$ at the points $x_0, x_0 + h, \ldots, x_0 + mh$. Note that the remainder term $R_m(h)$ given by

$$R_m(h) = \frac{f^{(m+1)}(\xi)}{(m+1)!} h^{m+1}$$

is typically small when $h$ is small, and thus can be ignored when $h$ is sufficiently small. This allows us to approximate $f(x_0 + h)$ by the Taylor polynomial $P_m(h)$.

In higher dimensions, the Taylor polynomial $P_m(h)$ generalizes to the Taylor expansion of a function $f(x)$ at a point $x_0$ as

$$f(x_0 + h) = f(x_0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_0) h_i + \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) h_i h_j + \sum_{i_1,\ldots,i_m=1}^{n} \frac{\partial^{m+1} f}{\partial x_{i_1} \cdots \partial x_{i_m}}(x_0) h_{i_1} \cdots h_{i_m} + R_m(h),$$

where $R_m(h)$ is the remainder term. For a symmetric matrix $Q$ and a vector $\xi$, the quadratic form $F(\xi) = (Q\xi \cdot \xi)$ is a special case of this expansion, and the bilinear form $G(\xi, \eta)$ is related to the quadratic form by $G(\xi, \eta) = (Q\xi \cdot \eta)$. The corresponding Taylor expansion for functions of the form $G(\xi, \eta)$ can be derived by substituting the Taylor expansion of $f(x)$ into the definition of $G(\xi, \eta)$ and using the linearity of the gradient operator.

In summary, the Taylor expansion provides a powerful tool for approximating functions and their derivatives in higher dimensions, and the quadratic and bilinear forms provide natural ways of representing the Taylor expansion and its remainder term. These concepts are fundamental in many areas of mathematics and physics, and understanding them is crucial for working with higher-dimensional functions and their approximations.
where $\xi$ is a point between $x_0$ and $x_0 + h$, and the Taylor polynomial is

$$P_m(h) = \sum_{k=0}^{m} \frac{f^{(k)}(x_0)}{k!} h^k.$$

The polynomial $P_m(h)$ was constructed so that for all $0 \leq k \leq m$ we have

$$\frac{d^k}{dh^k} f(x_0 + h) \bigg|_{h=0} = \frac{d^k}{dh^k} P_m(h) \bigg|_{h=0}.$$

We can do a similar procedure for functions on $\mathbb{R}^n$ but we will only do this in detail for the quadratic Taylor polynomials. We are given a function $f : B(x_0, r) \to \mathbb{R}$ that is twice differentiable in a ball $B(x_0, r)$ and such that its second derivative matrix $D^2 f$ is continuous in $B(x_0, r)$. Our goal is to find a vector $b$ and a matrix $Q$ with entries $q_{ij}$ such that we have

$$f(x_0 + h) = f(x_0) + (b \cdot h) + \frac{1}{2}(Qh \cdot h) + o(||h||^2). \quad (9.68)$$

Note that

$$\frac{1}{2}(Qh \cdot h) + o(||h||^2) = o(||h||),$$

hence we must have $b = Df(x_0)$, so that (9.68) becomes

$$f(x_0 + h) = f(x_0) + (Df(x_0) \cdot h) + \frac{1}{2}(Qh \cdot h) + o(||h||^2). \quad (9.69)$$

If we blindly differentiate both sides of (9.68) twice in the directions $e_i$ and $e_j$ and manage to argue that we may pass to the limit $h \to 0$, disregarding the derivative of the $o(||h||^2)$-term in (9.68), we would get that

$$\frac{\partial^2 f(x_0)}{\partial x_i \partial x_j} = q_{ij}, \quad (9.70)$$

that is, the matrix $Q$ is $Q = D^2 f(x_0)$. Now, the question is whether (9.68) does, indeed, hold with $b = Df(x_0)$ and $Q = D^2 f(x_0)$. Let us start with the following lemma.

**Lemma 9.28** Let $f$ be twice differentiable in a ball $B(x_0, r)$, and $D^2 f(x)$ be continuous in $B(x_0, r)$. Assume that $Df(x_0) = 0$ and $D^2 f(x_0) = 0$, then

$$f(x_0 + h) - f(x_0) = o(||h||^2), \quad as \; h \to 0. \quad (9.71)$$

**Proof.** Let us fix $h \in B(x_0, r)$ and consider the function

$$g(t) = f(x_0 + th) - f(x_0), \quad t \in [-1, 1],$$

of a scalar variable $t$. Note that $g(0) = 0$ and, using the definition of the partial derivatives of $f$ and also the partial derivatives of $\partial f / \partial x_j$, we can compute

$$g'(t) = (h \cdot Df(x_0 + th)) = \sum_{j=1}^{n} h_j \frac{\partial f(x_0 + th)}{\partial x_j},$$

$$g''(t) = \sum_{j=1}^{n} h_j \left( h \cdot D \frac{\partial f(x_0 + th)}{\partial x_j} \right) = \sum_{j=1}^{n} h_j \sum_{i=1}^{n} h_i \frac{\partial^2 f(x_0 + th)}{\partial x_i \partial x_j} = \sum_{i,j=1}^{n} h_i h_j \frac{\partial^2 f(x_0 + th)}{\partial x_i \partial x_j}. \quad (9.72)$$
The intermediate value theorem tells us that
\[ g(1) - g(0) = g'(c_1), \] (9.73)
with some \( c_1 \in (0, 1) \). Applying the intermediate value theorem again we conclude that
\[ g'(c_1) = g''(c_2)c_1, \] (9.74)
with some \( c_2 \in (0, c_1) \). Using this in (9.73) gives
\[ g(1) - g(0) = g''(c_2)c_2. \] (9.75)
Let us use the definition of \( g(t) \) and the fact that \( g(0) = 0 \) in (9.75):
\[ f(x_0 + h) = g''(c_2)c_2. \] (9.76)
Now, let us replace \( g''(c_2) \) by the corresponding expression in (9.72):
\[ f(x_0 + h) = \sum_{i,j=1}^{n} h_i h_j \frac{\partial^2 f(x_0 + c_2 h)}{\partial x_i \partial x_j} c_2. \] (9.77)
As \( D^2 f(x_0) = 0 \) and \( D^2 f(x) \) is continuous at \( x_0 \), given any \( \varepsilon > 0 \) we can find \( \delta > 0 \) so that
\[ \left| \frac{\partial^2 f(x_0 + z)}{\partial x_i \partial x_j} \right| < \varepsilon \] for all \( z \in \mathbb{R}^n \) such that \( \| z - x_0 \| < \delta. \) (9.78)
As \( |c_2| < 1 \), as soon as \( |h| < \delta \), we have
\[ |f(x_0 + h)| \leq \sum_{i,j=1}^{n} |h_i h_j| \left| \frac{\partial^2 f(x_0 + c_2 h)}{\partial x_i \partial x_j} \right| \leq \sum_{i,j=1}^{n} \|h\|^2 \varepsilon \leq n^2 \|h\|^2 \varepsilon. \] (9.79)
It follows that \( \|f(x_0 + h)\| = o(\|h\|^2) \), and we are done. □

Now we can prove the second order approximation theorem.

**Theorem 9.29** Let \( f \) be twice differentiable in a ball \( B(x_0, r) \), and \( D^2 f(x) \) be continuous in \( B(x_0, r) \). Assume that \( Df(x_0) = 0 \) and \( D^2 f(x_0) = 0 \), then
\[ f(x_0 + h) = f(x_0) + (Df(x_0) \cdot h) + \frac{1}{2}(D^2 f(x_0)(h \cdot h) + o(\|h\|^2), \text{ as } h \to 0. \] (9.80)

**Proof.** This is a simple consequence of Lemma 9.28. Define the function
\[ \tilde{f}(x) = f(x) - f(x_0) - (Df(x_0) \cdot (x - x_0)) - \frac{1}{2}(D^2 f(x_0)(x - x_0) \cdot (x - x_0)). \]
Note that the \( \tilde{f}(x_0) = 0, D\tilde{f}(x_0) = 0 \) and \( D^2 \tilde{f}(x_0) = 0 \). Moreover, \( D^2 \tilde{f}(x) \) is continuous at \( x_0 \). It follows from Lemma 9.28 that \( \tilde{f}(x) = o(\|x - x_0\|^2) \), which is exactly the claim of the present theorem. □
9.10 The critical points of a function and the second derivative test

Now, we are ready to discuss conditions for a function \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) to attain a local maximum or a local minimum at a point \( x_0 \). Recall that a function \( f \) defined on an open set \( U \) attains a local maximum at a point \( x_0 \in U \) if there is exists \( r > 0 \) so that

\[
f(x) \leq f(x_0) \quad \text{for all } x \in U \text{ such that } |x - x_0| < r.
\]

Similarly, \( f \) attains local minimum at a point \( x_0 \in U \) if there is exists \( r > 0 \) so that

\[
f(x) \geq f(x_0) \quad \text{for all } x \in U \text{ such that } |x - x_0| < r.
\]

Note that if \( f \) is differentiable at a point \( x_0 \) and attains a local maximum or a local minimum at \( x_0 \) then we must have \( D_k f = 0 \) for all \( 1 \leq k \leq n \), which means that \( Df(x_0) = 0 \). To see that, assume that \( r \) attains a local maximum at \( x_0 \), and for a given \( k \) define the function

\[
g(t) = f(x + te_j),
\]

of a real variable \( t \in (-r, r) \). Then \( g(t) \) attains its maximum on the interval \((-r, r)\) at the point \( t = 0 \), hence \( g'(0) = 0 \), which exactly means that \( D_k f(x_0) = 0 \). The question is whether we can determine if a given \( x_0 \) such that \( Df(x_0) = 0 \) is a local maximum, a local minimum or neither. Note that any of the three possibilities are possible: simply look at \( f(x) = x_1^2 + x_2^2 \), \( g(x) = -x_1^2 - x_2^2 \), and \( p(x) = x_1^2 - x_2^2 \) at \( x_0 = (0, 0) \in \mathbb{R}^2 \). All three functions have a gradient equal to zero at \( x_0 \), and \( f \) attains a local minimum at \( x_0 \), \( g \) attains a local maximum at \( x_0 \) but for \( p(x) \) the point \( x_0 \) is neither a local maximum nor a local minimum.

