A crash course in advanced calculus

Alex Degtyarev

September 24, 2023

These notes are distributed in the hope that they may be useful, but without any warranty. Proofs are mostly omitted: once a statement is given, a math student is supposed to be able to find a proof on one's own (*cf.* Appendix A); this is the only way to learn mathematics. However, I decided to include some of the proofs, merely as an illustration of a "typical" usage of the new concepts introduced.

A few more advanced theorems are dealt with in the appendices.

The goal of these notes is to serve as a prerequisite to my course in complex analysis, even though complex analysis *per se* is almost never mentioned.

This is a preliminary version. Any remarks concerning typos or inaccuracies (beyond the obvious intent to keep the exposition informal) are welcome.

Contents

1 Metric spaces **2**
1.1 Definition and examples . . . 3
1.2 Limits of sequences and maps . . . 4
1.3 Open and closed sets . . . 6
1.4 Digression: topology . . . 8
1.5 Continuous maps . . . 9
1.6 Uniform continuity . . . 11

2 Compactness **12**
2.1 Compact spaces and subsets . . . 12
2.2 Compactness and continuity . . . 14

3 Completeness **15**

4 Connectedness **17**
4.1 Connectedness 17
4.2 Connected components 19
4.3 Path connectedness 19

5 Uniform convergence **21**
5.1 Pointwise *vs.* uniform convergence 21
5.2 Weaker versions 23
5.3 Digression: function spaces 24

6 Series **25**
6.1 Numeric series 25
6.2 Functional series 28
6.3 Power series 29

A Proving theorems in advanced calculus **31**
A.1 Proof of Theorem 1.19 32
A.2 Continuity of line integrals 33

B Proof of Theorem 1.8 **35**

C Proof of Theorem 2.4 **36**

D Proof of Theorem 6.12 **38**

1 Metric spaces

Reviewing the multiple definitions of limit in your calculus textbook, you may notice that they repeat one another almost literally; the only thing that does change is the inequality at the very end:

$$|x-a| < \varepsilon \quad \text{in univariate calculus, or} \tag{1.1}$$

$$\left.\begin{array}{l} \sqrt{(x-a)^2+(y-b)^2} \\ |x-a|+|y-b| \\ \max\{|x-a|,|y-b|\} \end{array}\right\} < \varepsilon \quad \text{(depending on the book) in bivariate.} \tag{1.2}$$

This fact suggests that, in order to develop a unified theory and avoid repetition of all proofs, we need a new concept that would encompass all these cases.

1.1 Definition and examples

It is not difficult to observe that the left hand side of (1.1), (1.2) is some sort of distance between the variable point x, or (x, y), and fixed point a, or (a, b). Thus, all that we need to speak about limits and continuity is to know how to measure the distance between points, and a quick glance at the proofs found in the textbooks suggests that very few properties of this distance function are used.

Definition 1.3. A *metric space* is a set X equipped with a *metric*, or distance function, $\rho\colon X \times X \to \mathbb{R}_+$ such that

1. $\rho(x, y) = 0$ if and only if $x = y$ (metric distinguishes points),
2. $\rho(x, y) = \rho(y, x)$ (*symmetry*), and
3. $\rho(x, z) \leqslant \rho(x, y) + \rho(y, z)$ (the *triangle inequality*).

Example 1.4 (not very useful)**.** On any set X, one can consider the so-called *discrete metric* $\rho(x, x) = 0$ and $\rho(x, y) = 1$ whenever $x \neq y$. It is not very useful as only essentially constant sequences would converge (see §1.2 below).

Example 1.5. The left hand sides of (1.1) are the most commonly used metrics on $\mathbb{R}^n$:

$$L^1\colon \qquad \rho(x, y) = \sum_{k=1}^{n} |x_k - y_k|,$$

$$L^2\colon \qquad \rho(x, y) = \sqrt{\sum_{k=1}^{n} (x_k - y_k)^2},$$

$$L^\infty\colon \qquad \rho(x, y) = \max_{1 \leqslant k \leqslant n} |x_k - y_k|.$$

In fact, for any real number $1 \leqslant p \leqslant \infty$, one can consider the L^p*-metric*

$$\rho(x, y) = \left(\sum_{k=1}^{n} |x_k - y_k|^p \right)^{1/p};$$

for this one, proving the triangle inequality is a bit more challenging.

In fact, all these metrics are *equivalent* (see Theorem 1.8 below) in the sense that they give rise to the same notions of limit and continuity (see §1.2 below). That is why there is no harm in different definitions used in calculus textbooks.

If $n = 1$, Example 1.5 boils down to $\rho(x, y) = |x - y|$. If $X = \mathbb{C} \simeq \mathbb{R}^2$, the Euclidean L^2-metric is $\rho(x, y) = |x - y|$ (the modulus of a complex number).

Some infinite dimensional examples are found in §5.3 below.

Example 1.6. Any subset $Z \subset X$ of a metric space naturally has the *induced* (or restricted) metric; with this metric, Z is called a *subspace* of X. For example, the unit sphere

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

with the induced Euclidean metric is a metric space on its own.

Example 1.7. All metrics in Example 1.5 belong to a very special class of metrics, those compatible with the vector space structure. A *normed* space is a (real) vector space X equipped with a *norm* (or length) function $\|\cdot\|: X \to \mathbb{R}_+$ such that

1. $\|x\| = 0$ if and only if $x = 0$,
2. $\|rx\| = |r|\|x\|$ for any $r \in \mathbb{R}$, and
3. $\|x + y\| \leqslant \|x\| + \|y\|$ (the triangle inequality).

If $\|\cdot\|$ is a norm, then $\rho(x, y) := \|x - y\|$ is a metric.

Theorem 1.8 (see Appendix B)**.** *On a* **finite dimensional** *vector space X, any two norms $\|\cdot\|_1$, $\|\cdot\|_2$ are* equivalent: *there are constants $C > c > 0$ such that*

$$c\|x\|_2 \leqslant \|x\|_1 \leqslant C\|x\|_2$$

for any vector $x \in X$. In particular, any two norms result in the same notions of limit, continuity, closed and open sets, compactness, etc (*see below*).

Convention 1.9. To avoid a cultural shock by too high level of generality, and since our principal concern is $X = \mathbb{C}$ or $\mathbb{R}$ with the standard Euclidean metric, from now on I will use the notation $\rho(x, y) = |x - y|$, even though the "$-$" might not be defined at all. (As a typical example, if X is a *subspace* of $\mathbb{R}^n$ and $x, y \in X$, the difference $x - y$ does not need to lie in X.) Still, unless mentioned otherwise, all definitions and statements hold for an arbitrary metric space.

With this convention, some statements and proofs can be copied from calculus textbooks literally, even though they acquire a totally new meaning.

1.2 Limits of sequences and maps

Now, we are ready to copy (from our favorite calculus textbook) the definitions of limits. Below, X, Y, *etc* are metric spaces.

Definition 1.10. A point $a \in X$ is called the *limit* of a sequence $x_n \in X$,

$$\lim_{n\to\infty} x_n = a, \quad \text{or} \quad x_n \underset{n\to\infty}{\longrightarrow} a, \quad \text{or} \quad x_n \longrightarrow a,$$

if

$$\forall \varepsilon > 0\ \exists N \in \mathbb{N} \colon n > N \Rightarrow |x_n - a| < \varepsilon. \tag{1.11}$$

If limit exists, x_n is said to *converge*; otherwise, it *diverges*.

Comparing this to the standard definition of limit in $\mathbb{R}$, we conclude that

$$x_n \to a \quad \text{if and only if} \quad |x_n - a| \to 0. \tag{1.12}$$

Convention 1.13. Here and below, instead of introducing a new constant N in (1.11) (especially, if many such constants $N_1 \leqslant N_2 \leqslant \dots$ need to be introduced step-by-step), we often merely say "$|x_n - a| < \varepsilon$ for $n \gg 0$."

Example 1.14. In a finite dimensional space $\mathbb{R}^n$, with any norm, convergence of sequences is coordinatewise: a sequence converges if and only if so do all its coordinate sequences.

The definition of the limit of a map is a bit trickier. Originally, the notion of limit was designed to "assign" a value to a function that is *not* defined at all (the definition of derivative) or "wrongly" defined (removable discontinuity) at the limit point $x = a$. For this reason, we have to consistently exclude a itself from the consideration, to the extend that f is not even assumed to be defined at a.

