Department of Mathematics, London School of Economics

Real Analysis
(MA203)

Amol Sasane

Contents

Preface

What is Real Analysis?

First of all "Analysis" refers to the subdomain of Mathematics, which is roughly speaking an abstraction of the familiar subject of Calculus. Calculus arose as a box of tools enabling one to handle diverse problems in the applied sciences such as physics and engineering where quantities change (for example with time), and calculations based on "rates of change" were needed. It soon became evident that the foundations of Calculus needed to be made mathematically precise. This is roughly the subject of Mathematical Analysis, where Calculus is made rigorous. But another byproduct of this rigorization process is that mathematicians discovered that many of the things done in the set-up of usual calculus can be done in a much more general set up, enabling one to expand the domain of applications. We will study such things in this course.

Secondly, why do we use the adjective "Real"? We will start with the basic setting of making rigorous Calculus with *real* numbers, but we will also develop Calculus in more abstract settings, for example in $\mathbb{R}^n$. Using this adjective "Real" also highlights that the subject is different from "Complex Analysis" which is all about doing analysis in $\mathbb{C}$. (It turns out that Complex Analysis is a very specialized branch of analysis which acquires a somewhat peculiar character owing to the special geometric meaning associated with the multiplication of complex numbers in the complex plane.)

Amol Sasane

Chapter 1

$\mathbb{R}$, metric spaces and $\mathbb{R}^n$

We are familiar with concepts from calculus such as

(1) convergence of sequences of real numbers,

(2) continuity of a function $f : \mathbb{R} \to \mathbb{R}$,

(3) differentiability of a function $f : \mathbb{R} \to \mathbb{R}$.

Once these notions are available, one can prove useful results involving such notions. For example, we have seen the following:

Theorem 1.1. *If $f : [a, b] \to \mathbb{R}$ is continuous, then f has a minimizer on $[a, b]$.*

Theorem 1.2. *If $f : \mathbb{R} \to \mathbb{R}$ is such that $f''(x) \geq 0$ for all $x \in \mathbb{R}$ and $f'(x_0) = 0$, then x_0 is a minimizer of f.*

We will revisit these concepts in this course, and see that the same concepts can be defined in a much more general context, enabling one to prove results similar to the above in the more general set up. This means that we will be able to solve problems that arise in applications (such as optimization and differential equations) that we wouldn't be able to solve earlier with our limited tools. Besides these immediate applications, concepts and results from real analysis are fundamental in mathematics itself, and are needed in order to study almost any topic in mathematics.

In this chapter, we wish to emphasize that the key idea behind defining the above concepts is that of a distance between points. In the case when one works with real numbers, this distance is provided by the absolute value of the difference between the two numbers: thus the distance between $x, y \in \mathbb{R}$ is taken as $|x - y|$. This coincides with our geometric understanding of distance when the real numbers are represented on the "number line". For instance, the distance between -1 and 3 is 4, and indeed $4 = |-1 - 3|$.

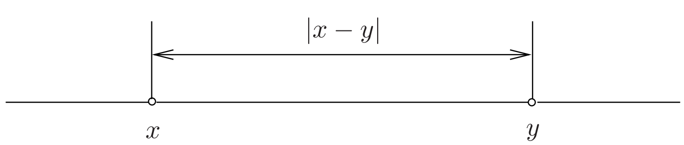


Figure 1. Distance between real numbers.

Recall for example, that a sequence $(a_n)_{n \in \mathbb{N}}$ is said to *converge with limit* $L \in \mathbb{R}$ if for every $\epsilon > 0$, there exists a $N \in \mathbb{N}$ such that whenever $n > N$, $|a_n - L| < \epsilon$. In other words, the sequence converges to L if no matter what *distance* $\epsilon > 0$ is given, one can guarantee that all the terms of the sequence beyond a certain index N are at a *distance* of at most ϵ away from L (this is the inequality $|a_n - L| < \epsilon$). So we notice that in this notion of "convergence of a sequence" indeed the notion of distance played a crucial role. After all, we want to say that the terms of the sequence get "close" to the limit, and to measure closeness, we use the distance between points of $\mathbb{R}$.

A similar thing happens with all the other notions listed at the outset. For example, recall that a function $f : \mathbb{R} \to \mathbb{R}$ is said to be *continuous at* $c \in \mathbb{R}$ if for every $\epsilon > 0$, there exists a $\delta > 0$ such that whenever $|x - c| < \delta$, $|f(x) - f(c)| < \epsilon$. Roughly, given any distance ϵ, I can find a distance δ such that whenever I choose an x not farther than a distance δ from c, I am guaranteed that $f(x)$ is not farther than a distance of ϵ from $f(c)$. Again notice the key role played by the distance in this definition.

1.1. Distance in $\mathbb{R}$

The distance between points $x, y \in \mathbb{R}$ is taken as $|x - y|$. Thus we have a map that associates to a pair $(x, y) \in \mathbb{R} \times \mathbb{R}$ of real numbers, the number $|x - y| \in \mathbb{R}$ which is the distance between x and y. We think of this map $(x, y) \mapsto |x - y| : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ as the "distance function" in $\mathbb{R}$.

Define $d : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ by $d(x, y) = |x - y|$ $(x, y \in \mathbb{R})$. Then it can be seen that this distance function d satisfies the following properties:

(D1) For all $x, y \in \mathbb{R}$, $d(x, y) \geq 0$. If $x \in \mathbb{R}$, then $d(x, x) = 0$. If $x, y \in \mathbb{R}$ are such that $d(x, y) = 0$, then $x = y$.

(D2) For all $x, y \in \mathbb{R}$, $d(x, y) = d(y, x)$.

(D3) For all $x, y, z \in \mathbb{R}$, $d(x, y) + d(y, z) \geq d(x, z)$.

It turns out that these are the key properties of the distance which are needed in developing analysis in $\mathbb{R}$. So it makes sense that when we want to generalize the situation with the set $\mathbb{R}$ being replaced by an arbitrary set X, we must define a distance function

$$d : X \times X \to \mathbb{R}$$

that associates a number (the distance!) to each pair of points $x, y \in X$, and which has the same properties (D1)-(D3) (with the obvious changes: $x, y, z \in X$). We do this in the next section.

1.2. Metric space

Definition 1.3. A *metric space* is a set X together with a function $d : X \times X \to \mathbb{R}$ satisfying the following properties:

(D1) (*Positive definiteness*) For all $x, y \in X$, $d(x, y) \geq 0$. For all $x \in X$, $d(x, x) = 0$. If $x, y \in X$ are such that $d(x, y) = 0$, then $x = y$.

(D2) (*Symmetry*) For all $x, y \in X$, $d(x, y) = d(y, x)$.

(D3) (*Triangle inequality*) For all $x, y, z \in X$, $d(x, y) + d(y, z) \geq d(x, z)$.

Such a d is referred to as a *distance function* or *metric*.

Let us consider some examples.

Example 1.4. $X := \mathbb{R}$, with $d(x, y) := |x - y|$ $(x, y \in \mathbb{R})$, is a metric space. ◊

Example 1.5. For any nonempty set X, define

$$d(x, y) = \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{if } x = y. \end{cases}$$

This d is called the *discrete metric*. Then d satisfies (D1)-(D3), and so X with the discrete metric is a metric space. $\Diamond$

Note that in particular $\mathbb{R}$ with the discrete metric is a metric space as well. So the above two examples show that the distance function in a metric space is not unique, and what metric is to be used depends on the application one has in mind. Hence whenever we speak of a metric space, we always need to specify not just the set X but also the distance function d being considered. So often we say "consider the metric space (X, d)", where X is the set in question, and $d : X \times X \to \mathbb{R}$ is the metric considered.

However, for some sets, there are some natural candidates for distance functions. One such example is the following one.

Example 1.6 (Euclidean space $\mathbb{R}^n$)**.** In $\mathbb{R}^2$ and $\mathbb{R}^3$, where we can think of vectors as points in the plane or points in the space, we can use the distance distance between two points as the length of the line segment joining these points. Thus (by Pythagoras's Theorem) in $\mathbb{R}^2$, we may use

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

as the distance between the points $x = (x_1, x_2)$ and $y = (y_1, y_2)$ in $\mathbb{R}^2$. Similarly, in $\mathbb{R}^3$, one may use

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2}$$

as the distance between the points $x = (x_1, x_2, x_3)$ and $y = (y_1, y_2, y_3)$ in $\mathbb{R}^3$. See Figure 2.

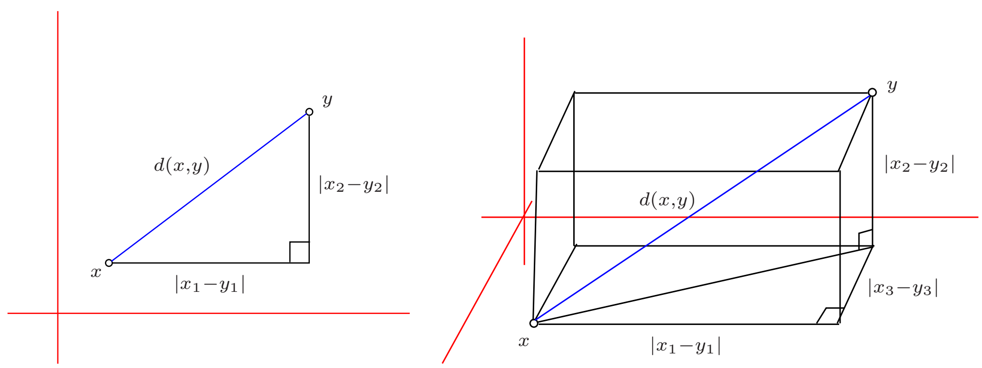


Figure 2. Distance in $\mathbb{R}^2$ and $\mathbb{R}^3$.

In an analogous manner to $\mathbb{R}^2$ and $\mathbb{R}^3$, more generally, for $x, y \in \mathbb{R}^n =: X$, we define the *Euclidean distance* by

$$d(x, y) = \sqrt{\sum_{k=1}^{n} (x_k - y_k)^2} = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}$$

for

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n, \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^n.$$

Then $\mathbb{R}^n$ is a metric space with the Euclidean distance, and is referred to as the *Euclidean space*. The verification of (D3) can be done by using the *Cauchy-Schwarz inequality*: For real numbers $x_1, \dots, x_n$ and $y_1, \dots, y_n$, there holds that

$$\left(\sum_{k=1}^{n} x_k^2\right)\left(\sum_{k=1}^{n} y_k^2\right) \geq \left(\sum_{k=1}^{n} x_k y_k\right)^2.$$

This last property (D3) is sometimes referred to as the triangle inequality. The reason behind this is that, for triangles in Euclidean geometry of the plane, we know that the sum of the lengths of two sides of a triangle is at least as much as the length of the third side. If we now imagine the points $x, y, z \in \mathbb{R}^2$ as the three vertices of a triangle, then this is what (D3) says; see Figure 3.

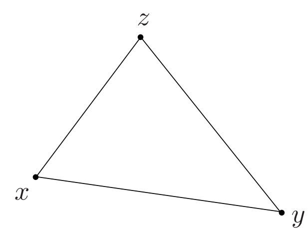


Figure 3. How the triangle inequality gets its name.

Throughout this course, whenever we refer to $\mathbb{R}^n$ as a metric space, unless specified otherwise, we mean that it is equipped with this Euclidean metric. Note that Example 1.4 corresponds to the case when $n = 1$. $\Diamond$

Exercise 1.7. Verify that the d given in Example 1.5 does satisfy (D1)-(D3).

Exercise 1.8. One can show the Cauchy-Schwarz inequality as follows: Let x, y be vectors in $\mathbb{R}^n$ with the components $x_1, \dots, x_n$ and $y_1, \dots, y_n$, respectively. For $t \in \mathbb{R}$, consider the function

$$f(t) = (x + ty)^\top (x + ty) = x^\top x + 2tx^\top y + t^2 y^\top y.$$

From the second expression, we see that f is a quadratic function of the variable t. It is clear from the first expression that $f(t)$, being the sum of squares

$$\sum_{k=1}^{n} (x_k + ty_k)^2,$$

is nonnegative for all $t \in \mathbb{R}$. This means that the discriminant of f must be ≤ 0, since otherwise, f would have two distinct real roots, and would then have negative values between these roots! Calculate the discriminant of the quadratic function and show that its nonpositivity yields the Cauchy-Schwarz inequality.

Normed space. Frequently in applications, one needs a metric not just in any old set X, but in a *vector space* X.

Recall that a (real) vector space X, is just a set X with the two operations of vector addition $+ : X \times X \to \mathbb{X}$ and scalar multiplication $\cdot : \mathbb{R} \times X \to X$ which together satisfy the vector space axioms.

But now if one wants to also do analysis in a vector space X, there is so far no ready-made available notion of distance between vectors. One way of creating a distance in a vector space is to equip it with a "norm" $\|\cdot\|$, which is the analogue of absolute value $|\cdot|$ in the vector space $\mathbb{R}$. The distance function is then created by taking the norm $\|x - y\|$ of the difference between pairs of vectors $x, y \in X$, just like in $\mathbb{R}$ the Euclidean distance between $x, y \in \mathbb{R}$ was taken as $|x - y|$.

Definition 1.9. A *normed space* is a vector space $(X, +, \cdot)$ together with a *norm*, namely, a function $\|\cdot\| : X \to \mathbb{R}$ satisfying the following properties:

(N1) (*Positive definiteness*) For all $x \in X$, $\|x\| \geq 0$. If $x \in X$ is such that $\|x\| = 0$, then $x = \mathbf{0}$ (the zero vector in X).

(N2) (*Positive homogeneity*) For all $\alpha \in \mathbb{R}$ and all $x \in X$, $\|\alpha \cdot x\| = |\alpha| \, \|x\|$.

(N3) (*Triangle inequality*) For all $x, y \in X$, $\|x + y\| \leq \|x\| + \|y\|$.

It is easy to check that if X is a normed space then

$$d(x,y) := \|x-y\| \quad (x,y\in X)$$

satisfies (D1)-(D3) and makes X a metric space. This distance is referred to as the *induced distance* in the normed space $(X, \|\cdot\|)$. Clearly then

$$\|x\| = \|x-\mathbf{0}\| = d(x,\mathbf{0}),$$

and so the norm of a vector is the induced distance to the zero vector in a normed space $(X, \|\cdot\|)$.

Example 1.10. $\mathbb{R}$ is a vector space with the usual operations of addition and multiplication. It is easy to see that the absolute value function

$$x \mapsto |x| \quad (x \in \mathbb{R})$$

satisfies (N1)-(N3), and so $\mathbb{R}$ is a normed space, and the induced distance is the usual Euclidean metric in $\mathbb{R}$.

More generally, in the vector space $\mathbb{R}^n$, with addition and scalar multiplication defined componentwise, we can introduce the 2-norm as

$$\|x\|_2 := \sqrt{x_1^2 + \cdots + x_n^2}$$

for vectors $x \in \mathbb{R}^n$ having components $x_1, \dots, x_n$. Then $\|\cdot\|_2$ satisfies (N1)-(N3) and makes $\mathbb{R}^n$ a normed space. The induced metric is then the usual Euclidean metric in $\mathbb{R}^n$. $\Diamond$

Exercise 1.11. Verify that the norm $\|\cdot\|_2$ given on $\mathbb{R}^n$ in Example 1.10 does satisfy (N1)-(N3).

Exercise 1.12. $(*)$ Let X be a metric space with a metric d. Define $d_1 : X \times X \to \mathbb{R}$ by

$$d_1(x,y) = \frac{d(x,y)}{1+d(x,y)} \quad (x,y \in X).$$

Note that $d_1(x,y) \le 1$ for all $x,y \in X$. Show that d_1 is a metric on X.

Exercise 1.13. Consider the vector space $\mathbb{R}^{m\times n}$ of matrices with m rows and n columns of real numbers, with the usual entrywise addition and scalar multiplication. Define for $M \in \mathbb{R}^{m\times n}$ with the entry in the ith row and jth column denoted by m_{ij}, for $1 \le i \le m$, $1 \le j \le n$, the number

$$\|M\|_\infty := \max_{1\le i\le m,\ 1\le j\le n} |m_{ij}|.$$

Show that $\|\cdot\|_\infty$ defines a norm on $\mathbb{R}^{m\times n}$.

Exercise 1.14. Let $C[a,b]$ denote the set of all continuous functions $f : [a,b] \to \mathbb{R}$. Then $C[a,b]$ is a vector space with addition and scalar multiplication defined pointwise. If $f \in C[a,b]$, define

$$\|f\|_\infty = \max_{x\in[a,b]} |f(x)|.$$

Note that the function $x \mapsto |f(x)|$ is continuous on $[a,b]$, and so by the Extreme Value Theorem, the above maximum exists.

(1) Show that $\|\cdot\|_\infty$ is a norm on $C[a,b]$.
(2) Let $f \in C[a,b]$ and let $\epsilon > 0$. Consider the set $B(f,\epsilon) := \{g \in C[a,b] : \|f-g\|_\infty < \epsilon\}$. Draw a picture to explain the geometric significance of the statement $g \in B(f,\epsilon)$.

Exercise 1.15. $C[a,b]$ can also be equipped with other norms. For example, prove that

$$\|f\|_1 := \int_a^b |f(x)|dx \quad (f \in C[a,b])$$

also defines a norm on $C[a,b]$.

Exercise 1.16. If (X,d) is a metric space, and if $Y \subset X$, then show that $(Y, d|_{Y\times Y})$ is a metric space. (Here $d|_{Y\times Y}$ denotes the restriction of d to the set $Y\times Y$, that is, $d|_{Y\times Y}(y_1,y_2) = d(y_1,y_2)$ for $y_1,y_2 \in Y$.) Hence every subset of a metric space is itself a metric space with the restriction of the original metric. The metric $d|_{Y\times Y}$ is referred to as the *induced metric on Y by d* or the *subspace metric of d with respect to Y*, and the metric space $(Y, d|_{Y\times Y})$ is called a *metric subspace* of (X,d).

Exercise 1.17. ($*$) The set of integers $\mathbb{Z}$ ($\subset \mathbb{R}$) inherits the Euclidean metric from $\mathbb{R}$, but it also carries a very different metric, called the *p-adic metric*. Given a prime number p, and an integer n, the p-adic "norm"[1] of n is $|n|_p := 1/p^k$, where k is the largest power of p that divides n. The norm of 0 is by definition 0. The more factors of p, the smaller the p-norm. The p-adic metric[2] on $\mathbb{Z}$ is $d_p(x, y) := |x - y|_p$ ($x, y \in \mathbb{Z}$).

(1) Prove that if $x, y \in \mathbb{Z}$, then $|x + y|_p \le \max\{|x|_p, |y_p|\}$.
(2) Show that d_p is a metric on $\mathbb{Z}$.

Exercise 1.18. ($*$) Let ℓ^2 denote the set of all "square summable" sequences of real numbers:

$$\ell^2 = \left\{ (a_n)_{n\in\mathbb{N}} : \sum_{n=1}^{\infty} |a_n|^2 < \infty \right\}.$$

(1) Show that ℓ^2 is a vector space with addition and scalar multiplication defined termwise.
(2) Let $\|(a_n)_{n\in\mathbb{N}}\|_2 := \sqrt{\sum_{n=1}^{\infty} |a_n|^2}$ for $(a_n)_{n\in\mathbb{N}} \in \ell^2$. Prove that $\|\cdot\|_2$ defines a norm on ℓ^2.

So ℓ^2 is an infinite-dimensional analogue of the Euclidean space $(\mathbb{R}^n, \|\cdot\|_2)$.

Exercise 1.19 (Hamming Distance)**.** Let $\mathbb{F}_2^n$ be the set of all ordered n-tuples of zeros and ones. For example, $\mathbb{F}_2^3 = \{000, 001, 010, 011, 100, 101, 110, 111\}$. For $x, y \in \mathbb{F}_2^n$, let

$$d(x, y) = \text{the number of places where } x \text{ and } y \text{ have different entries.}$$

For example, in $\mathbb{F}_2^3$, we have $d(110, 110) = 0$, $d(010, 110) = 1$ and $d(101, 010) = 3$. Show that $(\mathbb{F}_2^n, d)$ is a metric space. (This metric is used in the digital world, in coding and information theory.)

1.3. Neighbourhoods and open sets

Let (X, d) be a metric space. Now that we have a metric, we can describe "neighbourhoods" of points by considering sets which include all points whose distance to the given point is not too large.

Definition 1.20 (Open ball)**.** Let (X, d) be a metric space. If $x \in X$ and $r > 0$, we call the set

$$B(x, r) = \{y \in X : d(x, y) < r\}$$

the *open ball* centered at x with radius r.

The picture we have in mind is shown in Figure 4.

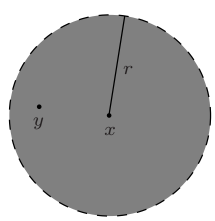


Figure 4. The open ball $B(x, r)$.

In the sequel, for example in our study of continuous functions, *open sets* will play an important role. Here is the definition.

Definition 1.21 (Open set)**.** Let (X, d) be a metric space. A set $U \subset X$ is said to be *open* if for every $x \in U$, there exists an $r > 0$ such that $B(x, r) \subset U$.

[1]Note that $\mathbb{Z}$ is not a real vector space and so this is not really a norm in the sense we have learnt.

[2]Some physicists propose this to be the appropriate metric for the fabric of space-time. Thus metrics are something we choose, depending on context. It is not something that falls out of the sky!

Note that the radius r can depend on the choice of the point x. See Figure 5. Roughly speaking, in a open set, no matter which point you take in it, there is always some "room" around it consisting only of points of the open set.

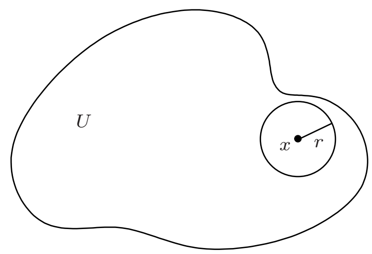


Figure 5. Open set.

Example 1.22. Let us show that the set (a, b) is open in $\mathbb{R}$. Given any $x \in (a, b)$, we have $a < x < b$. Motivated by Figure 6, let us take $r = \min\{x - a, b - x\}$. Then we see that $r > 0$ and whenever $|y - x| < r$, we have $-r < y - x < r$. So

$$a = x - (x - a) \leq x - r < y < x + r \leq x + (b - x) = b,$$

that is, $y \in (a, b)$. Hence $B(x, r) \subset (a, b)$. Consequently, (a, b) is open.

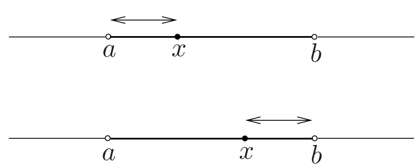


Figure 6. (a, b) is open in $\mathbb{R}$.

On the other hand, the interval $[a, b]$ is not open, because $x := a \in [a, b]$, but no matter how small an $r > 0$ we take, the set $B(a, r) = \{y \in \mathbb{R} : |y - a| < r\} = (a - r, a + r)$ contains points that do not belong to $[a, b]$: for example,

$$a - \frac{r}{2} \in B(a, r) \text{ but } a - \frac{r}{2} \notin [a, b].$$

Figure 7 illustrates this. ◊

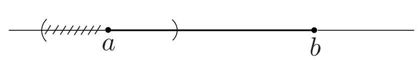


Figure 7. $[a, b]$ is not open in $\mathbb{R}$.

Example 1.23. The set X is open, since given an $x \in X$, we can take any $r > 0$, and notice that $B(x, r) \subset X$ trivially.

The empty set $\emptyset$ is also open ("vacuously"). Indeed, the reasoning is as follows: can one show an x for which there is no $r > 0$ such that $B(x, r) \subset \emptyset$? And the answer is no, because there is no x in the empty set (let alone an x which has the extra property that there is no $r > 0$ such that $B(x, r) \subset \emptyset$!). ◊

Exercise 1.24. Let (X, d) be a metric space, $x \in X$ and $r > 0$. Show that the open ball $B(x, r)$ is an open set.

Lemma 1.25. *Any finite intersection of open sets is open.*

Proof. It is enough to consider two open sets, as the general case follows immediately by induction on the number of sets. Let U_1, U_2 be two open sets. Let $x \in U_1 \bigcap U_2$. Then there exist $r_1 > 0$, $r_2 > 0$ such that $B(x, r_1) \subset U_1$ and $B(x, r_2) \subset U_2$. Take $r = \min\{r_1, r_2\}$. Then $r > 0$, and we claim that $B(x, r) \subset U_1 \bigcap U_2$. To see this, let $y \in B(x, r)$. Then $d(x, y) < r_1$ and $d(x, y) < r_2$. So $y \in B(x, r_1) \bigcap B(x, r_1) \subset U_1 \bigcap U_2$. □

Example 1.26. The finiteness condition in the above lemma cannot be dropped. Here is an example. Consider the open sets

$$U_n := \left(-\frac{1}{n}, \frac{1}{n}\right) \quad (n \in \mathbb{N})$$

in $\mathbb{R}$. Then we have $\displaystyle\bigcap_{n\in\mathbb{N}} U_n = \{0\}$, which is not open in $\mathbb{R}$. ◊

Lemma 1.27. *Any union of open sets is open.*

Proof. Let U_i $(i \in I)$ be a family of open sets indexed[3] by the set I. If

$$x \in \bigcup_{i\in I} U_i,$$

then $x \in U_{i_*}$ for some $i_* \in I$. But as U_{i_*} is open, there exists a $r > 0$ such that $B(x, r) \subset U_{i_*}$. Thus

$$B(x, r) \subset U_{i_*} \subset \bigcup_{i\in I} U_i,$$

and so we see that the union $\displaystyle\bigcup_{i\in I} U_i$ is open. □

Definition 1.28 (Closed set)**.** Let (X, d) be a metric space. A set F is *closed* if its complement $X \setminus F$ is open.

Example 1.29. $[a, b]$ is closed in $\mathbb{R}$. Indeed, its complement $\mathbb{R} \setminus [a, b]$ is the union of the two open sets $(-\infty, a)$ and (b, ∞). Hence $\mathbb{R} \setminus [a, b]$ is open, and $[a, b]$ is closed.

The set $(-\infty, b]$ is closed in $\mathbb{R}$. (Why?)

The sets $(a, b]$, $[a, b)$ are neither open nor closed in $\mathbb{R}$. (Why?) ◊

Example 1.30. X, $\emptyset$ are closed. ◊

Exercise 1.31. Show that arbitrary intersections of closed sets are closed. Prove that a finite union of closed sets is closed. Can the finiteness condition be dropped in the previous claim?

Exercise 1.32. We know that the segment $(0, 1)$ is open in $\mathbb{R}$. Show that the segment $(0, 1)$ considered as a subset of the plane, that is, the set $I = \{(x, y) \in \mathbb{R}^2 : 0 < x < 1,\ y = 0\}$ is not open in $\mathbb{R}^2$.

Exercise 1.33. Consider the following three metrics on $\mathbb{R}^2$: for $x = (x_1, x_2),\ y = (y_1, y_2) \in \mathbb{R}^2$,

$$\begin{aligned} d_1(x, y) &:= |x_1 - y_1| + |x_2 - y_2|, \\ d_2(x, y) &:= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}, \\ d_\infty(x, y) &:= \max\{|x_1 - y_1|, |x_2 - y_2|\}. \end{aligned}$$

We already know that d_2 defines a metric on $\mathbb{R}^2$: it is just the Euclidean metric induced by the norm $\|\cdot\|_2$.

(1) Verify that d_1 and d_∞ are also metrics on $\mathbb{R}^2$.
(2) Sketch the "unit balls" $B(\mathbf{0}, 1)$ in each of the metrics.
(3) Give a pictorial "proof without words" to show that a set U is open in $\mathbb{R}^2$ in the Euclidean metric if and only if it is open when $\mathbb{R}^2$ is equipped with the metric d_1 or the metric d_∞. *Hint:* Inside every square you can draw a circle, and inside every circle, you can draw a square!

[3]This means that we have a set I, and for each $i \in I$, there is a set U_i.

Exercise 1.34. Determine if the following statements are true or false.

(1) If a set is not open, then it is closed.
(2) If a set is open, then it is not closed.
(3) There are sets which are both open and closed.
(4) There are sets which are neither open nor closed.
(5) $\mathbb{Q}$ is open in $\mathbb{R}$.
(6) $(*)$ $\mathbb{Q}$ is closed in $\mathbb{R}$.
(7) $\mathbb{Z}$ is closed in $\mathbb{R}$.

Exercise 1.35. Show that the unit sphere with center $\mathbf{0}$ in $\mathbb{R}^3$, namely the set

$$\mathbb{S}^2 := \{x \in \mathbb{R}^3 : x_1^2 + x_2^2 + x_3^2 = 1\}$$

is closed in $\mathbb{R}^3$.

Exercise 1.36. Let (X, d) be a metric space. Show that a singleton (a subset of X containing precisely one element) is always closed. Conclude that every finite subset of X is closed.

Exercise 1.37. Let X be any nonempty set equipped with the discrete metric. Prove that every subset Y of X is both open and closed.

Exercise 1.38. $(*)$ A subset Y of a metric space (X, d) is said to be *dense* in X if for all $x \in X$ and all $\epsilon > 0$, there exists a $y \in Y$ such that $d(x, y) < \epsilon$. (That is, if we take any $x \in X$ and consider any ball $B(x, \epsilon)$ centered at x, it contains a point from Y.) Show that $\mathbb{Q}$ is dense in $\mathbb{R}$ by proceeding as follows.

If $x, y \in \mathbb{R}$ and $x < y$, then show that there is a $q \in \mathbb{Q}$ such that $x < q < y$. *Hint:* By the Archimedean property[4] of $\mathbb{R}$, there is a positive integer n such that $n(y - x) > 1$. Next there are positive integers m_1, m_2 such that $m_1 > nx$ and $m_2 > -nx$ so that $-m_2 < nx < m_1$. Hence there is an integer m such that $m - 1 \leq nx < m$. Consequently $nx < m \leq 1 + nx < ny$, which gives the desired result.

Conclude that $\mathbb{Q}$ is dense in $\mathbb{R}$.

Exercise 1.39. Is the set $\mathbb{R} \setminus \mathbb{Q}$ of irrational numbers dense in $\mathbb{R}$? *Hint:* Take any $x \in \mathbb{R}$. If x is irrational itself, then we may just take y to be x and we are done; whereas if x is rational, then take $y = x + \sqrt{2}/n$ with a sufficiently large n.

Exercise 1.40. A subset C of a normed space X is called *convex* if for all $x, y \in C$, and all $t \in (0, 1)$, $(1 - t)x + ty \in C$. (Geometrically, this means that for any pair of points in C, the "line segment" joining them also lies in C.)

(1) Show that the open ball $B(0, r)$ with center $0 \in X$ and radius $r > 0$ is convex.
(2) Is the unit circle $\mathbb{S}^1 = \{x \in \mathbb{R}^2 : \|x\|_2 = 1\}$ a convex set in $\mathbb{R}^2$?
(3) Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Prove that the "Linear Programming simplex"[5]
$$\Sigma := \{x \in \mathbb{R}^n : Ax = b,\ x_1 \geq 0,\ \ldots,\ x_n \geq 0\}$$
is a convex set in $\mathbb{R}^n$.

Exercise 1.41. Define $d_e : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ by

$$d_e(u, v) = \begin{cases} \|u\|_2 + \|v\|_2 & \text{if } u \neq v, \\ 0 & \text{if } u = v, \end{cases} \qquad (u, v \in \mathbb{R}^2).$$

We call this metric the "express railway metric". (For example in the British context, to get from A to B, travel via London, the origin.)