For the rest of this discussion, see Section 7 in Chapter 2 of the Simon book, and Sections 8.4.3.-8.4.5 in Zorich.

10 A crash course on integrals

10.1 Extension of a continuous function from a dense subset

Before we turn to the integral, let us consider the following question. Let \( X \) be a metric space, and \( S \) be a dense subset of \( X \). Suppose we have a real-valued continuous function \( f \) that is defined on \( S \), \( f : S \rightarrow \mathbb{R} \), the question is if we can extend \( f \) to all of \( X \) so that the extension is continuous. In other words, can we construct a function \( \bar{f} : X \rightarrow \mathbb{R} \) such that \( \bar{f} \) is continuous on \( X \) and for \( s \in S \) we have \( \bar{f}(s) = f(s) \)?

First, let us find a candidate for what \( \bar{f} \) should be. Given \( x \in X \), since \( S \) is a dense subset of \( X \), there exists a sequence \( s_n \rightarrow x \). If \( \bar{f} \) is continuous on \( X \), then we must have \( \bar{f}(s_n) \rightarrow \bar{f}(x) \) as \( n \rightarrow \infty \). Since \( \bar{f} = f \) on \( S \), we should have \( \bar{f}(s_n) = f(s_n) \). This brings the following possibility: let us define \( \bar{f}(x) \) for \( x \notin S \) as follows – take a sequence \( s_n \rightarrow x \) and set

\[
\bar{f}(x) = \lim_{n \rightarrow \infty} f(s_n).
\]

This brings two immediate questions: (i) does the limit in (10.1) exist, and (ii) if the limit exists, will it be the same for all sequences \( s_n \) that converge to a given point \( x \notin S \)?

Let us make a stronger assumption than that \( f \) is continuous on \( S \): let us assume that \( f \) is uniformly continuous on \( S \), so that for any \( \varepsilon > 0 \) there exists \( \delta_\varepsilon > 0 \) such that for any \( s_1, s_2 \in S \) with \( d(s_1, s_2) < \delta_\varepsilon \) we have \( |f(s_1) - f(s_2)| < \varepsilon \). Now, as the sequence \( s_n \) converges to \( x \), this sequence is Cauchy, hence for any \( \varepsilon > 0 \) there exists \( N_\varepsilon \) so that \( d(s_n, s_m) < \delta_\varepsilon \), with \( \delta_\varepsilon \) as above. Then we
have \(|f(s_n) - f(s_m)| < \varepsilon\) for all \(n > N_\varepsilon\). It follows that the sequence \(f(s_n)\) is Cauchy, hence the limit in \((10.1)\) exists. This answers (i). Furthermore, (ii) has also been answered — indeed, take another sequence \(s'_n\) that converges to \(x\) and consider the alternating sequence \(s_1, s'_1, s_2, s'_2, \ldots\), that is \(s_k'' = s_k\) if \(k\) is even and \(s_k'' = s'_k\) if \(k\) is odd. The sequence \(s_k''\) also converges to \(x\), hence, by what we have just shown, the limit \(f(s_k'')\) exists. But then the limits of \(f(s_k)\) and \(f(s'_k)\) must coincide.

**Exercise 10.1** Fill in the details in proof of the claim that

\[
\lim_{n \to \infty} f(s_n) = \lim_{n \to \infty} f(s'_n)
\]

above.

Now we know that \(\bar{f}(x)\) is well-defined for \(x \notin S\). It remains to show that \(\bar{f}(x)\) is a continuous on \(X\). Since \(f\) is uniformly continuous on \(S\), given \(\varepsilon > 0\), we can choose \(\delta > 0\) so that for all \(s_1, s_2 \in S\) such that \(d(s_1, s_2) < \delta\) we have \(|f(s_1) - f(s_2)| < \varepsilon/4\). Let us take \(x, y \in X\) such that \(d(x, y) < \delta/4\). As \(S\) is dense in \(X\), we can find two sequences \(s_n, s'_n \in S\) such that

\[
\lim_{n \to \infty} s_n = x, \quad \lim_{n \to \infty} s'_n = y,
\]

and then

\[
\bar{f}(x) = \lim_{n \to \infty} f(s_n), \quad \bar{f}(y) = \lim_{n \to \infty} f(s'_n).
\]

Let us choose \(N\) sufficiently large, so that all of the following hold:

\[
d(s_N, x) < \delta/4, \quad d(s'_N, y) < \delta/4, \quad |\bar{f}(x) - f(s_N)| < \varepsilon/4, \quad |\bar{f}(y) - f(s'_N)| < \varepsilon/4.
\]

\[(10.2)\]

The first two inequalities in \((10.2)\) imply that

\[
d(s_N, s'_N) < d(s_N, x) + d(x, y) + d(y, s'_N) < \frac{\delta}{4} + \frac{\delta}{4} + \frac{\delta}{4} < \delta,
\]

thus the way we have chosen \(\delta\) implies that \(|f(s_N) - f(s'_N)| < \varepsilon/4\). Now, the last two inequalities in \((10.2)\) imply that

\[
|\bar{f}(x) - \bar{f}(y)| \leq |\bar{f}(x) - f(s_N)| + |f(s_N) - f(s'_N)| + |\bar{f}(y) - f(s'_N)| < \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} < \varepsilon.
\]

It follows that the function \(\bar{f}\) is uniformly continuous on \(X\). Thus, we have proved the following theorem.

**Theorem 10.2** Let \(X\) be a metric space and \(S\) its dense subset. Suppose that a function \(f : S \to \mathbb{R}\) is uniformly continuous on \(S\). Then there exists a function \(\bar{f} : X \to \mathbb{R}\) that is uniformly continuous on \(X\) and such that \(\bar{f}(s) = f(s)\) for all \(s \in S\).

### 10.2 The integral of a continuous function

We now give a brief and short-cutty way to define the integral of a continuous function on an interval \([a, b]\) using the above strategy of extension. There is a class of functions on which one "knows" what the integral should be if we think of the integral as the area under the graph of a function. Namely, if \(f : [a, b] \to \mathbb{R}\) is an affine function, that is,

\[
f(t) = \alpha t + \beta
\]

with some \(\alpha, \beta \in \mathbb{R}\), then by the area of a triangle formula we should have

\[
\int_a^b f(x) \, dx = \left( \frac{\alpha a + \beta}{2} + \beta \right)(b - a).
\]

\[(10.3)\]

Here, \((b - a)\beta\) is the area of the rectangle under the graph, and \(\alpha(b - a)(a + b)/2\) is the area of the triangle – draw a picture to clarify this!
The integral of a piecewise affine continuous function

The next clear property the integral should have is that the integral is additive with respect to "interval splitting": for \(a<b<c\) we should have

\[
\int_a^c f(x)dx = \int_a^b f(x)dx + \int_b^c f(x)dx. \tag{10.4}
\]

This tells us what the integral must be for piecewise affine functions. A function is piecewise affine if there exists a partition \(a=t_0<t_1<\ldots<t_n=b\) of the interval \([a,b]\) so that \(f(t) = \alpha_i t + \beta_i\) on each interval \(t \in [t_{i-1}, t_i]\), with \(\alpha_i, \beta_i \in \mathbb{R}\). For piecewise affine functions we should have

\[
\int_a^b f(x)dx = \sum_{i=1}^n \left( \alpha_i \frac{t_i + t_{i-1}}{2} + \beta_i \right)(t_i - t_{i-1}). \tag{10.5}
\]

Note that we require that the piecewise affine functions are continuous: the values at \(t_i\) on the left and on the right must match. Our goal is to extend the notion of the integral from piecewise affine to all continuous functions.

It is not completely obvious that the above definition of the integral makes sense even for piecewise affine functions. The issue is that if \(f\) is piecewise affine, there may be many ways of choosing the partition points \(t_i\). For instance, if one such partition works, one could always refine it by adding more division points. One then has to show that the result is independent of the particular choice of division points, subject to \(f\) being affine on each subinterval.