Definition 1.15. A point $b \in Y$ is called the *limit* of a map $f \colon X \smallsetminus a \to Y$ as x approaches $a \in X$,

$$\lim_{x\to a} f(x) = b, \quad \text{or} \quad f(x) \underset{x\to a}{\longrightarrow} b,$$

if

$$\forall \varepsilon > 0\ \exists \delta > 0 \colon 0 < |x - a| < \delta \Rightarrow |f(x) - b| < \varepsilon. \tag{1.16}$$

Going over a calculus textbook, we can easily prove basic properties of limits:

- if a limit exists, it is unique;
- if x_n converges, it is bounded;
- any subsequence of a converging sequence converges to the same limit.

For maps, we should be more careful: the existence of a limit implies that f is bounded *in a neighborhood of* a, *i.e.*, for $0 < |x - a| < \delta$ for some $\delta > 0$.

More subtle properties of limits, like additivity or multiplicativity, need more work just because in a general metric spaces they make no sense (as there is no addition or multiplication). However, if X is a normed space (Example 1.7), then

$$\lim(x_n + y_n) = \lim x_n + \lim y_n \quad \text{and} \quad \lim r x_n = r \lim x_n \text{ for } r \in \mathbb{R}.$$

(For maps, the extra assumptions should be made about the target space Y; in particular, all familiar properties hold if $Y = \mathbb{C}$ or $\mathbb{R}$.) If X is a *normed algebra* (*i.e.*, there also is a multiplication $X \times X \to X$, which is bilinear and compatible with the norm, $\|xy\| \leqslant \|x\|\|y\|$), then also

$$\lim x_n y_n = (\lim x_n)(\lim y_n).$$

The "quotient rule" holds if $Y = \mathbb{C}$ or $\mathbb{R}$.

Example 1.17. In addition to $\mathbb{R}$ and $\mathbb{C}$, where $|xy| = |x||y|$, a useful example of a normed algebra is the algebra of square matrices (of a fixed size), or operators $A\colon \mathbb{R}^n \to \mathbb{R}^n$, with the *operator norm*

$$\|A\| := \sup_{|x|=1} |Ax|,$$

where $|\cdot|$ is the Euclidean norm of vectors. In view of Theorem 2.12 below, one can change $\sup$ to $\max$. Here, we only have the inequality $\|AB\| \leqslant \|A\|\|B\|$.

Limits of maps can alternatively be defined in terms of limits of sequences.

Definition 1.18. A point $b \in Y$ is called the *limit* of a map $f\colon X \smallsetminus a \to Y$ as x approaches $a \in X$ if $f(x_n) \to b$ whenever $x_n \in X \smallsetminus a$ and $x_n \to a$.

Theorem 1.19 (see §A.1)**.** *For a map $f\colon X \to Y$ between two metric spaces, definitions 1.15 and 1.18 are equivalent.*

1.3 Open and closed sets

Various special subsets of metric spaces are important in analysis.

Definition 1.20. The *open ball* and *closed ball* of radius $r \geqslant 0$ about a point $a \in X$ are, respectively, the sets

$$B_r(a) := \{x \in X \mid |x - a| < r\} \quad \text{and} \quad \bar{B}_r(a) := \{x \in X \mid |x - a| \leqslant r\}.$$

The notation $\bar{B}_r(a)$ is not very fortunate, as the bar (which often does stand for the closure) in complex analysis is reserved for complex conjugation. On the other hand, we do not use closed balls very often.

Definition 1.21. A subset $U \subset X$ is *open* if $\forall a \in U\ \exists \delta > 0\colon B_\delta(a) \subset U$. Given a point $a \in X$, any open subset $U \ni a$ (or, more generally, any subset *containing* an open subset $U \ni a$) is called a *neighborhood* of a.

The definitions of limits can be restated in terms of neighborhoods (thus avoiding an explicit reference to the metric): (1.11) and (1.16) become, respectively,

$$\forall \text{ n/hood } U \ni a\ \ \exists N \in \mathbb{N}\colon n > N \Rightarrow x_n \in U, \tag{1.22}$$

$$\forall \text{ n/hood } V \ni b\ \ \exists \text{ n/hood } U \ni a\colon U \smallsetminus a \subset f^{-1}(V). \tag{1.23}$$

Theorem 1.24. *Open sets have the following properties*:

1. $\varnothing$ *and* X *itself are open*;
2. *if all* U_α, $\alpha \in S$ *(any index set) are open, so is* $\bigcup_{\alpha \in S} U_\alpha$;
3. *if* U *and* V *are open, so is* $U \cap V$.

Example 1.25. According to Theorem 1.24, the collection of open sets is closed under **arbitrary** unions and **finite** intersections. The intervals $(-1/n, 1/n) \subset \mathbb{R}$, $n > 0$, are open, but their intersection $\{0\}$ is not.

Definition 1.26. A subset $A \subset X$ is closed if $\lim x_n \in A$ whenever a sequence $x_n \in A$ converges. (Note that we do *not* require that the limit should exist, but if it does, it must be in A.)

Theorem 1.27. *Closed sets have the following properties*:

1. $\varnothing$ *and* X *itself are closed*;
2. *if* A *and* B *are closed, so is* $A \cup B$;
3. *if all* A_α, $\alpha \in S$ *(any index set) are closed, so is* $\bigcap_{\alpha \in S} A_\alpha$.

Furthermore, a set $A \subset X$ *is closed if and only if* $X \smallsetminus A$ *is open.*

Example 1.28. The collection of closed sets is closed under **finite** unions and **arbitrary** intersections (*cf.* Theorem 1.24). The segments $[1/n, 1] \subset \mathbb{R}$, $n > 0$, are closed, but their union $(0, 1]$ is not.

The next theorem is *not* a tautology as, for now, "open ball" and "closed ball" are merely terms not referring to any particular properties of the sets.

Theorem 1.29. *Open balls are open. Closed balls are closed.*

Example 1.30. Intuitively, closed are the sets containing their boundary (whatever this means), whereas open are those not containing a single boundary point.

Alternatively, open subsets are those given by *strict* inequalities, *e.g.*, $\{x < y\}$ or $\{x^2 + y^2 < 1\}$ in $\mathbb{R}^2$, whereas closed are those given by *equations* and/or *non-strict* inequalities, *e.g.*, $\{x \leqslant y^2\}$ or $\{x + y = 1\}$. See Remark 1.44 below.

Warning 1.31. It should be emphasized that "closed" is not an antonym to "open": most subsets are neither open nor closed, *e.g.*, $[0, 1) \subset \mathbb{R}$.

Note also that "open" and "closed" are not properties of metric spaces but rather of subsets in an ambient space. Thus, $A := [0, 1)$ is both open and closed in itself; it is open but not closed in the ray $[0, \infty)$, closed but not open in the ray $(-\infty, 0)$, and neither open nor closed in the whole line $\mathbb{R}$.

A more meaningful example is the upper hemisphere $z > 0$ in the unit sphere $S^2 := \left\{(x, y, z) \in \mathbb{R}^3 \;\middle|\; x^2 + y^2 + z^2 = 1\right\}$: it is open in S^2 but not in R^3.

Remark 1.32. In general, a subset U of a subspace $Y \subset X$ is open in Y if and only if there is an open set $V \subset X$ such that $U = V \cap Y$. In other words, open in a subspace are the sets cut off by open subsets of the ambient space. A similar observation applies to closed sets. In topology (see §1.4 below), this is the *definition* of the induced topology on a subspace.

Remark 1.33. When we discuss the differentiability of functions, in order to make the derivative $f'(a)$ meaningful and capturing the local structure of f, we want to be able to step away from a "in any direction"; in other words, we want f to be *defined on a neighborhood of* a. Another desired property is the connectedness of the domain of f (see §4 below), as otherwise a point of the subset may not "know" anything about other points (to the extent that even $(f' \equiv 0) \Rightarrow (f = \text{const})$ no longer holds). An open connected subset of X is called a *domain* (usually this notion applies to $X = \mathbb{R}^n$), and functions are typically assumed to be defined on such when their differentiability is discussed.

1.4 Digression: topology

As mentioned above, limits (and then, further, continuity and such) can be defined in terms of open sets, without an explicit reference to the metric. On the other hand, it is not uncommon that different metrics give rise to the same collection of open sets, hence the same limits (*cf.* Theorem 1.8). This observation leads us to yet another level of abstraction.

Definition 1.34. A *topological space* is a set X equipped with a *topology*, *i.e.*, a collection τ of subsets of X, called *open*, satisfying the three properties stated in Theorem 1.24. *Closed* in X are the sets whose complement is open.

Using (1.22) and (1.23), we can define limits of sequences and maps in topological spaces. Some of the more advanced properties considered below (most notably, compactness, see §2, and connectedness, see §4) are of a purely topological nature, whereas some others (*e.g.*, completeness, see §3, or various "uniform" properties) are not.