On the other hand, consider $d_s : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ defined by

$$d_s(u, v) = \begin{cases} \|u - v\|_2 & \text{if } u, v \text{ are linearly independent}, \\ \|u\|_2 + \|v\|_2 & \text{otherwise}, \end{cases} \qquad (u, v \in \mathbb{R}^2).$$

We call this the "stopping railway metric". (To get from A to B travel via London, unless A and B are on the same London route.)

Show that the express railway metric and the stopping railway metric are indeed metrics on $\mathbb{R}^2$.

---

[4] The Archimedean property of $\mathbb{R}$ says that if $x, y \in \mathbb{R}$ and $x > 0$, then there exists an $n \in \mathbb{N}$ such that $y < nx$. See the notes for MA103.

[5] This set arises the "feasible set" in a certain optimization problem in $\mathbb{R}^n$, where the constraints are described by a bunch of linear inequalities.

1.4. Notes (not part of the course)

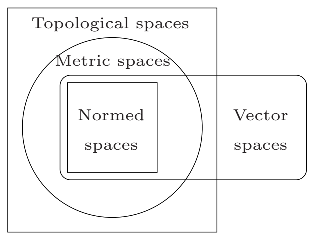
Topology. If we look at the collection $\mathcal{O}$ of open sets in a metric space (X, d), we notice that it has the following three properties:

(T1) $\emptyset, X \in \mathcal{O}$.

(T2) If U_i $(i \in I)$ is family of sets from $\mathcal{O}$ indexed by I, then $\bigcup_{i \in I} U_i \in \mathcal{O}$.

(T3) If $U_1, \dots, U_n$ is a finite collection of sets from $\mathcal{O}$, then $\bigcap_{i=1}^{n} U_i \in \mathcal{O}$.

More generally, if X is any set (not necessarily one equipped with a metric), then any collection $\mathcal{O}$ of subsets of X that satisfy the properties (T1), (T2), (T3) is called a *topology on* X and $(X, \mathcal{O})$ is called a *topological space*. So for a metric space X, if we take $\mathcal{O}$ to be family of open sets in X, then we obtain a topological space. More generally, if one has a topological space $(X, \mathcal{O})$ given by the topology $\mathcal{O}$, we call each element of $\mathcal{O}$ *open*.



It turns out that one can in fact extend some of the notions from Real Analysis (such as convergence of sequences and continuity of maps) in the even more general set up of topological spaces, devoid of any metric, where the notion of closeness is specified by considering arbitrary open neighbourhoods provided by elements of $\mathcal{O}$. In some applications this is exactly the right thing needed, but we will not go into such abstractions in this course. In fact, this is a very broad subdiscipline of mathematics called *Topology*.

Construction of the set of real numbers. In these notes, we treat the real number system $\mathbb{R}$ as a given. But one might wonder if we can take the existence of real numbers on faith alone. It turns out that a mathematical proof of its existence can be given. Roughly, we are already familiar with the natural numbers, the integers, and the rational numbers, and their rigorous mathematical construction is also relatively straightforward. However, the set $\mathbb{Q}$ of rational numbers has "holes" (for example in MA103 we have seen that this manifests itself in the fact that $\mathbb{Q}$ does not possess the least upper bound property). The set of real numbers $\mathbb{R}$ is obtained by "filling these holes". There are several ways of doing this. One is by a general method called "completion of metric spaces". Another way, which is more intuitive, is via "(Dedekind) cuts", where we view real numbers as places where a line may be cut with scissors. More precisely, a *cut* $A|B$ in $\mathbb{Q}$ is a pair of subsets A, B of $\mathbb{Q}$ such that $A \bigcup B = \mathbb{Q}$, $A \neq \emptyset$, $B \neq \emptyset$, $A \bigcap B = \emptyset$, if $a \in A$ and $b \in B$ then $a < b$, and A contains no largest element. $\mathbb{R}$ is then taken as the set of all cuts $A|B$. Here are two examples of cuts:

$$\begin{aligned} A|B &= \{r \in \mathbb{Q} : r < 1\} | \{r \in \mathbb{Q} : r \geq 1\} \\ A|B &= \{r \in \mathbb{Q} : r \leq 0 \text{ or } r^2 < 2\} | \{r \in \mathbb{Q} : r > 0 \text{ and } r^2 \geq 2\}. \end{aligned}$$

It turns out that $\mathbb{R}$ is a field containing $\mathbb{Q}$, and it possesses the least upper bound property. The interested reader is referred to the Appendix to Chapter 1 in the classic textbook by Walter Rudin [**R**].

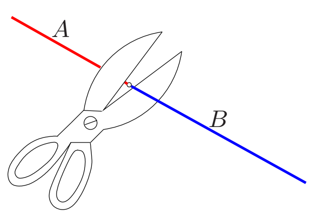


Figure 8. Dedekind cut.

Chapter 2

Sequences

In this chapter we study sequences in metric spaces. The notion of a convergent sequence is an important concept in Analysis. Besides its theoretical importance, it is also a natural concept arising in applications when one talks about better and better approximations to the solution of a problem using a numerical scheme. For example the method of Archimedes for finding the area of a circle by sandwiching it between the areas of a circumscribed and an inscribed regular polygon of ever increasing number of sides. There are also numerical schemes for finding a minimizer of a convex function (Newton's method), or for finding a solution to an ordinary differential equation (Euler's method), where convergence in more general metric spaces (such as $\mathbb{R}^n$ or $C[a,b]$) will play a role.

Before proceeding onto sequences in general metric spaces, let us first begin with (numerical) sequences in $\mathbb{R}$.

2.1. Sequences in $\mathbb{R}$

Let us recall the definition of a convergent sequence of real numbers.

Definition 2.1. A sequence $(a_n)_{n\in\mathbb{N}}$ of real numbers is said to be *convergent with limit* $L \in \mathbb{R}$ if for every $\epsilon > 0$, there exists a $N \in \mathbb{N}$ such that whenever $n > N$, $|a_n - L| < \epsilon$.

We have learnt that the limit of a convergent sequence $(a_n)_{n\in\mathbb{N}}$ is unique, and we denote it by

$$\lim_{n\to\infty} a_n.$$

We have also learnt the following important result[1]:

Theorem 2.2 (Bolzano-Weierstrass Theorem)**.** *Every bounded sequence has a convergent subsequence.*

An important consequence of this result is the fact that in $\mathbb{R}$, the set of convergent sequences coincides with the set of Cauchy sequences. Let us first recall the definition of a Cauchy sequence[2].

Definition 2.3. A sequence $(a_n)_{n\in\mathbb{N}}$ of real numbers is said to be a *Cauchy sequence* if for every $\epsilon > 0$, there exists a $N \in \mathbb{N}$ such that whenever $m, n > N$, $|a_n - a_m| < \epsilon$.

Roughly speaking, we can make the terms of the sequence arbitrarily close to each other provided we go far enough in the sequence.

[1] See Theorem 1.2.8 on page 25 of the lecture notes for the second part of MA103.
[2] See Exercise 6 on page 13 of the lecture notes for the second part of MA103.

Example 2.4. The sequence $(\frac{1}{n})_{n\in\mathbb{N}}$ is Cauchy. Indeed, we have $\left|\frac{1}{n}-\frac{1}{m}\right| \le \frac{1}{n}+\frac{1}{m} < \frac{1}{N}+\frac{1}{N}=\frac{2}{N}$ whenever $n,m>N$. Thus given $\epsilon>0$, we can choose $N\in\mathbb{N}$ larger than $\frac{2}{\epsilon}$ so that we then have $|\frac{1}{n}-\frac{1}{m}|<\frac{2}{N}<\epsilon$ for all $n,m>N$. Consequently, $(\frac{1}{n})_{n\in\mathbb{N}}$ is Cauchy. ♢

Exercise 2.5. Show that if $(a_n)_{n\in\mathbb{N}}$ is a Cauchy sequence, then $(a_{n+1}-a_n)_{n\in\mathbb{N}}$ converges to 0.

Example 2.6. This example shows that for a sequence $(a_n)_{n\in\mathbb{N}}$ to be Cauchy, it is not enough that $(a_{n+1}-a_n)_{n\in\mathbb{N}}$ converges to 0. Take $a_n:=\sqrt{n}$ $(n\in\mathbb{N})$. Then

$$a_{n+1}-a_n=\sqrt{n+1}-\sqrt{n}=\frac{1}{\sqrt{n+1}+\sqrt{n}}\stackrel{n\to\infty}{\longrightarrow}0,$$

but $(a_n)_{n\in\mathbb{N}}$ is not Cauchy, since for any $n\in\mathbb{N}$, $a_{4n}-a_n=\sqrt{4n}-\sqrt{n}=\sqrt{n}\ge 1$. ♢

The next result says that Cauchyness is a *necessary* condition for a sequence to be convergent.

Lemma 2.7. *Every convergent sequence is Cauchy.*

Proof. Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of real numbers that converges to L. Let $\epsilon>0$. Then there exists an $N\in\mathbb{N}$ such that $|a_n-L|<\frac{\epsilon}{2}$. Thus for $n,m>N$, we have

$$|a_n-a_m|=|a_n-L+L-a_m|\le|a_n-L|+|a_m-L|<\frac{\epsilon}{2}+\frac{\epsilon}{2}=\epsilon.$$

So the sequence $(a_n)_{n\in\mathbb{N}}$ is a Cauchy sequence. □

Now we will prove the remarkable fact in $\mathbb{R}$, Cauchyness turns out to be also a *sufficient* condition for the sequence to be convergent. In other words, in $\mathbb{R}$, every Cauchy sequence is convergent. This is a very useful fact since, in order to prove that a sequence is convergent using the definition, we would need to guess what the limit is. In contrast, checking whether or not a sequence is Cauchy needs only knowledge of the terms of the sequence, and no guesswork regarding the limit is needed. So this is a powerful technique for proving existence results.

Theorem 2.8. *Every Cauchy sequence in* $\mathbb{R}$ *is convergent.*

Proof. There are three main steps. First we show that every Cauchy sequence is bounded. Then we use the Bolzano-Weierstrass theorem to conclude that it must have a convergent subsequence. Finally we show that a Cauchy sequence having a convergent subsequence must itself be convergent.

Step 1. Suppose that $(a_n)_{n\in\mathbb{N}}$ is a Cauchy sequence. Choose any positive ϵ, say $\epsilon=1$. Then there exists an $N\in\mathbb{N}$ such that for all $n,m>N$, $|a_n-a_m|<\epsilon$. In particular, with $m=N+1>N$, and $n>N$, $|a_n-a_{N+1}|<\epsilon$. Hence by the triangle inequality, for all $n>N$,

$$|a_n|=|a_n-a_{N+1}+a_{N+1}|\le|a_n-a_{N+1}|+|a_{N+1}|<1+|a_{N+1}|.$$

On the other hand, for $n\le N$, $|a_n|\le\max\{|a_1|,\dots,|a_N|,|a_{N+1}|+1\}=:M$. Consequently, $|a_n|\le M$ $(n\in\mathbb{N})$, that is, the sequence $(a_n)_{n\in\mathbb{N}}$ is bounded.

Step 2.By the Bolzano-Weierstrass Theorem, $(a_n)_{n\in\mathbb{N}}$ has a convergent subsequence $(a_{n_k})_{k\in\mathbb{N}}$ that is convergent, to L, say.

Step 3. Finally we show that $(a_n)_{n\in\mathbb{N}}$ is also convergent with limit L. Let $\epsilon>0$. Then there exists an $N\in\mathbb{N}$ such that for all $n,m>N$,

$$|a_n-a_m|<\frac{\epsilon}{2}. \tag{2.1}$$

Also, since $(a_{n_k})_{k\in\mathbb{N}}$ converges to L, we can find a $n_K>N$ such that $|a_{n_K}-L|<\frac{\epsilon}{2}$. Taking $m=n_K$ in (2.1), we have for all $n>N$ that

$$|a_n-L|=|a_n-a_{n_K}+a_{n_K}-L|\le|a_n-a_{n_K}|+|a_{n_K}-L|<\frac{\epsilon}{2}+\frac{\epsilon}{2}=\epsilon.$$

Thus $(a_n)_{n\in\mathbb{N}}$ is also convergent with limit L, and this completes the proof. □

Exercise 2.9. Determine if the following statements are true or false.

(1) Every subsequence of a convergent real sequence is convergent.

(2) Every subsequence of a divergent real sequence is divergent.

(3) Every subsequence of a bounded real sequence is bounded.

(4) Every subsequence of an unbounded real sequence is unbounded.

(5) Every subsequence of a monotone real sequence is monotone.

(6) Every subsequence of a nonmonotone real sequence is nonmonotone.

(7) If every subsequence of a real sequence converges, the sequence itself converges.

(8) If for a real sequence $(a_n)_{n\in\mathbb{N}}$, the sequences $(a_{2n})_{n\in\mathbb{N}}$ and $(a_{2n+1})_{n\in\mathbb{N}}$ both converge, then $(a_n)_{n\in\mathbb{N}}$ converges.

(9) If for a real sequence $(a_n)_{n\in\mathbb{N}}$, the sequences $(a_{2n})_{n\in\mathbb{N}}$ and $(a_{2n+1})_{n\in\mathbb{N}}$ both converge to the same limit, then $(a_n)_{n\in\mathbb{N}}$ converges.

Exercise 2.10. Fill in the blanks in the following proof of the fact that *every bounded increaing sequence of real numbers converges.*

Let $(a_n)_{n\in\mathbb{N}}$ be a bounded increasing sequence of real numbers. Let M be the ______ upper bound of the set of upper bounds of $\{a_n : n \in \mathbb{N}\}$. The existence of M is guaranteed by the ______ of the set of real numbers. We show that M is the ______ of $(a_n)_{n\in\mathbb{N}}$. Taking $\epsilon > 0$, we must show that there exists a positive integer N such that ______ for all $n > N$. Since $M - \epsilon < M$, $M - \epsilon$ is not ______ of $\{a_n : n \in \mathbb{N}\}$. Therefore there exists N with ______ $\geq a_N >$ ______. Since $(a_n)_{n\in\mathbb{N}}$ is ______, $|a_n - M| < \epsilon$ for all $n \geq N$. □

Exercise 2.11. (*ast*) Consider the sequence

$$\left(1+\frac{1}{1}\right)^1, \left(1+\frac{1}{2}\right)^2, \left(1+\frac{1}{3}\right)^3, \dots .$$

(1) Show that the sequence is increasing.

(2) Prove that each term in the sequence is smaller than 3.

Hint: Use the Binomial Theorem.

From the previous exercise, one can conclude that the sequence converges. In fact it can be shown that

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = e := 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \dots .$$

Exercise 2.12. Circle the convergent sequences in the following list, and find the limit in case the sequence converges:

(1) $(\cos(\pi n))_{n\in\mathbb{N}}$

(2) $(1+n^2)_{n\in\mathbb{N}}$

(3) $\left(\frac{\sin n}{n}\right)_{n\in\mathbb{N}}$

(4) $\left(1-\frac{3n^2}{n+1}\right)_{n\in\mathbb{N}}$

(5) $\left(n^{\frac{1}{n}}\right)_{n\in\mathbb{N}}$

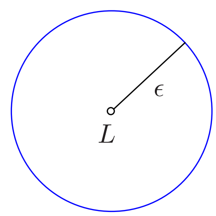
(6) $0.9,\ 0.99,\ 0.999,\ \dots$

2.2. Sequences in metric spaces

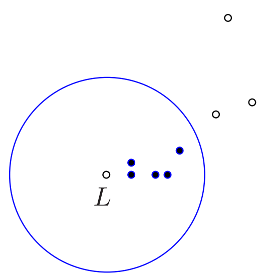
We now give the notion of convergence of a sequence in a general metric space. We will see that essentially the definition is the same as in $\mathbb{R}$, except that instead of having the distance between the nth term and the limit L given by $|a_n - L|$, now we will replace it by $d(a_n, L)$ in a general metric space with metric d.

Definition 2.13. A sequence $(a_n)_{n\in\mathbb{N}}$ of points in a metric space (X, d) is said to be *convergent with limit* $L \in X$ if for every $\epsilon > 0$, there exists a $N \in \mathbb{N}$ such that whenever $n > N$, $d(a_n, L) < \epsilon$.

Let us understand this definition pictorially. We have been given a sequence $(a_n)_{n\in\mathbb{N}}$ of points and a candidate L for its limit. We are allowed to say that this sequence converges to L if given any $\epsilon > 0$, that is, no matter how small a ball we consider around L,



we can find an index N such that all the terms of the sequence beyond this index lie inside the ball.



Lemma 2.14. *The limit of a convergent sequence in a metric space is unique.*

Proof. Suppose that $(a_n)_{n\in\mathbb{N}}$ is a convergent sequence, and let it have two distinct limits L_1 and L_2. Then $d(L_1, L_2) > 0$. Set

$$\epsilon = \frac{1}{2} d(L_1, L_2) > 0.$$

Then there exists an N_1 such that for all $n > N_1$, $d(a_n, L_1) < \epsilon$. Also, there exists an N_2 such that for all $n > N_2$, $d(a_n, L_2) < \epsilon$. Hence for any $n > \max\{N_1, N_2\}$, we have

$$d(L_1, L_2) \le d(L_1, a_n) + d(a_n, L_2) < \epsilon + \epsilon = d(L_1, L_2),$$

a contradiction. Thus the limit of $(a_n)_{n\in\mathbb{N}}$ is unique. □

If $(a_n)_{n\in\mathbb{N}}$ is a convergent sequence, then we will denote its limit by $\lim\limits_{n\to\infty} a_n$.

Exercise 2.15. Let $(a_n)_{n\in\mathbb{N}}$ be a sequence in the Euclidean space $\mathbb{R}^d$. Show that $(a_n)_{n\in\mathbb{N}}$ is convergent with limit L if and only if for every $k \in \{1, \dots, d\}$, the sequence $(a_n^{(k)})_{n\in\mathbb{N}}$ in $\mathbb{R}$ formed by the kth component of the terms of $(a_n)_{n\in\mathbb{N}}$ is convergent with limit $L^{(k)}$. (Here we use the notation $v^{(k)}$ to mean the kth component of a vector $v \in \mathbb{R}^d$.)

Exercise 2.16. Consider the sequence $(a_n)_{n\in\mathbb{N}}$ in the Euclidean space $\mathbb{R}^2$:

$$a_n := \begin{bmatrix} \dfrac{n}{4n+2} \\ \dfrac{n^2}{n^2+1} \end{bmatrix} \quad (n \in \mathbb{N}).$$

Show that $(a_n)_{n\in\mathbb{N}}$ is convergent. What is its limit?

Exercise 2.17. Let X be a nonempty set equipped with the discrete metric. Show that a sequence $(a_n)_{n\in\mathbb{N}}$ is convergent if and only if it is eventually a constant sequence (that is, there is a $c \in X$ and an $N \in \mathbb{N}$ such that for all $n > N$, $a_n = c$).

Exercise 2.18. (∗) Let (X, d) be a metric space and let $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ be convergent sequences in X with limits a and b, respectively. Prove that $(d(a_n, b_n))_{n\in\mathbb{N}}$ is a convergent sequence in $\mathbb{R}$ with limit $d(a, b)$. *Hint:* $d(a_n, b_n) \leq d(a_n, a) + d(a, b) + d(b, b_n)$.

Exercise 2.19. Let $v_1 = (x_1, y_1) \in \mathbb{R}^2$ be such that $0 < x_1 < y_1$. Define

$$v_{n+1} := (x_{n+1}, y_{n+1}) := \left(\sqrt{x_n y_n}, \frac{x_n + y_n}{2}\right) \quad (n \in \mathbb{N}).$$

(1) Show that $0 < x_n < x_{n+1} < y_{n+1} < y_n$ and that $y_{n+1} - x_{n+1} < y_{n+1} - x_n = \dfrac{y_n - x_n}{2}$.

(2) Conclude that $\lim\limits_{n\to\infty} v_n$ exists and equals (c, c) for some number $c \in \mathbb{R}$. This value c is called the arithmetic-geometric mean[3], of x_1 and y_1, and is denoted by $\text{agm}(x_1, y_1)$.

We can also define Cauchy sequences in a metric space analogous to the situation in $\mathbb{R}$.

Definition 2.20. A sequence $(a_n)_{n\in\mathbb{N}}$ of points in a metric space (X, d) is said to be a *Cauchy sequence* if for every $\epsilon > 0$, there exists a $N \in \mathbb{N}$ such that whenever $n > N$, $d(a_n, a_m) < \epsilon$.

Lemma 2.21. *Every convergent sequence is Cauchy.*

Proof. The proof is the same, mutatis mutandis[4], as the proof of Lemma 2.7. Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of points in X that converges to $L \in X$. Let $\epsilon > 0$. Then there exists an $N \in \mathbb{N}$ such that $d(a_n, L) < \frac{\epsilon}{2}$. Thus for $n, m > N$, we have $d(a_n, a_m) \leq d(a_n, L) + d(L, a_m) < \dfrac{\epsilon}{2} + \dfrac{\epsilon}{2} = \epsilon$. So the sequence $(a_n)_{n\in\mathbb{N}}$ is a Cauchy sequence. □

Just as in $\mathbb{R}$, it would be useful to know if also in arbitrary metric spaces, the set of convergent sequences coincides with the set of Cauchy sequences. Unfortunately, this is not always true. For example, consider the metric space $X = (0, 1]$ with the same Euclidean metric as used in $\mathbb{R}$. Then the sequence $(1/n)_{n\in\mathbb{N}}$ is easily seen to be Cauchy, but is not convergent in X, as there is a missing point in X, namely 0. However, in some other metric spaces, such as $\mathbb{R}$, the set of convergent sequences and the set of Cauchy sequences do coincide. So it makes sense to give such metric spaces a special name: they are called "complete".

Definition 2.22. A metric space in which every Cauchy sequence converges is called *complete*.

Example 2.23. $\mathbb{R}$ with the Euclidean metric is complete. ◊

Exercise 2.24. (∗) Show that $\mathbb{Q}$ with the Euclidean metric is not complete. *Hint:* Revisit the solution to part (6) of Exercise 1.34.

Exercise 2.25. Let X be a metric space. If $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in X which has a convergent subsequence $(x_{n_k})_{k\in\mathbb{N}}$ with limit L, then show that $(x_n)_{n\in\mathbb{N}}$ is convergent with the same limit L.

Theorem 2.26. $\mathbb{R}^d$ *is complete.*

Proof. (Essentially, this is because $\mathbb{R}$ is complete, and one has d copies of $\mathbb{R}$ in $\mathbb{R}^d$.) Suppose that $(a_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}^d$:

$$a_n = \begin{bmatrix} x_n^{(1)} \\ \vdots \\ x_n^{(d)} \end{bmatrix}.$$

---

[3]Gauss observed that $I(a, b) := \displaystyle\int_{-\infty}^{\infty} \frac{dx}{\sqrt{(x^2 + a^2)(y^2 + b^2)}}$ satisfies $I(a, b) = I\left(\frac{a+b}{2}, \sqrt{ab}\right)$ with the help of the substitution $t = \frac{1}{2}\left(x - \frac{ab}{x}\right)$, and using this, obtained the remarkable result that $\displaystyle\int_{-\infty}^{\infty} \frac{dx}{\sqrt{(x^2 + a^2)(y^2 + b^2)}} = \frac{\pi}{\text{agm}(a, b)}$.

[4]Latin phrase meaning "by changing those things which need to be changed".

We have the inequalities

$$|x_n^{(k)} - x_m^{(k)}| \le \|a_n - a_m\|_2 \quad (n, m \in \mathbb{N},\ k = 1, \dots, d),$$

from which it follows that each of the sequences $(x_n^{(k)})_{n\in\mathbb{N}}$, $k = 1, \dots, d$, is Cauchy in $\mathbb{R}$, and hence convergent, with respective limits, say $L^{(1)}, \dots, L^{(d)} \in \mathbb{R}$. So given $\epsilon > 0$, there exists a large enough N such that whenever $n > N$, we have

$$|x_n^{(k)} - L^{(k)}| < \frac{\epsilon}{\sqrt{d}} \quad (k = 1, \dots, d).$$

Set

$$L = \begin{bmatrix} L^{(1)} \\ \vdots \\ L^{(d)} \end{bmatrix} \in \mathbb{R}^d.$$

Thus for $n > N$,

$$\|a_n - L\|_2 = \sqrt{\sum_{k=1}^{d} |x_n^{(k)} - L^{(k)}|^2} < \sqrt{\sum_{k=1}^{d} \frac{\epsilon^2}{d}} = \epsilon.$$

Consequently, the sequence $(a_n)_{n\in\mathbb{N}}$ converges to L. □

Exercise 2.27. $\mathbb{R}^{n\times m}$ with the metric induced by $\|\cdot\|_\infty$ is complete. (See Exercise 1.13 for the definition of the norm $\|\cdot\|_\infty$ on the vector space $\mathbb{R}^{n\times m}$.)

The following theorem is an important result, and lies at the core of a result on the existence of solutions for Ordinary Differential Equations (ODEs). You can learn more about this in the course Differential Equations (MA209). (See Exercise 1.14 for the definition of the norm $\|\cdot\|_\infty$ on $C[a, b]$.)

Theorem 2.28. $(*)$ *$C[a, b]$ with the metric induced by $\|\cdot\|_\infty$ is complete.*

Proof. (*You may skip this proof.*) The idea behind the proof is similar to the proof of the completeness of $\mathbb{R}^n$. If $(f_n)_{n\in\mathbb{N}}$ is a Cauchy sequence, then we think of the $f_n(x)$ as being the "components" of f_n indexed by $x \in [a, b]$. We first freeze an $x \in [a, b]$, and show that $(f_n(x))_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$, and hence convergent to a number (which depends on x), and which we denote by $f(x)$. Next we show that the function $x \mapsto f(x)$ is continuous, and finally that $(f_n)_{n\in\mathbb{N}}$ does converge to f.

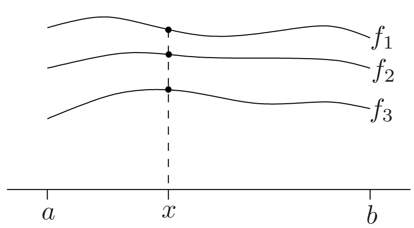


Figure 1. The Cauchy sequence $(f_n(x))_{n\in\mathbb{N}}$ obtained from the Cauchy sequence $(f_n)_{n\in\mathbb{N}}$ by freezing x.

Let $(f_n)_{n\in\mathbb{N}}$ be a Cauchy sequence. Let $x \in [a, b]$. We claim that $(f_n(x))_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$. Let $\epsilon > 0$. Then there exists an $N \in \mathbb{N}$ such that for all $n, m > N$, $\|f_n - f_m\|_\infty < \epsilon$. But

$$|f_n(x) - f_m(x)| \le \max_{y\in[a,b]} |f_n(y) - f_m(y)| = \|f_n - f_m\|_\infty < \epsilon,$$

for $n, m > N$. This shows that indeed $(f_n(x))_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$. But $\mathbb{R}$ is complete, and so the Cauchy sequence $(f_n(x))_{n\in\mathbb{N}}$ is in fact convergent, with a limit which depends on which

$x \in [a, b]$ we had frozen at the outset. To highlight this dependence on x, we denote the limit of $(f_n(x))_{n\in\mathbb{N}}$ by $f(x)$. (Thus $f(a)$ is the number which is the limit of the convergent sequence $(f_n(a))_{n\in\mathbb{N}}$, $f(b)$ is the number which is the limit of the convergent sequence $(f_n(b))_{n\in\mathbb{N}}$, and so on.) So we have a function

$$x \text{ is sent to the number which is the limit of the convergent sequence } (f_n(x))_{n\in\mathbb{N}}$$

from $[a, b]$ to $\mathbb{R}$. We call this function f. This will serve as the limit of the sequence $(f_n)_{n\in\mathbb{N}}$. But first we have to see if it belongs to $C[a, b]$, that is, we need to check that this f is continuous on $[a, b]$.

Let $x \in [a, b]$. We will show that f is continuous at x. Recall that in order to do this, we have to show that for each $\epsilon > 0$, there exists a $\delta > 0$ such that whenever $|y - x| < \delta$, we have $|f(y) - f(x)| < \epsilon$. Let $\epsilon > 0$. Choose N large enough so that for all $n, m > N$,

$$\|f_n - f_m\|_\infty < \frac{\epsilon}{3}.$$

Let $y \in [a, b]$. Then for $n > N$, $|f_n(y) - f_{N+1}(y)| \le \|f_n - f_{N+1}\|_\infty < \frac{\epsilon}{3}$. Now let $n \to \infty$:

$$|f(y) - f_{N+1}(y)| = \lim_{n\to\infty} |f_n(y) - f_{N+1}(y)| \le \frac{\epsilon}{3}.$$

As the choice of $y \in [a, b]$ was arbitrary, we have for all $y \in [a, b]$ that

$$|f(y) - f_{N+1}(y)| \le \frac{\epsilon}{3}.$$

Now $f_{N+1} \in C[a, b]$. So there exists a $\delta > 0$ such that whenever $|y - x| < \delta$, we have

$$|f_{N+1}(y) - f_{N+1}(x)| < \frac{\epsilon}{3}.$$

Thus whenever $|y - x| < \delta$, we have

$$\begin{aligned} |f(y) - f(x)| &= |f(y) - f_{N+1}(y) + f_{N+1}(y) - f_{N+1}(x) + f_{N+1}(x) - f(x)| \\ &\le |f(y) - f_{N+1}(y)| + |f_{N+1}(y) - f_{N+1}(x)| + |f_{N+1}(x) - f(x)| \\ &\le \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon. \end{aligned}$$

This shows that f is continuous at x. As the choice of $x \in [a, b]$ was arbitrary, f is continuous on $[a, b]$.

Finally, we show that $(f_n)_{n\in\mathbb{N}}$ does converge to f. Let $\epsilon > 0$. Choose N large enough so that for all $n, m > N$, $\|f_n - f_m\|_\infty < \epsilon$. Fix $n > N$. Let $x \in [a, b]$. Then for all $m > N$, $|f_n(x) - f_m(x)| \le \|f_n - f_m\|_\infty < \epsilon$. Thus

$$|f_n(x) - f(x)| = \lim_{m\to\infty} |f_n(y) - f_m(y)| \le \epsilon.$$

But $x \in [a, b]$ was arbitrary. Hence

$$\|f_n - f\|_\infty = \max_{x\in[a,b]} |f_n(x) - f(x)| \le \epsilon.$$

But we could have fixed any $n > N$ at the outset and obtained the same result. So we have that for all $n > N$, $\|f_n - f\|_\infty \le \epsilon$. Thus

$$\lim_{n\to\infty} f_n = f,$$

and this completes the proof. □

The norm $\|\cdot\|_\infty$ is special in that $C[a, b]$ is complete with the corresponding induced metric. It turns out that $C[a, b]$ with the other natural norm met earlier, namely the $\|\cdot\|_1$-norm is not complete. The objective in the following exercise is to demonstrate this.

Exercise 2.29. ($*$) Let $C[0,1]$ be equipped with the $\|\cdot\|_1$-norm given by

$$\|f\|_1 := \int_0^1 |f(x)|dx \quad (f \in C[0,1]).$$

Show that the corresponding metric space is not complete. For example, you may consider the sequence $(f_n)_{n\in\mathbb{N}}$ with the f_n as shown in Figure 2. Show that

$$\|f_n - f_m\|_1 = \int_{\frac{1}{2}}^{\frac{1}{2}+\max\{\frac{1}{n+1},\frac{1}{m+1}\}} |f_n(x) - f_m(x)|dx \le \frac{2}{N},$$

for $n, m > N$, and so $(f_n)_{\in\mathbb{N}}$ is Cauchy. Next, prove that if $(f_n)_{n\in\mathbb{N}}$ converges to $f \in C[0,1]$, then f must satisfy

$$f(x) = \begin{cases} 0 & \text{for } x \in [0, \frac{1}{2}], \\ 1 & \text{for } x \in (\frac{1}{2}, 1], \end{cases}$$

which does not belong to $C[0,1]$, a contradiction.