Exercise 10.3 Show that if \(f\) is a piecewise affine function then the integral defined by the right side of (10.5) does not depend on the choice of the partition. Hint: given two such partitions, take their common refinement, consisting of all the division points in either, and then show that the integrals defined using the original two partitions are equal to that defined using their common refinement, and thus to each other.

Once one knows that the integral (10.5) is well-defined on the set \(D = D([a,b])\) of piecewise affine functions, one can show the following basic properties of the integral for piecewise affine functions.

Proposition 10.4 (i) Linearity: let \(f_1, f_2 \in D([a,b])\) and \(c_1, c_2 \in \mathbb{R}\), then

\[
\int_a^b (c_1 f_1(x) + c_2 f_2(x))dx = c_1 \int_a^b f_1(x)dx + c_2 \int_a^b f_2(x)dx.
\]

(ii) Positivity: if \(f \in D([a,b])\), and \(f(x) \geq 0\) for all \(x \in [a,b]\), then

\[
\int_a^b f(x)dx \geq 0.
\]

(iii) Boundedness: if \(f \in D([a,b])\) then

\[
\left| \int_a^b f(x)dx \right| \leq \int_a^b |f(x)|dx \leq (b-a)\|f\|, \tag{10.6}
\]

where \(\|f\|\) is the norm on \(C(a,b)\):

\[
\|f\| = \sup\{|f(x)| : x \in [a,b]\}.
\]

(iv) Additivity: if \(a<b<c\) and \(f \in D([a,c])\) then

\[
\int_a^b f(x)dx + \int_b^c f(x)dx = \int_a^c f(x)dx.
\]
Consider now a partition $a \leq t_1 < t_2 < \cdots < t_n = b$. Therefore, given $n \in \mathbb{N}$ and $f \in C[a,b]$, there is a sequence of piecewise affine functions $f_n$. Consider a piecewise affine function $f$ of the interval $[a,b]$ where the first equality uses the linearity of the integral and the second the positivity of the integral of $f - g \geq 0$.

**Approximation of continuous functions by piecewise affine functions**

Next, we will extend the definition of the integral from piecewise affine functions on $[a,b]$ to all continuous functions on $[a,b]$. First, we need to show that a continuous function can be approximated by a sequence of piecewise affine functions.

**Lemma 10.6** Let $f \in C[a,b]$ be a continuous function. There is a sequence of piecewise affine functions $f_n \in C[a,b]$ such that $f_n \to f$ in $C[a,b]$.

**Proof.** As $f$ is continuous on a closed interval $[a,b]$, $f$ is uniformly continuous on that interval. Therefore, given $n \in \mathbb{N}$ we can find $\delta_n$ so that $|f(x) - f(y)| < 1/n$ for all $x, y$ such that $|x - y| < \delta_n$. Consider now a partition $a = t_0 < t_1 < \cdots < t_k = b$ of the interval $[a,b]$ such that $|t_j - t_{j-1}| < \delta_n$ for all $1 \leq j \leq k$. Consider a piecewise affine function $f_n$ so that $f_n(t)$ is linear on each interval $[t_{j-1}, t_j]$ and $f_n(t_j) = f(t_j)$ for all $0 \leq j \leq k$. Draw a picture to visualize $f$! Note that $f_n$ depends on $n$ through the partition. Then, for each $t \in [a,b]$, we can find $1 \leq j \leq k$ so that $t \in [t_{j-1}, t_j]$. Then, we have

$$|f_n(t) - f(t)| \leq |f_n(t) - f_n(t_j)| + |f_n(t_j) - f(t_j)| + |f(t_j) - f(t)|. \quad (10.7)$$

The basic properties of linear functions imply that the first term in (10.7) can be estimated as

$$|f_n(t) - f_n(t_j)| \leq |f_n(t_{j-1} - f_n(t_j))| = |f(t_{j-1}) - f(t_j)| \leq \frac{1}{n}, \quad (10.8)$$

since $|t_{j-1} - t_j| < \delta_n$ for all $j$. The second term in the right side of (10.7) vanishes since $f_n(t_j) = f(t_j)$, and the last one can be bounded by

$$|f(t_j) - f(t)| < \frac{1}{n}, \quad (10.9)$$

because $|t_j - t| < \delta_n$. It follows that for all $t \in [a,b]$ we have

$$|f_n(t) - f(t)| \leq \frac{2}{n}, \quad (10.10)$$

hence $\|f_n - f\| \to 0$ as $n \to +\infty$, thus $f_n \to f$ in $C[a,b]$. □
Defining the integral of a continuous function

Suppose \( f \in C([a,b]) \), and let \( f_n \in D[a,b] \) be a sequence such that \( f_n \to f \), in \( C[a,b] \). Note that by Lemma 10.6 at least one such sequence exists but of course it is not unique. We would like to show, first, that

\[
\lim_{n \to \infty} \int_a^b f_n(x)\,dx
\]

exists and, second, that it is independent of the choice of the particular sequence in \( D[a,b] \) that converges to \( f \). If we show both of these properties, then we can define

\[
\int_a^b f(x)\,dx = \lim_{n \to \infty} \int_a^b f_n(x)\,dx. \tag{10.11}
\]

**Lemma 10.7** Let \( f_n \in D[a,b] \) be a sequence of piecewise affine functions that converges to a function \( f \in C[a,b] \), then the limit

\[
\lim_{n \to \infty} \int_a^b f_n(x)\,dx
\]

exists.

**Proof.** As the sequence \( f_n \) converges in \( C[a,b] \), it is a Cauchy sequence in \( C[a,b] \). Thus, for all \( \varepsilon > 0 \) there is \( N \) such that \( n, m \geq N \) imply \( ||f_n - f_m|| < \varepsilon/(b-a) \). The linearity property (i) and the boundedness property (iii) in Proposition 10.4 imply that

\[
\left| \int_a^b f_n - \int_a^b f_m \right| = \left| \int_a^b (f_n - f_m) \right| \leq (b-a)||f_n - f_m|| < \varepsilon.
\]

It follows that the sequence of numbers

\[
I_n = \int_a^b f_n(x)\,dx
\]

is a Cauchy sequence in \( \mathbb{R} \), thus it converges. \( \square \)

**Lemma 10.8** Let \( f_n \in D[a,b] \) and \( g_n \in D[a,b] \) be two sequences of piecewise affine functions in \( D[a,b] \) that both converge to the same function \( f \in C[a,b] \), then

\[
\lim_{n \to \infty} \int_a^b f_n(x)\,dx = \lim_{n \to \infty} \int_a^b g_n(x)\,dx. \tag{10.12}
\]

**Proof.** Consider the alternating sequence \( f_1, g_1, f_2, g_2, f_3, g_3, \ldots \), that is, the sequence \( h_n \) with

\[
h_{2k-1} = f_k, \quad h_{2k} = g_k.
\]

This sequence also converges to \( f \), as follows easily from the definition of convergence. Thus, Lemma 10.7 implies that

\[
\lim_{n \to \infty} \int_a^b h_n(x)\,dx
\]

exists, but then all subsequences of

\[
\int_a^b h_n(x)\,dx
\]

converge to this very same limit, and (10.12) follows. \( \square \)
Putting these together, we can define
\[ \int_a^b f \]
for \( f \in C([a, b]) \) as follows: take any sequence \( f_n \) in \( D[a, b] \) such that \( f_n \to f \) in \( C[a, b] \), and set
\[ \int_a^b f = \lim_{n \to \infty} \int_a^b f_n. \] (10.13)

This argument remains valid if the the target space \( \mathbb{R} \) of \( f \) is replaced by any complete normed vector space \( V \), such as \( \mathbb{R}^m \). Complete normed vector spaces are called Banach spaces.

It is straightforward to check that the integral inherits the properties from \( D[a, b] \) in Proposition 10.4.