Unfortunately, without quite a few extra axioms, these new general concepts are not very well behaved: limit may not be unique, Theorem 1.19 does not hold, and neither does Definition 1.26 or most other definitions in terms of limits of sequences. (Of course, the definitions may still be given, but the resulting objects are useless.) For this reason, we will not pursue this line very actively.

1.5 Continuous maps

We are ready to introduce the most fundamental notion of calculus.

Definition 1.35. A map $f\colon X \to Y$ is called *continuous at a point* $a \in X$ if

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

(In particular, we imply that the limit must exist!) The map f is called *continuous* if it is continuous at each point of X.

In view of Theorem 1.19, we have the following restatement (which does *not* extend to topological spaces).

Theorem 1.36. *A map $f\colon X \to Y$ between* **metric** *spaces is continuous at a point $a \in X$ if and only if $f(x_n) \to f(a)$ whenever $x_n \in X$ and $x_n \to a$.*

Example 1.37. The identity map $X \to X$, $x \mapsto x$, and constant map $X \to Y$, $x \mapsto \text{const} \in Y$, are continuous. For any fixed $a \in X$, the distance map $X \to \mathbb{R}$, $x \mapsto |x - a|$ is continuous. For the latter, we use the triangle inequality twice:

$$|x-a| - |y-x| \leqslant |y-a| \leqslant |x-a| + |y-x|.$$

When speaking about the continuity, it is convenient to recast the definition of limit to avoid the weird exclusion of the limit point a. Thus, (1.16) and (1.23) take the following form: f is continuous at a if and only if

$$\forall \varepsilon > 0\, \exists \delta > 0\colon |x-a| < \delta \Rightarrow |f(x) - f(a)| < \varepsilon, \quad \text{or} \tag{1.38}$$

$$\forall \text{ n/hood } V \ni f(a)\ \exists \text{ n/hood } U \ni a\colon U \subset f^{-1}(V). \tag{1.39}$$

Using (1.38), we conclude the f is continuous if and only if

$$\forall a \in X\ \forall \varepsilon > 0\ \exists \delta > 0\ \forall x \in X\colon |x-a| < \delta \Rightarrow |f(x) - f(a)| < \varepsilon, \tag{1.40}$$

whereas (1.39) is recast into the following topological definition of continuity.

Theorem 1.41. *A map $f\colon X \to Y$ is continuous if and only if the pull-back $f^{-1}(V)$ of any open set $V \subset Y$ is open in X. Equivalently* (*see Theorem 1.27*)*, f is continuous if and only if the pull-back of any closed set is closed.*

Warning 1.42. In general, it is *not* true that the continuous **image** of a closed/open set is, respectively, closed/open. For example, the continuous map $\mathbb{R}^2 \to \mathbb{R}$, $(x, y) \mapsto x$, takes the closed hyperbola $\{xy = 1\}$ to $\mathbb{R} \smallsetminus 0$, which is not closed. Likewise, the map $\mathbb{R} \to \mathbb{R}$, $x \mapsto x^2$, takes the open interval $(-1, 1)$ to $[0, 1)$, which is not open. Though, there are situations where one can guarantee that $f(\text{closed}) = \text{closed}$ or $f(\text{open}) = \text{open}$, *cf.*, respectively, Theorem 2.13 below and a remark thereafter or the *open mapping theorem* (not included into these notes) for holomorphic functions.

A bijective continuous map $f\colon X \to Y$ *whose inverse $f^{-1}\colon Y \to X$ is also continuous* is called a *homeomorphism*. Theorem 1.41 suggests that homeomorphisms are indeed the isomorphisms in the category of topological spaces: two such spaces are "indistinguishable" from the point of view of topology if and only if they are related by a homeomorphism (*cf.* group isomorphisms, isomorphisms of vector spaces, *etc*). All topological properties of such spaces are identical.

Warning 1.43. Unlike algebra, where the set-theoretic inverse of a morphism is usually also a morphism, the inverse of a continuous map does *not* need to be continuous, *e.g.*,

$$[0, 2\pi) \to S^1 \subset \mathbb{C}, \quad t \mapsto \cos t + i \sin t.$$

The map is continuous and invertible, but it cannot be a homeomorphism since S^1 is compact whereas $[0, 2\pi)$ is not (see §2 and Theorem 2.5 below). In a sense, this example is the source of most problems in complex analysis: there is no "nice" continuous way to represent angles, or arguments, by numbers.

Remark 1.44. Theorem 1.41 makes the second paragraph in Example 1.30 more precise. In $\mathbb{R}$, the singleton $\{0\}$ and closed ray $(-\infty, 0]$ are closed, whereas the open ray $(-\infty, 0)$ is open. Hence, if $f, g, h, \ldots : X \to \mathbb{R}$ are continuous, the sets of the form

$$\{f(x) = 0\} = f^{-1}(0) \quad \text{or} \quad \{g(x) \leqslant 0\} = g^{-1}(-\infty, 0]$$

and arbitrary intersections thereof (see Theorem 1.27) are closed, whereas sets of the form

$$\{h(x) < 0\} = h^{-1}(-\infty, 0)$$

and finite intersections thereof (see Theorem 1.24) are open.

Continuous maps between metric spaces enjoy all familiar properties:

1. if f is continuous at a, it is bounded in a neighborhood of a;
2. *composition of continuous maps is continuous*: if $f\colon X \to Y$ is continuous (at a) and $g\colon Y \to Z$ is continuous (at $f(a)$), then $g \circ f$ is continuous (at a);
3. *continuous maps commute with limits*: if f is continuous (at $\lim x_n$), then

$$\lim_{n\to\infty} f(x_n) = f\left(\lim_{n\to\infty} x_n\right).$$

If the target Y is a normed space (or normed algebra), then the pointwise sums (and products) of continuous maps are continuous. If $Y = \mathbb{C}$ or $\mathbb{R}$, the "quotient rule" holds as well.

1.6 Uniform continuity

In (1.38), we can freely interchange the two universal quantifiers $\forall a$ and $\forall \varepsilon$ but, in general, we cannot swap $\forall a$ and $\exists \delta$: given an ε, we should find *its own* δ for each a separately. If we do change the two quantifiers, we arrive at a stronger property: a map $f\colon X \to Y$ is said to be *uniformly continuous* if

$$\forall \varepsilon > 0 \; \exists \delta > 0 \; \forall a, x \in X\colon |x - a| < \delta \Rightarrow |f(x) - f(a)| < \varepsilon. \tag{1.45}$$

Note that this definition is symmetric with respect to a and x.

Uniform continuity is a metric property; it has no topological counterpart.

Example 1.46. All maps in Example 1.37 are uniformly continuous. The function $(0, \infty) \to \mathbb{R}$, $x \to 1/x$, is continuous, but not uniformly. More examples of uniformly continuous functions are provided by Theorem 2.14 below, and the usage of this notion is briefly discussed in Appendix A.

2 Compactness

In the next three chapters we discuss the fundamental "three C's": compactness, completeness, and connectedness. This are certain special properties of metric spaces that are used extensively in analysis.

2.1 Compact spaces and subsets

The definition "compact = bounded closed" found in most calculus textbooks is totally meaningless and should be forgotten: it is merely a property of but a few normed spaces (see Theorem 2.5 below). Besides, it is completely unclear how this definition can be used. Historically, there are two "true" definitions.

Definition 2.1 (sequential compactness)**.** A space K (or, more generally, subset $K \subset X$) is called *compact* if any sequence $x_n \in K$ has a converging subsequence $x_{n_k} \to a \in K$. (Observe that, if K is a subset of a bigger space, we demand that the limit must also be in K.)

Definition 2.2 (compactness)**.** A space K (or, more generally, subset $K \subset X$) is called *compact* if any collection U_α, $\alpha \in S$ (any index set) of *open* sets such that $K \subset \bigcup_{\alpha \in S} U_\alpha$ (an *open covering* of K) has a **finite** subcollection $U_1, \ldots, U_n$ such that still $K \subset U_1 \cup \ldots \cup U_n$ (a *finite subcovering*).

The last definition has a practical restatement: one passes from open sets U_α to the closed sets $A_\alpha := X \smallsetminus U_\sigma$ and uses de Morgan's laws.

Theorem 2.3 (restatement of Definition 2.2)**.** *A space X is compact if and only if, for any collection of closed sets $A_\alpha \in X$, $\alpha \in S$, such that any* **finite** *intersection $A_{\alpha_1} \cap \ldots \cap A_{\alpha_n}$ is nonempty, the full intersection $\bigcap_{\alpha \in S} A_\alpha$ is nonempty.*

Unlike closedness (see Warning 1.31), compactness is an intrinsic property of K: it does not depend on the ambient space X of which K may be a subset. (For Definition 2.2, Remark 1.32 needs to be used.)