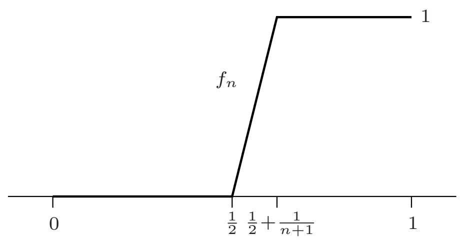


Figure 2. f_n.

Exercise 2.30. Show that any nonempty set X equipped with the discrete metric is complete.

Exercise 2.31. Prove that $\mathbb{Z}$ equipped with the Euclidean metric induced from $\mathbb{R}$ is complete.

2.3. Pointwise versus uniform convergence

Convergence in $(C[a,b], \|\cdot\|_\infty)$ is referred to as *uniform convergence*. More generally, we have the following definition.

Definition 2.32. Let X be any set and $f, f_n : X \to \mathbb{R}$ ($n \in \mathbb{N}$) be functions.

(1) The sequence $(f_n)_{n\in\mathbb{N}}$ is said to *converge uniformly to* f if

$$\forall \epsilon > 0,\ \exists N \in \mathbb{N} \text{ such that } \forall n > N,\ \forall x \in X,\ |f_n(x) - f(x)| < \epsilon.$$

(2) The sequence $(f_n)_{n\in\mathbb{N}}$ is said to *converge pointwise to* f if

$$\forall \epsilon > 0,\ \forall x \in X,\ \exists N \in \mathbb{N} \text{ such that } \forall n > N,\ |f_n(x) - f(x)| < \epsilon.$$

Consider the two statements:

$$\forall \epsilon > 0,\ \exists N \in \mathbb{N} \text{ such that } \forall n > N,\ \forall x \in X,\ |f_n(x) - f(x)| < \epsilon.$$

and

$$\forall \epsilon > 0,\ \forall x \in X,\ \exists N \in \mathbb{N} \text{ such that } \forall n > N,\ |f_n(x) - f(x)| < \epsilon.$$

Can you spot the difference? What has changed is the order of

$$\boxed{\forall x \in X}$$

and

$$\boxed{\exists N \in \mathbb{N} \text{ such that } \forall n > N.}$$

This seemingly small change makes a world of difference. Indeed, even in everyday language, the two statements:

There exists a faculty member B such that for every student A, B is a personal tutor of A.

and

For every student A, there exists a faculty member B such that B is a personal tutor of A.

clearly mean two quite different things, where the former says that the faculty member B is personal tutor to *all* students, while in the latter statement, the faculty member who is claimed to exist may *depend on* which student A is chosen.

This is the same sort of a difference between the uniform convergence requirement, namely:

$$\forall \epsilon > 0,\ \exists N \in \mathbb{N} \text{ such that } \forall n > N,\ \forall x \in X,\ |f_n(x) - f(x)| < \epsilon.$$

and the pointwise convergence requirement, namely

$$\forall \epsilon > 0,\ \forall x \in X,\ \exists N \in \mathbb{N} \text{ such that } \forall n > N,\ |f_n(x) - f(x)| < \epsilon.$$

In the former, the same N works for all $x \in X$, while in the latter, the N might depend on the x in question.

It is clear that if f_n converges uniformly to f, then f_n converges pointwise to f, but there are pointwise convergent sequences of functions which do not converge uniformly. Here is an example to illustrate this.

Example 2.33. Consider $f_n : (0,1) \to \mathbb{R}$ $(n \in \mathbb{N})$ given by $f_n(x) = x^n$ $(x \in (0,1),\ n \in \mathbb{N})$. Then $(f_n)_{n\in\mathbb{N}}$ converges pointwise to the zero function f, defined by $f(x) = 0$ $(x \in (0,1))$. Indeed, for each $x \in (0,1)$,

$$\lim_{n\to\infty} x^n = 0,$$

and so the following statement is true:

$$\forall \epsilon > 0,\ \forall x \in (0,1),\ \exists N \in \mathbb{N} \text{ such that } \forall n > N,\ |f_n(x) - f(x)| < \epsilon.$$

If fact, we can choose $N > \dfrac{\log \epsilon}{\log x}$ so that $x^N < \epsilon$, and if we have $n > N$, we are guaranteed that

$$|f_n(x) - f(x)| = |x^n - 0| = x^n < x^N < \epsilon.$$

It is clear that our choice of N in the above depends not only on ϵ, but also on the point $x \in (0,1)$. The closer x is to 1, the larger N is. The question arises: Is there an N so that for $n > N$, $|f_n(x) - f(x)| < \epsilon$ for *all* $x \in (0,1)$? We will show below that the answer is "no". In other words, $(f_n)_{n\in\mathbb{N}}$ does not converge uniformly to the zero function f.

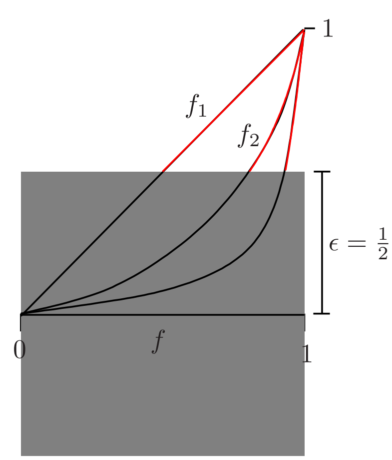


Figure 3. $(x^n)_{n\in\mathbb{N}}$ converges pointwise, but not uniformly, to the zero function on $(0,1)$.

For example, let $\epsilon = 1/2 > 0$. Let us suppose that there does exist an $N \in \mathbb{N}$ such that for all $n > N$,

$$\forall x \in (0,1),\ |f_n(x) - f(x)| = x^n < \frac{1}{2} = \epsilon.$$

In particular, we would have $\forall x \in (0,1)$, $x^{N+1} < \frac{1}{2}$. Take $x = 1 - \frac{1}{m}$, $m \in \mathbb{N}$, $m \geq 2$. Then

$$\left(1 - \frac{1}{m}\right)^{N+1} < \frac{1}{2}.$$

Letting $m \to \infty$, we obtain $1 \leq \frac{1}{2}$, a contradiction.

See Figure 3, which explains this visually. If $(f_n)_{n\in\mathbb{N}}$ were to converge to the zero function uniformly, then in particular, for all n large enough, the graph of f_n would lie in a strip of width $\epsilon = 1/2$ around the graph of the zero function. But no matter how large an n we take, some part of the graph of f_n always falls outside the strip. $\Diamond$

To test whether a sequence $(f_n)_{n\in\mathbb{N}}$ is uniformly convergent, first find its pointwise limit (if it exists), and then check to see whether the convergence is uniform.

Exercise 2.34. Suppose that X be a set and $f_n : X \to \mathbb{R}$ ($n \in \mathbb{N}$) be a sequence which is pointwise convergent to $f : X \to \mathbb{R}$. Let the numbers $a_n := \sup\{|f_n(x) - f(x)| : x \in X\}$ ($n \in \mathbb{N}$) all exist. Prove that $(f_n)_{n\in\mathbb{N}}$ converges uniformly to f if and only if $\lim_{n\to\infty} a_n = 0$.

Let $f_n : (0,\infty) \to \mathbb{R}$ be given by $f_n(x) = xe^{-nx}$ for $x \in (0,\infty)$ and $n \in \mathbb{N}$. Show that the sequence $(f_n)_{n\in\mathbb{N}}$ converges uniformly on $(0,\infty)$.

Exercise 2.35. Let $f_n : [0,1] \to \mathbb{R}$ be defined by

$$f_n(x) = \frac{x}{1+nx} \quad (x \in [0,1]).$$

Does $(f_n)_{n\in\mathbb{N}}$ converge uniformly on $[0,1]$?

Exercise 2.36. Let $f_n : \mathbb{R} \to \mathbb{R}$ be defined by

$$f_n(x) = 1 - \frac{1}{(1+x^2)^n} \quad (x \in \mathbb{R},\ n \in \mathbb{N}).$$

Show that the sequence $(f_n)_{n\in\mathbb{N}}$ of continuous functions converges pointwise to the function

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

which is discontinuous at 0.

Uniform convergence often implies that the limit function inherits properties possessed by the terms of the sequence. For instance, we will see later on that if a sequence of continuous functions converges uniformly to a function f, then f is also continuous; see Proposition 4.15. Morally, the reason nice things can happen with uniform convergence is that we can exchange two limiting processes, which is not allowed always when one just has pointwise convergence. The following exercises demonstrate the precariousness of exchanging limiting processes arbitrarily.

Exercise 2.37. For $n \in \mathbb{N}$ and $m \in \mathbb{N}$, set $a_{m,n} = \frac{m}{m+n}$.

Show that for each fixed n, $\lim_{m\to\infty} a_{m,n} = 1$, while for each fixed m, $\lim_{n\to\infty} a_{m,n} = 0$.

Is $\lim_{m\to\infty} \lim_{n\to\infty} a_{m,n} = \lim_{n\to\infty} \lim_{m\to\infty} a_{m,n}$?

Exercise 2.38. Let $f_n : \mathbb{R} \to \mathbb{R}$ be defined by

$$f_n(x) = \frac{\sin(nx)}{\sqrt{n}} \quad (x \in \mathbb{R},\ n \in \mathbb{N}).$$

Show that $(f_n)_{n\in\mathbb{N}}$ converges pointwise to the zero function f. However, show that $(f_n')_{n\in\mathbb{N}}$ does not converge pointwise to (the zero function) f'.

Exercise 2.39. Let $f_n : [0,1] \to \mathbb{R}$ $(n \in \mathbb{N})$ be defined by $f_n(x) = nx(1-x^2)^n$ $(x \in [0,1])$. Show that $(f_n)_{n\in\mathbb{N}}$ converges pointwise to the zero function f. However, show that

$$\lim_{n\to\infty} \int_0^1 f_n(x)dx = \frac{1}{2} \neq 0 = \int_0^1 \lim_{n\to\infty} f(x)dx.$$

2.4. Convergent sequences and closed sets

We have learnt that closed sets are ones whose complement is open. Here is another characterization of closed sets.

Theorem 2.40. *A set F is closed if and only if for every convergent sequence $(a_n)_{n\in\mathbb{N}}$ such that $a_n \in F$ $(n \in \mathbb{N})$, we have that $\lim\limits_{n\to\infty} a_n \in F$.*

Proof. Suppose that F is closed. Let $(a_n)_{n\in\mathbb{N}}$ be a convergent sequence such that $a_n \in F$ $(n \in \mathbb{N})$ and denote its limit by L. Assume that $L \notin F$. Then $L \in \complement F$, the complement of F, which is open. So there exists an $r > 0$ such that the open ball $B(L,r)$ with center L and radius $r > 0$ is contained in $\complement F$, that is, $B(L,r)$ contains no points from F. But as $(a_n)_{n\in\mathbb{N}}$ is convergent with limit L, we can choose a large enough n so that $d(a_n, L) < r/2$. This implies that $a_n \in B(L,r)$. But also $a_n \in F$, and so we have arrived at a contradiction. See Figure 4. This shows the "only if" part.

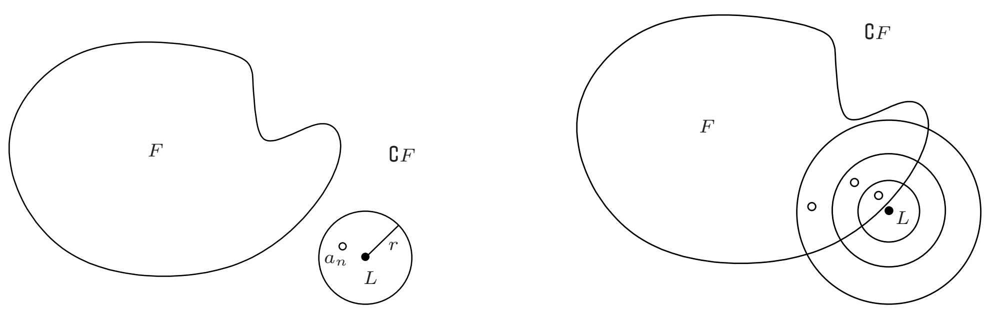


Figure 4. Proof of the theorem on the characterization of closed sets in terms of convergent sequences. The figure on the left is the one for the "only if" part, and the one on the right is for the "if" part.

Now suppose that the set is not closed. Then its complement $\complement F$ is not open. This means that there is a point $L \in \complement F$ such that for every $r > 0$, the open ball $B(L,r)$ has at least one point from F. Now take successively $r = 1/n$ $(n \in \mathbb{N})$, and choose a point $a_n \in F \cap B(L, 1/n)$. In this manner we obtain a sequence $(a_n)_{n\in\mathbb{N}}$ such that $a_n \in F$ for each n, and $d(a_n, L) < 1/n$. The property $d(a_n, L) < 1/n$ $(n \in \mathbb{N})$ implies that $(a_n)_{n\in\mathbb{N}}$ is a convergent sequence with limit L. So we have obtained the existence of a convergent sequence $(a_n)_{n\in\mathbb{N}}$ such that $a_n \in F$ $(n \in \mathbb{N})$, but

$$\lim_{n\to\infty} a_n = L \notin F.$$

See Figure 4. This completes the proof of the "if" part. □

Exercise 2.41.

(1) Let $0 \neq a \in \mathbb{R}^d$ and $\beta \in \mathbb{R}$. Show that the "hyperplane" $H = \{x \in \mathbb{R}^d : a^\top x = \beta\}$ is closed in $\mathbb{R}^d$.

(2) Prove that if $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$, then the set of solutions $S = \{x \in \mathbb{R}^n : Ax = b\}$ is a closed subset of $\mathbb{R}^n$.

(3) Show that the "Linear Programming simplex" $\Sigma := \{x \in \mathbb{R}^n : Ax = b,\ x_1 \geq 0,\ \ldots,\ x_n \geq 0\}$ is closed in $\mathbb{R}^n$.

2.5. Compact sets

In this section, we study an important class of subsets of a metric space, called compact sets. Before we learn the definition, let us give some motivation for this concept.

Of the different types of intervals in $\mathbb{R}$, perhaps the most important are those of the form $[a, b]$, where a, b are finite real numbers. Why are such intervals so important? This is not an easy question to answer, but we already know of one vital result, namely the Extreme Value Theorem, where such intervals play a vital role. Recall that the Extreme Value Theorem asserts that any continuous function $f : [a, b] \to \mathbb{R}$ attains a maximum and a minimum value on $[a, b]$. This result does not hold in general for continuous functions $f : I \to \mathbb{R}$ with $I = (a, b)$ or $I = [a, b)$ or $I = (a, \infty)$, and so on. Besides its theoretical importance in Analysis, the Extreme Value Theorem is also a fundamental result in Optimization Theory. It turns out that when we want to generalize this result, the notion of "compact sets" is pertinent, and we will learn (later on in Chapter 4) the following analogue of the Extreme Value Theorem: If K is a compact subset of a metric space X and $f : K \to \mathbb{R}$ is continuous, then f assumes a maximum and a minimum on K.

Here is the definition of a compact set.

Definition 2.42. Let (X, d) be a metric space. A subset K of X is said to be *compact* if every sequence in K has a convergent subsequence with limit in K, that is, if $(x_n)_{n\in\mathbb{N}}$ is a sequence such that $x_n \in K$ for each $n \in \mathbb{N}$, then there exists a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ which converges to some $L \in K$.

Example 2.43. The interval $[a, b]$ is a compact subset of $\mathbb{R}$. Indeed, every sequence $(a_n)_{n\in\mathbb{N}}$ contained in $[a, b]$ is bounded, and by the Bolzano-Weierstrass Theorem possesses a convergent subsequence, say $(a_{n_k})_{k\in\mathbb{N}}$, with limit L. But since

$$a \le a_{n_k} \le b,$$

by letting $k \to \infty$, we obtain $a \le L \le b$, that is, $L \in [a, b]$. Hence $[a, b]$ is compact.

On the other hand, (a, b) is not compact, since the sequence

$$\left(a + \frac{b-a}{2n}\right)_{n\in\mathbb{N}}$$

is contained in (a, b), but it has no convergent subsequence whose limit belongs to (a, b). Indeed this is because the sequence is convergent, with limit a, and so every subsequence of this sequence is also convergent with limit a, which doesn't belong to (a, b).

$\mathbb{R}$ is not compact since the sequence $(n)_{n\in\mathbb{N}}$ cannot have a convergent subsequence. Indeed, if such a convergent subsequence existed, it would also be Cauchy, but the distance between any two distinct terms, being distinct integers, is at least 1, contradicting the Cauchyness. $\Diamond$

In the above list of nonexamples, note that $\mathbb{R}$ is not bounded, and that (a, b) is not closed. On the other hand, in the example $[a, b]$, we see that $[a, b]$ is both bounded and closed. It turns out that **in** $\mathbb{R}^n$, having the property "closed and bounded[5]" is a characterization of compact sets, and we will show this below. But first let us show that in a general normed space, sets which are closed and bounded can fail to be compact.

Example 2.44. Consider the unit ball with center 0 in the normed space ℓ^2:

$$\mathbb{B} = \{x \in \ell^2 : \|x_2\|_2 \le 1\}.$$

Then $\mathbb{B}$ is bounded, it is closed (since its complement is easily seen to be open), but $\mathbb{B}$ is not compact, and this can be demonstrated as follows. Take the sequence $(e_n)_{n\in\mathbb{N}}$, where e_n is the

[5]A set S in a normed space X is said to be *bounded* if there exists an $R > 0$ such that for all $x \in S$, $\|x\| \le R$.

sequence with only the nth term equal to 1, and all other terms are equal to 0:

$$e_n := (\ 0\ ,\ \cdots\ ,\ 0\ ,\ \underbrace{1}_{n\text{th place}}\ ,\ 0\ ,\ \cdots\ ) \in \mathbb{B} \subset \ell^2$$

Then this sequence $(e_n)_{n\in\mathbb{N}}$ in $\mathbb{B} \subset \ell^2$ can have no convergent subsequence. Indeed, whenever $n \neq m$, $\|e_n - e_m\|_2 = \sqrt{2}$, and so no subsequence of $(e_n)_{n\in\mathbb{N}}$ can be Cauchy, much less convergent! $\Diamond$

We will now show the following important result.

Theorem 2.45. *A subset K of $\mathbb{R}^d$ is compact if and only if K is closed and bounded.*

Before showing this, we prove a technical result, which besides being interesting on its own, will also somewhat simplify the proof of the above theorem.

Lemma 2.46. *Every bounded sequence in $\mathbb{R}^d$ has convergent subsequence.*

Proof. We prove this using induction on d. Let us consider the case when $d = 1$. Then the statement is precisely the Bolzano-Weierstrass Theorem!

Now suppose that the result has been proved in $\mathbb{R}^d$ for some $d \geq 1$. We will show that it holds in $\mathbb{R}^{d+1}$. Let $(a_n)_{n\in\mathbb{N}}$ be a bounded sequence. We split each a_n into its first d components and its last component in $\mathbb{R}$:

$$a_n = \begin{bmatrix} \alpha_n \\ \beta_n \end{bmatrix},$$

where $\alpha_n \in \mathbb{R}^d$ and $\beta_n \in \mathbb{R}$. Clearly $\|\alpha_n\|_2 \leq \|a_n\|_2$, and so $(\alpha_n)_{n\in\mathbb{N}}$ is a bounded sequence in $\mathbb{R}^d$. By the induction hypothesis, it has a convergent subsequence, say $(\alpha_{n_k})_{k\in\mathbb{N}}$ which converges to, say $\alpha \in \mathbb{R}^d$. Consider now the sequence $(\beta_{n_k})_{k\in\mathbb{N}}$ in $\mathbb{R}$. Then $(\beta_{n_k})_{k\in\mathbb{N}}$ is bounded, and so by the Bolzano-Weierstrass Theorem, it has a convergent subsequence $(\beta_{n_{k_\ell}})_{\ell\in\mathbb{N}}$, with limit , say $\beta \in \mathbb{R}$. Then we have

$$a_{n_{k_\ell}} = \begin{bmatrix} \alpha_{n_{k_\ell}} \\ \beta_{n_{k_\ell}} \end{bmatrix} \stackrel{\ell\to\infty}{\longrightarrow} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} =: L \in \mathbb{R}^{d+1}.$$

Thus the bounded sequence $(a_n)_{n\in\mathbb{N}}$ has $(a_{n_{k_\ell}})_{\ell\in\mathbb{N}}$ as a convergent subsequence. $\square$

Now we return to the task of proving of Theorem 2.45.

Proof. ("If" part.) Let K be closed and bounded. Let $(a_n)_{n\in\mathbb{N}}$ be a sequence in K. Then $(a_n)_{n\in\mathbb{N}}$ is bounded, and so it has a convergent subsequence, with limit L. But since K is closed and since each term of the sequence belongs to K, it follows that also $L \in K$. Consequently, K is compact.

("Only if" part) Suppose K is compact.

Let $(a_n)_{n\in\mathbb{N}}$ be a sequence in K that converges to L. Then there is a convergent subsequence, say $(a_{n_k})_{k\in\mathbb{N}}$ that is convergent to a limit $L' \in K$. But as $(a_{n_k})_{k\in\mathbb{N}}$ is a subsequence of a convergent sequence with limit L, it is also convergent to L. By the uniqueness of limits, $L = L' \in K$. Thus K is closed.

Suppose K is not bounded. Then given any $n \in \mathbb{N}$, we can find an $a_n \in K$ such that $\|a_n\|_2 > n$. But this implies that no subsequence of $(a_n)_{n\in\mathbb{N}}$ is bounded. So no subsequence of $(a_n)_{n\in\mathbb{N}}$ can be convergent either. This contradicts the compactness of K. Thus our assumption was incorrect, that is, K is bounded. $\square$

Example 2.47. The intervals $(a, b]$, $[a, b)$ are not compact, since although they are bounded, they are not closed.

The intervals $(-\infty, b]$, $[a, \infty)$ are not compact, since although they are closed, they are not bounded. $\Diamond$

Let us consider an interesting compact subset of the real line, called the *Cantor set*.

Example 2.48 (Cantor set)**.** The Cantor set is constructed as follows. First, denote the closed interval $[0,1]$ by F_1. Next, delete from F_1 the open interval $(\frac{1}{3},\frac{2}{3})$ which is its middle third, and denote the remaining closed set by F_2. Clearly, $F_2=[0,\frac{1}{3}]\bigcup[\frac{2}{3},1]$. Next, delete from F_2 the open intervals $(\frac{1}{9},\frac{2}{9})$ and $(\frac{7}{9},\frac{8}{9})$, which are the middle thirds of its two pieces, and denote the remaining closed set by F_3. It is easy to see that $F_3=[0,\frac{1}{9}]\bigcup[\frac{2}{9},\frac{1}{3}]\bigcup[\frac{2}{3},\frac{7}{9}]\bigcup[\frac{8}{9},1]$. If we continue this process, at each stage deleting the open middle third of each closed interval remaining from the previous stage, we obtain a sequence of closed sets F_n, each of which contains all of its successors. Figure 5 illustrates this.

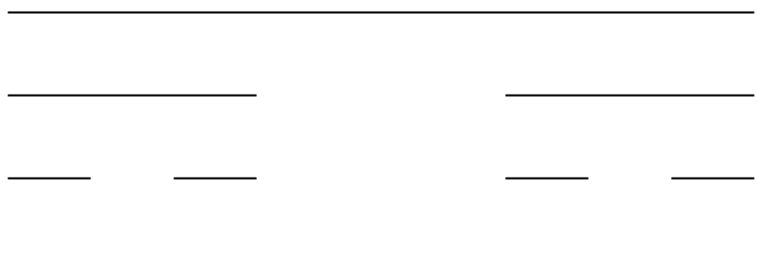


Figure 5. Construction of the Cantor set.

The Cantor set is defined by $F=\bigcap\limits_{n=1}^{\infty}F_n$.

As F is an intersection of closed sets, it is closed. Moreover it is contained in $[0,1]$ and so it is also bounded. Consequently it is compact. F consists of those points in the closed interval $[0,1]$ which "ultimately remain" after the removal of all the open intervals $(\frac{1}{3},\frac{2}{3})$, $(\frac{1}{9},\frac{2}{9})$, $(\frac{7}{9},\frac{8}{9})$, $\dots$. What points do remain? F clearly contains the end-points of the closed intervals which make up each set F_n:

$$0,1,\frac{1}{3},\frac{2}{3},\frac{1}{9},\frac{2}{9},\frac{7}{9},\frac{8}{9},\cdots.$$

Does F contain any other points? Actually, F contains many more points than the above list of end points. After all, the above list of endpoints is countable, but it can be shown that F is uncountable! It turns out that the Cantor set is a very intricate mathematical object, and is often a source of interesting examples/counterexamples in Analysis. (For example, as the sum of the lengths of the intervals removed is

$$\frac{1}{3}+2\frac{1}{3^2}+4\frac{1}{3^3}+\cdots=1,$$

(factor out $1/3$ and sum the resulting geometric series), the "(Lebesgue length) measure" of F is $1-1=0$. So this is an example of an uncountable set with "Lebesgue measure" 0.) $\Diamond$

Exercise 2.49. Let K be a compact subset of $\mathbb{R}^d$. Let F be a closed subset of $\mathbb{R}^d$. Show that $F\cap K$ is compact.

Exercise 2.50. Show that the unit sphere with center 0 in $\mathbb{R}^d$, namely

$$\mathbb{S}^{d-1}:=\{x\in\mathbb{R}^d:\|x\|_2=1\}$$

is compact.

Exercise 2.51. Show that $\left\{1,\frac{1}{2},\frac{1}{3},\dots\right\}\bigcup\{0\}$ is compact.

Exercise 2.52. $(*)$ Consider the metric space $(\mathbb{R}^{m\times m}, \|\cdot\|_\infty)$. Is the subset (the "General Linear" group[6]) $GL(m,\mathbb{R}) = \{A \in \mathbb{R}^{m\times m} : A \text{ is invertible}\}$ compact?

What about the set of orthogonal matrices $O(m,\mathbb{R}) = \{A \in \mathbb{R}^{m\times m} : AA^\top = I_m\}$?

Exercise 2.53. Consider the subset $H := \{(x_1, x_2) \in \mathbb{R}^2 : x_1x_2 = 1\}$ of $\mathbb{R}^2$. Show that H is not compact, but H is closed.

2.6. Notes (not part of the course)

Definition of compactness. The notion of a compact set that we have defined is really *sequential* compactness. In the context of the more general topological spaces, one defines the notion of compactness as follows.

Definition 2.54. Let X be a topological space with the topology given by the family of open sets $\mathcal{O}$. Let $Y \subset X$. A collection $\mathcal{C} = \{U_i : i \in I\}$ of open sets is said to be an *open cover of* Y if

$$Y \subset \bigcup_{i\in I} U_i.$$

$K \subset X$ is said to be a *compact set* if every open cover of K has a finite subcover, that is, given any open cover $\mathcal{C} = \{U_i : i \in I\}$ of K, there exist finitely many indices $i_1, \dots, i_n \in I$ such that

$$K \subset U_{i_1} \cup \cdots \cup U_{i_n}.$$

In the case of *metric spaces*, it can be shown that the set of compact sets *coincides* with the set of sequentially compact sets. But in general topological spaces, these may not be the same.

Uniform convergence, termwise differentiation and integration. Besides Proposition 4.15, one also has the following results associated with uniform convergence.

Proposition 2.55. *If $f_n : [a,b] \to \mathbb{R}$ $(n \in \mathbb{N})$ is a sequence of Riemann-integrable functions on $[a,b]$ which converges uniformly to $f : [a,b] \to \mathbb{R}$, then f is also Riemann-integrable on $[a,b]$, and moreover*

$$\int_a^b f(x)dx = \lim_{n\to\infty} \int_a^b f_n(x)dx.$$

Proposition 2.56. *Let $f_n : (a,b) \to \mathbb{R}$ $(n \in \mathbb{N})$ be a sequence of differentiable functions on (a,b), such that there exists a point $c \in (a,b)$ for which $(f_n(c))_{n\in\mathbb{N}}$ converges. If the sequence $(f'_n)_{n\in\mathbb{N}}$ converges uniformly to g on (a,b), then $(f_n)_{n\in\mathbb{N}}$ converges uniformly to a differentiable function f on (a,b), and moreover, $f'(x) = g(x)$ for all $x \in (a,b)$.*

We will see a proof of this last result later on when we study differentiation in Chapter 5.

[6]The General Linear group is so named because the columns of an invertible matrix are linearly independent, hence the vectors they define are in "general position" (linearly independent!), and matrices in the general linear group take points in general position to points in general position.

Chapter 3

Series

In this chapter we study series in normed spaces, but first we will begin with series in $\mathbb{R}$.

3.1. Series in $\mathbb{R}$

Given a sequence $(a_n)_{n\in\mathbb{N}}$, one can form a new sequence $(s_n)_{n\in\mathbb{N}}$ of its *partial sums*:

$$\begin{aligned} s_1 &:= a_1, \\ s_2 &:= a_1 + a_2, \\ s_3 &:= a_1 + a_2 + a_3, \\ &\vdots \end{aligned}$$

Definition 3.1. Let $(a_n)_{n\in\mathbb{N}}$ be a sequence and let $(s_n)_{n\in\mathbb{N}}$ be the sequence of its partial sums. If $(s_n)_{n\in\mathbb{N}}$ converges, we say that the *series* $\displaystyle\sum_{n=1}^{\infty} a_n$ *converges*, and we write $\displaystyle\sum_{n=1}^{\infty} a_n = \lim_{n\to\infty} s_n$.

If the sequence $(s_n)_{n\in\mathbb{N}}$ does not converge we say that the series $\displaystyle\sum_{n=1}^{\infty} a_n$ *diverges*.

Example 3.2. The series $\displaystyle\sum_{n=1}^{\infty}(-1)^n$ diverges. Indeed the sequence of partial sums is the sequence $-1, 0, -1, 0, \dots$ which is a divergent sequence.

The series $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$ converges. Its nth partial sum "telescopes":

$$s_n = \sum_{k=1}^{n} \frac{1}{k(k+1)} = \sum_{k=1}^{n} \left(\frac{1}{k} - \frac{1}{k+1}\right) = \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \cdots + \left(\frac{1}{n} - \frac{1}{n+1}\right) = 1 - \frac{1}{n+1}.$$

Since $\displaystyle\lim_{n\to\infty} s_n = 1 - 0 = 1$, we have $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 1.$ ◊

Exercise 3.3 (Tantalising $\tan^{-1}$)**.** Show that $\displaystyle\sum_{n=1}^{\infty} \tan^{-1} \frac{1}{2n^2} = \frac{\pi}{4}.$

Hint: Write $\displaystyle\frac{1}{2n^2} = \frac{(2n+1)-(2n-1)}{1+(2n+1)(2n-1)}$ and use $\displaystyle\tan(a-b) = \frac{\tan a - \tan b}{1 + \tan a \tan b}.$

Exercise 3.4. Show that for every real number $x > 1$, the series

$$\frac{1}{1+x} + \frac{2}{1+x^2} + \frac{4}{1+x^4} + \cdots + \frac{2^n}{1+x^{2^n}} + \dots$$

converges. *Hint:* Add $\frac{1}{1-x}$.

Exercise 3.5. Consider the Fibonnaci sequence $(F_n)_{n\in\mathbb{N}}$ with $F_0 = F_1 = 1$ and $F_{n+1} = F_n + F_{n-1}$ for $n \in \mathbb{N}$. Show that
$$\sum_{n=2}^{\infty} \frac{1}{F_{n-1}F_{n+1}} = 1.$$

Note that in the above example of the divergent series $\sum_{n=1}^{\infty}(-1)^n$, the sequence
$$(a_n)_{n\in\mathbb{N}} = ((-1)^n)_{n\in\mathbb{N}}$$
was not convergent. In fact, we have the following necessary condition for convergence of a series.