**Proposition 10.9** (i) **Linearity:** let \( f_1, f_2 \in C([a, b]) \) and \( c_1, c_2 \in \mathbb{R} \), then
\[ \int_a^b (c_1 f_1(x) + c_2 f_2(x))dx = c_1 \int_a^b f_1(x)dx + c_2 \int_a^b f_2(x)dx. \] (10.14)
(ii) **Positivity:** if \( f \in C([a, b]) \), and \( f(x) \geq 0 \) for all \( x \in [a, b] \), then
\[ \int_a^b f(x)dx \geq 0. \] (10.15)
(iii) **Boundedness:** if \( f \in C([a, b]) \) then
\[ \left| \int_a^b f(x)dx \right| \leq \int_a^b |f(x)|dx \leq (b - a)\|f\|, \] (10.16)
where \( \|f\| \) is the norm on \( C(a, b) \): \( \|f\| = \sup\{|f(x)|: x \in [a, b]\} \).
(iv) **Additivity:** if \( a < b < c \) and \( f \in C([a, c]) \) then
\[ \int_a^b f(x)dx + \int_b^c f(x)dx = \int_a^c f(x)dx. \] (10.17)

**Proof.** Linearity follows from the linearity of the limit and of the integral on \( D[a, b] \): if \( h_n \to f_1 \) and \( g_n \to f_2 \), with \( h_n, g_n \in D[a, b] \), then
\[
\begin{align*}
&c_1 \int_a^b f_1(x)dx + c_2 \int_a^b f_2(x)dx = c_1 \lim_{n \to \infty} \int_a^b h_n(x)dx + c_2 \lim_{n \to \infty} \int_a^b g_n(x)dx \\
&= \lim_{n \to \infty} \left( c_1 \int_a^b h_n(x)dx + c_2 \int_a^b g_n(x)dx \right) = \lim_{n \to \infty} \int_a^b (c_1 h_n(x) + c_2 g_n(x))dx \\
&= \int_a^b (c_1 f_1(x) + c_2 f_2(x))dx.
\end{align*}
\]
The first equality uses the definition of the integral, the second is the linearity of the limit in \( \mathbb{R} \), the third is the linearity of the integral on \( D[a, b] \), and the fourth uses the linearity of the limit in the vector space \( C([a, b]) \), so that \( c_1 h_n + c_2 g_n \) converges to \( c_1 f_1 + c_2 f_2 \) in \( C[a, b] \). The proof of additivity is completely similar.

To show the boundedness property, note that if \( f_n \to f \) in \( C[a, b] \) with \( f_n \in D[a, b] \), then also \( |f_n| \to |f| \) in \( C[a, b] \). This is because
\[ ||f_n(x)| - |f(x)|| \leq |f_n(x) - f(x)|, \] (10.18)
so that if the right side converges uniformly on \([a, b]\) to zero as \( n \to \infty \), then so does the left side.
Exercise 10.10 Check this!

Now, the first inequality in (10.16) is easy: by boundedness in $D$, we have

$$\left| \int_a^b f_n(x)dx \right| \leq \int_a^b |f_n(x)|dx,$$  \hspace{1cm} (10.19)

and the left side in (10.19) converges to

$$\left| \int_a^b f_n(x)dx \right| \to \left| \int_a^b f(x)dx \right|$$  \hspace{1cm} (10.20)

by the definition of the integral. As $|f_n| \to |f|$ in $C[a, b]$ and $|f_n| \in D[a, b]$, the right side of (10.19) converges to

$$\int_a^b |f(x)|dx,$$  \hspace{1cm} (10.21)

and the first inequality in (10.16) now follows from (10.16), (10.19) and (10.20).

To show the second inequality in (10.16), let $\varepsilon > 0$. Then there exists $N$ such that

$$\sup_{x \in [a,b]} |f_n(x)| \leq \sup_{x \in [a,b]} |f(x)| + \varepsilon,$$

for all $n \geq N$, and $|f_n| \in D[a, b]$, so the second inequality in (10.6), the boundedness on $D[a, b]$, shows that

$$\int_a^b |f_n(x)|dx \leq (b - a) \sup_{x \in [a,b]} |f_n(x)| \leq (b - a)(\sup |f| + \varepsilon)$$

for $n \geq N$. The integral in the left side above converges to

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

while the right side does not depend on $n$, hence

$$\int_a^b |f(x)|dx \leq (b - a)(\sup |f| + \varepsilon).$$  \hspace{1cm} (10.22)

As $\varepsilon > 0$ was arbitrary, this shows that

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

as claimed.

Finally positivity follows the same way as the bound we just observed: if $f \geq 0$ and $f_n \to f$, then for $\varepsilon > 0$ there exists $N$ such that for $n \geq N$, $f_n > -\varepsilon$, and thus

$$\int_a^b f_n(x)dx \geq -\varepsilon \int_a^b 1 \cdot dx = -(b - a)\varepsilon.$$

Note that $1 \in D$, and

$$\int_a^b 1 \cdot dx = b - a,$$
from the definition. As a consequence, we have
\[ \int_a^b f(x)dx \geq -(b-a)\varepsilon. \]
Since \( \varepsilon > 0 \) was arbitrary, we deduce that
\[ \int_a^b f(x)dx \geq 0, \]
finishing the proof. \( \square \)

In summary, we have shown the existence part of the following theorem.

**Theorem 10.11** Given \( a < b \), there exists a unique linear map \( I_a^b : C[a,b] \to \mathbb{R} \) such that for all \( f \in D[a,b] \) we have
\[ I_a^b(f) = \int_a^b f(x)dx, \] (10.23)
and such that there is a constant \( K_0 \) such that
\[ |I_a^b f| \leq K_0\|f\|, \ f \in C([a,b]). \] (10.24)
Furthermore, this unique linear map satisfies properties (10.14)-(10.17).

One writes
\[ \int_a^b f(x)dx = I_a^b(f). \]

**Proof.** As mentioned, we have already proved existence. To show uniqueness, assume that \( I_a^b \) is a linear map from \( C[a,b] \) to \( \mathbb{R} \) that satisfies (10.24), and has the property that for \( f_n \in D \) we have (10.23). Take any \( f \in C[a,b] \) and a sequence \( f_n \in D[a,b] \) such that \( f_n \to f \) as \( n \to \infty \). Then, by linearity of \( I_a^b \), we have
\[ I_a^b(f) - I_a^b(f_n) = I_a^b(f - f_n). \]
The estimate (10.24) implies then that
\[ |I_a^b(f) - I_a^b(f_n)| \leq K_0\|f - f_n\| \to 0, \]
as \( n \to +\infty \). We conclude that
\[ I_a^b f = \lim \int_a^b f_n, \]
thus \( I_a^b \) is uniquely determined by its values on \( D[a,b] \). \( \square \)

**10.3 The fundamental theorem of calculus**

We now prove the Newton-Leibniz theorem, also known as the fundamental theorem of calculus. For this purpose it is convenient to define
\[ \int_a^a f = 0. \]
The first part of the theorem says that the integral gives an antiderivative.
**Theorem 10.12**  *(The fundamental theorem of calculus, part I)* Suppose \( f : [a, b] \to \mathbb{R} \) is continuous, and let

\[
F(t) = \int_a^t f(x) \, dx, \quad t \in [a, b].
\]

Then the function \( F(x) \) is continuously differentiable on \((a, b)\), and \( F'(x) = f(x) \) for all \( x \in (a, b) \). In addition, the right derivative of \( F \) at \( x = a \) and the left derivative of \( F \) at \( b \) exist and are \( F'_r(a) = f(a) \) and \( F'_l(b) = f(b) \).

**Proof:** It suffices to show that \( F \) is differentiable with \( F' = f \), since the continuity of \( F' \) then follows from that of \( f \). Let \( t \in [a, b] \) be fixed and consider \( h > 0 \), the case of \( t \in (a, b) \) and \( h < 0 \) being identical similar. Then, for \( h < b - t \), we have, using the basic properties of the integral:

\[
F(t + h) - F(h) = \int_a^{t+h} f(x) \, dx - \int_a^t f(x) \, dx = \int_t^{t+h} f(x) \, dx
\]

so that

\[
F(t + h) - F(h) = h f(t) + \int_t^{t+h} (f(x) - f(t)) \, dx,
\]

and

\[
F(t + h) - F(h) - h f(t) = \int_t^{t+h} (f(x) - f(t)) \, dx.
\]

**Exercise 10.13** Check carefully which properties of the integral we are using in each step in (10.25).

Using the boundedness property (10.16) of the integral, we deduce from (10.26) that

\[
|F(t + h) - F(t) - h f(t)| \leq h \sup_{t \leq x \leq t+h} |f(x) - f(t)|.
\]

By the continuity of \( f \) at \( t \), given \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that \( |x - t| < \delta \) implies \( |f(x) - f(t)| < \varepsilon \). Thus, for \( 0 < h < \min(\delta, b - t) \) we have

\[
|F(t + h) - F(t) - h f(t)| \leq \varepsilon h.
\]

Together with the analogous argument for \( h < 0 \), this is exactly the definition of differentiability at \( t \), with \( F'(t) = f(t) \), proving the theorem. \( \square \)

In order to prove the second part of the fundamental theorem of calculus, we need an observation.

**Proposition 10.14**  *If \( f \in C([a, b]) \) is differentiable on \((a, b)\) and \( f'(x) = 0 \) for all \( x \in (a, b) \) then \( f \) is constant: \( f(x) = f(a) \) for all \( x \in [a, b] \).*

**Proof.** Let \( a < x_1 < x_2 < b \), then by the mean value theorem on \([x_1, x_2]\), there is \( c \in (x_1, x_2) \) such that

\[
f'(c) = \frac{f(x_2) - f(x_1)}{x_2 - x_1}.
\]

As, by assumption, we have \( f'(c) = 0 \), we conclude that \( f(x_1) = f(x_2) \). Thus, for all \( x \in [a, b] \), we have \( f(x) = f(a) \), so that \( f \) is a constant function. \( \square \)

**Theorem 10.15**  *(Fundamental theorem of calculus, part II)* If \( f \in C^1([a, b]) \) then

\[
f(b) - f(a) = \int_a^b f'(x) \, dx.
\]
This is the part of the theorem that is actually used to evaluate integrals explicitly: to find the integral
\[ \int_a^b g(x) dx, \]
for given a function \( g \in C([a,b]) \), one looks for \( f \in C^1([a,b]) \) such that \( f' = g \), and then
\[ \int_a^b g(x) dx = f(b) - f(a). \]

Note that part I of the fundamental theorem says that the indefinite integral of \( g \) is such a function, but does not give an explicit evaluation – you need a different method of finding the antiderivative for the explicit calculation. For instance, if \( g(t) = t^n \), then you may check that \( f(t) = t^{n+1}/(n+1) \) satisfies \( f'(t) = g(t) \) using the product rule for differentiation, and then apply the second part of the fundamental theorem of calculus to find the integral of \( g(x) \).