Theorem 2.4. *For metric spaces, Definitions 2.1 and 2.2 are equivalent.*

This statement is somewhat more involved than most theorems in calculus: one needs to go through several steps (see Appendix C below). In general, for topological spaces, neither of the two definitions implies the other. In topology, the compactness is given by Definition 2.2, whereas the *sequential compactness* introduced by Definition 2.1 is hardly ever used.

Any finite set is compact and, in fact, compactness is some sort of "finiteness" for bigger sets (so that, *e.g.*, we can use $\min$ and $\max$ instead of $\inf$ and $\sup$, *cf.* Theorem 2.12 below). Clearly, a compact set must be closed and bounded in any ambient metric space (why?), but the converse is almost never true. Luckily for us, the next theorem describes almost all compact sets that we will need.

Theorem 2.5 (Heine–Borel)**.** *In a* **finite dimensional** *normed space $\mathbb{R}^n$, a subset is compact if and only if it is bounded and closed.*

Idea of the proof. Essentially the statement boils down to the assertion that, in $\mathbb{R}^n$, any bounded sequence has a converging subsequence. This is first established for $\mathbb{R}$ (Bolzano–Weierstraß theorem, see Wikipedia; this is a fundamental property of $\mathbb{R}$ distinguishing it from $\mathbb{Q}$, and the proof depends on your favorite definition of $\mathbb{R}$) and then extended to any finite dimensional space by induction, using the fact that the convergence is coordinatewise (see Example 1.14). $\square$

Corollary 2.6. *Any open subset $U \subset \mathbb{R}^n$ is* locally compact*,* i.e.*, each point $a \in U$ has a compact neighborhood* (e.g.*, a sufficiently small closed ball*)*.*

A closed subset $A \subset X$ of a compact space X is compact, and a *discrete space* (one in which any set is open, equivalently closed) is compact if and only if it is finite. (These observations apply to both definitions of compactness.)

Corollary 2.7. *In a compact space X, any discrete subset is finite.*

We conclude with a few technical statements used quite often in the proofs. In Lemma 2.11, we need local compactness; that is why it is stated for $\mathbb{R}^n$ only.

Lemma 2.8 (Lebesgue)**.** *Let K be a compact metric space and U_α, $\alpha \in S$, an open covering of K. Then, there exists $\delta > 0$ such that, whenever $|y - x| < \delta$, the two points x, y are contained in a common set U_α of the covering.*

Proof (to illustrate the usage of Definition 2.1). Assume the contrary and pick a sequence $\delta_n \searrow 0$. (This notation is a shorthand for "a decreasing sequence of positive real numbers converging to 0".) By the assumption, for each δ_n, there is a pair x_n, y_n such that $|y_n - x_n| < \delta_n$ but x_n, y_n do not belong to any common element U_α of the covering.

Remark 2.9. At this point, we are supposed to say that, "by the compactness, the sequence x_n has a convergent subsequence $x_{n_k} \to a$; consider the corresponding subsequence y_{n_k} of y_n." However, especially if this step needs to be used several times (as in the inductive proof of Theorem 2.5), this purist's approach leads to multi-layered indices making the notation quite messy. For this reason, since x_n is **a** sequence picked by us, we usually just say "assume that we were lucky and x_n itself converges." Usually it is clear whether this inaccuracy is safe (*e.g.*, this is why we chose **a** sequence $\delta_n \searrow 0$ rather than just $\delta_n = 1/n$).

Thus, assume that $x_n \to a \in K$ converges. Since $|y_n - x_n| < \delta_n \to 0$, the other sequence $y_n \to a$ also converges to the same limit. On the other hand, since $\{U_\alpha\}$ is a covering, there is an element $U_\alpha \ni a$. By definition (U_α is open), it is a neighborhood of a, and by (1.22), we have $x_n, y_n \in U_\alpha$ for $n \gg 0$, which contradicts to our choice of the sequences. □

Lemma 2.10. *Assume that $K \subset X$ is compact, $A \subset X$ is closed, and $K \cap A = \varnothing$. Then, there exists $\delta >$ such that $|x - a| \geqslant \delta$ for all $x \in K$ and $a \in A$.*

The compactness is crucial here, *cf.* $\{xy = 1\}$ and $\{y = 0\}$ in $\mathbb{R}^2$.

Lemma 2.11. *Given $K \subset U \subset \mathbb{R}^n$ with K compact and U open, the set K has a compact neighbourhood $\bar{V}$ in U, so that $K \subset \bar{V} \subset U$. Besides, by Lemma 2.10, there exists $\delta > 0$ such that $x \in \bar{V}$ whenever $\operatorname{dist}(x, K) < \delta$.*

2.2 Compactness and continuity

Continuous functions on compact sets have particularly nice properties.

Theorem 2.12 (Weierstraß). *If K is compact, any continuous function $f\colon K \to \mathbb{R}$ is bounded and, moreover, takes its minimal and maximal values.*

Proof (a simpler illustration of Definition 2.1). Assume that f is unbounded from above and consider a sequence $M_n \nearrow \infty$. For each n, we can find $x_n \in K$ such that $f(x_n) > M_n$. By the compactness, we can assume that $x_n \to a \in K$. Then $f(x_n) \to f(a)$, which contradicts to the fact that $f(x_n)$ is unbounded.

Now, f is bounded, so that we have $M := \sup_{x \in K} f(x)$. Pick a sequence $M_n \nearrow M$ and ... (the rest is left to the reader). □

Theorem 2.13. *If K is compact and $f\colon K \to Y$ is continuous, then the image $f(K)$ is also compact.*

Try to prove this theorem using both Definition 2.1 and 2.2. Theorem 2.12 is an immediate consequence of Theorem 2.13: the image $f(K) \subset \mathbb{R}$ is compact, hence it is bounded (and so is f) and closed (thus containing its $\inf$ and $\sup$). Note also that any continuous map $f\colon K \to Y$ is *closed*, *i.e.*, takes closed sets to closed sets (*cf.* Warning 1.42): if $A \subset Y$ is closed, it is also compact; then the image $f(A)$ is compact and, hence, closed. Thus, *in this particular case of a compact domain*, if f is invertible, then f^{-1} is also continuous (see Theorem 1.41: the images under f are the pull-backs under f^{-1}).

The next important theorem is also based on Definition 2.1.

Theorem 2.14. *If K is compact, any continuous map $f\colon K \to Y$ is uniformly continuous.*

Idea of the proof. Assuming the contrary, fix an offending $\varepsilon > 0$, pick a sequence $\delta_n \searrow 0$, and construct a pair of sequences $x_n, y_n \in K$ such that $|y_n - x_n| < \delta_n$ but $|f(y_n) - f(x_n)| \geqslant \varepsilon$. Then, proceed as in the proof of Lemma 2.8. $\square$

3 Completeness

Using Definition 1.10 or various restatements thereof, it is impossible to establish the convergence of a sequence *unless its limit is known in advance*. Often, *e.g.*, in the proofs of various existence theorems, we manage to construct a sequence that would *allegedly* converge to the point we are looking for, and we need a means to establish this convergence *without knowing the limit point*. In $\mathbb{R}$, this tool is the so-called *Cauchy criterion*, and the validity of this criterion on a more general metric space is called *completeness*. Note that this is a purely metric property; it has no topological counterpart.

Definition 3.1. A sequence $x_n \in X$ is called *Cauchy* if

$$\forall \varepsilon > 0\, \exists N \in N\colon m, n > N \Longrightarrow |x_n - x_m| < \varepsilon.$$

Note that this definition does not refer to any extra data, only to x_n itself.

Example 3.2. Any converging sequence is Cauchy. (Prove this!) The converse is not true: the sequence $x_n = 1/n \in X := \mathbb{R} \smallsetminus 0$ is Cauchy (as it converges in the bigger space $\mathbb{R}$) but not convergent (as its “would be limit” is not in X).

Definition 3.3. A metric space X is *complete* if any Cauchy sequence converges. A complete normed space is called *Banach*.

This property has no topological counterpart.

A closed subset $X \subset Y$ of a complete space Y is also complete. Conversely, a complete space X is closed in any ambient space (*cf.* Example 3.2); due to this fact, complete metric spaces can be regarded as the “universally closed” ones. (In the category of “good”, *i.e.*, so-called *Hausdorff*, topological spaces, universally closed are the compact ones.) Any compact metric space is complete, but the converse is not true: see, *e.g.*, the next theorem that gives us a plethora of examples of complete spaces covering all our needs.