Proposition 3.6. *If the series* $\sum_{n=1}^{\infty} a_n$ *converges, then* $\lim_{n\to\infty} a_n = 0$.

Proof. Let $s_n := a_1 + \cdots + a_n$. Since the series converges we have
$$\lim_{n\to\infty} s_n = L$$
for some $L \in \mathbb{R}$. But as $(s_{n+1})_{n\in\mathbb{N}}$ is a subsequence of $(s_n)_{n\in\mathbb{N}}$, it follows that
$$\lim_{n\to\infty} s_{n+1} = L.$$
By the algebra of limits, $\lim_{n\to\infty} a_{n+1} = \lim_{n\to\infty}(s_{n+1} - s_n) = \lim_{n\to\infty} s_{n+1} - \lim_{n\to\infty} s_n = L - L = 0.$ ☐

Exercise 3.7. Does the series $\sum_{n=1}^{\infty} \cos\left(\frac{1}{n}\right)$ converge?

In Theorem 3.9, we will see an instance of a series which shows that although this condition is *necessary* for the convergence of a series, it is not *sufficient*. But first, let us see an important example of a convergent series. In fact, it lies at the core of most of the convergence results in Real Analysis.

Theorem 3.8. *Let* $r \in \mathbb{R}$. *The geometric series* $\sum_{n=0}^{\infty} r^n$ *converges if and only if* $|r| < 1$.

Proof. Let $|r| < 1$. First we will observe that $\lim_{n\to\infty} r^n = 0$. As $|r| < 1$,
$$|r| = \frac{1}{1+h}$$
for some $h > 0$. (Just take $h = 1/|r| - 1$.) Then $(1+h)^n = 1 + \binom{n}{1}h + \cdots + h^n > nh$. Thus
$$0 \le |r|^n = \frac{1}{(1+h)^n} < \frac{1}{nh},$$
and so by the Sandwich Theorem, $\lim_{n\to\infty} |r|^n = 0$. As $-|r|^n \le r^n \le |r|^n$, it follows again from the Sandwich Theorem that
$$\lim_{n\to\infty} r^n = 0.$$
Let $s_n := 1 + r + r^2 + \cdots + r^n$. Then $rs_n = r + r^2 + \cdots + r^n + r^{n+1}$, and so
$$(1-r)s_n = s_n - rs_n = 1 - r^{n+1}.$$
As $\lim_{n\to\infty} r^{n+1} = 0$, it follows that $\lim_{n\to\infty}(1-r)s_n = 1$. Hence $\sum_{n=1}^{\infty} r^n = \lim_{n\to\infty} s_n = \frac{1}{1-r}$.

Now suppose that $|r| \geq 1$. If $r = 1$, then

$$\lim_{n\to\infty} r^n = 1 \neq 0,$$

and so by Proposition 3.6, the series diverges. Similarly if $r = -1$, then $(r^n)_{n\in\mathbb{N}} = ((-1)^n)_{n\in\mathbb{N}}$ diverges, and so the series is divergent. Also if $|r| > 1$, then the sequence $(r^n)_{n\in\mathbb{N}}$ has the subsequence $(r^{2n})_{n\in\mathbb{N}}$ which is not bounded, and hence not convergent. Consequently $(r^n)_{n\in\mathbb{N}}$ diverges, and hence the series diverges. □

Theorem 3.9. *The harmonic series*[1] $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n}$ *diverges.*

Proof. Let $s_n := 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$. We have

$$s_{2n} - s_n = \frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n} \geq n\frac{1}{2n} = \frac{1}{2}. \tag{3.1}$$

If the series converges, then $\lim\limits_{n\to\infty} s_n = L$ for some L. But then also $\lim\limits_{n\to\infty} s_{2n} = L$, and so

$$\lim_{n\to\infty} (s_{2n} - s_n) = L - L = 0,$$

which contradicts (3.1). □

Note that the nth term of the above series satisfies

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} \frac{1}{n} = 0,$$

showing that the condition given in Proposition 3.6 is necessary but not sufficient for the convergence of the series.

Theorem 3.10. *Let $s \in \mathbb{R}$. The series*[2] $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^s}$ *converges if and only if $s > 1$.*

Proof. Let

$$S_n = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \cdots + \frac{1}{n^s}.$$

Clearly $S_1 \leq S_2 \leq S_3 \leq \dots$, so that $(S_n)_{n\in\mathbb{N}}$ is an increasing sequence.

Let $s > 1$. We have

$$\begin{aligned} S_{2n+1} &= 1 + \left(\frac{1}{2^s} + \frac{1}{4^s} + \cdots + \frac{1}{(2n)^s}\right) + \left(\frac{1}{3^s} + \frac{1}{5^s} + \cdots + \frac{1}{(2n+1)^s}\right) \\ &< 1 + 2\left(\frac{1}{2^s} + \frac{1}{4^s} + \cdots + \frac{1}{(2n)^s}\right) \\ &= 1 + \frac{2}{2^s}\left(1 + \frac{1}{2^s} + \cdots + \frac{1}{n^s}\right) \\ &= 1 + 2^{1-s} S_n \\ &< 1 + 2^{1-s} S_{2n+1}. \end{aligned}$$

[1] Its name derives from the concept of overtones, or harmonics in music: the wavelengths of the overtones of a vibrating string are $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$, and so on, of the string's fundamental wavelength.

[2] The function $s \mapsto \displaystyle\sum_{n=1}^{\infty} \frac{1}{n^s}$ is called the *Riemann-zeta function*, which is an important function in number theory; the connection with number theory is brought out by Euler's identity, which says that $\zeta(s) := \displaystyle\sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ prime}} \frac{1}{1-p^{-s}}$.

As $2^{1-s} < 1$ (because $s > 1$!), this last inequality yields

$$S_{2n+1} < \frac{1}{1-2^{1-s}} \quad (n \in \mathbb{N})$$

and

$$S_{2n} < S_{2n+1} < \frac{1}{1-2^{1-s}} \quad (n \in \mathbb{N})$$

So $(S_n)_{n\in\mathbb{N}}$ is bounded. But we know that an increasing sequence which is bounded above is convergent (to the supremum of its terms). Thus

$$\sum_{n=1}^{\infty} \frac{1}{n^s}$$

converges for $s > 1$.

If on the other hand $s \le 1$, the proof is similar to that of showing that the harmonic series diverges. Indeed, if the series converged, then

$$\lim_{n\to\infty} (S_{2n} - S_n) = 0,$$

while on the other hand,

$$S_{2n} - S_n = \frac{1}{(n+1)^s} + \frac{1}{(n+2)^s} + \cdots + \frac{1}{(2n)^s} \ge n\frac{1}{(2n)^s} \ge n\frac{1}{2n} = \frac{1}{2},$$

where we have used the fact that $s \le 1$ in order to obtain the last inequality. □

For a sequence $(a_n)_{n\in\mathbb{N}}$ with nonnegative terms, we sometimes write

$$\sum_{n=1}^{\infty} a_n < +\infty$$

to mean the series $\sum_{n=1}^{\infty} a_n$ converges.

Exercise 3.11. $(*)$ Prove that if $a_1 \ge a_2 \ge a_3 \ldots$ is a sequence of nonnegative numbers, and if

$$\sum_{n=1}^{\infty} a_n < +\infty,$$

then a_n approaches 0 faster than $1/n$, that is, $\lim_{n\to\infty} na_n = 0$.

Hint: Consider the inequalities $a_{n+1} + \cdots + a_{2n} \ge n \cdot a_{2n}$ and $a_{n+1} + \cdots + a_{2n+1} \ge n \cdot a_{2n+1}$.

Show that the assumption $a_1 \ge a_2 \ge a_3 \ldots$ above cannot be dropped by considering the *lacunary series* whose n^2th term is $1/n^2$ and all other terms are zero.

Proposition 3.12. *If* $\sum_{n=1}^{\infty} |a_n|$ *converges, then* $\sum_{n=1}^{\infty} a_n$ *converges.*

Proof. Let $s_n := a_1 + \cdots + a_n$. We will show that $(s_n)_{n\in\mathbb{N}}$ is a Cauchy sequence. We have for $n > m$:

$$\begin{aligned} |s_n - s_m| &= |(a_1 + \cdots + a_n) - (a_1 + \cdots + a_m)| = |a_{m+1} + \cdots + a_n| \\ &\le |a_{m+1}| + \cdots + |a_n| = (|a_1| + \cdots + |a_n|) - (|a_1| + \cdots + |a_m|) = \sigma_n - \sigma_m, \end{aligned}$$

where $\sigma_k := |a_1| + \cdots + |a_k|$ $(k \in \mathbb{N})$. Since

$$\sum_{n=1}^{\infty} |a_n| < +\infty,$$

its sequence of partial sums $(\sigma_n)_{n\in\mathbb{N}}$ is convergent, and in particular, Cauchy. This shows, in light of the above inequality $|s_n - s_m| \le \sigma_n - \sigma_m$, that $(s_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$ and hence it is convergent. □

Definition 3.13. If the series $\displaystyle\sum_{n=1}^{\infty}|a_n|$ converges, then we say that $\displaystyle\sum_{n=1}^{\infty}a_n$ *converges absolutely.*

Exercise 3.14. Does the series $\displaystyle\sum_{n=1}^{\infty}\frac{\sin n}{n^2}$ converge?

Example 3.15. The series $\displaystyle\sum_{n=1}^{\infty}\frac{(-1)^n}{n}$ does not converge absolutely, since

$$\sum_{n=1}^{\infty}\left|\frac{(-1)^n}{n}\right| = \sum_{n=1}^{\infty}\frac{1}{n},$$

and we have seen that the harmonic series diverges.

A series of the form

$$\sum_{n=1}^{\infty}(-1)^n a_n$$

with $a_n \geq 0$ for all $n \in \mathbb{N}$ is called an *alternating series.* We note that the series considered above, namely

$$\sum_{n=1}^{\infty}\frac{(-1)^n}{n}$$

is an alternating series $\displaystyle\sum_{n=1}^{\infty}(-1)^n a_n$ with $a_n := \dfrac{1}{n}$ $(n \in \mathbb{N})$.

We will now learn a result below, called the Leibniz Alternating Series Theorem, will enable us to conclude that in fact this alternating series is in fact convergent (since the sufficiency conditions for convergence in the Leibniz Alternating Series Theorem are satisfied:

$$a_1 = 1 \geq a_2 = \frac{1}{2} \geq a_3 = \frac{1}{3} \geq \dots$$

and $\displaystyle\lim_{n\to\infty} a_n = \lim_{n\to\infty}\frac{1}{n} = 0$). ◊

Theorem 3.16 (Leibniz Alternating Series Theorem)**.** *Let $(a_n)_{n\in\mathbb{N}}$ be a sequence such that*

(1) *it has nonnegative terms ($a_n \geq 0$ for all n),*

(2) *it decreasing ($a_1 \geq a_2 \geq a_3 \geq \dots$), and*

(3) $\displaystyle\lim_{n\to\infty} a_n = 0$.

Then the series $\displaystyle\sum_{n=1}^{\infty}(-1)^n a_n$ converges.

A pictorial "proof without words" is shown in Figure 1. The sum of the lengths of the disjoint dark intervals is at most the length of $(0, a_1)$.

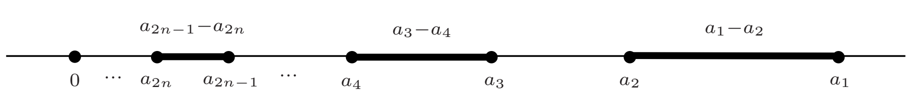


Figure 1. Pictorial proof of the Leibniz Alternating Series Theorem.

Proof. We may just as well prove the convergence of $\displaystyle\sum_{n=1}^{\infty}(-1)^{n+1}a_n \left(= -\sum_{n=1}^{\infty}(-1)^n a_n\right)$.

Let $s_n = a_1 - a_2 + a_3 - + \cdots + (-1)^{n-1}a_n$. Clearly

$$\begin{aligned} s_{2n+1} &= s_{2n-1} - a_{2n} + a_{2n+1} \leq s_{2n-1}, \\ s_{2n+2} &= s_{2n} + a_{2n+1} - a_{2n+2} \geq s_{2n}, \end{aligned}$$

and so the sequence $s_2, s_4, s_6, \dots$ is increasing, while the sequence $s_3, s_5, s_7, \dots$ is decreasing. Also,

$$s_{2n} \leq s_{2n} + a_{2n+1} = s_{2n+1} \leq s_{2n-1} \leq \cdots \leq s_3.$$

So $(s_{2n})_{n\in\mathbb{N}}$ is a bounded ($s_2 \leq s_{2n} \leq s_3$ for all n), increasing sequence, and hence it is convergent. But as $(a_{2n+1})_{n\in\mathbb{N}}$ is also convergent with limit 0, it follows that

$$\lim_{n\to\infty} s_{2n+1} = \lim_{n\to\infty}(s_{2n} + a_{2n+1}) = \lim_{n\to\infty} s_{2n}.$$

Hence $(s_n)_{n\in\mathbb{N}}$ is convergent, and so the series converges. □

Exercise 3.17. Let $s > 0$. Show that $\displaystyle\sum_{n=1}^{\infty}\frac{(-1)^n}{n^s}$ converges.

Exercise 3.18. Prove that $\displaystyle\sum_{n=1}^{\infty}(-1)^n\frac{\sqrt{n}}{n+1}$ converges.

Exercise 3.19. Prove that $\displaystyle\sum_{n=1}^{\infty}(-1)^n\sin\frac{1}{n}$ converges.

3.1.1. Comparison, Ratio, Root. We will now learn three important tests for the convergence of a series:

(1) the comparison test (where we compare with a series whose convergence status is known)
(2) the ratio test (where we look at the behaviour of the ratio of terms a_{n+1}/a_n)
(3) the root test (where we look at the behaviour of $\sqrt[n]{|a_n|}$)

Theorem 3.20 (Comparison test)**.**

(1) *If $(a_n)_{n\in\mathbb{N}}$ and $(c_n)_{n\in\mathbb{N}}$ are such that there exists an $N \in \mathbb{N}$ such that $|a_n| \leq c_n$ for all $n \geq N$, and $\displaystyle\sum_{n=1}^{\infty}c_n$ converges, then $\displaystyle\sum_{n=1}^{\infty}a_n$ converges absolutely.*
(2) *If $(a_n)_{n\in\mathbb{N}}$ and $(d_n)_{n\in\mathbb{N}}$ are such that there exists an $N \in \mathbb{N}$ such that $a_n \geq d_n \geq 0$ for all $n \geq N$, and $\displaystyle\sum_{n=1}^{\infty}d_n$ diverges, then $\displaystyle\sum_{n=1}^{\infty}a_n$ diverges.*

Proof. Let

$$\begin{aligned} s_n &:= |a_1| + \cdots + |a_n| \\ \sigma_n &:= c_1 + \cdots + c_n. \end{aligned}$$

We have for $n > m$

$$|s_n - s_m| = |a_{m+1}| + \cdots + |a_n| \leq c_{m+1} + \cdots + c_n = |\sigma_n - \sigma_m|,$$

and since $(\sigma_n)_{n\in\mathbb{N}}$ is Cauchy, $(s_n)_{n\in\mathbb{N}}$ is also Cauchy. Hence $(s_n)_{n\in\mathbb{N}}$ is absolutely convergent.

The second claim follows from the first one. For if $\displaystyle\sum_{n=1}^{\infty}a_n$ converges, so must $\displaystyle\sum_{n=1}^{\infty}d_n$. □

Theorem 3.21 (Ratio test)**.** *Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of nonzero terms.*

(1) *If $\exists r \in (0,1)$ and $\exists N \in \mathbb{N}$ such that $\forall n > N$, $\left|\dfrac{a_{n+1}}{a_n}\right| \le r$, then $\displaystyle\sum_{n=1}^{\infty} a_n$ converges absolutely.*

(2) *If $\exists N \in \mathbb{N}$ such that $\forall n > N$, $\left|\dfrac{a_{n+1}}{a_n}\right| \ge 1$, then $\displaystyle\sum_{n=1}^{\infty} a_n$ diverges.*

Proof. (1) We have

$$\begin{array}{rcl} |a_{N+1}| & \le & r|a_N|, \\ |a_{N+2}| & \le & r|a_{N+1}| \le r^2|a_N|, \\ |a_{N+3}| & \le & r|a_{N+2}| \le r^3|a_N|, \\ & \vdots & \end{array}$$

Since the geometric series $\displaystyle\sum_{n=1}^{\infty} r^n$ converges, we obtain

$$\sum_{n=N+1}^{\infty} |a_n| < +\infty$$

by the Comparison Test. By adding the finitely sum $|a_1| + \cdots + |a_N|$ to each partial sum of this last series, we see also that $\displaystyle\sum_{n=1}^{\infty} |a_n|$ converges. This completes the proof of the claim in (1).

(2) The given condition implies that

$$\cdots \ge |a_{N+3}| \ge |a_{N+2}| \ge |a_{N+1}|, \tag{3.2}$$

If the series $\displaystyle\sum_{n=1}^{\infty} a_n$ was convergent, then $0 = \displaystyle\lim_{n\to\infty} a_n = \lim_{k\to\infty} a_{N+k}$. Hence $\displaystyle\lim_{k\to\infty} |a_{N+k}| = 0$ as well. But by the (3.2), we see that $\displaystyle\lim_{k\to\infty} |a_{N+k}| \ge |a_{N+1}| > 0$, a contradiction. □

It does not suffice for convergence of the series that for all sufficiently large n, $\left|\dfrac{a_{n+1}}{a_n}\right| < 1$.

For example, for the harmonic series $\left|\dfrac{a_{n+1}}{a_n}\right| = \left|\dfrac{\frac{1}{n+1}}{\frac{1}{n}}\right| = \dfrac{n}{n+1} < 1$, but $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n}$ diverges.

Corollary 3.22. *Let $(a_n)_{n\in\mathbb{N}}$ have nonzero terms. If $\displaystyle\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| < 1$, then $\displaystyle\sum_{n=1}^{\infty} a_n$ converges absolutely.*

Proof. Let $L := \displaystyle\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right|$. Then $\epsilon := \dfrac{1-L}{2} > 0$. Choose $N \in \mathbb{N}$ such that for $n > N$,

$$\left|\frac{a_{n+1}}{a_n}\right| - L \le \left|\left|\frac{a_{n+1}}{a_n}\right| - L\right| < \epsilon = \frac{1-L}{2},$$

and so $\left|\dfrac{a_{n+1}}{a_n}\right| < \dfrac{1+L}{2} =: r < \dfrac{1+1}{2} = 1$. Then we use (1) from Theorem 3.21. □

Example 3.23 (The exponential series)**.** Let $a \in \mathbb{R}$. The series $e^a := \sum_{n=0}^{\infty} \frac{1}{n!} a^n$ converges.

If $a = 0$, clearly $e^a = e^0 = 1$. If $a \neq 0$, the convergence follows from the ratio test. Indeed,

$$\left| \frac{\frac{a^{n+1}}{(n+1)!}}{\frac{a^n}{n!}} \right| = \frac{|a|}{n+1} \stackrel{n\to\infty}{\longrightarrow} 0.$$

◊

Theorem 3.24 (Root test)**.**

(1) *If* $\exists r \in (0, 1)$ *and* $\exists N \in \mathbb{N}$ *such that* $\forall n > N$, $\sqrt[n]{|a_n|} \le r$, *then* $\sum_{n=1}^{\infty} a_n$ *converges absolutely.*

(2) *If for infinitely many* n, $\sqrt[n]{|a_n|} \ge 1$, *then* $\sum_{n=1}^{\infty} a_n$ *diverges.*

Proof. (1) We have $|a_n| \le r^n$ for all $n > N$, so that by the Comparison Test, $\sum_{n=N+1}^{\infty} |a_n|$ converges.

(2) Suppose that for the subsequence $(a_{n_k})_{k\in\mathbb{N}}$, we have $\sqrt[n_k]{|a_{n_k}|} \ge 1$. Then $|a_{n_k}| \ge 1$. If the series was convergent, then $\lim_{n\to\infty} a_n = 0$, and so also $\lim_{n\to\infty} |a_{n_k}| = 0$, a contradiction. Thus the series cannot converge. □

It does not suffice for convergence of the series that for sufficiently large n, $\sqrt[n]{|a_n|} < 1$.

For example, for the harmonic series $\sqrt[n]{|a_n|} = \frac{1}{\sqrt[n]{n}} < 1$, but $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges.

Exercise 3.25. Determine if the following series are convergent or not.

(1) $\sum_{n=1}^{\infty} \frac{n^2}{2^n}$.

(2) $\sum_{n=1}^{\infty} \frac{(n!)^2}{(2n)!}$.

(3) $\sum_{n=1}^{\infty} \left(\frac{4}{5}\right)^n n^5$.

Exercise 3.26. Prove that $\sum_{n=1}^{\infty} \frac{n}{n^4 + n^2 + 1}$ converges. (∗) Can you also find its value? *Hint:* Write the denominator as $(n^2 + 1)^2 - n^2$.

Exercise 3.27. (∗) Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of real numbers such that the series $\sum_{n=1}^{\infty} a_n^4$ converges. Show that $\sum_{n=1}^{\infty} a_n^5$ converges. *Hint:* First conclude that for large n, $|a_n| < 1$.

Exercise 3.28. (∗) We have seen that the series $\sum_{n=1}^{\infty} \frac{1}{n^p}$ converges if and only if $p > 1$. Suppose that we sum the series for all n that can be written without using the numeral 9. (Imagine that the key for 9 on the keyboard is broken.) Call the resulting summation $\sum'$. Prove that the series $\sum' \frac{1}{n^p}$ converges if and only if $p > \log_{10} 9$. *Hint:* First show that there are $8 \cdot 9^{k-1}$ numbers without 9 between 10^{k-1} and 10^k.

Exercise 3.29. Determine if the following statements are true or false.

(1) If $\sum_{n=1}^{\infty} |a_n|$ is convergent, then so is $\sum_{n=1}^{\infty} a_n^2$.

(2) If $\sum_{n=1}^{\infty} a_n$ is convergent, then so is $\sum_{n=1}^{\infty} a_n^2$.

(3) If $\lim_{n\to\infty} a_n = 0$, then $\sum_{n=1}^{\infty} a_n$ converges.

(4) If $\lim_{n\to\infty} (a_1 + \cdots + a_n) = 0$, then $\sum_{n=1}^{\infty} a_n$ converges.

(5) $\sum_{n=1}^{\infty} \log \frac{n+1}{n}$ converges.

(6) If $a_n > 0$ ($n \in \mathbb{N}$) and the partial sums of $(a_n)_{n\in\mathbb{N}}$ are bounded above, then $\sum_{n=1}^{\infty} a_n$ converges.

(7) If $a_n > 0$ ($n \in \mathbb{N}$) and $\sum_{n=1}^{\infty} a_n$ converges, then $\sum_{n=1}^{\infty} \frac{1}{a_n}$ diverges.

Exercise 3.30 (Fourier series)**.** In order to understand a complicated situation, it is natural to try to break it up into simpler things. For example, from Calculus we learn that an analytic function can be expanded into a Taylor series, where we break it down into the simplest possible analytic functions, namely monomials $1, x, x^2, \dots$ as follows:

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \dots .$$

The idea behind the Fourier series is similar. In order to understand a complicated *periodic* function, we break it down into the simplest periodic functions, namely sines and cosines. Thus if $T \geq 0$ and $f : \mathbb{R} \to \mathbb{R}$ is T*-periodic*, that is, $f(x) = f(x+T)$ ($x \in \mathbb{R}$), then one tries to find coefficients $a_0, a_1, a_2, \dots$ and $b_1, b_2, b_3, \dots$ such that

$$f(x) = a_0 + \sum_{n=1}^{\infty} \left(a_n \cos\left(\frac{2\pi n}{T}x\right) + b_n \sin\left(\frac{2\pi n}{T}x\right)\right). \tag{3.3}$$

(1) Suppose that the Fourier series given in (3.3) converges pointwise to f on $\mathbb{R}$. Show that if

$$\sum_{n=1}^{\infty} (|a_n| + |b_n|) < \infty,$$

then in fact the series converges uniformly.

(2) The aim of this part of the exercise is to give experimental evidence for two things. Firstly, the plausibility of the Fourier expansion, and secondly, that the uniform convergence might fail if the condition in the previous part of this exercise does not hold. To this end, let us consider the *square wave* $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = \begin{cases} 1 & \text{if } x \in [n, n+1) \text{ for } n \text{ even}, \\ -1 & \text{if } x \in [n, n+1) \text{ for } n \text{ odd}. \end{cases}$ Then f is a 2-periodic signal. From the theory of Fourier Series, which we will not discuss in this course, the coefficients can be calculated, and they happen to be $0 = a_0 = a_1 = a_2 = a_3 = \dots$ and $b_n = \begin{cases} \frac{4}{n\pi} & \text{if } n \text{ is odd}, \\ 0 & \text{if } n \text{ is even}. \end{cases}$ Write a Maple program to plot the graphs of the partial sums of the series in (3.3) with, say, 3, 33, 333 terms. Discuss your observations.

Exercise 3.31. Let $(a_n)_{n\in\mathbb{N}}$ be a sequence with nonnegative terms. Show that if $\sum_{n=1}^{\infty} a_n$ converges, then so does $\sum_{n=1}^{\infty} \sqrt{a_n a_{n+1}}$.

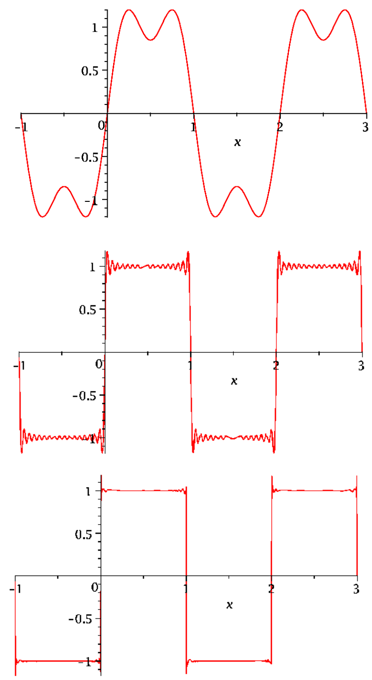


Figure 2. Partial sums of the Fourier series for the square wave considered in Exercise 3.30.

Exercise 3.32. (*) Let $(a_n)_{n\in\mathbb{N}}$ be a sequence with nonnegative terms. Prove that $\sum_{n=1}^{\infty} a_n$ converges if and only if $\sum_{n=1}^{\infty} \frac{a_n}{1+a_n}$ converges.

Exercise 3.33. Let ℓ^1 be defined by $\ell^1 = \left\{ (a_n)_{n\in\mathbb{N}} : \sum_{n=1}^{\infty} |a_n| < \infty \right\}$. Show that $\ell^1 \subset \ell^2$. Is $\ell^1 = \ell^2$?

Exercise 3.34. As the harmonic series $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges, the reciprocal $1/s_n$ of the nth partial sum

$$s_n := 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$$

approaches 0 as $n \to \infty$. So the necessary condition for the convergence of the series

$$\sum_{n=1}^{\infty} \frac{1}{s_n}$$

is satisfied. But we don't know yet whether or not it actually converges. It is clear that the harmonic series diverges very slowly, which means that $1/s_n$ decreases very slowly, and this prompts the guess that this series diverges. Show that in fact our guess is correct. *Hint:* $s_n < n$.

Exercise 3.35. Show that the series $\sum_{n=1}^{\infty} \frac{1}{n^n}$ converges.

Exercise 3.36. Consider the Fibonnaci sequence $(F_n)_{n\in\mathbb{N}}$ with $F_0 = F_1 = 1$ and $F_{n+1} = F_n + F_{n-1}$ for $n \in \mathbb{N}$. Show that

$$\sum_{n=0}^{\infty} \frac{1}{F_n} < +\infty.$$

Hint: $F_{n+1} = F_n + F_{n-1} > F_{n-1} + F_{n-1} = 2F_{n-1}$. Using this, show that both $\frac{1}{F_0} + \frac{1}{F_2} + \frac{1}{F_4} + \ldots$ and $\frac{1}{F_1} + \frac{1}{F_3} + \frac{1}{F_5} + \ldots$ converge.

3.1.2. Power series. Let $(c_n)_{n\in\mathbb{N}}$ be a real sequence (thought of as a sequence of "coefficients"). An expression of the type

$$\sum_{n=0}^{\infty} c_n x^n$$

is called a *power series* in the variable $x \in \mathbb{R}$.

Example 3.37. All polynomial expressions are (finite) power series, with the coefficients being eventually all zeros.

$\displaystyle\sum_{n=0}^{\infty} x^n$, $\displaystyle\sum_{n=0}^{\infty} \frac{1}{n!}x^n$ are also examples of power series. $\Diamond$

Note that we have not said anything about the set of $x \in \mathbb{R}$ where the power series converges. Of course the power series always converges for $x = 0$. It turns out that there is a maximal open interval $B(0, r) = (-r, r)$ centered at 0 where the power series converges absolutely, and we call the radius r of this interval $B(0, r) = (-r, r)$ as the *radius of convergence* of the power series. If the power series converges for all $x \in \mathbb{R}$, that is, if the above maximal interval is $(-\infty, \infty)$, we say that the power series has *infinite radius of convergence*.

Example 3.38. The radius of convergence of $\displaystyle\sum_{n=0}^{\infty} x^n$ is 1. Indeed, the geometric series converges for $x \in (-1, 1)$ and diverges whenever $|x| \geq 1$. $\Diamond$

Example 3.39. The radius of convergence of $\displaystyle\sum_{n=0}^{\infty} \frac{1}{n!}x^n$ is infinite, since e^x converges for every $x \in \mathbb{R}$. $\Diamond$

Example 3.40. The radius of convergence of $\displaystyle\sum_{n=0}^{\infty} n^n x^n$ is zero. Indeed, whenever $x \neq 0$,

$$\sqrt[n]{|n^n x^n|} = n|x| > 1$$

for all n large enough. By the Root test, it follows that the power series diverges for all nonzero real numbers. $\Diamond$

Theorem 3.41. *A power series*

$$\sum_{n=0}^{\infty} c_n x^n$$

is either absolutely convergent for all $x \in \mathbb{R}$, or there is a unique nonnegative real number ρ such that the power series is absolutely convergent for all $x \in \mathbb{R}$ with $|x| < \rho$ and is divergent for all $x \in \mathbb{R}$ with $|x| > \rho$.

Proof. Let $S := \left\{ y \in \mathbb{R} : \text{there exists an } x \in \mathbb{R} \text{ such that } y = |x| \text{ and } \displaystyle\sum_{n=0}^{\infty} c_n x^n \text{ converges} \right\}$. Clearly $0 \in S$. We consider the following two possible cases.

<u>1</u>° S is not bounded above. Then given $x \in \mathbb{R}$, we can find a $|x_0| \in S$ such that $|x| < |x_0|$ and $\displaystyle\sum_{n=0}^{\infty} c_n x_0^n$ converges. It follows that its nth term goes to 0 as $n \to \infty$, and in particular, it is bounded: $|c_n x_0^n| \leq M$. Then with $r := \dfrac{|x|}{|x_0|}$ (< 1), we have

$$|c_n x^n| = |c_n x_0^n| \left(\frac{|x|}{|x_0|} \right)^n \quad (n \in \mathbb{N}),$$

But the Geometric Series $\sum_{n=0}^{\infty} Mr^n$ converges, and so by the Comparison Test, the series $\sum_{n=0}^{\infty} c_n x^n$ is absolutely convergent.