**Proof.** Let us set
\[ F(t) = \int_a^t f'(x) dx. \]

By the first part of the fundamental theorem of calculus proven above, \( F \) is continuously differentiable with \( F'(t) = f'(t) \) for all \( t \in [a,b] \). Thus, the function \( g = f - F \) satisfies \( g \in C^1([a,b]) \) and \( g'(t) = 0 \) for all \( t \in [a,b] \). Proposition 10.14 implies that \( g(b) = g(a) \), so that
\[ f(b) - F(b) = f(a) - F(a), \]
which is exactly
\[ f(b) - f(a) = F(b) - F(a) = \int_a^b f'(x) dx, \]
completing the proof. □

**A comment on the integration by parts**

Note that integration by parts is a simple consequence of the fundamental theorem of calculus and the product rule for differentiation:
\[ (fg)' = f'g + fg'. \]

Indeed, for \( f, g \in C^1([a,b]) \), we have
\[
\begin{align*}
(fg)(b) - (fg)(a) &= \int_a^b (f(x)g(x))' dx = \int_a^b (f'(x)g(x) + f(x)g'(x)) dx \\
&= \int_a^b f'(x)g(x) dx + \int_a^b f(x)g'(x) dx,
\end{align*}
\]
and now a simple rearrangement gives the integration by parts formula:
\[
\int_a^b f(x)g'(x) dx = fg|_a^b - \int_a^b f(x)g'(x) dx, \quad fg|_a^b = f(b)g(b) - f(a)g(a).
\]

Consider, for example, the integral
\[ \int_a^b xe^x dx. \]
If we know that \((e^x)' = e^x\), but cannot guess the anti-derivative of \(x e^x\), we can integrate it by parts as follows:

\[
\int_a^b x e^x \, dx = \int_a^b x (e^x)' \, dx = x e^x \big|_a^b - \int_a^b 1 \cdot e^x \, dx = be^b - ae^a - e^b + e^a.
\]

Essentially, this discovers the fact that

\[
x e^x = \frac{d}{dx} \left[ (x - 1)e^x \right].
\]

Or we can use the fact that \((\log x)' = 1/x\), to compute, for \(0 < a < b\):

\[
\int_a^b x \log x \, dx = \int_a^b (\log x) \left(\frac{x^2}{2}\right)' \, dx = \frac{b^2}{2} \log b - \frac{a^2}{2} \log a - \int_a^b \frac{1}{x} \cdot \frac{x^2}{2} \, dx
\]

\[
= \frac{b^2}{2} \log b - \frac{a^2}{2} \log a - \frac{b^2}{4} + \frac{a^2}{4}.
\]

It may also help to understand convergence of indefinite integrals. Recall that if a function \(f(x)\) is continuous on \([a, +\infty)\), then we set

\[
\int_a^\infty f(x) \, dx = \lim_{R \to \infty} \int_a^R f(x) \, dx,
\]

provided that the limit in the right side exists. Let us look at

\[
\int_1^\infty \frac{\sin x}{x^p} \, dx.
\]

Note that if \(p > 1\), then we have

\[
\left| \int_{R_1}^{R_2} \frac{\sin x}{x^p} \, dx \right| \leq \int_{R_1}^{R_2} \left| \frac{\sin x}{x^p} \right| \, dx \leq \int_{R_1}^{R_2} \frac{1}{x^p} \, dx = \frac{1}{p + 1} \left( \frac{1}{R_1^{p+1}} - \frac{1}{R_2^{p+1}} \right) \leq \frac{1}{p + 1} \frac{1}{R_1^{p+1}},
\]

thus the limit

\[
\lim_{R \to \infty} \int_1^R \frac{\sin x}{x^p} \, dx
\]

exists, and the integral

\[
\int_1^\infty \frac{\sin x}{x^p} \, dx
\]

is well-defined.

**Exercise 10.16** Make the above argument precise: use (10.28) to show that if \(p > 1\) then for any sequence \(R_n \to +\infty\) the limit

\[
\lim_{n \to +\infty} \int_1^{R_n} \frac{\sin x}{x^p} \, dx
\]

exists and does not depend on the choice of the sequence \(R_n \to +\infty\). Hint: show that the sequence

\[
I_n = \int_1^{R_n} \frac{\sin x}{x^p} \, dx
\]

is Cauchy.

The question we would like to answer is whether the integral converges for \(p \in (0, 1)\). Note that the integral has to diverge for \(p \leq 0\) since the integrand does not tend to zero as \(n \to +\infty\).
Exercise 10.17 Let the function $f$ be continuous on $[a, +\infty)$, and assume that the integral
\[ \int_{a}^{\infty} f(x) \, dx \]
exists, that is, the limit in the right side of (10.27) exists. Show that then
\[ \lim_{x \to +\infty} f(x) = 0. \]

Let us then consider the case $0 < p < 1$. We have to use some cancellation because of the sign changes in $\sin x$, because of the following.

Exercise 10.18 Show that the integral
\[ \int_{1}^{\infty} \frac{|\sin x|}{x^p} \, dx \]
does not converge for $p \in (0, 1)$.

Consider next
\[ I(R) = \int_{1}^{R} \frac{\sin x}{x^p} \, dx, \]
with the idea to let $R \to +\infty$ eventually. Note that $\sin x = (-\cos x)'$, $(1/x^p)' = -p/x^{p+1}$, hence we can integrate by parts as follows:
\[ I(R) = \int_{1}^{R} \frac{\sin x}{x^p} \, dx = \int_{1}^{R} \frac{(-\cos x)'}{x^p} \, dx = -\frac{\cos R}{R^p} + \cos 1 - \int_{1}^{R} \left( \frac{1}{x^p} \right)' (-\cos x) \, dx \]
\[ = -\frac{\cos R}{R^p} + \cos 1 - p \int_{1}^{R} \frac{\cos x}{x^{p+1}} \, dx. \tag{10.29} \]

As in Exercise 10.16, the integral
\[ \int_{1}^{\infty} \frac{\cos x}{x^m} \, dx = \lim_{R \to +\infty} \int_{1}^{R} \frac{\cos x}{x^m} \, dx \]
exists for all $m > 1$, thus the limit
\[ \lim_{R \to +\infty} \int_{1}^{R} \frac{\cos x}{x^{p+1}} \, dx \]
exists for all $p > 0$ and equals
\[ \int_{1}^{\infty} \frac{\cos x}{x^{p+1}} \, dx. \]

Going back to (10.29) and passing to the limit $R \to +\infty$, we conclude that the integral
\[ \int_{1}^{\infty} \frac{\sin x}{x^p} \, dx = \cos 1 - p \int_{1}^{\infty} \frac{\cos x}{x^{p+1}} \, dx \]
also exists. Note how integration by parts allowed us to increase the power of $x$ in the denominator, and reduce a problem that involves the integral of a function that is decays slowly so the integral is not absolutely convergent (this terminology is similar to the absolute convergence to a series) to an integrand that decays faster by one power of $x$ and the integral converges absolutely.
11 The contraction mapping principle

The contraction mapping principle is the most basic tool that can be used to prove existence of solutions in many situations in analysis. Many problems can be formulated in the form

\[ f(x) = y_0 , \tag{11.1} \]

where \( y_0 \) is an element of some metric space \( X \), \( f \) is a mapping from \( X \) to \( X \), and \( x \) is the unknown that we need to find. We can reformulate it as

\[ f(x) + x - y_0 = x. \]

The advantage of the latter formulation is that now we have what is known as a fixed point problem. These are equations of the form

\[ F(x) = x , \tag{11.2} \]

where \( F \) is a mapping from a metric space \( X \) to itself, and \( x \) is an unknown point \( x \in X \). A solution of (11.2) is known as a fixed point of the mapping \( F \). In other words, (11.1) is equivalent to (11.2) with

\[ F(x) = f(x) + x - y_0 . \tag{11.3} \]

Of course, there is a serious difference: the notion of a fixed point as a solution to (11.2) requires only that \( F \) maps \( X \) to itself, and the notion of a solution of (11.1) also requires only that \( f \) maps \( X \) to itself and \( y_0 \in X \). However, to pass from (11.1) to (11.2) we need to introduce the mapping \( F \) in (11.3) which requires an addition structure on \( X \): we need \( X \) to be a vector space to be able to do that. However, often \( X \) is a vector space, so that issue is not a problem.