Theorem 3.4. *A* **finite dimensional** *normed space* $\mathbb{R}^n$ *is complete.*

Idea of the proof. Any Cauchy sequence is bounded; hence, by Theorem 2.5, it has a converging subsequence. Then, by the Cauchy property, the whole sequence converges to the same limit. □

Remark 3.5. According to Example 3.2, the space $X := \mathbb{R} \smallsetminus 0$ is not complete, essentially due to the fact that the limits of some "converging" sequences have been "artificially removed." This situation is typical: it is not very difficult to show that any metric space X can be embedded to a complete one. The "minimal" complete space containing X as a subspace is essentially unique; it is called the *completion* of X. Thus, $\mathbb{R}$ is the completion of $\mathbb{R} \smallsetminus 0$, and $\mathbb{R}$ is also the completion of $\mathbb{Q}$ (in a sense, this is precisely how and why $\mathbb{R}$ is constructed).

We conclude with a typical example of an existence theorem based on the completeness of the space involved. (This theorem is crucial in the proof of many other similar results, *e.g.*, the famous *inverse* and *implicit function theorems* in calculus.) A *contraction* is a transformation (self-map) $f\colon X \to X$ with the following property: there is a constant $q < 1$ such that $|f(y) - f(x)| \leqslant q|y - x|$ for all $x, y \in X$. Clearly, any contraction is continuous, even uniformly.

Theorem 3.6 (Banach contraction principle)**.** *Any contraction* $f\colon X \to X$ *of a* **complete** *metric space* X *has a unique fixed point.*

Proof. The uniqueness is obvious: if $f(x) = x$ and $f(y) = y$, then, by the definition, we have $|y - x| \leqslant q|y - x|$, implying $|y - x| = 0$.

The proof of the existence is a very typical application of completeness.

First, we construct a certain sequence in the hope that its limit is the point that we are looking for. Take for $x_0 \in X$ any point and define $x_n := f^n(x_0)$ for $n \geqslant 1$. In other words, we have $x_{n+1} = f(x_n)$: if $x_n \to a$, passing to the limit in the above recurrence relation would result in $a = f(a)$.

Next, we establish that the sequence x_n is Cauchy. Iterating the definition of contraction, we have $|f^n(y) - f^n(x)| \leqslant q^n|y - x|$. Hence, for any $n \geqslant 1$,

$$|x_{n+1} - x_n| = |f^n(x_1) - f^n(x_0)| \leqslant q^n|x_1 - x_0|.$$

Acting as suggested in Appendix A below, for $n \geqslant m$ we have

$$|x_n - x_m| \leqslant \sum_{k=m}^{n-1} |x_{k+1} - x_k| \leqslant |x_1 - x_0| \sum_{k=m}^{n-1} q^k < \frac{q^m}{1-q} |x_1 - x_0| =: b_m.$$

(The geometric series converges to the sum indicated due to the assumption that $q < 1$ and, implicitly, $q > 0$.) Since $0 < q < 1$, we have $b_m \to 0$ and the last inequality establishes the Cauchy property.

Finally, referring to the completeness of X, we conclude that x_n converges, $x_n \to a$, and prove that the limit a is the/a point searched for. Often, as in our case, this last step is combined with the first one, as an intuition behind the construction of the original sequence x_n in the first place. □

Note that the completeness of X in Theorem 3.6 is essential: if $X = \mathbb{R} \smallsetminus 0$, the map $x \mapsto \frac{1}{2}x$ is a contraction, but it has no fixed points. The whole idea behind the concept of completeness is ruling out pathological examples like this.

4 Connectedness

Connectedness is a formal framework for working with the intuitive notion of a metric (or topological) space made out of a single chunk as opposed to spaces broken into several unrelated pieces.

4.1 Connectedness

Recall that Definition 1.34 (or Theorems 1.24 and 1.27) dictates that each topological/metric space X has at least two subsets, *viz.* $\varnothing$ and X itself, that are both open and closed.

Definition 4.1. A space X is *connected* if it satisfies any of the following equivalent conditions:

1. X has no open-closed subsets other than $\varnothing$ and X itself;
2. there is no pair of disjoint open subsets $U, V \neq \varnothing$ such that $X = U \cup V$;
3. there is no pair of disjoint closed subsets $A, B \neq \varnothing$ such that $X = A \cup B$.

A pair (U, V) as in (2) or (A, B) as in (3) is called a *partition* of X. In fact, (2) and (3) are literally the same: *e.g.*, since $U = X \smallsetminus V$ and V is open, then U is also closed, see Theorem 1.27.

Theorem 4.2. *A closed segment $[a, b] \subset \mathbb{R}$ is connected.*

Idea of the proof. Let $[a, b] = A \cup B$ be a partition. Assume that $a \in A$ and, since $B \neq \varnothing$, consider $c := \inf B$. Using the closedness of A and B, prove that $c \in A \cap B$, contradicting to the assumption that A and B are disjoint. $\square$

Theorem 4.3. *Let $X = \bigcup X_\alpha$, $\alpha \in S$ (any index set) and assume that each X_α is connected and all X_α have a common point $a \in \bigcap X_\alpha$. Then X is connected.*

Proof. Consider a partition $X = A \cup B$ and assume that $a \in A$. Since $B \neq \varnothing$, it intersects at least one of X_α. Then, at least this particular X_α becomes partitioned *via* $X_\alpha = (A \cap X_\alpha) \cup (B \cap X_\alpha)$, see Remark 1.32. $\square$

Corollary 4.4. *The connected subsets of $\mathbb{R}$ are the intervals (open, closed, semi-open, finite or infinite) of the form $\langle a, b \rangle$, $-\infty \leqslant a \leqslant b \leqslant \infty$.*

Proof. Closed intervals are connected by Theorem 4.2. All others can be represented as unions of such, *e.g.*,

$$(-1, 1) = \bigcup_{n \geqslant 0} \left[-1 + \frac{1}{n}, 1 - \frac{1}{n}\right], \qquad [0, \infty) = \bigcup_{n \geqslant 0} [0, n],$$

etc., so that Theorem 4.3 applies.

Conversely, if $X \subset \mathbb{R}$ is connected, then, for any pair $a < b$ of points in X and any $a < c < b$ we must have $c \in X$, as otherwise

$$X = \big((-\infty, c) \cap X\big) \cap \big((c, \infty) \cap X\big)$$

would be a partition of X (see Remark 1.32). $\square$

In $\mathbb{R}^n$, even in $\mathbb{R}^2 = \mathbb{C}$, connected subspaces are mach more diverse. In a sense, that is why we had to introduce and study the notion of connectedness.

In view of Corollary 4.4, the following statement is a generalization of the *intermediate value theorem* in univariate calculus.

Theorem 4.5. *If X is connected and $f : X \to Y$ is continuous, then the image $f(X)$ is also connected.*

Corollary 4.6. *If a continuous function $f : X \to \mathbb{R}$ on a connected space X takes values $a < b$, it also takes any intermediate value $c \in [a, b]$.*

4.2 Connected components

The (*connected*) *component* of a point $a \in X$ is the union of all connected subspaces $a \in Y \subset X$. By Theorem 4.3, the component of a is connected, so that it is the largest connected subspace of X containing a.

Theorem 4.7. *Connected components constitute a partition of X* (*in the sense of combinatorial mathematics*: *"to be in the same component" is an equivalence relation*)*, so that X is disjoint union of its components.*

Intuitively, components are the "pieces" that X is made of. For example, the components of $\mathbb{R} \smallsetminus 0$ are $(-\infty, 0)$ and $(0, \infty)$, whereas the components of $\mathbb{Q}$ are single points. (Spaces with this latter property are called *totally disconnected*; not that they are not necessarily discrete, as the example of $\mathbb{Q}$ shows!)

Theorem 4.8. *The connected components of X are closed in X. They are not necessarily open* (cf. $\mathbb{Q}$ *above*)*; however, if X has but finitely many components, each of them is open* (*see Theorem 1.27*).

4.3 Path connectedness

The other version of connectedness is geometrically more intuitive but more involved technically. It may be regarded as part of the so-called *homotopy theory*. In this section, we reserve the notation $I := [0, 1] \subset \mathbb{R}$ for the unite segment; it is very common in homotopy theory in general.

Definition 4.9. A *path* in a (topological) space X is a continuous map $\gamma \colon I \to X$. The space X is *path connected* if any two points $x, y \in X$ can be connected by a path, *i.e.*, there is a path $\gamma \colon I \to X$ such that $\gamma(0) = x$ and $\gamma(1) = y$.

Example 4.10. Any interval $\langle a, b\rangle \subset \mathbb{R}$ (*cf.* Corollary 4.4) is path connected. More generally, any *convex* subset $X \subset \mathbb{R}^n$ is path connected: a path connecting $x, y \in X$ is $t \mapsto (1-t)x + ty$. In particular, both open and closed balls in $\mathbb{R}^n$ are path connected (and, hence, connected due to Theorem 4.11 below).