<u>2</u>° S is bounded above. Let $\rho := \sup S$.

If $x \in \mathbb{R}$ and $|x| < \rho$, then by the definition of supremum, it follows that there exists a $|x_0| \in S$ such that $|x| < |x_0|$. Then we repeat the proof in 1° above to conclude that the series $\sum_{n=0}^{\infty} c_n x^n$ is absolutely convergent.

On the other hand, if $x \in \mathbb{R}$ and $|x| > \rho$, then $|x| \notin S$, and by the definition of S, the series $\sum_{n=0}^{\infty} c_n x^n$ diverges.

If ρ, ρ' have the property described in the theorem and $\rho < \rho'$, then $\rho < r := \frac{\rho + \rho'}{2} < \rho'$. Because $0 < r < \rho'$, $\sum_{n=1}^{\infty} c_n r^n$ converges, while as $\rho < r$, it follows that $\sum_{n=1}^{\infty} c_n r^n$ diverges, a contradiction. □

If ρ is the radius of convergence of a power series, then $(-\rho, \rho)$ is called the *interval of convergence* of that power series. We note that the interval of convergence is the empty set if $\rho = 0$, and we set the interval of convergence to be $\mathbb{R}$ when the radius of convergence is infinite. See Figure 3.

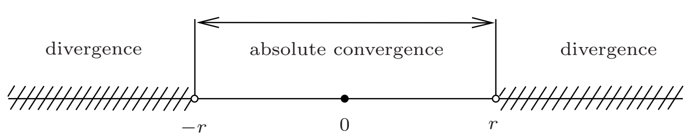


Figure 3. The interval of convergence of a power series.

The calculation of the radius of convergence is facilitated in some cases by the following result.

Theorem 3.42. *Consider the power series*

$$\sum_{n=0}^{\infty} c_n x^n.$$

If $L := \lim_{n\to\infty} \left|\frac{c_{n+1}}{c_n}\right|$ *exists, then* $\rho = 1/L$ *if* $L \neq 0$*, and the radius of convergence is infinite if* $L = 0$.

Proof. Let $L \neq 0$. We have that for all nonzero x such that $|x| < \rho = 1/L$ that there exists a $q < 1$ and a N large enough such that

$$\frac{|c_{n+1}x^{n+1}|}{|c_n x^n|} = \left|\frac{c_{n+1}}{c_n}\right| |x| \leq q < 1$$

for all $n > N$. (This is because

$$\left|\frac{c_{n+1}}{c_n}x\right| \stackrel{n\to\infty}{\longrightarrow} L|x| < 1.$$

So we may take for example $q = (L|x|+1)/2 < 1$.) Thus by the Ratio Test, the power series converges absolutely for such x.

If $L = 0$, then for any nonzero $x \in \mathbb{R}$, we can guarantee that

$$\frac{|c_{n+1}x^{n+1}|}{|c_n x^n|} = \left|\frac{c_{n+1}}{c_n}\right||x| \le q < 1$$

for all $n > N$. (This is because

$$\left|\frac{c_{n+1}}{c_n}x\right| \stackrel{n\to\infty}{\longrightarrow} 0|x| = 0 < 1.$$

So we may take for example $q = 1/2 < 1$.) Thus again by the Ratio Test, the power series converges absolutely for such x.

On the other hand, if $L \neq 0$ and $|x| > 1/L$, then there exists a N large enough such that

$$\frac{|c_{n+1}x^{n+1}|}{|c_n x^n|} = \left|\frac{c_{n+1}}{c_n}\right||x| > 1$$

for all $n > N$. This is because

$$\left|\frac{c_{n+1}}{c_n}x\right| \stackrel{n\to\infty}{\longrightarrow} L|x| > 1.$$

So again by the Ratio Test, the power series diverges. □

Exercise 3.43. $(*)$ Consider the power series $\sum_{n=0}^{\infty} c_n x^n$. If $L := \lim_{n\to\infty} \sqrt[n]{|c_n|}$ exists, then $\rho = 1/L$ if $L \neq 0$, and the radius of convergence is infinite if $L = 0$.

Note that whether or not the power series converges at $x = \rho$ and $x = -\rho$ is not answered by Theorem 3.41. In fact this is a delicate issue, and either convergence or divergence can take place at these points, as demonstrated by the following examples.

Example 3.44. We have the following:

Power series	Radius of convergence	Set of x's for which the power series converges
$\sum_{n=1}^{\infty} x^n$	1	$(-1,1)$
$\sum_{n=1}^{\infty} \frac{x^n}{n^2}$	1	$[-1,1]$
$\sum_{n=1}^{\infty} \frac{x^n}{n}$	1	$[-1,1)$
$\sum_{n=1}^{\infty} (-1)^n \frac{x^n}{n}$	1	$(-1,1]$

◊

Exercise 3.45. Check all the claims in Example 3.44.

Exercise 3.46. Find the radius of convergence for each of the following power series:

$$\sum_{n=1}^{\infty}(-1)^{n-1}\frac{x^{2n-1}}{(2n-1)!}=x-\frac{x^3}{3!}+\frac{x^5}{5!}-+\cdots \quad\text{and}\quad \sum_{n=0}^{\infty}(-1)^{n}\frac{x^{2n}}{(2n)!}=1-\frac{x^2}{2!}+\frac{x^4}{4!}-+\cdots.$$

3.2. Series in normed spaces

We can't define series in a general metric space, since we need to *add* terms. But in the setting of a normed space, addition of vectors is available, and so we can define the notion of convergence of a series in a normed space.

Definition 3.47. Let $(a_n)_{n\in\mathbb{N}}$ be a sequence in a normed space $(X,\|\cdot\|)$. The sequence $(s_n)_{n\in\mathbb{N}}$ of *partial sums* is defined as follows:

$$s_n = a_1+\cdots+a_n\in X \quad (n\in\mathbb{N}).$$

The series

$$\sum_{n=1}^{\infty}a_n$$

is called *convergent* if $(s_n)_{n\in\mathbb{N}}$ converges in $(X,\|\cdot\|)$. Then we write

$$\sum_{n=1}^{\infty}a_n=\lim_{n\to\infty}s_n.$$

If the sequence $(s_n)_{n\in\mathbb{N}}$ does not converge we say that the series $\sum\limits_{n=1}^{\infty}a_n$ *diverges.*

It turns out the convergence in a *complete* normed space is guaranteed by the convergence of an associated real series (of the norms of its terms). We first introduce the notion of *absolute convergence* of a series in a normed space, analogous to the absolute convergence of a real series.

Definition 3.48. Let $(a_n)_{n\in\mathbb{N}}$ be a sequence in a normed space $(X,\|\cdot\|)$. We say that the series

$$\sum_{n=1}^{\infty}a_n$$

converges absolutely if the (real) series $\sum\limits_{n=1}^{\infty}\|a_n\|$ converges.

Theorem 3.49. *Let $(a_n)_{n\in\mathbb{N}}$ be a sequence in a complete normed space $(X,\|\cdot\|)$ such that $\sum\limits_{n=1}^{\infty}\|a_n\|<+\infty$. Then $\sum\limits_{n=1}^{\infty}a_n$ converges in X.*

Proof. The proof is the same, mutatis mutandis, as the proof of the fact that absolutely convergent real series converge. The only change is we use norms instead of absolute values, and use the completeness of X in order to conclude that when the partial sums form a Cauchy sequence, they converge to a limit in X.

Let $s_n := a_1+\cdots+a_n$. We will show that $(s_n)_{n\in\mathbb{N}}$ is a Cauchy sequence. We have for $n>m$:

$$\begin{aligned}\|s_n-s_m\| &= \|(a_1+\cdots+a_n)-(a_1+\cdots+a_m)\| = \|a_{m+1}+\cdots+a_n\|\\ &\le \|a_{m+1}\|+\cdots+\|a_n\| = (\|a_1\|+\cdots+\|a_n\|)-(\|a_1\|+\cdots+\|a_m\|)=\sigma_n-\sigma_m,\end{aligned}$$

where $\sigma_k := \|a_1\|+\cdots+\|a_k\|$. Since the series

$$\sum_{n=1}^{\infty}\|a_n\|$$

converges (given!), its sequence of partial sums is convergent, and in particular, Cauchy. This shows, in light of the above inequality $\|s_n - s_m\| \le \sigma_n - \sigma_m$, that $(s_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in X. As X is complete, $(s_n)_{n\in\mathbb{N}}$ is convergent. □

Example 3.50. Consider the sequence $(f_n)_{n\in\mathbb{N}}$ in the normed space $(C[0,1], \|\cdot\|_\infty)$, where

$$f_n(x) = \left(\frac{x}{2}\right)^n \quad (x \in [0,1],\ n \in \mathbb{N}).$$

The series $\sum\limits_{n=1}^{\infty} f_n$ converges in $(C[0,1], \|\cdot\|_\infty)$ since

$$\|f_n\|_\infty = \max_{x\in[0,1]} \left|\left(\frac{x}{2}\right)^n\right| = \frac{1}{2^n},$$

and the series $\sum\limits_{n=1}^{\infty} \|f_n\|_\infty = \sum\limits_{n=1}^{\infty} \frac{1}{2^n}$ converges.

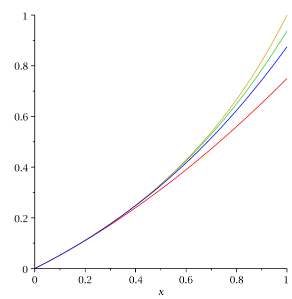


Figure 4. Plots of the graphs of the partial sums and the limit f.

In fact, one can see directly that since

$$s_n(x) = f_1(x) + \cdots + f_n(x) = \frac{\frac{x}{2}\left(1 - \left(\frac{x}{2}\right)^{n+1}\right)}{1 - \left(\frac{x}{2}\right)} = \frac{x}{2-x}\left(1 - \frac{x^n}{2^n}\right),$$

with f defined by $f(x) = \dfrac{x}{2-x}$, we have that

$$\|s_n - f\|_\infty = \max_{x\in[0,1]} \frac{x}{2-x}\frac{x^n}{2^n} \le \frac{1}{2^n} \stackrel{n\to 0}{\longrightarrow} 0,$$

and so the series $\sum\limits_{n=1}^{\infty} f_n$ converges to f in $(C[0,1], \|\cdot\|_\infty)$.

Figure 3.50 shows plots of the partial sums and their limit f. ◊

The following result plays a central role in Differential Equation theory.

Theorem 3.51. *Let $A \in \mathbb{R}^{n\times n}$. Then the exponential series*

$$e^A := I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \cdots$$

converges in $(\mathbb{R}^{n\times n}, \|\cdot\|_\infty)$.

Proof. It is easy to see that if $A, B \in \mathbb{R}^{n\times n}$, then $\|AB\|_\infty \le n\|A\|_\infty \|B\|_\infty$. This follows immediately from:

$$|(AB)_{ij}| = \left|\sum_{k=1}^{n} A_{ik}B_{kj}\right| \le \sum_{k=1}^{n} |A_{ik}||B_{kj}| \le \sum_{k=1}^{n} \|A\|_\infty \|B\|_\infty = n\|A\|_\infty \|B\|_\infty.$$

Hence by induction, we have $\|A^k\|_\infty \le n^{k-1}\|A\|_\infty^k \le (n\|A\|_\infty)^k$. Thus

$$\left\|\frac{1}{k!}A^k\right\|_\infty \le \frac{1}{k!}(n\|A\|_\infty)^k.$$

As the series $e^{n\|A\|_\infty} = \displaystyle\sum_{k=1}^{\infty} \frac{1}{k!}(n\|A\|_\infty)^k$ converges, by the Comparison Test, the real series

$$\sum_{k=1}^{\infty} \left\|\frac{1}{k!}A^k\right\|_\infty$$

converges. Since $(\mathbb{R}^{n\times n}, \|\cdot\|_\infty)$ is complete, it follows that

$$e^A := I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \cdots$$

converges in $(\mathbb{R}^{n\times n}, \|\cdot\|_\infty)$. □

Exercise 3.52. Calculate $e^{\mathbf{0}}$ (here $\mathbf{0}$ denotes the $n \times n$ matrix with all entries 0) and e^I.

Exercise 3.53. $(*)$ Let X be a normed space in which every series $\displaystyle\sum_{n=1}^{\infty} a_n$ for which there holds

$$\sum_{n=1}^{\infty} \|a_n\| < +\infty,$$

is convergent in X. Prove that X is complete. *Hint:* Given a Cauchy sequence $(x_n)_{n\in\mathbb{N}}$, construct a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ satisfying $\|x_{n_{k+1}} - x_{n_k}\| < \frac{1}{2^k}$. Then take $a_1 = x_{n_1}$, $a_2 = x_{n_2} - x_{n_1}$, $a_3 = x_{n_3} - x_{n_2}$, and so on, and use the fact that a Cauchy sequence possessing a convergent subsequence must itself be convergent, which was a result established in Exercise 2.25.

3.3. Notes (not part of the course)

In connection with the divergence of the harmonic series, we mention the following

Erdös conjecture on arithmetic progressions (APs) : If the sums of the reciprocals of the numbers of a set A of natural numbers divergesm then A contains arbitrarily long arithmetic progressions. That is, if $\displaystyle\sum_{n\in A} \frac{1}{n}$ diverges, then A contains APs of any given length.

We know that $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n}$ diverges, and in this case the claim is trivially true. It can be shown that

$$\sum_{p \text{ is prime}} \frac{1}{p}$$

diverges. So one may ask: Does the claim hold in this special case? The answer is "Yes", and this is the Green-Tao Theorem proved in 2004. Terrence Tao was awarded the Fields Medal in 2006, among other things, for this result of his.

Chapter 4

Continuous functions

4.1. Continuity of functions from $\mathbb{R}$ to $\mathbb{R}$

Recall that continuity is a "local" concept, and we have the following notion of the continuity of a function at a point.

Definition 4.1. Let I be an interval and let $c \in I$. $f : I \to \mathbb{R}$ is said to be *continuous at c* if for every $\epsilon > 0$, there exists a $\delta > 0$ such that whenever $x \in I$ satisfies $|x - c| < \delta$, we have $|f(x) - f(c)| < \epsilon$.

f is said to be *continuous on I* if for every $c \in I$, f is continuous at c.

We have seen that if $f, g : \mathbb{R} \to \mathbb{R}$ are continuous on $\mathbb{R}$, then their *composition* $f \circ g : \mathbb{R} \to \mathbb{R}$ defined by

$$(f \circ g)(x) = f(g(x)) \quad (x \in \mathbb{R}),$$

is also continuous on $\mathbb{R}$.

Recall also that we had learnt the following important properties possessed by continuous functions: they preserve convergent sequences, the Intermediate Value Theorem and the Extreme Value Theorem.

Theorem 4.2. *Let I be an interval, $c \in I$, and $f : I \to \mathbb{R}$. The following are equivalent:*

(1) *f is continuous at c.*

(2) *For every sequence $(x_n)_{n\in\mathbb{N}}$ contained in I such that $(x_n)_{n\in\mathbb{N}}$ converges to c, the sequence $(f(x_n))_{n\in\mathbb{N}}$ converges to $f(c)$.*

In other words, f is continuous at c if and only if f "preserves" convergent sequences.

Exercise 4.3. Show that the statement (2) in Theorem 4.2 can be weakened to the following:

($2'$) For every sequence $(x_n)_{n\in\mathbb{N}}$ contained in I such that $(x_n)_{n\in\mathbb{N}}$ converges to c, the sequence $(f(x_n))_{n\in\mathbb{N}}$ converges.

Theorem 4.4 (Intermediate Value Theorem)**.** *Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$. If $y \in \mathbb{R}$ is such that $f(a) \le y \le f(b)$ or $f(b) \le y \le f(a)$ (that is, if y lies between $f(a)$ and $f(b)$), then there exists a $c \in [a, b]$ such that $f(c) = y$.*

In other words, a continuous function attains all real values between the values of the function attained at the endpoints.

Finally, we recall the Extreme Value Theorem.

Theorem 4.5 (Extreme Value Theorem)**.** *Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$. There exists a $c \in [a, b]$ and there exists a $d \in [a, b]$ such that*

$$\begin{aligned} f(c) &= \sup\{f(x) : x \in [a, b]\}, \\ f(d) &= \inf\{f(x) : x \in [a, b]\}. \end{aligned}$$

We note that since $c, d \in [a, b]$, we have $f(c), f(d) \in \{f(x) : x \in [a, b]\}$, and so the supremum and infimum are in fact maximum and minimum, respectively:

$$\begin{aligned} f(c) &= \sup\{f(x) : x \in [a, b]\} = \max\{f(x) : x \in [a, b]\}, \\ f(d) &= \inf\{f(x) : x \in [a, b]\} = \min\{f(x) : x \in [a, b]\}. \end{aligned}$$

We observe that in our definition of continuity of a function at a point, they key idea is that:

"We are guaranteed that $f(x)$ stays close to $f(c)$ for all x close enough to c."

But "closeness" is something we know not just in $\mathbb{R}$ but in the context of general metric spaces! We will now learn that indeed continuity can in fact be defined in a quite abstract setting, when we have maps between metric spaces. We will also gain insights into the above properties of continuous functions when we study analogues of the above results in our more general setting.

Exercise 4.6. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} x & \text{if } x \text{ is rational,} \\ -x & \text{if } x \text{ is irrational.} \end{cases}$$

Prove that f is continuous only at 0. *Hint:* For every rational number, there is a sequence of irrational numbers that converges to it, and for every irrational number, there is a sequence of rational numbers that converges to it, by the results in Exercise 1.38, 1.39.

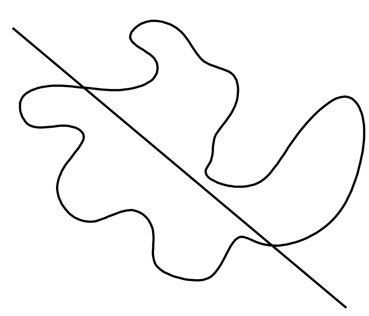
Exercise 4.7. ($*$) Every nonzero rational number x can be uniquely written as $x = n/d$, where n, d denote integers without any common divisors and $d > 0$. When $r = 0$, we take $d = 1$ and $n = 0$. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 0 & \text{if } x \text{ is irrational,} \\ \dfrac{1}{d} & \text{if } x \left(= \dfrac{n}{d}\right) \text{ is rational.} \end{cases}$$

Prove that f is discontinuous at every rational number, and continuous at every irrational number.

Hint: For an irrational number x, given any $\epsilon > 0$, and any interval $(N, N + 1)$ containing x, show that there are just finitely many rational numbers r in $(N, N + 1)$ for which $f(r) \geq \epsilon$. Use this to show the continuity at irrationals.

Exercise 4.8. Consider a flat pancake of arbitrary shape. Show that there is a straight line cut that divides the pancake into two parts having equal areas. Can the direction of the straight line cut be chosen arbitrarily?



4.2. Continuity of maps between metric spaces

Definition 4.9. Let (X, d_X), (Y, d_Y) be metric spaces, $c \in X$ and $f : X \to Y$ be a map. f is said to be *continuous at* c if for every $\epsilon > 0$, there exists a $\delta > 0$ such that whenever $x \in X$ satisfies $d_X(x, c) < \delta$, we have $d_Y(f(x), f(c)) < \epsilon$. See Figure 1.

f is said to be *continuous on* X if for every $c \in X$, f is continuous at c.

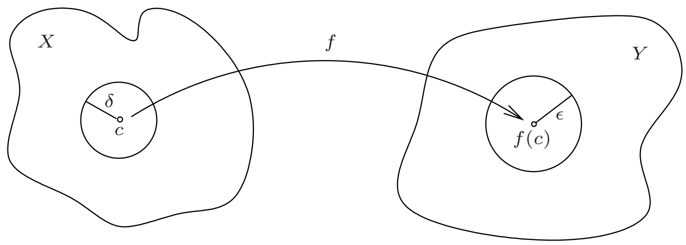


Figure 1. Continuity of f at c.

First of all we notice that if I is an interval in $\mathbb{R}$, and we take $X = I$, $Y = \mathbb{R}$, both equipped with the Euclidean metric, then the above definition of continuity of a function $f : I \to \mathbb{R}$ at a $c \in \mathbb{R}$ coincides with our earlier Definition 4.1.

We remark that although X may be equal to Y, they might be equipped with different metrics; see as an extreme example, Exercise 4.11 below.

Exercise 4.10. Show that $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x) = \left\{ \begin{array}{ll} \dfrac{x_1 x_2}{\sqrt{x_1^2 + x_2^2}} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0 \end{array} \right\}$ is continuous at 0.

Exercise 4.11. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ 1 & \text{if } x > 0. \end{cases}$

(1) If both the domain $X = \mathbb{R}$ and the codomain $Y = \mathbb{R}$ are equipped with the Euclidean metric, then show that f is not continuous at 0.

(2) On the other hand, if we equip the domain $X = \mathbb{R}$ of f with the discrete metric, and the codomain $Y = \mathbb{R}$ of f with the Euclidean metric, then prove that f is continuous at 0.

Exercise 4.12. Let (X, d) be a metric space, and let $p \in X$. Show that the distance to p is a continuous map, that is, prove that the function $f : X \to \mathbb{R}$ defined by $f(x) := d(x, p)$ $(x \in X)$ is continuous.

Exercise 4.13. Show that addition $(x, y) \mapsto x + y$ and multiplication $(x, y) \mapsto xy$ are continuous maps from $\mathbb{R}^2$ to $\mathbb{R}$ with the usual Euclidean metrics.

Exercise 4.14. $(*)$ Consider the normed space $(C[0, 1], \|\cdot\|_\infty)$, and let $S : C[0, 1] \to C[0, 1]$ be defined by

$$(S(f))(x) = (f(x))^2 \quad (x \in [0, 1],\ f \in C[0, 1]).$$

Show that S is continuous.

Proposition 4.15. *Let (X, d_X) be a metric space, and let $f_n : X \to \mathbb{R}$ $(n \in \mathbb{N})$ be a sequence of continuous functions that converges uniformly to $f : X \to \mathbb{R}$. Then f is continuous.*

Proof. Let $c \in X$ and $\epsilon > 0$. Choose an $N \in \mathbb{N}$ such that that for all $x \in X$, $|f_N(x) - f(x)| < \epsilon/3$. As f_N is continuous, we can find a $\delta > 0$ such that for all $x \in X$ satisfying $d_X(x, c) < \delta$, we have $|f_N(x) - f_N(c)| < \epsilon/3$. So for all $x \in X$ satisfying $d_X(x, c) < \delta$, we have using the triangle inequality that

$$|f(x) - f(c)| \leq |f(x) - f_N(x)| + |f_N(x) - f_N(c)| + |f_N(c) - f(c)| < \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon.$$

Hence f is continuous at c. Since the choice of $c \in X$ was arbitrary, it follows that f is continuous on X. $\square$

Exercise 4.16. A subset S of $\mathbb{R}^n$ is said to be *path connected* if for any two points $x, y \in \mathbb{R}^n$, there exists a continuous function $\gamma : [0,1] \to S$ such that $f(0) = x$ and $f(1) = y$. (We think of γ as a "path" beginning at x and ending at y.)

(1) Show that every convex set[1] C ($\subset \mathbb{R}^n$) is path connected.

(2) We define the relation R on S by setting $x\,R\,y$ if there is a path $\gamma : [0,1] \to S$ such that $\gamma(0) = x$ and $\gamma(1) = y$. Prove that R is an equivalence relation on S. The equivalence classes of S under R are called the *path components* of S. Thus if S is path connected, then it has a unique path component, namely S itself.

(3) Which of the following subsets of $\mathbb{R}^2$ are path connected? If the set is not path connected, then determine its path components. $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$, $\{(x, y) \in \mathbb{R}^2 : xy = 0\}$, $\{(x, y) \in \mathbb{R}^2 : xy = 1\}$.

4.3. Continuous maps and open sets

We will now learn an important property of continuous functions, namely that "inverse images" of open sets under a continuous map are open. In fact, we will see that this property is a characterization of continuity.

But first we fix some standard notation. Let $f : X \to Y$ be a map, and let $V \subset Y$. Then we set $f^{-1}(V) := \{x \in X : f(x) \in V\}$, and call it the *inverse image of* V *under* f. See Figure 2. Clearly $f^{-1}(Y) = X$ and $f^{-1}(\emptyset) = \emptyset$.

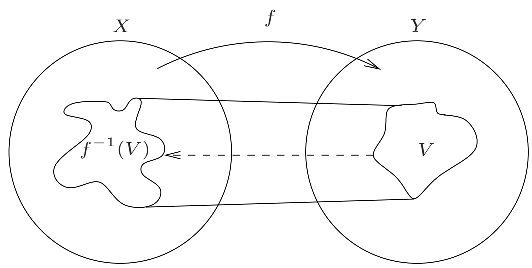


Figure 2. Inverse image of V under f.

Exercise 4.17. Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = \cos x$ ($x \in \mathbb{R}$). Find $f^{-1}(V)$, where $V = \{-1, 1\}$, $V = \{1\}$, $V = [-1, 1]$, $V = \mathbb{R}$, $V = (-\frac{1}{2}, \frac{1}{2})$.

On the other hand if $U \subset X$, then we set $f(U) := \{f(x) \in Y : x \in U\}$, and call it the *image of* U *under* f. See Figure 3.

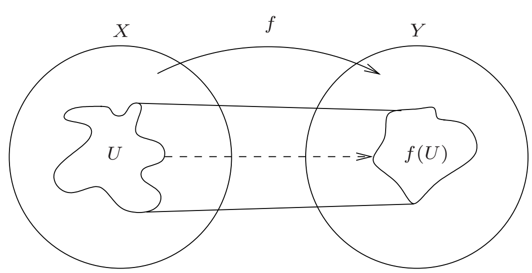


Figure 3. Image of U under f.

Exercise 4.18. Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = \cos x$ ($x \in \mathbb{R}$). Find $f(U)$, where $U = \mathbb{R}$, $U = [0, 2\pi]$, $U = [\delta, \delta + 2\pi]$ where δ is any positive number.

[1] See Exercise 1.40 for the definition of convex sets.

Theorem 4.19. *Let (X, d_X), (Y, d_Y) be metric spaces and $f : X \to Y$ be a map. Then f is continuous on X if and only if for every V open in Y, $f^{-1}(V)$ is open in X.*

Proof. (If) Let $c \in X$, and let $\epsilon > 0$. Consider the open ball $B(f(c), \epsilon)$ with center $f(c)$ and radius ϵ in Y. We know that this open ball $V := B(f(c), \epsilon)$ is an open set in Y. Thus we also know that $f^{-1}(V) = f^{-1}(B(f(c), \epsilon))$ is an open set in X. But the point $c \in f^{-1}(B(f(c), \epsilon))$, because $f(c) \in B(f(c), \epsilon)$ ($d_Y(f(c), f(c)) = 0 < \epsilon$!). So by the definition of an open set, there is a $\delta > 0$ such that $B(c, \delta) \subset f^{-1}(B(f(c), \epsilon))$. In other words, whenever $x \in X$ satisfies $d_X(x, c) < \delta$, we have that $x \in f^{-1}(B(f(c), \epsilon))$, that is, $f(x) \in B(f(c), \epsilon)$, which implies $d_Y(f(x), f(c)) < \epsilon$. Hence f is continuous at c. But the choice of $c \in X$ was arbitrary. Consequently f is continuous on X. See the picture on the left side of Figure 4.

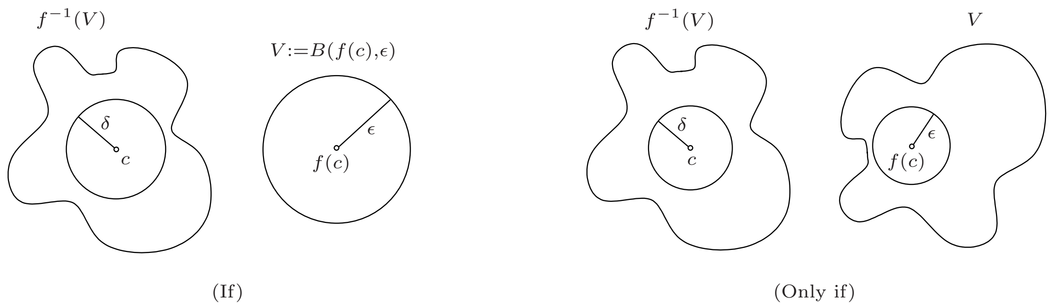


Figure 4. Continuity and open sets: Proof of Theorem 4.19.

(Only if) Now suppose that f is continuous, and let V be an open subset of Y. We would like to show that $f^{-1}(V)$ is open. So let $c \in f^{-1}(V)$. Then $f(c) \in V$. As V is open, there is a small open ball $B(f(c), \epsilon)$ with center $f(c)$ and radius ϵ that is contained in V. By the continuity of f at c, there is a $\delta > 0$ such that whenever $d_X(x, c) < \delta$, we have $d_Y(f(x), f(c)) < \epsilon$, that is, $f(x) \in V$. But this means that $B(c, \delta) \subset f^{-1}(V)$. Indeed, if $x \in B(c, \delta)$, then $d_X(x, c) < \delta$ and so by the above, $f(x) \in V$, that is, $x \in f^{-1}(V)$. Consequently, $f^{-1}(V)$ is open in X. See the picture on the right side of Figure 4. ☐

Note that the theorem does *not* claim that for every U open in X, $f(U)$ is open in Y. Consider for example $X = Y = \mathbb{R}$ equipped with the Euclidean metric, and the constant function $f(x) = c$ ($x \in \mathbb{R}$). Then $X = \mathbb{R}$ is open in $X = \mathbb{R}$, but $f(X) = \{c\}$ is not open in $Y = \mathbb{R}$.

Corollary 4.20. *Let (X, d_X), (Y, d_Y) be metric spaces and $f : X \to Y$ be a map. Then f is continuous on X if and only if for every F closed in Y, $f^{-1}(F)$ is closed in X.*

Proof. If $F \subset Y$, then $f^{-1}(Y \setminus F) = X \setminus (f^{-1}(F))$. ☐

Exercise 4.21. Fill in the details of the proof of Corollary 4.20.

Theorem 4.22. *Let $(X, d_X), (Y, d_Y)$, (Z, d_Z) be metric spaces, $f : X \to Y$ and $g : Y \to Z$ be continuous maps. Then the composition map $g \circ f : X \to Z$, defined by $(g \circ f)(x) := g(f(x))$ ($x \in X$), is continuous.*

Proof. Let W be open in Z. Then since g is continuous, $g^{-1}(W)$ is open in Y. Also, since f is continuous, $f^{-1}(g^{-1}(W))$ is open in X. Finally, we note that $(g \circ f)^{-1}(W) = f^{-1}(g^{-1}(W))$. Consequently, $g \circ f$ is continuous. ☐

Exercise 4.23. In the proof of Theorem 4.22, we used the fact that $(g\circ f)^{-1}(W) = f^{-1}(g^{-1}(W))$. Check this.

Exercise 4.24. Let X be a metric space and $f : X \to \mathbb{R}$ be a continuous map. Determine if the following statements are true or false.

(1) $\{x \in X : f(x) < 1\}$ is an open set.
(2) $\{x \in X : f(x) > 1\}$ is an open set.
(3) $\{x \in X : f(x) = 1\}$ is an open set.
(4) $\{x \in X : f(x) \le 1\}$ is a closed set.
(5) $\{x \in X : f(x) = 1\}$ is a closed set.
(6) $\{x \in X : f(x) = 1 \text{ or } f(x) = 2\}$ is a closed set.
(7) $\{x \in X : f(x) = 1\}$ is a compact set.

Analogous to Theorem 4.2, we have the following characterization of continuous maps in terms of convergence of sequences.