An important class of mappings of a metric space \( X \) onto itself are contractions. We say that a mapping \( f : X \to X \) is a contraction if there exists a number \( q \in (0,1) \) so that for an \( x_1, x_2 \in X \) we have

\[ d(f(x_1), f(x_2)) \leq q d(x_1, x_2) . \tag{11.4} \]

We also need the following definition: a metric space \( X \) is complete if any Cauchy sequence in \( X \) converges. The spaces \( \mathbb{R}^n \) are complete. Before we proceed further with the contraction mapping theorem, let us show that the space \( C[0,1] \) with the norm

\[ \| f \| = \sup_{0 \leq x \leq 1} |f(x)| \]

is a complete metric space.

**Theorem 11.1** The metric space \( C[0,1] \) is complete.

**Proof.** Let \( f_n \) be a Cauchy sequence in \( C[0,1] \). This means that for any \( \varepsilon > 0 \) there exists \( N \) so that for any \( n, m \geq N \) we have

\[ \| f_n - f_m \| < \varepsilon . \]

In other words, we have

\[ \sup_{0 \leq x \leq 1} |f_n(x) - f_m(x)| < \varepsilon , \]

so that

\[ |f_n(x) - f_m(x)| < \varepsilon , \]

for all \( n, m \geq N \) and all \( x \in [0,1] \). This means that the sequence \( f_n \) is uniformly Cauchy on \([0,1]\). Recall that Theorem 8.16 implies that then \( f_n \) is uniformly convergent, and by Theorem 8.17 the limit is a continuous function. Thus, the space \( C[0,1] \) is a complete metric space. □
Theorem 11.2 Let $X$ be a complete metric space, and $f : X \rightarrow X$ be a contraction, then $f$ has a unique fixed point $a$ in $X$. Moreover, for any $x_0 \in X$, the sequence defined recursively by $x_{k+1} = f(x_k)$, with $x_1 = f(x_0)$, converges to $a$ as $n \rightarrow +\infty$. The rate of convergence can be estimated by

$$d(x_n, a) \leq \frac{q^n}{1-q}d(x_1, x_0).$$

(11.5)

Note that the theorem provides an algorithm to compute the unique fixed point, and that the rate of convergence in (11.5) depends on how close $q$ is to 0 or 1: it gets faster for $q$ close to 0 and slower for $q$ close to 1.

Proof. We will show that the sequence $x_k$ is a Cauchy sequence. Note that

$$d(x_{n+1}, x_n) = d(f(x_n), f(x_{n-1})) \leq qd(x_n, x_{n-1}),$$

(11.6)

so that an induction argument shows that

$$d(x_{k}, x_{k-1}) \leq q^n d(x_1, x_0).$$

(11.7)

Now, by the triangle inequality and (11.7), we have

$$d(x_{n+k}, x_{n}) \leq d(x_{n+k}, x_{n+k-1}) + d(x_{n+k-1}, x_{n+k-2}) + \cdots + d(x_{n+1}, x_{n})$$

$$\leq (q^{n+k-1} + q^{n+k-2} + \cdots + q^n)d(x_1, x_0) \leq \frac{q^n}{1-q}d(x_1, x_0).$$

(11.8)

It follows that if $0 < q < 1$, then the sequence $x_n$ is Cauchy. Since the space $X$ is complete, the limit of $x_n$ exists, and we set

$$a = \lim_{n \rightarrow \infty} x_n.$$ 

Now, as $f$ is a continuous map, passing to the limit $n \rightarrow \infty$ in the recursion relation $x_n = f(x_{n-1})$, we arrive at $a = f(a)$, hence $a$ is a fixed point of $f$.

The reason the fixed point is unique is that $f$ is a contraction. Indeed, if $a_1$ and $a_2$ are two fixed points, so that $a_1 = f(a_1)$ and $a_2 = f(a_2)$, then by the contraction property we have

$$d(f(a_1), f(a_2)) \leq qd(a_1, a_2).$$

However, as both $a_1$ and $a_2$ are fixed points of $f$, the left side above equals $d(a_1, a_2)$. Since $q \in (0, 1)$, we deduce that $d(a_1, a_2) = 0$ and $a_1 = a_2$. \(\Box\)

Existence theorem for ordinary differential equations

Let us consider an ordinary differential equation (ODE) for an unknown function $y(x)$

$$y' = f(x, y)$$

(11.9)

supplemented by the initial condition

$$y(x_0) = y_0.$$ 

(11.10)

with some given $x_0 \in \mathbb{R}$ and $y_0 \in \mathbb{R}$. We assume that the function $f(x, y)$ is continuous in $(x, y)$ and Lipschitz in $y$: there exists a constant $M > 0$ so that

$$|f(x, y_1) - f(x, y_2)| \leq M|y_1 - y_2|$$

for all $x \in \mathbb{R}$ and $y_1, y_2 \in \mathbb{R}$.

(11.11)

Theorem 11.3 Under the above assumptions on $f(x, y)$, there exists an interval $(x_0 - \delta_0, x_0 + \delta_0)$, so that the problem (11.9)-(11.10) has a unique solution $y(x)$ on the interval $(x_0 - \delta_0, x_0 + \delta_0)$.
Proof. We can write (11.9)-(11.10) together, using the fundamental theorem of calculus as

\[ y(x) = y_0 + \int_{x_0}^{x} f(t, y(t))dt. \]  \hfill (11.12)

Let us define the map \( A \) that maps a function \( y(x) \) to a function \( A[y] \) via

\[ (A[y])(x) = y_0 + \int_{x_0}^{x} f(t, y(t))dt. \]  \hfill (11.13)

Then (11.12) can be written as

\[ y(x) = A[y](x), \]  \hfill (11.14)

so that \( y(x) \) is a solution to (11.12), or, equivalently, to (11.9)-(11.10) if and only if the function \( y \) is a fixed point of the mapping \( A \). The first step is to show that \( A \) maps the space \( C[x_0 - \delta, x_0 + \delta] \) to itself.

**Lemma 11.4** If \( y \in C[x_0 - \delta, x_0 + \delta] \) for some \( \delta > 0 \), then \( A[y] \) is also in \( C[x_0 - \delta, x_0 + \delta] \).

**Proof of Lemma.** Let \( y(x) \) be a continuous function on the closed interval \([x_0 - \delta, x_0 + \delta]\) for some given \( \delta > 0 \). Then for any \( x_1, x_2 \in [x_0 - \delta, x_0 + \delta] \), we have

\[ A[y](x_1) - A[y](x_2) = y_0 + \int_{x_0}^{x_1} f(t, y(t))dt - y_0 - \int_{x_0}^{x_2} f(t, y(t))dt = \int_{x_0}^{x_1} f(t, y(t))dt - \int_{x_0}^{x_2} f(t, y(t))dt. \]  \hfill (11.15)

The function \( y \) is continuous on \([x_0 - \delta, x_0 + \delta]\), hence it is bounded on that interval: there exists \( K \) such that \( |y(t)| \leq K \) for all \( x_0 - \delta \leq t \leq x_0 + \delta \). As the function \( f \) is continuous on \([x_0 - \delta, x_0 + \delta] \times \mathbb{R}\), there exists \( M > 0 \) so that \( |f(x, y)| \leq M \) for all \( x \in [x_0 - \delta, x_0 + \delta] \) and \( y \in [-K, K] \). It follows that \( |f(t, y(t))| \leq M \) for all \( t \in [x_0 - \delta, x_0 + \delta] \). Using this in (11.15) gives

\[ |A[y](x_1) - A[y](x_2)| \leq \int_{x_0}^{x_1} |f(t, y(t))|dt \leq M|x_1 - x_2|. \]  \hfill (11.16)

It follows that the function \( A[y] \) is continuous. \( \Box \)

We return to the proof of the theorem. Our goal is to show that if \( \delta \) is sufficiently small, then \( A \) is a contraction on \( C[x_0 - \delta, x_0 + \delta] \). To this end, let us take two functions \( y_1, y_2 \in C[x_0 - \delta, x_0 + \delta] \) and write

\[ A[y_1](x) - A[y_2](x) = y_0 + \int_{x_0}^{x} f(t, y_1(t))dt - y_0 - \int_{x_0}^{x} f(t, y_2(t))dt = \int_{x_0}^{x} [f(t, y_1(t)) - f(t, y_2(t))]dt. \]  \hfill (11.17)

We will now use the Lipschitz property (11.11) of the function \( f(x, y) \):

\[ |A[y_1](x) - A[y_2](x)| \leq \int_{x_0}^{x} |f(t, y_1(t)) - f(t, y_2(t))|dt \leq \int_{x_0}^{x} M|y_1(t) - y_2(t)|dt. \]  \hfill (11.18)

Note that for all \( t \in [x_0, x] \) we have

\[ |y_1(t) - y_2(t)| \leq \sup_{x_0 - \delta \leq t \leq x_0 + \delta} |y_1(x) - y_2(x)| = \|y_1 - y_2\|. \]