The *path component* of a point $x \in X$ is the set of all points $y \in X$ that can be connected to x by a path. Path components also constitute a partition of X (*cf.* Theorem 4.7), so that "to be in the same path component" is an equivalence relation. A formal justification of this claim relies on certain constructions involving paths and used throughout in homotopy theory:

- *reflexivity* ($a \sim a$): the *constant path* $t \mapsto a$;

- *symmetry* ($a \sim b \Rightarrow b \sim a$): the *inverse path* $\gamma^{-1}\colon t \mapsto \gamma(1-t)$ (which is not to be mixed with the same notation for the inverse function);
- *transitivity* ($a \sim b\ \&\ b \sim c \Rightarrow a \sim c$): the *path product*

  $$\gamma \cdot \delta\colon t \mapsto \begin{cases} \gamma(2t), & \text{if } t \in [0, 1/2], \\ \delta(2t-1), & \text{if } t \in [1/2, 1] \end{cases}$$

  (not to be mixed with composition), which is defined whenever $\gamma(1) = \delta(0)$. Here, we need to travel twice the distance in the same unit time; hence, we have to double the speed!

In general, path components are neither closed nor open (*cf.* Example 4.13 below).

Path connectedness is stronger than connectedness: essentially, this follows from Theorems 4.2 and 4.5. We also have a counterpart of Theorem 4.5.

Theorem 4.11. *If X is path connected, it is connected.*

Theorem 4.12. *If X is path connected and $f\colon X \to Y$ is continuous, then the image $f(X)$ is also path connected.*

Example 4.13. In general, the converse of Theorem 4.11 does not hold. A classical example is the subspace

$$X := \big(\{0\} \times [-1, 1]\big) \cup \left\{ y = \sin\frac{1}{x} \,\middle|\, x \in (0, 1] \right\} \subset \mathbb{R}^2$$

(which is, in fact, the closure in $\mathbb{R}^2$ of the graph on the right). It is connected (*cf.* Theorem 4.7) but not path connected (an easy exercise in calculus).

Luckily, in all cases that we are concerned about path connectedness is equivalent to connectedness.

Theorem 4.14. *An* **open** *subset $X \in \mathbb{R}^n$ is path connected if and only if it is connected.*

Proof. The "only if" part is given by Theorem 4.11, and the "if" part follows from the fact that X is *locally path connected*, *i.e.*, each point $a \in X$ has a path connected neighborhood (*e.g.*, a sufficiently small open ball, see Example 4.10, which must exist due to the assumption that X is open). This property implies (using the concept of path product) that each path component of X is open. Hence, if there were more than one, then

$$X = \{\text{one of the path components}\} \cup \{\text{union of the others}\}$$

would be a partition as in Definition 4.1(2). $\square$

5 Uniform convergence

In this section, we discuss briefly the appropriate version of convergence of a sequence of *maps* $f_n : X \to Y$.

5.1 Pointwise *vs.* uniform convergence

Given a sequence of maps $f_n : X \to Y$, it seems natural to declare f_n convergent, $f_n \to f$, if the sequence $f_n(x) \in Y$ converges at each point $x \in X$. Clearly, in this case the limit $\lim f_n$ is again a map $f : X \to Y$, $x \mapsto \lim f_n(x)$. This kind of convergence is called *pointwise*. Spelling (1.11) out, $f_n \to f$ if and only if

$$\forall x \in X\; \forall \varepsilon > 0\; \exists N \in \mathbb{N} : n > N \Rightarrow |f_n(x) - f(x)| < \varepsilon. \tag{5.1}$$

Example 5.2. Unfortunately, the pointwise limit does not need to retain any nice properties of the original maps, even the continuity. For example, the sequence $f_n : [0, 1] \to \mathbb{R}$, $x \mapsto x^n$, of continuous functions converges to

$$f : x \mapsto \begin{cases} 0, & \text{if } x \in [0, 1), \\ 1, & \text{if } x = 1, \end{cases}$$

which is no longer continuous.

The reason for this pathological behaviour is the fact that we do not control the "speed" of convergence, which may change dramatically from point to point: in (5.1), we need to find its own N for each point $x \in X$. Thus, as in §1.6, it may make sense to swap the two quantifiers and consider a stronger uniform notion.

Definition 5.3. A sequence $f_n : X \to Y$ converges to $f : X \to Y$ *uniformly* on X, $f_n \rightrightarrows f$ (there seems to be no common notation; we will use this one), if

$$\forall \varepsilon > 0\; \exists N \in \mathbb{N}\; \forall x \in X : n > N \Rightarrow |f_n(x) - f(x)| < \varepsilon.$$

Warning 5.4. There is no analogue of Theorem 2.14: the fact that the convergence is uniform has to be checked on a case-by-case basis. Though, most our needs are covered by a few general results like Theorems 6.15 and 6.17 below.

Theorem 5.5 (uniform Cauchy criterion)**.** *If the target Y is complete, a sequence $f_n : X \to Y$ is uniformly convergent if and only if it is* uniformly Cauchy*, i.e.,*

$$\forall \varepsilon > 0\; \exists N \in \mathbb{N}\; \forall x \in X : m, n > N \Rightarrow |f_n(x) - f_m(x)| < \varepsilon.$$

Uniform convergence is a metric notion; it has no topological counterpart. We illustrate the importance of this notion by proving two important theorems.

Theorem 5.6. *Uniform limit of a sequence of continuous maps is continuous.*

Proof. Thus, we assume that $f_n \rightrightarrows f$ and each f_n is continuous, and we need to show that f is continuous at a point $a \in X$. In other words, we need to estimate the distance $|f(x) - f(a)| < \varepsilon$. Following the guidelines of Appendix A, we bound it by a few other distances over which we have more control:

$$|f(x) - f(a)| \leqslant |f(x) - f_n(x)| + |f_n(x) - f_n(a)| + |f_n(a) - f(a)|.$$

Now, the timing is crucial. First, by the uniform convergence, we find and fix some $n \gg 0$ such that $|f(y) - f_n(y)| < \varepsilon$ *for each point* $y \in K$. (At this point, we do not know yet what x should be.) Then, by the continuity of this particular term f_n, we find $\delta > 0$ such that $|f_n(x) - f_n(a)|$ for $|x - a| < \delta$. We conclude that $|x - a| < \delta \Rightarrow |f_n(x) - f_n(a)| < 3\varepsilon$, which is good enough. (A purist would use $\varepsilon/3$ in all three auxiliary inequalities, but in longer proofs it would be too difficult to keep track of all extra constants along the way; hence, we never do that.) □

Theorem 5.7. *If a sequence $f_n \colon [a, b] \to \mathbb{R}$ of continuous functions converges uniformly to a function f, then*

$$\int_a^b f_n(x)dx \to \int_a^b f(x)dx.$$

Proof. Here is the estimate (see (A.4) below):

$$\left|\int_a^b f_n(x)dx - \int_a^b f(x)dx\right| \leqslant \int_a^b |f_n(x) - f(x)|dx < (b - a)\varepsilon;$$

you fill in the quantifiers. Do not forget to refer to Theorem 2.14 first to make sure that the integrals dealt with exist in the first place! □

Remark 5.8. Theorem 5.7 and its proof extend to all other types of integrals: double, triple, line of any kind, *etc.*; we will take all such statements for granted. Important is that the integrals should be **proper**, *i.e.*, over a compact region: we need finite length/area/volume. Convergence of sequences of improper integrals is more involved and has to be established on a case-by-case basis.

Warning 5.9. Integrating in Theorem 5.7 over $[a, x]$ rather than $[a, b]$, we see that the sequence of anti-derivatives (with appropriate constants of integration) also converges: $F_n(x) \to F(x)$. Moreover, this convergence is uniform on $[a, b]$ as the extra constant $|x - a|$ in the proof is uniformly bounded by $|b - a|$.

This observation does not work in the opposite direction and the convergence $f_n \rightrightarrows f$ does not tell us anything about the derivatives: *e.g.*,

$$\sqrt{x^2 + \frac{1}{n}} \rightrightarrows |x|,$$

which is not differentiable. For a more drastic example, the *Weierstraß function* in Wikipedia is continuous but *nowhere* differentiable; still, by Theorem 5.10 below, on any bounded segment $[a, b]$ it is a uniform limit of polynomials.

The next technical theorem will occasionally be used to simplify contours of integration and such. For proof, see, *e.g.*, *Bernstein polynomials* in Wikipedia.