Theorem 4.25. *Let $(X, d_X), (Y, d_Y)$ be metric spaces, $c \in X$, and let $f : X \to Y$ be a map. Then following two statements are equivalent:*

(1) *f is continuous at c.*
(2) *For every sequence $(x_n)_{n\in\mathbb{N}}$ in X such that $(x_n)_{n\in\mathbb{N}}$ converges to c, $(f(x_n))_{n\in\mathbb{N}}$ converges to $f(c)$.*

Proof. (1) $\Rightarrow$ (2): Suppose that f is continuous at c. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in X such that $(x_n)_{n\in\mathbb{N}}$ converges to c. Let $\epsilon > 0$. Then there exists a $\delta > 0$ such that for all $x \in X$ satisfying $d_X(x, c) < \delta$, we have $d_Y(f(x), f(c)) < \epsilon$. As the sequence $(x_n)_{n\in\mathbb{N}}$ converges to c, for this $\delta > 0$, there exists an $N \in \mathbb{N}$ such that whenever $n > N$, $d_X(x_n, c) < \delta$. But then by the above, $d_Y(f(x_n), f(c)) < \epsilon$. So we have shown that for every $\epsilon > 0$, there is an $N \in \mathbb{N}$ such that for all $n > N$, $d_Y(f(x_n), f(c)) < \epsilon$. In other words, the sequence $(f(x_n))_{n\in\mathbb{N}}$ converges to $f(c)$.

(2) $\Rightarrow$ (1): Suppose that f is not continuous at c. This means that there is an $\epsilon > 0$ such that for every $\delta > 0$, there is an $x \in X$ such that $d_X(x, c) < \delta$, but $d_Y(f(x), f(c)) > \epsilon$. We will use this statement to construct a sequence $(x_n)_{n\in\mathbb{N}}$ for which the conclusion in (2) does not hold. Choose $\delta = 1/n$, and denote the corresponding x as x_n: thus, $d_X(x_n, c) < \delta = 1/n$, but $d_Y(f(x_n), f(c)) > \epsilon$. Clearly the sequence $(x_n)_{n\in\mathbb{N}}$ is convergent with limit c, but $(f(x_n))_{n\in\mathbb{N}}$ does not converge to $f(c)$ since $d_Y(f(x_n), f(c)) > \epsilon$ for all $n \in \mathbb{N}$. Consequently if (1) does not hold, then (2) does not hold. In other words, we have shown that (2) $\Rightarrow$ (1). □

Exercise 4.26. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function such that for all $x, y \in \mathbb{R}$, $f(x+y) = f(x)+f(y)$. Show that there exists a real number a such that for all $x \in \mathbb{R}$, $f(x) = ax$. *Hint:* Show first that for natural numbers n, $f(n) = nf(1)$. Extend this to integers n, and then to rational numbers n/d. Finally use the density of $\mathbb{Q}$ in $\mathbb{R}$ to prove the claim.

Exercise 4.27. Find all continuous functions $f : \mathbb{R} \to \mathbb{R}$ such that for all $x \in \mathbb{R}$, $f(x) + f(2x) = 0$.
Hint: Show that $f(x) = -f(\frac{x}{2}) = f(\frac{x}{4}) = -f(\frac{x}{8}) = \cdots$.

Exercise 4.28. Show that the multiplication function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x) = x_1x_2$ ($x \in \mathbb{R}^2$) is continuous on $\mathbb{R}^2$ using the characterization of continuous functions in terms of preservation of convergent sequences. Compare this with Exercise 4.13.

Exercise 4.29. (∗) Two metric spaces are called *homeomorphic* if there exists a continuous bijection $f : X \to Y$ such that $f^{-1} : Y \to X$ is also continuous. For example, circles and triangles in $\mathbb{R}^2$ are homeomorphic; see Figure 5.

Similarly, the an open interval in $\mathbb{R}$ is homeomorphic to a circle missing one point in $\mathbb{R}^2$: we just "identify" the two ends of the interval, as shown in the picture on the left hand side of Figure 6. Using this, one can

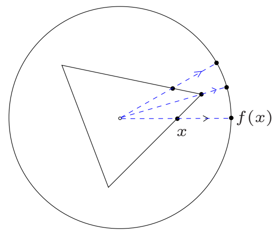


Figure 5. Homeomorphism between the triangle and the circle in $\mathbb{R}^2$.

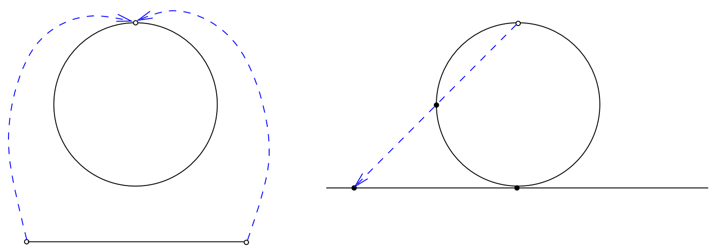


Figure 6. $\mathbb{R}$ is homeomorphic to an open interval in $\mathbb{R}$.

also see pictorially that the open interval in $\mathbb{R}$ is homeomorphic to $\mathbb{R}$, as shown in the picture on the right hand side of Figure 6.

A natural question is whether the continuity of f^{-1} is actually implied by the continuity of a bijection f. It is not, and the aim of this exercise is show such an instructive example. Look at Figure 7, where a continuous bijection f is produced from the set $X := [0, 1)$ onto the triangle in $\mathbb{R}^2$ by "bending" the interval at the three points B, C, D, stretching the parts between the segments BC and CD, and sending points A, B, C, D to the points $f(A) = (1, 0), f(B), f(C), f(D)$ as shown. But now prove that f^{-1} can't be continuous by looking at the sequence $(1, (-1)^n/n)$.

Figure 7. The triangle in $\mathbb{R}^2$ and the interval $[0, 1)$.

Exercise 4.30. ($*$) A "manifold" is a topological space that locally resembles the Euclidean space. More precisely, we will call a subset M of $\mathbb{R}^n$ a *manifold* (*of dimension*[2] k) if for every $x \in M$, there is an open set O_x containing x such that O_x is homeomorphic to an open subset U of $\mathbb{R}^k$. Locally the surface of the earth, which is a sphere, looks flat, and so we expect that the sphere in $\mathbb{R}^3$ is a manifold of dimension 2. Give an argument, based on pictures, that the unit sphere

$$\mathbb{S}^2 := \{x \in \mathbb{R}^3 : \|x\|_2 = 1\}$$

is indeed a manifold of dimension 2. (This explains why one uses the superscript "2" on top of $\mathbb{S}$ in the (standard) notation for the unit sphere in $\mathbb{R}^3$. Similarly the circle $\mathbb{S}^1 := \{x \in \mathbb{R}^2 : \|x\|_2 = 1\}$ in $\mathbb{R}^2$ is a manifold of dimension 1, and is denoted by $\mathbb{S}^1$. More generally, it can be shown that the unit sphere $\mathbb{S}^{n-1} := \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$ in $\mathbb{R}^n$ is a manifold of dimension $n - 1$ in $\mathbb{R}^n$.)

[2] It can be shown that this is a well-defined notion.

Exercise 4.31. (∗) Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x) = \begin{cases} \dfrac{x_1^2 - x_2^2}{x_1^2 + x_2^2} & \text{if } x \neq 0, \\ c & \text{if } x = 0. \end{cases}$$

Show that no matter what $c \in \mathbb{R}$ we take in the above, f is not continuous at 0.

Exercise 4.32. (∗) Show that the determinant function $M \mapsto \det M$ from $(\mathbb{R}^{n\times n}, \|\cdot\|_\infty)$ to $(\mathbb{R}, |\cdot|)$ is continuous. Prove that the set of invertible matrices is open in $(\mathbb{R}^{n\times n}, \|\cdot\|_\infty)$. *Hint:* Consider $\det^{-1}(\{0\})$.

Exercise 4.33. Give an example of a continuous function $f : X \to Y$, where X, Y are metric spaces, and a Cauchy sequence $(x_n)_{n\in\mathbb{N}}$ for which $(f(x_n))_{n\in\mathbb{N}}$ is not a Cauchy sequence in Y.

Exercise 4.34. Let X, Y be metric spaces. A map $f : X \to Y$ is said to be *open* if for every open subset U of X, $f(U)$ is open in Y.

Now take $X = \mathbb{R}$ with the usual Euclidean metric, and take $Y = \mathbb{R}$ with the discrete metric. Consider the identity map $f : X \to Y$ defined by $f(x) = x$ $(x \in \mathbb{R})$. Show that f is open, but for all $x \in \mathbb{R}$, f is not continuous at x.

Exercise 4.35. (∗) Define f, g on $\mathbb{R}^2$ by $f(0,0) = g(0,0) = 0$, and if $(x,y) \neq (0,0)$, then we set

$$f(x,y) = \frac{xy^2}{x^2+y^4}, \quad g(x,y) = \frac{xy^2}{x^2+y^6}.$$

(1) Show that f is bounded on $\mathbb{R}^2$, that is, $\exists M \in \mathbb{R}$ such that for all $(x,y) \in \mathbb{R}^2$, $|f(x,y)| \leq M$.
(2) Prove that g is unbounded in every ball centered at $(0,0)$.
(3) Show that f is not continuous at $(0,0)$.
(4) Prove that g is not continuous at $(0,0)$.
(5) If $Y \subset X$ and if Φ is a function defined on X, the *restriction* of Φ to Y is the function φ whose domain is Y, and such that $\varphi(y) = \Phi(y)$ $(y \in Y)$. Show that the restrictions of both f and g to every straight line in $\mathbb{R}^2$ are continuous!

For functions from the Euclidean space $\mathbb{R}^n$ to the Euclidean space $\mathbb{R}^m$, we have the following simplification.

Proposition 4.36. *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous if and only if each of its components $f_1, \dots, f_m : \mathbb{R}^n \to \mathbb{R}$ are continuous. (Here $f_k(x) := e_k^\top x$ $(x \in \mathbb{R}^n)$, and e_k, $k = 1, \dots, d$ are the standard basis vectors.)*

Proof. We have $|f_k(x) - f_k(y)| \leq \sqrt{\sum_{i=1}^{n} |f_i(x) - f_i(y)|^2} = \|f(x) - f(y)\|_2$. □

Another case when checking continuity becomes considerably simpler is in the case of *linear transformations* between *normed* spaces.

Proposition 4.37. *Let $(X, \|\cdot\|_X)$, $(Y, \|\cdot\|_Y)$ be normed spaces and let $T : X \to Y$ be a linear transformation. Then the following are equivalent:*

(1) *T is continuous.*
(2) *T is continuous at 0.*
(3) *There exists an $M > 0$ such that for all $x \in X$, $\|Tx\|_Y \leq M\|x\|_X$.*

Proof. (1) ⇒ (2) follows from the definition. Let us show that (2) ⇒ (3). As T is continuous at 0, we have that given $\epsilon := 1 > 0$, there is a $\delta > 0$ such that whenever $\|x - 0\|_X = \|x\|_X < \delta$, we have $\|Tx - T0\|_Y = \|Tx - 0\|_Y = \|Tx\|_Y < 1$. Define $M := 2/\delta$. Then if $x = 0$, we have

$\|Tx\|_Y = \|T0\|_Y = \|0\|_Y = 0 \leq (2/\delta)\cdot 0 = M \cdot \|0\|_X = M\|x\|_X$. If on the other hand, $x \neq 0$,then define

$$y := \frac{\delta}{2\|x\|_X} x.$$

Then $\|y\| = \delta/2 < \delta$ and so $\|Ty\|_Y < 1$. We have

$$1 > \|Ty\|_Y = \left\| T\left(\frac{\delta}{2\|x\|_X} x\right)\right\|_Y = \frac{\delta}{2\|x\|_X}\|Tx\|_Y,$$

and so upon rearranging, we obtain

$$\|Tx\|_Y < \frac{2}{\delta}\|x\|_X = M\|x\|_X.$$

Consequently (3) holds.

Finally, we show that (3) ⇒ (1). Let $c \in X$, and $\epsilon > 0$. Choose $\delta = \epsilon/M > 0$. Then whenever $\|x - c\|_X < \delta = \epsilon/M$, we have

$$\|Tx - Tc\|_Y = \|T(x-c)\|_Y \leq M\|x - c\|_X < M\delta = M\frac{\epsilon}{M} = \epsilon.$$

Hence f is continuous at c. But the choice of $c \in X$ was arbitrary, and so f is continuous. □

Example 4.38. Consider the map $I : C[a, b] \to \mathbb{R}$ from the normed space $(C[a, b], \|\cdot\|_\infty)$ to $\mathbb{R}$ given by

$$I(f) = \int_a^b f(x)dx \quad (f \in C[a, b]).$$

Then clearly I is a linear transformation. Moreover, since for every $f \in C[a, b]$ we have

$$|I(f)| = \left|\int_a^b f(x)dx\right| \leq \int_a^b |f(x)|dx \leq \int_a^b \|f\|_\infty dx = \|f\|_\infty (b - a),$$

it follows that I is continuous. ◊

Example 4.39. Let $A \in \mathbb{R}^{n\times m}$. Consider the map T_A from the Euclidean space $\mathbb{R}^n$ to the Euclidean space $\mathbb{R}^m$, given by matrix multiplication:

$$T_A x = Ax \quad (x \in \mathbb{R}^n).$$

Then T_A is a linear transformation, and it is continuous, since

$$\|T_A x\|_2 = \|Ax\|_2 = \sqrt{\sum_{i=1}^m \left(\sum_{j=1}^n a_{ij}x_j\right)^2} \leq \sqrt{\sum_{i=1}^m n\|A\|_\infty^2 \|x\|_2^2} = \sqrt{mn}\|A\|_\infty \|x\|_2.$$

Hence T_A is continuous. ◊

Exercise 4.40. Show that if $A \in \mathbb{R}^{n\times m}$, then $\ker A = \{x \in \mathbb{R}^m : Ax = 0\}$ is a closed subspace of $\mathbb{R}^n$.

Exercise 4.41. (∗) Prove that *every* subspace of $\mathbb{R}^n$ is closed. *Hint:* Construct a linear transformation whose kernel is the given subspace.

Exercise 4.42. A metric space X is said to be *connected* if X is not the union of two disjoint nonempty sets.

Prove that a continuous image of a connected set is connected, that is, if X and Y are metric spaces such that X is connected, and if $f : X \to Y$ is continuous, then $f(X)$ is connected as a metric space (with the metric induced from Y).

4.3.1. Limits and continuity.

Definition 4.43. Let U be an open subset of $\mathbb{R}^n$, $c \in U$, $L \in \mathbb{R}^m$ and $f : U \to \mathbb{R}^m$. Then we write

$$f(x) \stackrel{x \to c}{\longrightarrow} L$$

or

$$\lim_{x \to c} f(x) = L$$

if for every $\epsilon > 0$, there exists a $\delta > 0$ such that whenever $x \in U$ satisfies $0 < \|x - c\|_2 < \delta$, we have $\|f(x) - L\|_2 < \epsilon$. We then say that f *has a limit at* c, and call L its *limit*.

We can recast this definition in terms of sequences.

Theorem 4.44. *Let U be an open subset of $\mathbb{R}^n$, $c \in U$, $L \in \mathbb{R}^m$ and $f : U \to \mathbb{R}^m$. Then*

$$\lim_{x \to c} f(x) = L$$

if and only if for every sequence $(x_n)_{n \in \mathbb{N}}$ contained in U such that $x_n \neq c$ $(n \in \mathbb{N})$, $\lim_{n \to \infty} x_n = c$,

$$\lim_{n \to \infty} f(x_n) = L.$$

Proof. (If) Suppose that $\lim_{x \to c} f(x) = L$ does not hold. Then there is an $\epsilon > 0$ such that for every $\delta > 0$, there is a point $x \in U$ (depending on δ), for which $0 < \|x - c\|_2 < \delta$, but $\|f(x) - L\|_2 \geq \epsilon$. Taking δ successively to be $1/n$ $(n \in \mathbb{N})$, we can thus find a sequence $x_n \in U$ such that $x_n \neq c$ $(n \in \mathbb{N})$ and $\|f(x_n) - L\|_2 \geq \epsilon$. But this last condition means that $\lim_{n \to \infty} f(x_n) = L$ does not hold. This completes the proof of the "if part".

(Only if) Suppose now that $\lim_{x \to c} f(x) = L$, and that $(x_n)_{n \in \mathbb{N}}$ is a sequence contained in U such that $x_n \neq c$ $(n \in \mathbb{N})$, $\lim_{n \to \infty} x_n = c$. Let $\epsilon > 0$. Then there exists a $\delta > 0$ such that whenever $x \in U$ satisfies $0 < \|x - c\|_2 < \delta$, we have $\|f(x) - L\|_2 < \epsilon$. Also, there exists an $N \in \mathbb{N}$ such that for all $n > N$, $0 < \|x_n - c\| < \delta$. Consequently, for $n > N$ we have $\|f(x_n) - L\|_2 < \epsilon$. Hence $\lim_{n \to \infty} f(x_n) = L$. □

Corollary 4.45. *Let U be an open subset of $\mathbb{R}^n$, $c \in U$ and $f : U \to \mathbb{R}^m$. If f has a limit at c, then it is unique.*

This result follows from the above theorem and the fact that convergent sequences have a unique limit.

Moreover, using the algebra of limits for real sequences, it follows also that the same sort of results carry over to limits of real-valued functions.

Theorem 4.46. *Let U be an open subset of $\mathbb{R}^n$, $c \in U$. Suppose that $f, g : U \to \mathbb{R}$ and*

$$\lim_{x \to c} f(x) = L_f, \quad \lim_{x \to c} g(x) = L_g,$$

where $L_f, L_g \in \mathbb{R}$. Define $f + g, fg : U \to \mathbb{R}$ by

$$(f + g)(x) = f(x) + g(x), \quad (fg)(x) = f(x)g(x) \quad (x \in U).$$

Then we have

(1) $\lim_{x \to c}(f + g)(x) = L_f + L_g = \lim_{x \to c} f(x) + \lim_{x \to c} g(x)$.

(2) $\lim_{x \to c}(fg)(x) = L_f L_g = \left(\lim_{x \to c} f(x)\right)\left(\lim_{x \to c} g(x)\right)$.

The following result is clear from the definitions.

Theorem 4.47. *Let U be an open subset of $\mathbb{R}^n$, $c \in U$. Then $f : U \to \mathbb{R}^m$ is continuous at c if and only if*

$$\lim_{x \to c} f(x) = f(c).$$

4.4. Compactness and continuity

In this section we will learn about a very useful result in Optimization Theory, on the existence of global minimizers of real-valued continuous functions on compact sets.

Theorem 4.48. *Let K be a compact subset of a metric space X, and let Y be a metric space. Suppose that $f : K \to Y$ is a continuous function. Then $f(K)$ is a compact subset of Y.*

Proof. Suppose that $(y_n)_{n\in\mathbb{N}}$ is a sequence in contained in $f(K)$. Then for each $n \in \mathbb{N}$, there exists an $x_n \in K$ such that $y_n = f(x_n)$. Thus we obtain a sequence $(x_n)_{n\in\mathbb{N}}$ in the set K. As K is compact, there exists a convergent subsequence, say $(x_{n_k})_{k\in\mathbb{N}}$, with limit $L \in K$. As f is continuous, it preserves convergent sequences. So $(f(x_{n_k}))_{k\in\mathbb{N}} = (y_{n_k})_{k\in\mathbb{N}}$ is convergent with limit $f(L) \in f(K)$. Consequently, $f(K)$ is compact. □

Now we prove the aforementioned result which turns out to be very useful in Optimization Theory, namely that a real-valued continuous function on a compact set attains its maximum/minimum. This is a generalization of the Extreme-Value Theorem we had learnt earlier, where the compact set in question was just the interval $[a, b]$.

Theorem 4.49 (Weierstrass's theorem)**.** *Let K be a nonempty compact subset of a metric space X, and let $f : K \to \mathbb{R}$ be a continuous function. Then there exists a $c \in K$ such that*

$$f(c) = \sup\{f(x) : x \in K\}.$$

We note that since $c \in K$, $f(c) \in \{f(x) : x \in K\}$, and so the supremum above is actually a maximum:

$$f(c) = \sup\{f(x) : x \in K\} = \max\{f(x) : x \in K\}.$$

Also, under the same hypothesis of the above result, there exists a minimizer in K, that is, there exists a $d \in K$ such that

$$f(d) = \inf\{f(x) : x \in K\} = \min\{f(x) : x \in K\}.$$

This follows from the above result by just looking at $-f$, that is by applying the above result to the continuous function $g : K \to \mathbb{R}$ given by

$$g(x) = -f(x) \quad (x \in K).$$

Proof of Theorem 4.49. We know that the image of K under f, namely the set $f(K)$ is compact and hence bounded. So $\{f(x) : x \in K\}$ is bounded. It is also nonempty since K is nonempty. But by the least upper bound property of $\mathbb{R}$, a nonempty bounded subset of $\mathbb{R}$ has a least upper bound. Thus $M := \sup\{f(x) : x \in K\} \in \mathbb{R}$. Now consider $M - 1/n$ ($n \in \mathbb{N}$). This number cannot be an upper bound for $\{f(x) : x \in K\}$. So we there must be an $x_n \in K$ such that $f(x_n) > M - 1/n$. In this manner we get a sequence $(x_n)_{n\in\mathbb{N}}$ in K. As K is compact, $(x_n)_{n\in\mathbb{N}}$ has a convergent subsequence $(x_{n_k})_{k\in\mathbb{N}}$ with limit, say c, belonging to K. As f is continuous, $(f(x_{n_k}))_{k\in\mathbb{N}}$ is convergent as well with limit $f(c)$. But from the inequalities $f(x_n) > M - 1/n$ ($n \in \mathbb{N}$), it follows that $f(c) \geq M$. On the other hand, from the definition of M, we also have that $f(c) \leq M$. Hence $f(c) = M$. □

Example 4.50. Since the set $K = \{x \in \mathbb{R}^3 : x_1^2 + x_2^2 + x_3^2 = 1\}$ is compact in $\mathbb{R}^3$ and since the function $x \mapsto x_1 + x_2 + x_3$ is continuous on $\mathbb{R}^3$, it follows that the optimization problem

$$\left\{ \begin{array}{ll} \text{minimize} & x_1 + x_2 + x_3 \\ \text{subject to} & x_1^2 + x_2^2 + x_3^2 = 1 \end{array} \right\},$$

has a minimizer. ◊

Remark 4.51. (∗) In Optimization Theory, one often meets *necessary* conditions for an optimal solution, that is, results of the following form:

> If $\widehat{x}$ is an optimal solution to the optimization problem $\left\{ \begin{array}{ll} \text{maximize} & f(x) \\ \text{subject to} & x \in \mathcal{F}\ (\subset \mathbb{R}^n) \end{array} \right\}$,
>
> then $\widehat{x}$ satisfies $\boxed{***}$.

(Where $\boxed{***}$ are certain mathematical conditions, such as the Lagrange multiplier equations.) Now such a result has limited use as such since even if we find all $\widehat{x}$(s) which satisfy $\boxed{***}$, we can't conclude that there is one that is optimal. But now suppose that we know that $f : \mathcal{F} \to \mathbb{R}$ is continuous and that $\mathcal{F}$ is compact. Then we know that an optimal solution exists, and so we know that among the $\widehat{x}$(s) that satisfy $\boxed{***}$, there is at least one which is an optimal solution.

Exercise 4.52. (∗) Let $N : \mathbb{R}^d \to \mathbb{R}$ be any norm on $\mathbb{R}^d$. The aim of this exercise is to show that N is "equivalent to[3]" $\|\cdot\|_2$, that is, there are constants $M, m > 0$ such that

$$\text{for all } x \in \mathbb{R}^d,\ m\|x\|_2 \le N(x) \le M\|x\|_2.$$

(1) Let $e_1 = (1, 0, \dots, 0), \dots, e_d = (0, \dots, 0, 1)$ be the standard basis vectors in $\mathbb{R}^d$. Thus the vector $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ is the linear combination $x_1 e_1 + \cdots + x_d e_d$. Show, using the Cauchy-Schwarz inequality, that there is a constant $M > 0$ such that for all $x \in \mathbb{R}^d$, $N(x) \le M\|x\|_2$.

(2) Prove using the triangle inequality for N that for all $x, y \in \mathbb{R}^d$, $|N(x) - N(y)| \le N(x - y)$. Conclude that the map $N : (\mathbb{R}^d, \|\cdot\|_2) \to \mathbb{R}$ is continuous.

(3) Consider the compact set $K := \{x \in \mathbb{R}^d : \|x\|_2 = 1\}$ and use Weierstrass's theorem to prove the existence of $m > 0$ such that for all $x \in \mathbb{R}^d$, $m\|x\|_2 \le N(x)$.

Exercise 4.53. In each case, give an example of a continuous function $f : S \to T$, such that $f(S) = T$ or else explain why there can be no such f. (We use the usual topologies, for example $(0, 1)$ in the first part has the Euclidean topology of $\mathbb{R}$.)

(1) $S = (0, 1)$, $T = (0, 1]$.

(2) $S = (0, 1)$, $T = (0, 1) \cup (1, 2)$.

(3) $S = \mathbb{R}$, $T = \mathbb{Q}$.

(4) $S = [0, 1] \bigcup [2, 3]$, $T = \{0, 1\}$.

(5) $S = [0, 1] \times [0, 1]$, $T = \mathbb{R}^2$.

(6) $S = [0, 1] \times [0, 1]$, $T = (0, 1) \times (0, 1)$.

(7) $S = (0, 1) \times (0, 1)$, $T = \mathbb{R}^2$.

Exercise 4.54. (∗) Let (X, d) be a metric space and let $f : X \to X$ be a function that satisfies

$$\text{for all } x, y \in X \text{ such that } x \neq y,\ \ d(f(x), f(y)) < d(x, y). \tag{4.1}$$

(1) Prove that f has at most one fixed point (that is, a point $c \in X$ such that $f(c) = c$).

(2) Let $X = (0, \frac{1}{2})$ with the usual topology and consider the function $f : X \to X$ given by $f(x) = x^2$ for $x \in (0, \frac{1}{2})$. Show that f satisfies (4.54), but it has no fixed point.

(3) Show that the function $g : X \to \mathbb{R}$ given by $g(x) = d(x, f(x))$ $(x \in X)$ is continuous.

(4) Prove that if X is compact, then f has exactly one fixed point.
Hint: g attains a minimum on X.

[3]The fuss about equivalent norms on a vector space X is that whenever two norms N_1, N_2 are equivalent, the open sets in (X, N_1) coincide with the ones in (X, N_2), and so as topological spaces, they are the same!

Exercise 4.55. Let X be a compact metric space and let $f : X \to \mathbb{Z}$ be a continuous function. (Here $\mathbb{Z}$ has the Euclidean topology induced from $\mathbb{R}$.) Prove that f can assume only finitely many values.

Exercise 4.56. ($*$)

(1) Show that $[0,1]$ and $(0,1)$ are homeomorphic when both spaces are equipped with the discrete metric.

(2) Show that $[0,1]$ and $(0,1)$ are not homeomorphic when both spaces are equipped with the Euclidean metric.

4.5. Uniform continuity

Definition 4.57. Let $(X, d_X), (Y, d_Y)$ be metric spaces. $f : X \to Y$ is said to be *uniformly continuous* if for every $\epsilon > 0$, there exists a $\delta > 0$ such that for all $x, y \in X$ satisfying $d_X(x, y) < \delta$, there holds that $d_Y(f(x), f(y)) < \epsilon$.

Note that in the definition we are introducing the notion of uniform continuity of a function *on a set*, and not *at a point*.

While checking *continuity* of a function *on a set*, we check its continuity at every point as follows. Given a point c in X and an $\epsilon > 0$, we need to find a ball of radius δ around c such that the image of that ball under f is contained in a ball of radius ϵ around $f(c)$. But in general, the radius δ of the ball around c may depend on the c we choose, and it may not be possible to use the same δ everywhere in X. For a *uniformly continuous* function, it is possible to choose a δ that depends only on ϵ and *not* on the point c.

Proposition 4.58. *Let $(X, d_X), (Y, d_Y)$ be metric spaces. If $f : X \to Y$ is uniformly continuous, then f is continuous.*

Proof. Let $c \in X$. Suppose that $\epsilon > 0$. By the uniform continuity of f, there exists a $\delta > 0$ such that for all $x, y \in X$ satisfying $d_X(x, y) < \delta$, there holds that $d_Y(f(x), f(y)) < \epsilon$. In particular, if $x \in X$ satisfies $d_X(x, c) < \delta$, we have $d_Y(f(x), f(c)) < \epsilon$. Thus f is continuous at c. But the choice of c was arbitrary, and so f is continuous on X. □

The following example shows that uniform continuity is a strictly stronger notion than continuity, that is, there are continuous functions that are not uniformly continuous.

Example 4.59. The function $f : (0,1) \to \mathbb{R}$ given by

$$f(x) = 1/x \quad (0 < x < 1)$$

is continuous on $(0,1)$. Indeed, if $(x_n)_{\mathbb{N}}$ is a convergent sequence in $(0,1)$ with limit L, then $(f(x_n))_{n\in\mathbb{N}} = (1/x_n)_{\mathbb{N}}$ is a convergent sequence with limit $1/L = f(L)$.

However, f is not uniformly continuous. Suppose it is. Then given $\epsilon > 0$, there exists a $\delta > 0$ such that whenever $|x - y| < \delta$, we have

$$\left|\frac{1}{x} - \frac{1}{y}\right| < \epsilon.$$

Consider $x = \dfrac{1}{n}$ and $y = \dfrac{1}{2n}$. Then $|x - y| = \dfrac{1}{2n}$, and so $|x - y| < \delta$ for all n large enough, but

$$\left|\frac{1}{x} - \frac{1}{y}\right| = n > \epsilon,$$

for n large enough. ◊

Exercise 4.60. ($*$) Show that $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ ($x \in \mathbb{R}$) is continuous, but not uniformly continuous. *Hint:* Consider $x = n$ and $y = n + \frac{1}{n}$ for large n.

Proposition 4.61. *Let $(X, d_X), (Y, d_Y)$ be metric spaces, and suppose that X is compact. If $f : X \to Y$ is continuous, then f is also uniformly continuous.*

Proof. We give an argument by contradiction. So let us suppose that f is not uniformly continuous. Then there exists an $\epsilon > 0$ such that for every $\delta > 0$, there are some $x, y \in X$ such that $d_X(x, y) < \delta$, but $d_Y(f(x), f(y)) \geq \epsilon$. In particular, if we take $\delta = 1/n$, then there exist $x_n, y_n \in X$ such that $d_X(x_n, y_n) < 1/n$ but $d_Y(f(x_n), f(y_n)) \geq \epsilon$. By using the compactness of X, and considering subsequences if necessary, we may assume that $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ are convergent, with limits say $x, y \in X$, respectively[4]. Since $d_X(x_n, y_n) < 1/n$, we obtain $d_X(x, y) \leq 0$, and so $x = y$. Also, by the continuity of f, we have $(f(x_n))_{n\in\mathbb{N}}, (f(y_n))_{n\in\mathbb{N}}$ converge respectively to $f(x), f(y)$. Hence $d_Y(f(x_n), f(y_n))$ converges to $d_Y(f(x), f(y)) = 0$ (as $x = y$!). But on the other hand, from $d_Y(f(x_n), f(y_n)) \geq \epsilon$, we obtain $d_Y(f(x), f(y)) \geq \epsilon > 0$, a contradiction. □

Exercise 4.62. Prove that the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = |x|$ ($x \in \mathbb{R}$) is uniformly continuous.