Using this in (11.18), we arrive at

\[ |A[y_1](x) - A[y_2](x)| \leq \int_{x_0}^{x} M|y_1(t) - y_2(t)|dt \leq \int_{x_0}^{x} M\|y_1 - y_2\||dt = \int_{x_0}^{x} M|x - x_0||y_1 - y_2||dt \leq M\delta\|y_1 - y_2\|. \]  \hfill (11.19)
Taking supremum over all $x \in [x_0 - \delta, x_0 + \delta]$ gives
\[\|A[y_1] - A[y_2]\| \leq \sup_{x \in [x_0 - \delta, x_0 + \delta]} |A[y_1](x) - A[y_2](x)| \leq M\delta\|y_1 - y_2\|.\] (11.20)

Therefore, if $M\delta < 1$ then $A$ is a contraction on $C[x_0 - \delta, x_0 + \delta]$. It follows that $A$ has a unique fixed point $y$ in $C[x_0 - \delta, x_0 + \delta]$. This means that the function $y(x)$ satisfies (11.12):
\[y(x) = y_0 + \int_{x_0}^{x} f(t, y(t))dt.\] (11.21)

It follows immediately that $y(x_0) = y_0$. Moreover, as the function $y(t)$ is continuous, and $f(t,y)$ is continuous in both variables, it follows that $p(t) = f(t, y(t))$ is continuous in $t$. The fundamental theorem of calculus implies then that $y(x)$ is differentiable and
\[y'(x) = f(x, y(x)).\] (11.22)

This finishes the proof. \(\square\)

Let us now combine the existence theorem for ODEs with the construction of the fixed point of a contraction mapping in the proof of the existence theorem of a fixed point. Consider an ODE
\[y'(t) = y(t), \quad y(0) = 1,\]
and write it, as in (11.12), in the form
\[y(t) = 1 + \int_{0}^{t} y(s)ds.\]

The mapping $A$ is now defined via
\[A[y](t) = 1 + \int_{0}^{t} y(s)ds.\]

Consider the recursive sequence $y_n(t)$, with $y_0(t) = 1$, and
\[y_{n+1}(t) = 1 + \int_{0}^{t} y_n(s)ds.\]

Exercise 11.5 Show by induction that
\[y_n(t) = \sum_{k=1}^{n} \frac{t^k}{k!}.\]

We see that the unique solution is $y(t) = e^t$.

12 The implicit function theorem

The implicit function theorem addresses the question of when an equation of the form $f(x, y) = 0$ uniquely defines $x$ in terms of $y$. Let us consider a very simple situation: the equation
\[x^2 - y = 0.\]
Then, of course, for \( y > 0 \) this equation has two solutions \( x = \pm \sqrt{y} \), for \( y = 0 \) it has one solution \( x = 0 \) and for \( y < 0 \) it has no real solutions. Let us change our perspective somewhat. Suppose we know a solution \((x_0, y_0)\) — that is, \( x_0^2 = y_0 \) and we ask: given a \( y \) close to \( y_0 \) can we find a unique \( x \) close to \( x_0 \) so that \( x^2 = y \)? That is, if we perturb \( y \) slightly, can we still find a unique solution? Note that in this case \( y_0 \geq 0 \) automatically, simply because \( x_0^2 = y_0 \). The answer to our question is that if \( y_0 > 0 \), and, say, \( x_0 > 0 \) then, indeed, for \( y \) close to \( y_0 \) we still have a solution to \( x^2 = y \) that is close to \( x_0 \): \( x = \sqrt{y} \). Similarly, if \( x_0 < 0 \), then we still have a solution \( x^2 = y \) that is close to \( x_0 \): \( x = -\sqrt{y} \). On the other hand, if \( x_0 = 0 \) so that \( y_0 = 0 \), then in any interval \( y \in (-\delta, \delta) \) around \( y_0 = 0 \) and any interval \((\delta', \delta')\) around \( x_0 = 0 \) we can find \( y < 0 \) for which the equation \( x^2 = y \) has no solutions and \( y > 0 \) for which \( x^2 = y \) has two solutions in the interval \((\delta', \delta')\) — we just need to take \( y < (\delta')^2 \). Thus, there is a difference between \( x_0 = 0, y_0 = 0 \) and other points on the graph of \( y = x^2 \) — we can invert the relationship around the latter but not the former. The implicit function theorem generalizes this trivial observation.

**12.1 The inverse function theorem on \( \mathbb{R} \)**

We begin with the inverse function theorem, that looks not at an "implicit" equation \( f(x, y) = 0 \) but simply \( f(x) = y \). In one dimension the situation is quite simple.

**Proposition 12.1** Let \( f(x) \) be continuously differentiable on an interval \([a, b]\), and set

\[
m = \inf_{a \leq x \leq b} f(x), \quad M = \sup_{a \leq x \leq b} f(x).
\]

(i) Then \( f \) is a one-to-one map from \([m, M]\) to \([a, b]\) if and only if \( f \) is monotonic on \([a, b]\).

(ii) In addition, if \( f \) is monotonic on \([a, b]\) then the inverse function \( g = f^{-1} : [m, M] \to [a, b] \) is continuously differentiable on \([m, M]\) if and only if \( f'(x) \neq 0 \) for all \( x \in [a, b] \). In that case, \( g'(y) = 1 / f'(f^{-1}(y)) \).

**Exercise 12.2** Prove the first statement (i) in the above proposition.

To prove (ii), assume first that \( x_0 \in (a, b) \) and \( f'(x_0) \neq 0 \). Without loss of generality, we may assume that \( f'(x_0) > 0 \). As the function \( f' \) is continuous at \( x_0 \), there exists \( \delta > 0 \) so that

\[
f'(x) > f'(x_0)/2 \quad \text{for all} \quad x \in (x_0 - \delta, x_0 + \delta),
\]

(12.1)

thus \( f \) is strictly increasing on that interval. Let us set \( \alpha = f(x_0 - \delta), \beta = f(x_0 + \delta) \) and take some \( y \in (\alpha, \beta) \), so that \( y = f(x) \) for some \( x \in (x_0 - \delta, x_0 + \delta) \), that is, \( x = g(y) \) — recall that \( g = f^{-1} \). As \( f \) is differentiable at \( x \), we have

\[
f(x_0 + h) = f(x_0) + f'(x_0)h + \zeta(h),
\]

with \( \zeta(h) = o(h) \) as \( h \to 0 \). Let us set \( h = x - x_0 \), so that \( x_0 + h = x \), then this is

\[
y = y_0 + f'(x_0)(x - x_0) + \zeta(x - x_0).
\]

This can be re-written as

\[
x - x_0 = \frac{1}{f'(x_0)}(y - y_0) - \frac{\zeta(x - x_0)}{f'(x_0)}.
\]

(12.2)

Observe that there exists \( c \) between \( x \) and \( x_0 \) so that

\[
y - y_0 = f'(c)(x - x_0),
\]

(12.3)
and, because of (12.1), we know that \( f'(c) > f'(x_0)/2 > 0 \). Therefore, for any \( \gamma > 0 \) there exists \( \delta > 0 \) so that

\[
\text{if } |y - y_0| < \delta, \text{ then } |x - x_0| < \gamma. \tag{12.4}
\]

Finally, let \( \varepsilon > 0 \) and find \( \gamma \) so that if \( |x - x_0| < \gamma \), then

\[
\left| \frac{\zeta(x - x_0)}{x - x_0} \right| < \varepsilon. \tag{12.5}
\]

Given \( \gamma > 0 \), find the corresponding \( \delta > 0 \) in (12.4). Now, if \( |y - y_0| < \delta \), then \( x \) that appears in (12.2) satisfies (12.5). Then, (12.2) implies that

\[
\left| g(y) - g(y_0) - \frac{1}{f'(x_0)} - \frac{\zeta(x - x_0)}{f'(x_0)(y - y_0)} \right| < \varepsilon. \tag{12.6}
\]

and thus

\[
\left| \frac{g(y) - g(y_0)}{y - y_0} - \frac{1}{f'(x_0)} \right| = \left| \frac{\zeta(x - x_0)}{f'(x_0)(y - y_0)} \right| < \frac{\varepsilon}{f'(x_0)f'(c)}. \tag{12.7}
\]

It follows that \( g'(y_0) \) exists and \( g'(y_0) = 1/f'(x_0) \).

**Exercise 12.3** Use the above argument to show that when \( f'(x_0) = 0 \), the function \( g = f^{-1} \) is not differentiable at \( y_0 = f(x_0) \). Hint: aim to show that

\[
\left| \frac{g(y) - g(y_0)}{y - y_0} \right| \to +\infty \text{ as } y \to y_0.
\]

### 12.2 The inverse function theorem for maps \( \mathbb{R}^n \to \mathbb{R}^n \)

We now consider the inverse function theorem for maps from \( \mathbb{R}^n \) to \( \mathbb{R}^n \). We will need the following lemmas.