Theorem 5.10 (Weierstraß approximation theorem)**.** *Any continuous real function* $f\colon [a, b] \to \mathbb{R}$ *is the uniform limit of a sequence of polynomials.*

If f is $m < \infty$ times continuously differentiable, we can apply Theorem 5.10 to $f^{(m)}$ and refer to Remark 5.8, concluding that f can be uniformly approximated by polynomials *together with the first* m *derivatives*. Due to Theorem A.8 below, line integrals can be approximated by those parametrized by polynomials.

5.2 Weaker versions

The uniformness of the convergence $f_n \to f$ on the whole domain X may be quite difficult to establish, often does not hold, and usually is not needed. For this reason, we introduce two related weaker versions of this notion.

Assume that $f_n \to f\colon X \to Y$. The convergence is said to be

- *locally uniform* if it is uniform on a neighborhood of each point $a \in X$;
- *uniform on compacta* if it is uniform on each compact subset $K \subset X$.

For example, it is clear that for Theorem 2.14 locally uniform convergence would suffice (as continuity is a local property), whereas for Theorem 5.7 (or for the convergence of the anti-derivatives, see Remark 5.8) we only need convergence uniform on compacta (as we integrate over compact segments).

In view of the next theorem, in complex analysis, following (some) textbooks, we will usually speak about convergence uniform on compacta. It is this kind of convergence that typically holds for power series (see Theorem 6.17 below).

Theorem 5.11. *Locally uniform convergence is uniform on compacta. On an* **open** *subset $X \subset \mathbb{R}^n$, convergence uniform on compacta is locally uniform.*

Proof. The first part is an illustration of the usage of Definition 2.2. Let $K \subset X$ be compact. For each point $a \in K$, there is an open neighborhood $U_a \ni a$ on which the convergence is uniform, and K is covered by these neighborhoods (*cf.* Remark C.6 below): $K \subset \bigcup_{a \in K} U_a$. By Definition 2.2, we can find a finite subcovering, $K \subset U_1 \cup \ldots \cup U_n$, and, the convergence being uniform on each of U_k, it is uniform on their (finite!) union. (Indeed, in Definition 5.3, in the self-explanatory notation, it suffices to take $N = \max_k N_k$. That would not work for an infinite union as the maximum may be infinite.)

The converse follows from the local compactness of X, see Corollary 2.6. $\square$

5.3 Digression: function spaces

Fix a compact space K and denote by $C(K, Y)$ the set of all continuous maps $f\colon K \to Y$. (If $Y = \mathbb{R}$ or $\mathbb{C}$, depending on the convention in place, it is usually omitted from the notation.) By Theorem 2.12 applied to the continuous (why?) function $K \to \mathbb{R}$, $x \mapsto |f(x) - g(x)|$, we have a well-defined maximum

$$|f - g| := \max_{x \in K} |f(x) - g(x)|$$

which makes $C(K, Y)$ a metric space. It is immediate that a sequence $f_n\colon K \to Y$ of continuous maps converges to $f\colon K \to Y$ *uniformly* if and only if $f_n \to f$ in $C(K, Y)$ with respect to the metric just defined. In this respect, the uniform convergence is more natural than the pointwise one, as it fits into our general framework. *E.g.*, Theorem 5.5 states that, if Y is complete, so is $C(K, Y)$. Thus, if Y is Banach, so is $C(K, Y)$, which is typically infinite dimensional.

As mentioned, uniform convergence is a purely metric notion. In topology, there are (more than one) ways to introduce a "natural" topological structure on $C(K, Y)$, and the convergence with respect to this structure could be referred to as "uniform convergence." However, this language is hardly ever used as in topology sequences do not play an important rôle.

There is a great variety of other useful function spaces. For example, one can consider m times continuously differentiable functions $[a, b] \to \mathbb{R}$ and use

$$|f - g| + |f' - g'| + \ldots + |f^{(m)} - g^{(m)}|$$

for the metric (to control the convergence of the derivatives, *cf.* Warning 5.9). Or, changing sums to integrals, one can consider L^p-metrics (see Example 1.5).

The latter are usually incomplete; for the completion, one needs to add some *measurable functions* and their *generalized derivatives*.

We leave details to courses in functional analysis.

6 Series

At first sight, a series is merely a special kind of sequence. The reason why series deserve a separate discussion is the fact that the questions asked and tools used are quite different from those for sequences. For example, we almost never try to find the sum of a series, confining ourselves to establishing its convergence.

Throughout this section, Y is a normed space, so that we can use sums.

6.1 Numeric series

Formally, a *series* is merely a sequence $a_n \in Y$, but, being a series, it is written in the form of a formal sum

$$\sum_{n=1}^{\infty} a_n. \tag{6.1}$$

In addition to a_n, associated to (6.1) is the sequence

$$A_n := \sum_{k=1}^{n} a_k \tag{6.2}$$

of its *partial sums*, and (6.1) is said to *converge* or *diverge* of so does (6.2), as a sequence. In the former case, $A := \lim A_n$ is called the *sum* of (6.1); we also say that (6.1) *converges to* A.

Theorem 6.3 (the n-th term test)**.** *If* (6.1) *converges, then* $a_n \to 0$.

The converse is certainly wrong, as the harmonic series shows.

Sometimes (*e.g.*, for *power series*, see §6.3 below), the summation in (6.1) starts from $n = 0$. It is clear that adding or dropping any **finite** number of terms does not affect the fact of convergence of the series, even though it would almost surely change its sum.

Spelling out the difference $A_n - A_{m-1}$, we arrive at the following restatement of completeness (*cf.* Definition 3.1).

Theorem 6.4 (Cauchy criterion)**.** *In a* **Banach** *space Y, a series (6.1) converges if and only if*

$$\forall \varepsilon > 0 \, \exists N \in N \colon m, n > N \Longrightarrow \left| \sum_{k=m}^{n} a_k \right| < \varepsilon.$$

A series (6.1) is said to *converge absolutely* (note that this is a single term which, for the moment, does not imply the convergence *per se*) if the series

$$\sum_{n=1}^{\infty} |a_n| \tag{6.5}$$

converges. Note that (6.5) is a series of positive real numbers, for which there are a great deal of convergence tests found in calculus textbooks and beyond.

Theorem 6.6. *In a* **Banach** *space Y, an absolutely convergent series converges.*

Proof. The beautiful but limited to $\mathbb{R}$ proof found in calculus textbooks does not work. We merely observe that

$$\left| \sum_{k=m}^{n} a_k \right| \leqslant \sum_{k=m}^{n} |a_k|$$

and apply Theorem 6.4 to both (6.5) and (6.1). □

The converse of Theorem 6.6 is not true, the best-known example being the alternating harmonic series.

Thus, in a Banach space Y, we have the following trichotomy:

- *absolute convergence*: both (6.1) and (6.5) converge;
- *conditional convergence*: (6.1) converges, but (6.5) diverges;
- *divergence*: both (6.1) and (6.5) diverge.

Proofs of most statements below are based on Theorem 6.4 and on the fact that the partial sums of (6.5) constitute an increasing sequence of real numbers.

For (6.5) there are but two options: it either converges of diverges to $+\infty$. The following restatement is often handy in establishing the convergence.

Theorem 6.7. *A series* (6.1) *is absolutely convergent if and only if the sequence*

$$\sum_{k=1}^{n} |a_k|$$

of partial sums of (6.5) *is bounded.*

Corollary 6.8. *In a* **Banach** *space Y, if a series* (6.1) *converges absolutely, then so does any subseries.*

Corollary 6.9. *In a* **Banach** *space Y, the terms of an absolutely convergent series can be shuffled: the resulting series still converges absolutely to the same sum.*

It is worth emphasizing that neither Corollary 6.8 nor Corollary 6.9 hold for conditionally convergent series. Most calculus textbooks explain that, in $\mathbb{R}$, the terms of a conditionally convergent series can be shuffled to achieve any sum.

Clearly, series commute with linear operations:

$$\sum_{n=1}^{\infty} (\alpha a_n + \beta b_n) = \alpha \sum_{n=1}^{\infty} a_n + \beta \sum_{n=1}^{\infty} b_n.$$

As usual, we implicitly state here that, if both series in the right hand side are (absolutely) convergent, then so is the left hand side.

Products are slightly less straightforward. Assume that Y is a Banach algebra (*i.e.*, we can also do products, which are bilinear and commute with limits; you can think that $Y = \mathbb{R}$ or $\mathbb{C}$, *cf.* also Example 1.17). The product of two series

$$\sum_{n=0}^{\infty} a_n, \quad \sum_{n=0}^{\infty} b_n \tag{6.10}$$

(it is more convenient to start from $n = 0$) is defined as

$$\sum_{n=0}^{\infty} (a_n b_0 + a_1 b_{n-1} + \ldots + a_{n-1} b_1 + a_0 b_n). \tag{6.11}$$

In other words, we declare a_n and b_n to be the terms of "degree n"; then, we do the "usual" each-by-each multiplication of the two "sums" and collect similar terms of the same "degree." This convention is particularly meaningful for power series.