Exercise 4.63. Let (X, d) be a metric space and let $c \in X$. Define $f : X \to \mathbb{R}$ by $f(x) = d(x, c)$. Prove that f is uniformly continuous on X.

Exercise 4.64. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is said to be (*globally*) *Lipschitz* if there exists a number $L > 0$ such that for all $x, y \in \mathbb{R}^n$, $\|f(x) - f(y)\|_2 \leq L\|x - y\|_2$. Prove that every Lipschitz function is uniformly continuous.

Exercise 4.65. Let X, Y be metric spaces, and let $f : X \to Y$ be uniformly continuous. Show that if $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in X, then $(f(x_n))_{n\in\mathbb{N}}$ is a Cauchy sequence in Y. Compare this with Exercise 4.33.

---

[4]Since X is compact, the sequence $(x_n)_{n\in\mathbb{N}}$ has a convergent subsequence, say $(x_{n_k})_{k\in\mathbb{N}}$, converging to, say $x \in X$. Also, the sequence $(y_{n_k})_{k\in\mathbb{N}}$ has a convergent subsequence, say $(y_{n_{k_\ell}})_{\ell\in\mathbb{N}}$, converging to, say $y \in Y$.

Chapter 5

Differentiation

Given a function $f : (a,b) \to \mathbb{R}$, and a point $c \in (a,b)$, consider the *difference quotient* for $x \in (a,b)$, $x \neq c$:

$$\frac{f(x) - f(c)}{x - c}.$$

Geometrically, this number represents the slope of the chord passing through the points $(c, f(c))$ and $(x, f(x))$ on the graph of f; see Figure 1.

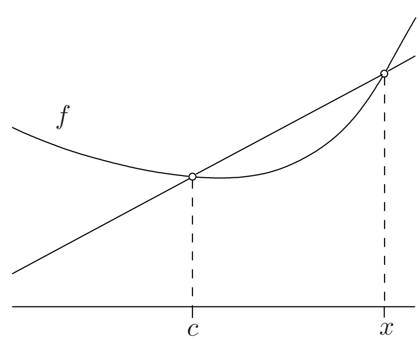


Figure 1. The difference quotient.

Suppose that as x goes to c, the difference quotients approach a number, say L, that is,

$$\lim_{x \to c} \frac{f(x) - f(c)}{x - c} = L.$$

In other words, for every $\epsilon > 0$, there exists a $\delta > 0$ such that whenever $x \in (a,b)$ satisfies $0 < |x - c| < \delta$, we have

$$\left| \frac{f(x) - f(c)}{x - c} - L \right| < \epsilon.$$

Then we say that f is *differentiable at the point* c. See Figure 2, where we see the geometric interpretation of L: it is the slope of the tangent to the graph of f at the point c. Notice also that if we "zoom into" the graph of f around the point $(c, f(c))$, the graph seems to coincide with the tangent line. In other words, the tangent line is a "linear approximation" of f near the point c.

The number L is unique, and we denote this unique number by $f'(c)$, and call it the *derivative of* f at c. If f is differentiable at every $c \in (a,b)$, then we say that f is *differentiable on* (a,b).

Theorem 5.1. *Let $f : (a,b) \to \mathbb{R}$ be differentiable at $c \in (a,b)$. Then f is continuous at c.*

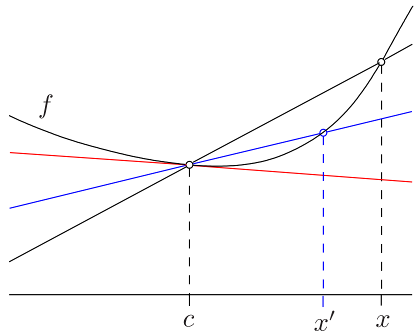


Figure 2. The chords approach the tangent line as x approaches c.

Proof. Let $\epsilon > 0$. First choose a $\delta' > 0$ such that for all $x \in (a, b)$ such that $0 < |x - c| < \delta'$, we have

$$\left| \frac{f(x) - f(c)}{x - c} - f'(c) \right| < 1.$$

Then rearrangement (using the triangle inequality) gives

$$|f(x) - f(c)| < (1 + |f'(c)|)|x - c|.$$

Now define $\delta := \min \left\{ \delta', \frac{\epsilon}{1 + |f'(c)|} \right\}$. Then for all $x \in (a, b)$ such that $0 < |x - c| < \delta$, we have

$$|f(x) - f(c)| < (1 + |f'(c)|)|x - c| \leq (1 + |f'(c)|) \frac{\epsilon}{1 + |f'(c)|} = \epsilon.$$

Consequently f is continuous at c. □

The converse of the theorem is not true, and the following example demonstrates this.

Example 5.2. The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = |x|$ $(x \in \mathbb{R})$ is (uniformly) continuous since

$$\Big| f(x) - f(y) \Big| = \Big| |x| - |y| \Big| \leq |x - y|.$$

But let us now show that f is not differentiable at 0. If it were, then given an $\epsilon > 0$, there exists a $\delta > 0$ such that whenever $0 < |x| < \delta$, we have

$$\left| \frac{|x|}{x} - f'(0) \right| < \epsilon.$$

In particular, if we take $x = \delta/2$, then we obtain $|1 - f'(0)| < \epsilon$. If we take $x = -\delta/2$, we also get $| - 1 - f'(0)| < \epsilon$. As the choice of $\epsilon > 0$ was arbitrary, we obtain $|1 - f'(0)| \leq 0$ and $| - 1 - f'(0)| \leq 0$. But this means that $f'(0) = 1$ and $f'(0) = -1$, and so $1 = -1$, a contradiction. ♢

One can show the following result about rules for differentiating the sum and product of differentiable functions.

Proposition 5.3. *Let $f, g : (a, b) \to \mathbb{R}$ be differentiable at $c \in (a, b)$. Then:*

(1) *The sum $f + g : (a, b) \to \mathbb{R}$ defined by $(f + g)(x) = f(x) + g(x)$ $(x \in (a, b))$ is differentiable at c, and $(f + g)'(c) = f'(c) + g'(c)$.*

(2) *The product $fg : (a, b) \to \mathbb{R}$ defined by $(fg)(x) = f(x) \cdot g(x)$ $(x \in (a, b))$ is differentiable at c, and $(fg)'(c) = f'(c)g(c) + f(c)g'(c)$.*

Proof. These claims follow from the algebra of limits, namely Theorem 4.46. Indeed we have

$$\lim_{x\to c}\frac{(f+g)(x)-(f+g)(c)}{x-c}=\lim_{x\to c}\frac{f(x)-f(c)}{x-c}+\lim_{x\to c}\frac{g(x)-g(c)}{x-c}=f'(c)+g'(c),$$

which proves (1). Also, (2) follows from the following:

$$\begin{aligned}\lim_{x\to c}\frac{(fg)(x)-(fg)(c)}{x-c} &= \lim_{x\to c}\frac{f(x)g(x)-f(c)g(x)+f(c)g(x)-f(c)g(c)}{x-c}\\ &= \lim_{x\to c}\frac{f(x)-f(c)}{x-c}\cdot g(x)+\lim_{x\to c}f(c)\cdot\frac{g(x)-g(c)}{x-c}\\ &= \lim_{x\to c}\frac{f(x)-f(c)}{x-c}\cdot\lim_{x\to c}g(x)+f(c)\lim_{x\to c}\frac{g(x)-g(c)}{x-c}\\ &= f'(c)g(c)+f(c)g'(c).\end{aligned}$$

This completes the proof. □

Example 5.4. The derivative of the constant function is clearly zero. It is also easy to see that if f is defined by $f(x)=x$ ($x\in\mathbb{R}$), then $f'(x)=1$. Repeated application of (2) above shows that the derivative of x^n ($n\in\mathbb{N}$) is nx^{n-1}. Thus every polynomial function is differentiable. ◊

Henceforth, we will take for granted the standard results on differentiating elementary functions such as $\sin x$ that the student is familiar from ordinary calculus.

Exercise 5.5.

(1) Show that if f,g are real valued differentiable functions on an open interval I, then $(fg)'(x)=f'(x)g(x)+f(x)g'(x)$, $x\in I$.

(2) Prove that if f,g are infinitely many times differentiable functions on an open interval I, then

$$(fg)^{(n)}(x)=\sum_{k=0}^{n}\binom{n}{k}f^{(k)}(x)g^{(n-k)}(x),\quad x\in I.$$

(3) For a real number x and n a nonnegative integer, define

$$x^{[n]}:=x(x-1)\cdots(x-n+1).$$

Show that if $x,y\in\mathbb{R}$, $(x+y)^{[n]}=\displaystyle\sum_{k=0}^{n}\binom{n}{k}x^{[k]}y^{[n-k]}$.

Hint: Differentiate t^{x+y} n times with respect to $t\in I:=(0,\infty)$.

5.1. Mean Value Theorem

We say that $f:(a,b)\to\mathbb{R}$ has a *local minimum* at $c\in(a,b)$ if there exists a $\delta>0$ such that whenever $x\in(a,b)$ satisfies $|x-c|<\delta$, we have $f(x)\geq f(c)$. In other words, "locally" around c, the value assumed by f at c is the smallest. See Figure 3. Local maximizers are defined likewise.

Theorem 5.6. *Let $f:(a,b)\to\mathbb{R}$ be such that f has a local minimum at $c\in(a,b)$, and f is differentiable at c. Then $f'(c)=0$.*

An analogous result holds for a local maximizer.

Proof. Let $\delta>0$ be such that $a<c-\delta<c<c+\delta<b$ and $f(x)\geq f(c)$ for x satisfying $|x-c|<\delta$. Given an $\epsilon>0$, we can also ensure (by making δ smaller if required) that for all x satisfying $0<|x-c|<\delta$, we have

$$\left|\frac{f(x)-f(c)}{x-c}-f'(c)\right|<\epsilon.$$

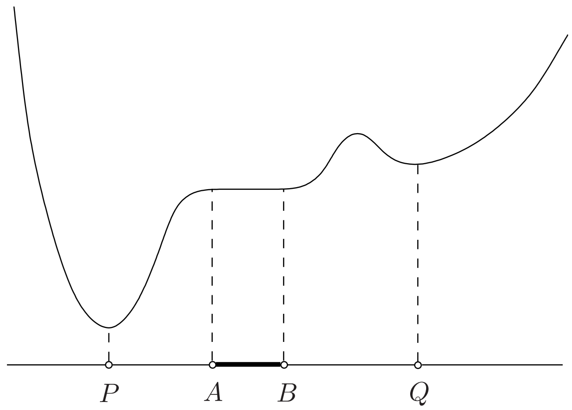


Figure 3. The points P, Q and all points in the interior of the line segment AB are all local minimizers.

Hence we have for all x satisfying $c < x < c + \delta$ that

$$0 - f'(c) \le \frac{f(x) - f(c)}{x - c} - f'(c) \le \left| \frac{f(x) - f(c)}{x - c} - f'(c) \right| < \epsilon.$$

(In order to obtain the first inequality, we have used the fact that $x - c > 0$ and $f(x) \ge f(c)$.) Similarly, for x satisfying $c - \delta < x < c$,we have

$$0 + f'(c) \le -\frac{f(x) - f(c)}{x - c} + f'(c) \le \left| \frac{f(x) - f(c)}{x - c} - f'(c) \right| < \epsilon.$$

Consequently, we have $|f'(c)| < \epsilon$. But the choice of $\epsilon > 0$ was arbitrary, and hence $f'(c) = 0$. ☐

Theorem 5.7 (Mean-Value Theorem)**.** *Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). Then there is a point $c \in (a, b)$ such that*

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

The above result has a simple geometric interpretation. If we look at the chord AB in the plane which joins the end points $A \equiv (a, f(a))$ and $B \equiv (b, f(b))$ of the graph of f, then there is a point $c \in (a, b)$, where the tangent to f at the point $C \equiv (c, f(c))$ is parallel to the chord AB. See Figure 4.

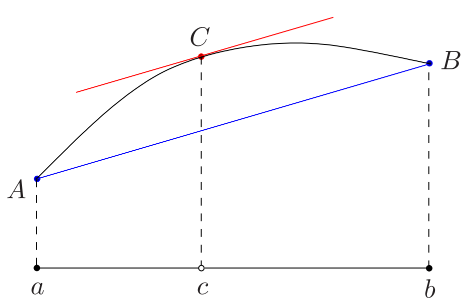


Figure 4. Geometric meaning of the Mean Value Theorem.

Proof of Theorem 5.7. Define $\varphi : [a, b] \to \mathbb{R}$ by

$$\varphi(x) = (f(b) - f(a))x - (b - a)f(x), \quad x \in (a, b).$$

Then φ is continuous on $[a, b]$, differentiable on (a, b) and

$$\varphi(a) = (f(b) - f(a))a - (b - a)f(a) = f(b)a - bf(a) = (f(b) - f(a))b - (b - a)f(b) = \varphi(b).$$

Moreover, for $x \in (a, b)$, we have $\varphi'(x) = f(b) - f(a) - (b - a)f'(x)$, and so in order to prove the theorem, it suffices to show that $\varphi'(c) = 0$ for some $c \in (a, b)$.

If φ is constant, this holds for all $x \in (a, b)$. If $\varphi(x) < \varphi(a) = \varphi(b)$ for some $x \in (a, b)$, let c be a point in $[a, b]$ where φ attains its minimum (Extreme Value Theorem!). Then since $\varphi(b) = \varphi(a)$, we conclude that $c \in (a, b)$. By the necessary condition for a local minimizer, we have $\varphi'(c) = 0$. A similar argument applies if we choose for c a point on $[a, b]$ where φ attains its maximum. □

Corollary 5.8 (Rolle's theorem)**.** *Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). If $f(a) = f(b)$, then there exists $c \in (a, b)$ such that $f'(c) = 0$.*

Exercise 5.9 (Cauchy's theorem)**.** If $f, g : [a, b] \to \mathbb{R}$ are continuous on $[a, b]$ and differentiable on (a, b), then show that there is a point $c \in (a, b)$ such that $(f(b) - f(a))g'(c) = (g(b) - g(a))f'(c)$.

Hint: Apply Rolle's Theorem to φ given by $\varphi(x) = \det \begin{bmatrix} f(x) & g(x) & 1 \\ f(a) & g(a) & 1 \\ f(b) & g(b) & 1 \end{bmatrix}$ $(x \in [a, b])$.

Corollary 5.10. *Suppose that $f : (a, b) \to \mathbb{R}$ is differentiable on (a, b). Then:*

(1) *If $f'(x) \geq 0$ for all $x \in (a, b)$, then f is increasing.*
(2) *If $f'(x) = 0$ for all $x \in (a, b)$, then f is constant.*
(3) *If $f'(x) \leq 0$ for all $x \in (a, b)$, then f is decreasing.*

Proof. For each pair of numbers $x_1, x_2 \in (a, b)$, by the Mean Value Theorem, we have

$$f(x_1) - f(x_2) = f'(x)(x_1 - x_2)$$

for some x between x_1 and x_2. All the above conclusions can be read off from here. □

The Mean Value Theorem can be used to prove interesting inequalities; here is an example.

Example 5.11. Let us show that for all $x > 0$, $\sqrt{1 + x} < 1 + \frac{1}{2}x$.

Consider the function $f : [0, \infty) \to \mathbb{R}$ defined by $f(x) = \sqrt{1 + x}$. Then f is continuous on $[0, \infty)$ and differentiable on $(0, \infty)$. If $x > 0$, then applying the Mean Value Theorem to f on the interval $[0, x]$, we obtain the existence of a c such that $0 < c < x$ and

$$\frac{f(x) - f(0)}{x - 0} = \frac{\sqrt{1 + x} - 1}{x} = f'(c) = \frac{1}{2\sqrt{1 + c}} < \frac{1}{2}.$$

Rearranging, we obtain the desired inequality. ♢

If f has a derivative $f'(x)$ at each $x \in (a, b)$, then we can consider the *derivative function*, namely the map f' given by $x \mapsto f'(x)$ on the interval (a, b). Suppose now that f' is itself differentiable on (a, b). Then we may consider the derivative function f'' of f'. One can continue in this manner (provided of course that each successive function obtained is again differentiable), and obtain the functions

$$f',\ f'',\ f^{(3)}, \dots,\ f^{(n)},$$

each of which is the derivative of the previous one. $f^{(n)}$ is called the nth derivative, or the derivative or order n, of f.

Example 5.12. All polynomials have derivatives of all orders, and eventually all high order derivatives are the zero function. ♢

Exercise 5.13. Let $f : \mathbb{R} \to \mathbb{R}$ be such that for all $x, y \in \mathbb{R}$, $|f(x) - f(y)| \leq (x - y)^2$. Prove that f is constant.

Exercise 5.14. Let $f : (a, b) \to \mathbb{R}$ be differentiable on (a, b) and suppose that there is number M such that for all $x \in (a, b)$, $|f'(x)| \leq M$. Show that f is uniformly continuous on (a, b).

Exercise 5.15. Show that the cubic $2x^3 + 3x^2 + 6x + 10$ has exactly one real zero.

Exercise 5.16. Let $f : \mathbb{R} \to \mathbb{R}$. We call $x \in \mathbb{R}$ a *fixed point* of f if $f(x) = x$.

(1) If f is differentiable, and for all $x \in \mathbb{R}$, $f'(x) \neq 1$, then prove that f has at most one fixed point.

(2) $(*)$ Show that if there is an $M < 1$ such that for all $x \in \mathbb{R}$, $|f'(x)| \leq M$, then there is a fixed point x_* of f, and that $x_* = \lim_{n\to\infty} x_n$, where x_1 is arbitrary and $x_{n+1} = f(x_n)$ $(n \in \mathbb{N})$.

(3) Visualize the process described in the part (2) above via the zig-zag path

$$(x_1, x_2) \longrightarrow (x_2, x_2) \longrightarrow (x_2, x_3) \longrightarrow (x_3, x_3) \longrightarrow (x_3, x_4) \longrightarrow \dots.$$

(4) Prove that the function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = x + \frac{1}{1 + e^x} \quad (x \in \mathbb{R})$$

has no fixed point, although $0 < f'(x) < 1$ for all $x \in \mathbb{R}$. Is this a contradiction to the result in part (2) above? Explain.

Exercise 5.17. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(0) = 0$ and for $x \neq 0$,

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

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

Exercise 5.18. $(*)$ Find all functions $f : \mathbb{R} \to \mathbb{R}$ such that f is differentiable on $\mathbb{R}$ and for all $x \in \mathbb{R}$ and all $n \in \mathbb{N}$,

$$f'(x) = \frac{f(x + n) - f(x)}{n}.$$

Hint: Conclude that f must be twice differentiable and calculate $f''(x)$.

Exercise 5.19. Show that for every real $a, b \in \mathbb{R}$, $|\cos a - \cos b| \leq |a - b|$.

Exercise 5.20. $(*)$ A function $f : \mathbb{R} \to \mathbb{R}$ is said to be *convex* if for all $x, y \in \mathbb{R}$ and all $t \in (0, 1)$, $f((1 - t)x + ty) \leq (1 - t)f(x) + tf(y)$.

(1) Draw a picture and explain the geometric meaning of the inequality above.

(2) Show that if f is twice differentiable, then f is convex if $f''(x) \geq 0$ for all $x \in \mathbb{R}$. *Hint:* If $x < y$, then apply the Mean Value Theorem for f on the interval $[x, (1-t)x + ty]$ and $[(1-t)x + ty, y]$.

(3) Prove that if f is a differentiable and convex function, then f' is increasing. *Hint:* If $x < u < y$, then using the convexity, derive the inequalities

$$\frac{f(u) - f(x)}{u - x} \leq \frac{f(y) - f(x)}{y - x} \leq \frac{f(y) - f(u)}{y - u}.$$

and pass the limits $u \searrow x$ and $u \nearrow y$.

(Combining this result with the result in part (2), we conclude that a twice differentiable function is convex *if and only if* $f''(x) \geq 0$ for all $x \in \mathbb{R}$.)

(4) Prove that if f is a differentiable, convex function and $f'(x_0) = 0$ for a $x_0 \in \mathbb{R}$, then x_0 is a minimizer of f.

Exercise 5.21. Prove that not all the zeros of the polynomial $x^4 - \sqrt{7}x^3 + 4x^2 - \sqrt{22}x + 15$ are real.
Hint: Repeated application of Rolle's Theorem.

5.2. Uniform convergence and differentiation

When we studied uniform convergence, we had mentioned that interchanging limits is facilitated by uniform convergence. An instance of this is the possibility of differentiating a uniformly convergent series termwise; as shown in the Corollary 5.23 below. This relies on the following result, which in turn is an application of the Mean Value Theorem.

Proposition 5.22. *Let $f_n : (a,b) \to \mathbb{R}$ ($n \in \mathbb{N}$) be a sequence of differentiable functions on (a,b), such that there exists a point $c \in (a,b)$ for which $(f_n(c))_{n\in\mathbb{N}}$ converges. If the sequence $(f_n')_{n\in\mathbb{N}}$ converges uniformly to g on (a,b), then $(f_n)_{n\in\mathbb{N}}$ converges uniformly to a differentiable function f on (a,b), and moreover, $f'(x) = g(x)$ for all $x \in (a,b)$.*

Proof. ($*$) (*You may skip reading this proof.*) Let $\epsilon > 0$. Choose $N_1 \in \mathbb{N}$ such that for all $m, n > N_1$,

$$|f_m'(x) - f_n'(x)| < \min\left\{\frac{\epsilon}{3}, \frac{\epsilon}{3(b-a)}\right\} \text{ and } |f_m(c) - f_n(c)| < \frac{\epsilon}{3}.$$

Let $x \in (a,b)$. Apply the Mean Value Theorem to the function $f_m - f_n$ on the interval with the endpoints x, c to get

$$f_m(x) - f_n(x) = f_m(c) - f_n(c) + (x-c)(f_m'(y) - f_n'(y))$$

for some y between x, c. Hence

$$|f_m(x) - f_n(x)| \le |f_m(c) - f_n(c)| + (b-a)|f_m'(y) - f_n'(y)| < \frac{2}{3}\epsilon < \epsilon.$$

Consequently, $(f_n)_{n\in\mathbb{N}}$ is uniformly convergent on (a,b). Let $f : (a,b) \to \mathbb{R}$ be its limit. As each f_n is continuous, so is f.

To show that f is differentiable at a point $x_0 \in (a,b)$, we apply the Mean Value Theorem once again to $f_m - f_n$ on the interval with endpoints x_0 and $x \in (a,b)$, $x \neq x_0$. We obtain

$$(f_m(x) - f_n(x)) - (f_m(x_0) - f_n(x_0)) = (x - x_0)(f_m'(y) - f_n'(y))$$

for some y between x and x_0. Dividing by $x - x_0$ and taking absolute values, we get

$$\left|\frac{f_m(x) - f_m(x_0)}{x - x_0} - \frac{f_n(x) - f_n(x_0)}{x - x_0}\right| \le |f_m'(y) - f_n'(y)| < \frac{\epsilon}{3}.$$

Passing the limit $m \to \infty$ yields

$$\left|\frac{f(x) - f(x_0)}{x - x_0} - \frac{f_n(x) - f_n(x_0)}{x - x_0}\right| \le \frac{\epsilon}{3}.$$

Now choose $N_2 \in \mathbb{N}$ such that $|f_n'(x_0) - g(x_0)| < \epsilon/3$ for all $n > N_2$. Let $N := \max\{N_1, N_2\}$ and choose $\delta > 0$ such that $0 < |x - x_0| < \delta$ implies

$$\left|\frac{f_N(x) - f_N(x_0)}{x - x_0} - f_N'(x_0)\right| < \frac{\epsilon}{3}.$$

Then combining these inequalities, we get for $0 < |x - x_0| < \delta$ that

$$\left|\frac{f(x) - f(x_0)}{x - x_0} - g(x_0)\right| < \epsilon.$$

As the choice of $\epsilon > 0$ was arbitrary, it follows that f is differentiable at x_0 and $f'(x_0) = g(x_0)$. $\square$

Corollary 5.23. *Let $f_n : (a,b) \to \mathbb{R}$, $n \in \mathbb{N}$, be a sequence of differentiable functions such that*

(1) $\displaystyle\sum_{n=1}^{\infty} f_n(c)$ *converges for some $c \in (a,b)$, and*

(2) $\displaystyle g := \sum_{n=1}^{\infty} f_n'$ *is uniformly convergent on (a,b).*

Then $\displaystyle f := \sum_{n=1}^{\infty} f_n$ is uniformly convergent on (a,b) and $f' = g$ on (a,b).

Example 5.24. We will show that the exponential function $f(x) := e^x$ $(x \in \mathbb{R})$ satisfies the differential equation

$$\frac{d}{dx} f(x) = f(x) \quad (x \in \mathbb{R}).$$

We have

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

converges uniformly on each bounded interval. Thus, by Corollary 5.23, we obtain $f'(x) = f(x)$ as required. $\Diamond$

Exercise 5.25. $(*)$ What is the sum of the series $1 + 2x + 3x^2 + 4x^3 + \dots$ for $x \in (-1, 1)$?

5.3. Derivative of maps from $\mathbb{R}^n$ to $\mathbb{R}^m$

In order to differentiate functions whose domain is $\mathbb{R}^n$ (or an open subset of $\mathbb{R}^n$) and takes values in $\mathbb{R}^m$, let us take another look at the familiar case when $n = m = 1$, and let us recast the old definition in a manner that will naturally lend itself for extension to the case when n or m is > 1.

We has defined $f : (a, b) \to \mathbb{R}$ to be differentiable at $c \in (a, b)$ if

$$f'(c) := \lim_{x \to c} \frac{f(x) - f(c)}{x - c}$$

exists. Or equivalently,

$$\lim_{x \to c} \left| \frac{f(x) - f(c)}{x - c} - f'(c) \right| = 0.$$

This implies that the "remainder"

$$r(x) := f(x) - f(c) - f'(c)(x - c)$$

is small, in the sense that

$$\lim_{x \to c} \frac{|r(x)|}{|x - c|} = \lim_{x \to c} \frac{|f(x) - f(c) - f'(c)(x - c)|}{|x - c|} = \lim_{x \to c} \left| \frac{f(x) - f(c)}{x - c} - f'(c) \right| = 0.$$

Now consider the linear transformation $L : \mathbb{R} \to \mathbb{R}$ given by

$$L(h) = f'(c)h \quad (h \in \mathbb{R}).$$

Then we can write the remainder term as

$$r(x) = f(x) - f(c) - f'(c)(x - c) = f(x) - f(c) - L(x - c),$$

and we see that $r(x)$ measures how different $f(x) - f(c)$ is from the action of the linear transformation L on $x - c$. So we may view the number $f'(c)$ as describing a linear transformation L so that the remainder r defined by $r(x) := f(x) - f(c) - f'(c)(x - c)$ has the property that

$$\lim_{x \to c} \frac{|r(x)|}{|x - c|} = 0.$$

This motivates the following definition.

Definition 5.26. Let $U \subset \mathbb{R}^n$ be open, $c \in U$ and $f : U \to \mathbb{R}^m$. Then we say that f is *differentiable at* c if there exists a linear transformation $L : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{x \to c} \frac{\|f(x) - f(c) - L(x - c)\|_2}{\|x - c\|_2} = 0, \tag{5.1}$$

that is, for every $\epsilon > 0$, there exists a $\delta > 0$ such that whenever $x \in U$ satisfies $0 < \|x - c\|_2 < \delta$, there holds that

$$\frac{\|f(x) - f(c) - L(x - c)\|_2}{\|x - c\|_2} < \epsilon.$$

Then L is called the *derivative of* f *at* c, and we write $f'(c) = L$.

The relation (5.1) can be expressed by saying that

$$f(x) - f(c) = f'(c)(x - c) + r(x),$$

where the remainder r satisfies

$$\lim_{x \to c} \frac{\|r(x)\|_2}{\|x - c\|_2} = 0.$$

So we can interpret this as saying that $f'(c)$ is that linear transformation which has the property that for x close to c, $f(x) - f(c)$ is approximately equal to its action on $x - c$. Next we show that in fact there can be only one such linear transformation.

Lemma 5.27. *Let $U \subset \mathbb{R}^n$ be open, $c \in U$ and $f : U \to \mathbb{R}^m$ be differentiable at c. Then the derivative of f at c is unique.*

Proof. Suppose that L_1, L_2 are two linear transformations such that

$$\lim_{x \to c} \frac{\|f(x) - f(c) - L_1(x - c)\|_2}{\|x - c\|_2} = 0 = \lim_{x \to c} \frac{\|f(x) - f(c) - L_2(x - c)\|_2}{\|x - c\|_2}.$$

Thus given $\epsilon > 0$, we can choose a $\delta > 0$ such that whenever $0 < \|x - c\|_2 < \delta$, we have

$$\begin{aligned} \frac{\|f(x) - f(c) - L_1(x - c)\|_2}{\|x - c\|_2} &< \epsilon, \\ \frac{\|f(x) - f(c) - L_2(x - c)\|_2}{\|x - c\|_2} &< \epsilon. \end{aligned}$$

Hence by the triangle inequality, it follows for whenever $0 < \|x - c\|_2 < \delta$,

$$\frac{\|L_2(x - c) - L_1(x - c)\|_2}{\|x - c\|_2} < 2\epsilon,$$

that is, $\|L_2(x - c) - L_1(x - c)\|_2 \le 2\epsilon \|x - c\|_2$. Now given any nonzero $h \in \mathbb{R}^n$, define

$$x = c + \frac{\delta}{2\|h\|_2} h.$$

Then $0 < \|x - c\|_2 = \delta/2 < \delta$, and so $\|L_2 h - L_1 h\|_2 \le 2\epsilon \|h\|_2$. But the choice of $\epsilon > 0$ was arbitrary, and so $L_2 h = L_1 h$ for all nonzero $h \in \mathbb{R}^n$. Thus $L_1 = L_2$. $\square$

Example 5.28. Let $A \in \mathbb{R}^{m \times n}$. Consider the map $T_a : \mathbb{R}^n \to \mathbb{R}^m$ given by $T_A x = Ax$ $(x \in \mathbb{R}^n)$. If $c \in \mathbb{R}^n$, then is T_A differentiable at c? If so, then what is its derivative? The answers turn out to be very simple. We note that for $x \in \mathbb{R}^n$, we have

$$T_A(x) - T_A(c) = T_A(x - c),$$

and so

$$\lim_{x \to c} \frac{\|T_A(x) - T_A(c) - T_A(x - c)\|_2}{\|x - c\|_2} = \lim_{x \to c} \frac{0}{\|x - c\|_2} = \lim_{x \to c} 0 = 0.$$

So T_A is differentiable at $c \in \mathbb{R}^n$, and $T_A'(c) = T_A$! $\Diamond$

Exercise 5.29. $(*)$ Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at $c \in \mathbb{R}^n$. Show that the new function $g : \mathbb{R}^n \to \mathbb{R}$, defined by

$$g(x) := (f(x))^2 \quad (x \in \mathbb{R}^n)$$

is also differentiable at c. *Hint:* $(f(x))^2 - (f(c))^2 = (f(x) + f(c))(f(x) - f(c)) \approx 2f(c)f'(c)(x - c)$ for x near c.

Exercise 5.30. Let $Q \in \mathbb{R}^{n \times n}$ be a symmetric matrix, that is, $Q = Q^\top$. Define $q : \mathbb{R}^n \to \mathbb{R}$ by $q(x) = x^\top Q x$ $(x \in \mathbb{R}^n)$. Prove that q is differentiable at each $c \in \mathbb{R}^n$ and that $q'(c) : \mathbb{R}^n \to \mathbb{R}$ is given by $q'(c)v = 2c^\top Q v$ $(v \in \mathbb{R}^n)$.