**Lemma 12.4** Let \( A \) be an \( n \times n \) matrix, \( U \subset \mathbb{R}^n \) be an open set, and a map \( f : U \to \mathbb{R}^n \) be continuously differentiable on \( U \). Set \( g(x) = Af(x) \), then \( g : U \to \mathbb{R}^n \) is a continuously differentiable map with

\[
Dg(x) = ADf(x). \tag{12.8}
\]

**Proof.** Recall that the entries of the matrix \( Dg \) are

\[
[Dg(x)]_{ij} = \frac{\partial g_i(x)}{\partial x_j},
\]

and

\[
g_i(x) = \sum_{k=1}^{n} A_{ik} f_k(x).
\]

It follows that

\[
\frac{\partial g_i(x)}{\partial x_j} = \sum_{k=1}^{n} A_{ik} \frac{\partial f_k(x)}{\partial x_j} = \sum_{k=1}^{n} A_{ik} [Df(x)]_{kj} = (ADf(x))_{ij},
\]

thus \( Dg(x) = ADf(x) \), and we are done. \( \square \)

**Lemma 12.5** Let \( A \) be an \( n \times n \) matrix with entries such that \( A_{ij} \leq \varepsilon \) for all \( 1 \leq i, j \leq n \), then for any vector \( v \in \mathbb{R}^n \) we have \( \|Av\| \leq n\varepsilon\|v\| \).
**Proof.** First, we recall the inequality

\[(x_1 + \cdots + x_n)^2 \leq n(x_1^2 + \cdots + x_n^2). \tag{12.9}\]

To see that (12.9) holds, write

\[(x_1 + \cdots + x_n)^2 = x_1^2 + \cdots + x_n^2 + 2 \sum_{1 \leq i < j \leq n} x_ix_j \leq x_1^2 + \cdots + x_n^2 + \sum_{1 \leq i < j \leq n} (x_i^2 + x_j^2) = x_1^2 + \cdots + x_n^2 + (n-1)(x_1^2 + \cdots + x_n^2) = n(x_1^2 + \cdots + x_n^2).\]

Note that for each $1 \leq i \leq n$ we have

\[|(Av)_i| = \left|\sum_{j=1}^{n} A_{ij}v_j\right| \leq \sum_{j=1}^{n} |A_{ij}| |v_j| \leq \varepsilon \sum_{j=1}^{n} |v_j|,\]

so that, using (12.9), we get

\[|(Av)_i|^2 \leq \varepsilon^2 \left(\sum_{j=1}^{n} |v_j|\right)^2 \leq n\varepsilon^2 \sum_{j=1}^{n} |v_j|^2 = n\varepsilon^2 \|v\|^2.\]

Next, summing over $i$, we get

\[\|Av\|^2 = \sum_{i=1}^{n} |(Av)_i|^2 \leq n^2 \varepsilon^2 \|v\|^2,\]

and the claim of the lemma follows. \(\square\)

**Theorem 12.6** Let $U \subset \mathbb{R}^n$ be an open set, and $x_0 \in U$. Let $f : U \to \mathbb{R}^n$ be a continuously differentiable function, and set $y_0 = f(x_0)$. Suppose that the derivative matrix $Df(x_0)$ is not singular, that is, the inverse matrix $[Df(x_0)]^{-1}$ exists. Then there exist an open set $V \subset U$ such that $x_0 \in V$, and an open set $W \subset \mathbb{R}^n$ such that $y_0 \in W$, such that $f$ is a one-to-one map from $V$ to $W$. Moreover, the inverse map $g = f^{-1} : W \to V$ is also continuously differentiable and for $y \in W$ we have $Dg(y) = [Df(g(y))]^{-1}$.

**Proof.** **Step 1. Reduction to the case $Df(x_0) = I$.** We first note that it suffices to prove the theorem under an additional assumption that

\[Df(x_0) = I, \tag{12.10}\]

the $n \times n$ identity matrix. Indeed, if $Df(x_0) \neq I$, we consider the function

\[\tilde{f}(x) = [Df(x_0)]^{-1}f(x).\]

The gradient matrix of $\tilde{f}$ is identity

\[D\tilde{f}(x_0) = [Df(x_0)]^{-1}Df(x_0) = I,\]

as follows from Lemma 12.4. Since the matrix $Df(x_0)$ is invertible, the function $f$ is one-to-one from a neighborhood $V$ of $x_0$ to a neighborhood $W$ of $y_0 = f(x_0)$ if and only if the function $\tilde{f}$ is a one-to-one map from $V$ to $\tilde{W} = [Df(x_0)]^{-1}W$, and $\tilde{W}$ is a neighborhood of the point $\tilde{y}_0 = \tilde{f}(x_0)$. Again, as the matrix $Df(x_0)$ is invertible, the function $\tilde{f}$ is continuously differentiable if and only
We compute the derivative \( g \) for \( 0 \leq \varepsilon < 1 \). It is continuously differentiable. Hence, we may assume without any loss of generality that \( f \) satisfies (12.10), and this is what we will do for the rest of the proof.

As the map \( f(x) \) is continuously differentiable in \( U \), and \( f \) satisfies (12.10), for any \( \varepsilon > 0 \) there exists \( r > 0 \) so that
\[
\left| \frac{\partial f_i}{\partial x_j} - \delta_{ij} \right| < \varepsilon \tag{12.11}
\]
for all \( x \in B(x_0, r) \). Here, \( \delta_{ij} \) is the Kronecker delta: \( \delta_{ij} = 1 \) if \( i = j \) and \( \delta_{ij} = 0 \) if \( i \neq j \). In other words, for each \( x \in B(x_0, r) \), we have
\[
Df(x) = I + E(x), \tag{12.12}
\]
with the matrix \( E(x) \) such that
\[
|E_{ij}(x)| \leq \varepsilon \text{ for all } 1 \leq i, j \leq n. \tag{12.13}
\]
It follows that for every \( v \in \mathbb{R}^n \) and all \( x \in B(x_0, r) \) we have
\[
\|Df(x)v\| = \|(I + E(x))v\| = \|v + E(x)v\| \geq \|v\| - \|E(x)v\|.
\]
Now, Lemma 12.5, with \( A = E(x) \), implies that for every \( v \in \mathbb{R}^n \) and all \( x \in B(x_0, r) \) we have
\[
\|Df(x)v\| \geq \|v\| - \varepsilon n\|v\| \geq \frac{1}{2}\|v\|
\]
as long as \( \varepsilon < 1/(2n) \). In particular, it follows that the kernel of \( Df(x) \) is \( \{0\} \), and the matrix \( Df(x) \) is invertible for all \( x \in B(x_0, r) \).

**Step 2. Reformulation as a fixed point problem for a contraction mapping.** Now, we turn to solving
\[
f(x) = y \tag{12.14}
\]
for \( y \) close to \( y_0 = f(x_0) \): we assume that \( y \in B(y_0, \rho) \) and will later see how small \( \rho \) needs to be. In order to turn this into a contraction mapping question, let us reformulate (12.14) as a fixed point problem by setting
\[
F(x) = x - f(x) + y, \tag{12.15}
\]
again, with \( y \in B(y_0, \rho) \) fixed. Our goal is to show that, provided that \( \rho \) is sufficiently small, and \( y \in B(y_0, \rho) \), we can find \( r_1 < r \) such that \( F \) maps \( B(x_0, r_1) \) to itself and is a contraction. This will show that there exists a unique \( x \in B(x_0, r_1) \) such that \( F(x) = x \), which means that there is a unique solution to (12.14) in \( B(x_0, r_1) \) for each \( y \in B(y_0, \rho) \), so that \( f \) is a one-to-one map from \( B(x_0, r_1) \) to \( B(y_0, \rho) \), hence the inverse map \( f^{-1} : B(y_0, \rho) \to B(x_0, r_1) \) is well-defined.

To verify that \( F \) is a contraction, let us take \( z, w \in B(x_0, r) \) and define
\[
g(t) = F(z + t(w - z)),
\]
for \( 0 \leq t \leq 1 \), so that \( F(z) = g(0) \), and \( F(w) = g(1) \). The fundamental theorem of calculus implies that
\[
F_k(w) - F_k(z) = \int_0^1 g'_k(t)dt, \quad 1 \leq k \leq n.
\]
We compute the derivative \( g'(t) \) using the chain rule:
\[
\frac{dg_k(t)}{dt} = \sum_{j=1}^n \frac{\partial F_k(z + t(w - z))}{\partial x_j} (w_j - z_j). \tag{12.16}
\]
The definition (12.15) of \( F(x) \) implies that
\[
\frac{\partial F_k(x)}{\partial x_j} = \delta_{kj} - \frac{\partial f_k(x)}{\partial x_j}.
\]

Recalling (12.12)-(12.13), we see that the derivative matrix \( DF(x) = E(x) \), and thus the entries of the matrix \( DF(x) \) satisfy
\[
\left| \frac{\partial F_i}{\partial x_j} \right| \leq \varepsilon \text{ for all } 1 \leq i, j \leq n. \tag{12.17}
\]

This is why we modified \( f(x) \) to make sure that \( Df(x_0) = I \) at the beginning of the proof.