Exercise 5.31. Consider the map $f : \mathbb{R}^n \to \mathbb{R}$ given by

$$f(x) = \|x\|_2^4 \quad (x \in \mathbb{R}^n).$$

Calculate $f'(c)$ for $c \in \mathbb{R}^n$. *Hint:* Use the results in Exercises 5.29 and 5.30.

5.4. Partial derivatives

Suppose that U is an open subset of $\mathbb{R}^n$, and let $f : U \to \mathbb{R}^m$ be a function. Let the components of f be denoted by $f_1, \dots, f_m$. Thus for $i = 1, \dots, m$,

$$f_i(x) := e_i^\top f(x) \quad (x \in U),$$

where $e_1, \dots, e_m$ denote the standard basis vectors in $\mathbb{R}^m$, that is,

$$e_1 := \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \dots, e_m := \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

Let $c \in U$. If

$$\frac{\partial f_i}{\partial x_j}(c) := \lim_{x_j \to c_j} \frac{f_i(c_1, \dots, c_{j-1}, x_j, c_{j+1}, \dots, c_n) - f_i(c_1, \dots, c_{j-1}, c_j, c_{j+1}, \dots, c_n)}{x_j - c_j}$$

exists, then we call $\frac{\partial f_i}{\partial x_j}(c)$ the (i, j)th partial derivative f at c. Thus, we look only at the ith component $f_i : U \to \mathbb{R}$, keep all the variables $x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n$ as fixed, with values $c_1, \dots, c_{j-1}, c_{j+1}, \dots, c_n$, respectively, and differentiate the function

$$x_j \mapsto f_i(c_1, \dots, c_{j-1}, x_j, c_{j+1}, \dots, c_n)$$

with respect to x_j at c_j.

Example 5.32. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$f(x_1, x_2) = \begin{bmatrix} x_1^2 + x_2^2 \\ x_1 x_2 \end{bmatrix}.$$

Then we have

$$\begin{aligned} \frac{\partial f_1}{\partial x_1}(c_1, c_2) &= 2c_1, & \frac{\partial f_1}{\partial x_2}(c_1, c_2) &= 2c_2, \\ \frac{\partial f_2}{\partial x_1}(c_1, c_2) &= c_2, & \frac{\partial f_2}{\partial x_2}(c_1, c_2) &= c_1. \end{aligned}$$

◊

Theorem 5.33. *Let U be an open subset of $\mathbb{R}^n$, and $c \in U$. If $f : U \to \mathbb{R}^m$ is differentiable at c, then all the partial derivatives of f at c, namely,*

$$\frac{\partial f_i}{\partial x_j}(c) \quad (i = 1, \dots, m;\ j = 1, \dots, n)$$

exist, and the matrix $[f'(c)]$ of the linear transformation $f'(c)$ with respect to the standard bases for $\mathbb{R}^n$ and $\mathbb{R}^m$ is given by

$$[f'(c)] = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(c) & \dots & \frac{\partial f_1}{\partial x_n}(c) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(c) & \dots & \frac{\partial f_m}{\partial x_n}(c) \end{bmatrix}.$$

Proof. Let $\epsilon > 0$. Then since f is differentiable at c, there exists a $\delta > 0$ such that for all $x \in U$ such that $0 < \|x - c\|_2 < \delta$, we have

$$\frac{\|f(x) - f(c) - f'(c)(x - c)\|_2}{\|x - c\|_2} < \epsilon.$$

Let x_j be such that $0 < |x_j - c_j| < \delta$. Define

$$x := (c_1, \dots, c_{j-1}, x_j, c_{j+1}, \dots, c_n).$$

Then $x - c = (0, \dots, 0, x_j - c, 0, \dots, 0) = (x_j - c)e_j$, and so $\|x - c\|_2 = |x_j - c|$. Consequently, for these x:s,

$$\frac{\|f(x) - f(c) - f'(c)(x - c)\|_2}{\|x - c\|_2} < \epsilon.$$

Moreover, $f_i(x) - f_i(c) - (e_i^\top [f'(c)]e_j)(x_j - c_j) = e_i^\top (f(x) - f(c) - f'(c)(x - c))$, and so

$$|f_i(x) - f_i(x) - (e_i^\top [f'(c)]e_j)(x_j - c_j)| \le \|f(x) - f(c) - f'(c)(x - c)\|_2.$$

Hence for x_j:s satisfying $0 < |x_j - c_j| < \delta$, we have

$$\left| \frac{f_i(c_1, \dots, c_{j-1}, x_j, c_{j+1}, \dots, c_n) - f_i(c_1, \dots, c_{j-1}, c_j, c_{j+1}, \dots, c_n)}{x_j - c_j} - e_i^\top [f'(c)]e_j \right| < \epsilon.$$

This completes the proof. □

The above result says that for the derivative to exist, it is necessary that the partial derivatives exist. Surprisingly, this is not a sufficient condition.

Example 5.34. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(0,0) = 0$ and for $(x_1, x_2) \ne (0,0)$,

$$f(x_1, x_2) = \frac{x_1 x_2}{x_1^2 + x_2^2}.$$

We will show that although $\dfrac{\partial f}{\partial x_1}(0,0)$ and $\dfrac{\partial f}{\partial x_2}(0,0)$ exist, $f'(0,0)$ doesn't, that is, f is not differentiable at $(0,0)$.

For $x_1 \ne 0$, we have $f(x_1, 0) - f(0,0) = 0 - 0 = 0$, and so

$$\frac{\partial f}{\partial x_1}(0,0) = \lim_{x_1 \to 0} \frac{f(x_1, 0) - f(0,0)}{x_1 - 0} = \lim_{x_1 \to 0} 0 = 0.$$

Similarly, $\dfrac{\partial f}{\partial x_2}(0,0) = 0$.

Thus all the partial derivatives of f exist at $(0,0)$. However, we will now show that $f'(0,0)$ does not exist. Suppose that $f'(0,0)$ does exist. Then by Theorem 5.33,

$$[f'(0,0)] = \begin{bmatrix} \dfrac{\partial f}{\partial x_1}(0,0) & \dfrac{\partial f}{\partial x_2}(0,0) \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}.$$

Let $\epsilon > 0$. Then there exists a $\delta > 0$ such that for all $x \in \mathbb{R}^2$ satisfying $0 < \|x - 0\|_2 < \delta$, we have

$$\frac{\|f(x) - f(0) - f'(0)(x - 0)\|_2}{\|x - 0\|_2} < \epsilon,$$

that is,

$$\frac{|x_1 x_2|}{(x_1^2 + x_2^2)\sqrt{x_1^2 + x_2^2}} < \epsilon.$$

For all $n \in \mathbb{N}$ large enough, with $x_1 := x_2 := 1/n$, we have that $0 < \|(x_1, x_2) - (0,0)\|_2 = \sqrt{2}/n < \delta$, and so there must hold

$$\frac{n}{2\sqrt{2}} = \frac{\dfrac{1}{n^2}}{\dfrac{2}{n^2}\dfrac{\sqrt{2}}{n}} = \frac{|x_1 x_2|}{(x_1^2 + x_2^2)\sqrt{x_1^2 + x_2^2}} < \epsilon,$$

for all large n, a contradiction. ◊

Let U be an open subset of $\mathbb{R}^n$. We say that $f : U \to \mathbb{R}$ has a *local minimum* at $c \in U$ if there exists a $\delta > 0$ such that whenever $x \in U$ satisfies $\|x - c\|_2 < \delta$, we have $f(x) \geq f(c)$. Local maximizers are defined analogously.

Corollary 5.35. *Let U be an open subset of $\mathbb{R}^n$. Let $f : U \to \mathbb{R}$ be such that f has a local minimum at $c \in U$, and f is differentiable at c. Then $f'(c) = 0$.*

Proof. It is clear that if c is a local minimizer for f, then each of the functions

$$x_i \mapsto f(c_1, \dots, c_{i-1}, x_i, c_{i+1}, \dots, c_n)$$

has a local minimum at c_i, and so by the one variable result, we have

$$\frac{\partial f}{\partial x_i}(c) = 0$$

for each $i = 1, \dots, m$. Consequently, Theorem 5.33 yields $f'(c) = 0$. □

An analogous result holds for a local maximizer.

Exercise 5.36. Let τ, ξ be real numbers such that $\tau - \xi^2 = 0$. Show that $f(x,t) := e^{\tau t + \xi x}$, $x, t \in (0, \infty)$, satisfies the diffusion equation

$$\frac{\partial f}{\partial t}(x,t) - \frac{\partial^2 f}{\partial x^2}(x,t) = 0.$$

(Here $\dfrac{\partial^2 f}{\partial x^2}$ is used to denote the partial derivative with respect to x of the map $(x,t) \mapsto \dfrac{\partial f}{\partial x}(x,t)$.)

Exercise 5.37. In the subject of "Calculus of Variations", the following type of optimization problem is studied:

$$\text{minimize } f(x) := \int_a^b F(x(t), x'(t), t)dt$$

where the integrand in the cost function is described by a function $F : \mathbb{R}^3 \to \mathbb{R}$ of three real variables (α, β, γ) taking values in $\mathbb{R}^3$:

$$(\alpha, \beta, \gamma)\ (\in \mathbb{R}^3) \quad \longmapsto \quad F(\alpha, \beta, \gamma)\ (\in \mathbb{R}),$$

and the (variable) function $x : [a,b] \to \mathbb{R}$ is such that $x(a) = y_a$ and $x_b = y_b$. Hence we observe that this is an optimization problem in which the domain of the cost function is itself a set of functions. Figure 5 illustrates this.

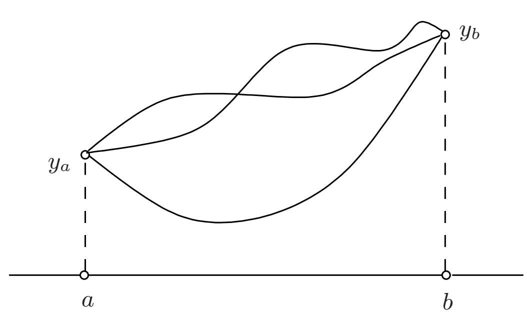


Figure 5. Possible xs.

The fundamental result in Calculus of Variations is that if x_* is an optimal solution to this problem, then it must satisfy the following "Euler-Lagrange equation":

$$\frac{\partial F}{\partial \alpha}(x(t), x'(t), t) - \frac{d}{dt}\left(\frac{\partial F}{\partial \beta}(x(t), x'(t), t)\right) = 0 \quad (t \in [a,b]).$$

Consider for example the following optimization problem which occurs in resource management. One wants to maximize the profit

$$f(x) := \int_0^T (P - ax(t) - bx'(t))x'(t)dt$$

associated with a possible choice of operation $x : [0, T] \to \mathbb{R}$ over the time interval $[0, T]$ satisfying $x(0) = 0$ and $x(T) = Q$. Here T, P, a, b, Q are given positive constants. Assuming that an optimal operation x_* exists, find it using the Euler-Lagrange equation.

Exercise 5.38. $(*)$ Find all (global) minimizers of $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x_1, x_2) = x_1^4 - 12x_1x_2 + x_2^4$, $(x_1, x_2) \in \mathbb{R}^2$.

Exercise 5.39. Verify that

$$\frac{\partial^2 f}{\partial x \partial y}(x, y) = \frac{\partial^2 f}{\partial y \partial x}(x, y) \tag{5.2}$$

for all $(x, y) \in \mathbb{R}^2$, where the function f is given by

$$f(x, y) = 3x^3 + 9y^2 - 9x^3 y \quad ((x, y) \in \mathbb{R}^2).$$

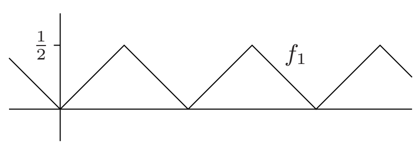
Show that (5.2) does not hold at $(0, 0)$ if f is given by

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

Exercise 5.40. Find the derivative of the multiplication map $(x, y) \mapsto xy : \mathbb{R}^2 \to \mathbb{R}$ at (x_0, y_0) in $\mathbb{R}^2$.

5.5. Notes (not part of the course)

In connection with Theorem 5.1, we might wonder how badly behaved continuous functions might be with respect to the notion of differentiability. It turns out that there are functions that are continuous *everywhere*, but differentiable *nowhere*. One construction is that of the so-called *blancmange function* obtained by taking the basic *sawtooth* function f_1,



and constructing $f_2, f_3, \dots$ by setting

$$\begin{aligned} f_2(x) &= \frac{1}{2} f_1(2x), \\ f_3(x) &= \frac{1}{4} f_1(4x), \\ f_4(x) &= \frac{1}{8} f_1(8x), \\ &\vdots \\ f_n(x) &= \left(\frac{1}{2}\right)^{n-1} f_1(2^{n-1}x), \\ &\vdots \end{aligned}$$

and adding these:

$$b(x) = \sum_{n=1}^{\infty} f_n(x), \quad x \in \mathbb{R}.$$

Then it can be shown that b is continuous on $\mathbb{R}$, but not differentiable at any $x \in \mathbb{R}$.

We have seen in Example 5.34 that even though all the partial derivatives exist at a point, the function may not be differentiable at that point. However, the following result says that if the partial derivatives are *continuous* in a neighbourhood of the point, then the function is differentiable in that neighbourhood. Here is the precise statement of the result, a proof of which can be found in Rudin [**R**].

Theorem 5.41. *Suppose that f maps an open set $U \subset \mathbb{R}^n$ into $\mathbb{R}^m$. If all the partial derivatives*

$$\frac{\partial f_i}{\partial x_j}(x) \quad (i = 1, \ldots, m;\ j = 1, \ldots, n)$$

exist at each $x \in U$ and the maps

$$x \mapsto \frac{\partial f_i}{\partial x_j}(x) : U \to \mathbb{R} \quad (i = 1, \ldots, m;\ j = 1, \ldots, n)$$

are continuous on U, then f is continuously differentiable on U.

(In the above the phrase "continuously differentiable on U" means that at each $x \in U$, $f'(x)$ exists and the map $x \mapsto f'(x) : U \to \mathcal{L}(\mathbb{R}^n; \mathbb{R}^m)$ is continuous on U. Here $\mathcal{L}(\mathbb{R}^n; \mathbb{R}^m)$ denotes the set of all linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$.)

Also, in Exercise 5.39, we saw a real function for which the order of taking partial derivatives mattered. The following result gives a sufficient condition for it to be irrelevant.

Theorem 5.42. *If $U \subset \mathbb{R}^n$ and $f : U \to \mathbb{R}$ is twice continuously differentiable (that is, $f''(x)$ exists at each $x \in U$ and the map $x \mapsto f''(x)$ is continuous on U), then*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x) \quad (1 \le i, j \le n,\ x \in U).$$

Epilogue: integration

One traditional topic in real analysis that we haven't covered in these notes is Integration Theory. There are two important types of integrals: Riemann integration and Lebesgue integration. Riemann integration will be covered in the course MA200 *Further Mathematical Methods* (Calculus), and it has the advantage that it is intuitive and easy to follow. However, it turns out that it is not amenable to certain natural limiting processes. For example, it turns out that the functional analogue of the Euclidean space $\mathbb{R}^n$, namely the space $C[a, b]$ equipped with the following norm

$$\|f\|_2 := \sqrt{\int_a^b |f(x)|^2 dx} \quad (f \in C[a, b])$$

is not complete, which turns out to be awkward when one wants to deal with applications. However, one can introduce a more general integral called the Lebesgue integral, which rescues this situation. The interested reader is referred to the book by Rudin [**R**] for these matters. However, in this last chapter, we study the foundations of Riemann integration in the case of a function $f : [a, b] \to \mathbb{R}$. We study the definition, elementary properties and finish with establishing the Fundamental Theorem of Integral Calculus.

5.6. Motivation and definition

It is a basic problem in geometry to calculate the area under a curve, which arises from concrete practical applications.

Given a "nice" function $f : [a, b] \to \mathbb{R}$, determine the area under the graph of f. Pictorially, find the area of the shaded region in Figure 6.

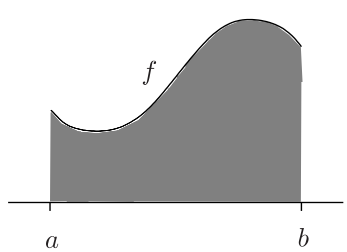


Figure 6. What is the area under the graph of f?

Of course if the function were a simple function, like a constant function taking value c everywhere, it would be easy to calculate the area. See Figure 7.

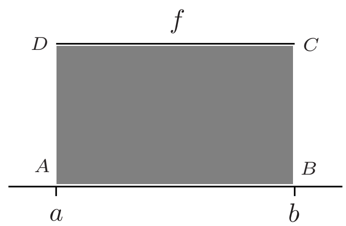


Figure 7. Area under the graph of f is $c \cdot (b - a)$.

Indeed it would just be the area of the rectangle $ABCD$, that is,

$$f(a) \cdot (b - a) = f(b) \cdot (b - a) = c \cdot (b - a).$$

But how do we define the area in the case when f is not constant and the graph of f has a shape which curves? Well, if there were numbers m, M such that $m \leq f(x) \leq M$ for all $x \in [a, b]$, surely we must have that the area under the graph of f, say $A(f)$, is flanked by the areas of the two rectangles $ABCD$ and $ABC'D'$, as shown in Figure 8, that is,

$$m \cdot (b - a) \leq A(f) \leq M \cdot (b - a).$$

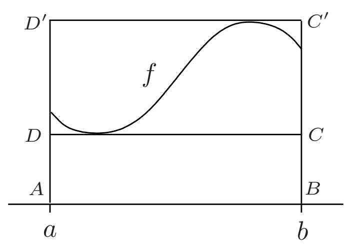


Figure 8. Bounds/estimates for the area under the graph of f.

This gives us the idea that we can estimate the area $A(f)$ by breaking it into little rectangles; see Figure 9.

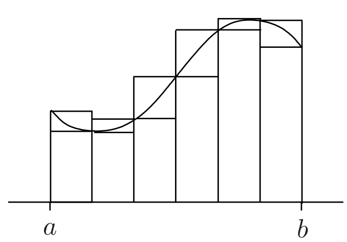


Figure 9. Calculation of the area under the graph of f.

In order to make this rigorous, we introduce the notion of a *partition* and that of *upper/lower sum* associated with a partition and a *bounded* function $f : [a, b] \to \mathbb{R}$.

Definition 5.43 (Partition)**.** A partition P of $[a, b]$ is a finite set $P = \{x_0, x_1, \ldots, x_{n-1}, x_n\}$ such that $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$. The set of all partitions P of $[a, b]$ will be denoted by $\mathcal{P}$.

Definition 5.44 (Upper/lower sum)**.** Let $f : [a, b] \to \mathbb{R}$ be a bounded function and let P be a partition of $[a, b]$.

An *upper sum* $U(f, P)$ *of* f *associated with* P is defined to be

$$U(f,P) := \sum_{k=0}^{n-1} \left(\sup_{x\in[x_k,x_{k+1}]} f(x) \right) \cdot (x_{k+1} - x_k).$$

(In other words, it is the area of a rectangle with the segment $[x_k, x_{k+1}]$ as the base and which is the shortest rectangle which contains the graph of the function f in the interval $[x_k, x_{k+1}]$.)

An *lower sum* $L(f, P)$ *of* f *associated with* P is defined to be

$$L(f,P) := \sum_{k=0}^{n-1} \left(\inf_{x\in[x_k,x_{k+1}]} f(x) \right) \cdot (x_{k+1} - x_k).$$

(In other words, it is the area of a rectangle with the segment $[x_k, x_{k+1}]$ as the base and which is the tallest rectangle which is contained in the region under the graph of the function f in the interval $[x_k, x_{k+1}]$.)

Clearly we expect that the actual area $A(f)$ is bounded above by *any* upper sum, and hence by the infimum of all upper sums. Similarly, $A(f)$ is bounded below by *any* lower sum, and hence by the supremum of all lower sums. Thus we expect that

$$\sup_{P\in\mathcal{P}} L_P(f) \le A(f) \le \sup_{P\in\mathcal{P}} U_P(f).$$

Of course as the partions get finer, we expect that for a nice function f the two bound get close to each other, since they both approximate $A(f)$ rather well. This motivates the following definition.

Exercise 5.45.

(1) Show that for any bounded function $f : [a, b] \to \mathbb{R}$, and any partitions P, P' of $[a, b]$, there holds that $U(f, P) \ge L(f, P')$.
Hint: Consider a "refinement" partition that contains both the partitions P and P'.

(2) Prove that $\inf_{P\in\mathcal{P}} U(f,P) \ge \sup_{P\in\mathcal{P}} L(f,P)$.

Definition 5.46. A bounded function $f : [a, b] \to \mathbb{R}$ is called *Riemann integrable* if

$$\sup_{P\in\mathcal{P}} L_P(f) = \inf_{P\in\mathcal{P}} U_P(f).$$

We denote the common value by $\displaystyle\int_a^b f(x)dx$, and call it the *Riemann integral of* f.

Here is an example of a Riemann integrable function.

Example 5.47. Consider the bounded function $f : [0, 1] \to \mathbb{R}$ given by $f(x) = x^2$, $x \in [0, 1]$. We will show that f is Riemann integrable and that

$$\int_0^1 f(x)dx = \frac{1}{3}.$$

Rather than considering all partitions, it turns out that one can be somewhat more efficient by just considering the special partitions

$$P_* := \left\{0, \frac{1}{n}, \frac{2}{n}, \frac{3}{n}, \dots, \frac{n-1}{n}, 1\right\},$$

which will be used to prove our claim. Clearly,

$$U(f,P_*) = \sum_{k=0}^{n-1} \left(\frac{k+1}{n}\right)^2 \frac{1}{n} = \frac{1^2+2^2+3^2+\cdots+n^2}{n^3} = \frac{n\cdot(n+1)\cdot(2n+1)}{6n^3} = \frac{(1+\frac{1}{n})(2+\frac{1}{n})}{6}.$$

Thus,

$$\inf_{P\in\mathcal{P}} U(f,P) \leq \inf_{n\in\mathbb{N}} U(f,P_*) = \lim_{n\to\infty} \frac{(1+\frac{1}{n})(2+\frac{1}{n})}{6} = \frac{1}{3}.$$

Similarly,

$$L(f,P_*) = \sum_{k=0}^{n-1} \left(\frac{k}{n}\right)^2 \frac{1}{n} = \frac{1^2+2^2+3^2+\cdots+(n-1)^2}{n^3} = \frac{(1-\frac{1}{n})(2-\frac{1}{n})}{6}.$$

Hence,

$$\sup_{P\in\mathcal{P}} L(f,P) \geq \sup_{n\in\mathbb{N}} L(f,P_*) = \lim_{n\to\infty} \frac{(1-\frac{1}{n})(2-\frac{1}{n})}{6} = \frac{1}{3}.$$

But these estimates yield that

$$\frac{1}{3} \geq \inf_{P\in\mathcal{P}} U(f,P) \geq \sup_{P\in\mathcal{P}} L(f,P) \geq \frac{1}{3},$$

showing that

$$f(b)-f(a) \inf_{P\in\mathcal{P}} U(f,P) = \sup_{P\in\mathcal{P}} L(f,P) = \frac{1}{3}.$$

Hence f is Riemann integrable and $\displaystyle\int_0^1 f(x)dx = \frac{1}{3}$. ◊

Here is an example of a function that is *not* Riemann integrable.

Example 5.48. Consider the indicator function f of rationals in $[0,1]$, that is, $f:[0,1]\to\mathbb{R}$ is given by

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

Clearly f is bounded by 1. We will show that

$$\inf_{P\in\mathcal{P}} U(f,P) = 1 > \sup_{P\in\mathcal{P}} L(f,P) = 0,$$

showing that f is not Riemann integrable.

Consider any partition $P = \{0, x_1, \dots, x_{n-1}, 1\}$ of $[0,1]$. Then each subinterval $[x_k, x_{k+1}]$ contains a rational number α_k and an irrational number β_k, and so

$$\sup_{x\in[x_k,x_{k+1}]} f(x) \geq f(\alpha_k) = 1 \text{ and } \inf_{x\in[x_k,x_{k+1}]} f(x) \leq f(\beta_k) = 0.$$

Consequently,

$$\begin{aligned} U(f,P) &= \sum_{k=0}^{n-1} \left(\sup_{x\in[x_k,x_{k+1}]} f(x) \right) \cdot (x_{k+1}-x_k) \geq \sum_{k=0}^{n-1} 1\cdot(x_{k+1}-x_k) \\ &= (x_1-x_0)+(x_2-x_1)+\cdots+(x_n-x_{n-1}) = x_n - x_0 = 1-0 = 1. \end{aligned}$$

Similarly,

$$L(f,P) = \sum_{k=0}^{n-1} \left(\inf_{x\in[x_k,x_{k+1}]} f(x) \right) \cdot (x_{k+1}-x_k) \leq \sum_{k=0}^{n-1} 0\cdot(x_{k+1}-x_k) = 0.$$

Thus,

$$\inf_{P\in\mathcal{P}} U(f,P) \geq 1 > 0 \geq \sup_{P\in\mathcal{P}} L(f,P).$$

Hence f is not Riemann integrable. ◊

All continuous functions are Riemann integrable.

Theorem 5.49. *All functions continuous on $[a,b]$ are Riemann integrable.*

Proof. We know that $[a, b]$ is compact, and so the continuity of f implies that f is bounded and also uniformly continuous. Given any $\epsilon > 0$, there exists $\delta > 0$ such that whenever $x, y \in [a, b]$ are such that $|x - y| < \delta$, we have $|f(x) - f(y)| < \epsilon$. Consider any partition $P_* = \{a, x_1, \dots, x_{n-1}, b\}$ such that $\max\limits_{k=0,\dots,n-1} |x_{k+1} - x_k| < \delta$. Then

$$U(f, P_*) - L(f, P_*) = \sum_{k=0}^{n-1} \left(\sup_{x \in [x_k, x_{k+1}]} f(x) - \inf_{x \in [x_k, x_{k+1}]} f(x) \right) (x_{k+1} - x_k) \le \sum_{k=0}^{n-1} \epsilon (x_{k+1} - x_k) = \epsilon(b - a).$$

Thus

$$0 \le \inf_{P \in P} U(f, P) - \sup_{P \in \mathcal{P}} L(f, P) \le \inf_{P \in P} U(f, P) - L(f, P_*) \le U(f, P_*) - L(f, P_*) \le \epsilon(b - a).$$

As the choice of $\epsilon > 0$ was arbitrary, it follows that

$$\inf_{P \in P} U(f, P) = \sup_{P \in \mathcal{P}} L(f, P),$$

and so f is Riemann integrable. □

One can show that the Riemann integral has the following elementary properties, which are left as an exercise to the student.

Exercise 5.50. Let f, g be Riemann integrable on $[a, b]$ and $\alpha \in \mathbb{R}$. Then the following hold:

(1) $\displaystyle\int_a^b f(x) + g(x) dx = \int_a^b f(x) dx + \int_a^b g(x) dx.$

(2) $\displaystyle\int_a^b \alpha \cdot f(x) dx = \alpha \int_a^b f(x) dx.$

(3) If for all $x \in [a, b]$, $f(x) \ge 0$, $\displaystyle\int_a^b f(x) dx \ge 0.$

(4) If for all $x \in [a, b]$, $f(x) \le g(x)$, then $\displaystyle\int_a^b f(x) dx \le \int_a^b g(x) dx.$

(5) $|f|$ is also Riemann integrable and $\displaystyle\left| \int_a^b f(x) dx \right| \le \int_a^b |f(x)| dx.$

(6) If f is continuous and nonnegative on $[a, b]$ and $\displaystyle\int_a^b f(x) dx = 0$, then f is identically equal to 0.

5.7. Fundamental theorem of integral calculus

We will now prove the following result.

Theorem 5.51. *Let $f : [a, b] \to \mathbb{R}$ be continuously differentiable on $[a, b]$. Then*

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

Proof. f' is continous on $[a, b]$ and so uniformly continuous there. Let $\epsilon > 0$. Then there exists a $\delta > 0$ such that whenever $x, y \in [a, b]$ are such that $|x - y| < \delta$, we have $|f'(x) - f'(y)| < \epsilon$. Consider any partition $P_* = \{x_0 = a,\ x_1, \dots,\ x_{n-1},\ x_n = b\}$ with

$$\max_{k=0,1,\dots,n} (x_{k+1} - x_k) < \delta.$$

By the Mean Value Theorem, for each $k = 0, \dots, n - 1$, there is a $c_k \in (x_k, x_{k+1})$ such that

$$\frac{f(x_{k+1}) - f(x_k)}{x_{k+1} - x_k} = f'(c_k).$$

Hence

$$\begin{aligned} U(f,P_*) - (f(b)-f(a)) &= \sum_{k=0}^{n-1}\left(\sup_{x\in[x_k,x_{k+1}]} f'(x)\right)(x_{k+1}-x_k) - \sum_{k=0}^{n-1}(f(x_{k+1})-f(x_k)) \\ &= \sum_{k=0}^{n-1}\left(\sup_{x\in[x_k,x_{k+1}]} f'(x) - f'(c_k)\right)(x_{k+1}-x_k) \quad \text{(note this is } \geq 0) \\ &\leq \sum_{k=0}^{n-1}\epsilon(x_{k+1}-x_k) = \epsilon(b-a). \end{aligned}$$

Similarly,

$$\begin{aligned} (f(b)-f(a)) - L(f,P_*) &= \sum_{k=0}^{n-1}\left(f'(c_k) - \inf_{x\in[x_k,x_{k+1}]} f'(x)\right)(x_{k+1}-x_k) \quad \text{(note this is } \geq 0) \\ &\leq \sum_{k=0}^{n-1}\epsilon(x_{k+1}-x_k) = \epsilon(b-a). \end{aligned}$$

Thus

$$\int_a^b f'(x)dx - (f(b)-f(a)) = \inf_{P\in\mathcal{P}} U(f,P) - (f(b)-f(a)) \leq U(f,P_*) - (f(b)-f(a)) \leq \epsilon(b-a),$$

and

$$\int_a^b f'(x)dx - (f(b)-f(a)) = \sup_{P\in\mathcal{P}} L(f,P) - (f(b)-f(a)) \geq L(f,P_*) - (f(b)-f(a)) \geq -\epsilon(b-a).$$

Consequently, $\left|\int_a^b f'(x)dx - (f(b)-f(a))\right| < \epsilon(b-a)$. As the choice of $\epsilon > 0$ was arbitrary, it follows that

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

This completes the proof. □

Exercise 5.52. Redo Example 5.47 using the Fundamental Theorem of Integral Calculus and the fact that $\frac{d^2}{dx^2}x^2 = 2$.

Bibliography

[B1] K.G. Binmore. *The Foundations of Analysis: a Straightforward Introduction. Book 1. Logic, Sets and Numbers.* Cambridge University Press, Cambridge-New York, 1980.
(This book is an informal introduction to logic and set theory. As the title suggests, the properties of the real numbers are emphasized in order to prepare the student for courses in analysis. Examples, exercises, and diagrams are abundant.)

[B2] K.G. Binmore. *The Foundations of Analysis: a Straightforward Introduction. Book 2. Topological Ideas.* Cambridge University Press, Cambridge-New York, 1981.
(This is a leisurely introduction to the basic notions of analysis. Clear motivation is provided for the results to follow.)

[R] W. Rudin. *Principles of Mathematical Analysis.* Third edition. McGraw-Hill, Singapore, 1976.
(This book is tough to read, but having ploughed through it, with the exercises, gives a solid foundation.)

[S] G.E. Shilov. *Elementary Real and Complex Analysis.* Revised English edition translated from the Russian and edited by Richard A. Silverman. Corrected reprint of the 1973 English edition. Dover Publications, Mineola, NY, 1996.
(This book is easier to read than Rudin's book, and offers the reader a leisurely but systematic introduction to Analysis. The Dover edition is affordable.)

Index