Theorem 6.12 (see Appendix D)**.** *In a* **Banach** *space Y, if both* (6.10) *converge* **absolutely** *to A and B, respectively, then* (6.11) *converges absolutely to AB.*

6.2 Functional series

Everything said above about "numeric" series applies literally to series whose terms are maps: we consider

$$\sum_{n=1}^{\infty} f_n, \qquad f_n \colon X \to Y, \quad \text{and} \tag{6.13}$$

$$\sum_{n=1}^{\infty} |f_n|, \qquad |f_n| \colon X \to \mathbb{R}, \quad x \mapsto |f_n(x)|, \tag{6.14}$$

and the (pointwise) convergence of (6.14) is referred to as the (pointwise) *absolute convergence* of (6.13). The novelty is the concept of uniform convergence (see §5), which is applied to the sequences of partial sums.

When the two are combined, we arrive at a slightly ambiguous term "uniform absolute (or absolute uniform) convergence." Since the principal source of "good" series is Theorem 6.15 below, we agree that this terms means that (6.14) (and hence also (6.13), assuming Y complete) converges uniformly.

Theorem 6.15 (Weierstraß M-test)**.** *Assume that Y is complete and that there is a* **convergent** *numeric series $\sum M_n$, $M_n \in \mathbb{R}$, such that*

$$|f_n(x)| \leqslant M_n$$

for all $x \in X$ and $n \geqslant 1$. Then (6.14) and, hence, (6.13) converge uniformly.

As already mentioned, there is a great deal of tests to establish the convergence of a numeric series $\sum M_n$ with positive real term.

Proof. Observe that

$$\sum_{k=m}^{n} |f_n(x)| \leqslant \sum_{k=m}^{n} M_n$$

for all x and apply Theorem 6.4 to $\sum M_n$ first, and then to (6.14). $\square$

Everything said in §5.1 applies to uniformly convergent series:

- if all f_n are continuous, so is the sum,
- whenever appropriate, the series can be integrated termwise,
- the product of two uniformly absolutely convergent series is also uniformly absolutely convergent (in Appendix D, we prepend $\forall x$ to all bounds),

etc. In general, a series cannot be differentiated!

6.3 Power series

In this section we assume that $X = Y = \mathbb{C}$. A *power series* (*centered* at $a \in \mathbb{C}$ and with *coefficients* $c_n \in \mathbb{C}$) is a functional series of the form

$$\sum_{n=0}^{\infty} c_n(z-a)^n. \tag{6.16}$$

Theorem 6.17 (disk of convergence)**.** *If* (6.16) *converges at a point* $z_0 \in \mathbb{C}$*, it converges absolutely uniformly on any disk* $D_r := \big\{|z-a| \leqslant r\big\}$*,* $r < |z_0 - a|$*.*

Proof. By Theorem 6.3, the convergence at $z = z_0$ implies $c_n(z_0-a)^n \to 0$. In fact, for the conclusion of the theorem, it suffices to assume that this sequence is bounded, $|c_n||z_0-a|^n \leqslant M$. Then,

$$|c_n(z-a)^n| = |c_n||z_0-a|^n \left|\frac{z-a}{z_0-a}\right|^n \leqslant Mq^n, \quad q := \frac{r}{|z_0-a|} < 1,$$

for all $z \in D_r$, and Theorem 6.15 applies. □

The standard implication of this theorem is the assertion that each series (6.16) has a certain *radius of convergence* $R \in [0, \infty]$ with the following properties:

1. (6.16) converges absolutely uniformly **on compacta** on the **open** *disk of convergence* $D := \{|z-a| < R\}$ (if $R > 0$);
2. (6.16) diverges outside the closed disk, for $|z-a| > R$ (if $R < \infty$);
3. the sum of (6.16) is a continuous function $f\colon D \to \mathbb{C}$;
4. (6.16) can be integrated termwise along any (compact) curve in D;
5. any two power series with a common center can be added, subtracted, and multiplied inside the intersection of their disks of convergence. The radius of convergence of the result is *at least* $\min\{R_1, R_2\}$.

Remark 6.18. For (1), one should observe, in addition, that any compact $K \subset D$ is contained in some closed disk $\bar{B}_r(a)$, $r < R$ (due to, *e.g.*, Lemma 2.10).

The behaviour of (6.16) on the boundary $\{|z-a| = R\}$ differs from point to point and may be quite erratic.

It can be (and is in some textbooks) shown that (6.16) can also be differentiated termwise (which does not follow from our general theory). The estimates needed are straightforward but tedious. We omit this part since, first, we did not consider complex differentiation at all and, second, this would follow from the general statement on sequences of holomorphic functions. Though, see Remark 6.20.

Power series can also be *divided*, provided that the coefficient c_0 of the divisor is nonzero, but the radius of convergence of the result is unpredictable (up to the closest zero of the divisor, which is not easily found in terms of the coefficients). For example, 1 and $1-z$ are power series with $R=\infty$, but their quotient

$$\frac{1}{1-z}=\sum_{n=0}^{\infty} z^n$$

(the geometric series) has radius of convergence $R=1$.

Remark 6.19. It will follow from the general theory of holomorphic functions that the disk of convergence of (6.16) is the *largest disk* to which the sum admits an analytic continuation. For example, the radius of convergence of $\sum(-1)^n z^{2n}$ is 1 because its sum $1/(1+z^2)$ has poles at $\pm i$.

In terms of the coefficients, the radius of convergence is computed *via*

$$\frac{1}{R}=\limsup_{n\to\infty}\sqrt[n]{|c_n|},$$

as given by Cauchy's radical test (*aka* n-th root test). The division reduces to equating same degree terms on both sides and finding the unknown coefficients of the quotient one-by-one.

Remark 6.20. A possible quick fix for the differentiability of (6.16) would be to consider the derivative

$$\sum_{n=0}^{\infty} nc_n(z-a)^{n-1} \tag{6.21}$$

and modify the estimate in the proof of Theorem 6.17,

$$|nc_n(z-a)^{n-1}|\leqslant\frac{M}{|z_0-a|}nq^{n-1}\leqslant\tilde{M}\tilde{q}^{n-1},\qquad q<\tilde{q}<1,\quad n\gg 0,$$

to conclude that the radius of convergence of (6.21) is at least that of (6.16). Then, (6.21), as a power series, can be integrated termwise back to (6.16).

Assuming that the differentiability of any power series has been established, we conclude that the sum $f\colon D\to\mathbb{C}$ of (6.16) is infinitely differentiable and that (6.16) is nothing but the Taylor series of f:

$$c_n=\frac{f^{(n)}(a)}{n!}.$$

A Proving theorems in advanced calculus

A great deal of theorems in calculus are proved by estimating a difference,

$$|A - B| < \varepsilon, \quad \forall \varepsilon > 0, \tag{A.1}$$

under certain extra conditions ($\exists \delta > 0$, $\forall n \gg 0$, *etc*). The key to success is breaking the difference into two or more and applying the triangle inequality,

$$|A - B| \leqslant |A - C| + |C - B|,$$

so that each of the new differences either is known or can easily be shown to be small. At the end, instead of (A.1), we usually obtain $|A - B| < 2\varepsilon$, 3ε, $M\varepsilon$ (for some, possibly huge or unknown, **constant** $M \in \mathbb{R}_+$), but this does not matter, as ε is **any** positive number. (A purist would go back and insert extra constants to all intermediate inequalities used, but this hardly makes sense.)

How do we show that an individual difference $|A - C|$ is small? Usually, this either is given in the hypotheses (in the form of a certain convergence) or follows from the continuity of a map,

$$|a - c| \Rightarrow |f(a) - f(c)|,$$

modulo the smallness of $|a-c|$ which is in turn to be established at the next step of the proof (*cf.* the chain of arguments in the first paragraph of §A.1). If a universal, independent of a point $x \in X$ bound is needed (*e.g.*, to be used further in (A.4) below), the uniform versions of these notions should be employed. In these cases, the uniform convergence is usually part of the hypotheses, whereas the uniform continuity is given by Theorem 2.14, and it is our responsibility to make sure that the function in question is restricted to a compact domain (*cf.* the beginning of the proof of Theorem A.8 below).

Often, one can make use of taking out a common factor:

$$|A - C| \leqslant (\text{bounded}) \cdot (\text{small}) \tag{A.2}$$

(see, *e.g.*, (A.4) below, where, strange as it seems, the "common factor" is f); a common trick is arguing along the following line (*cf.* (A.9) below):

$$|aA - bB| \leqslant |aA - aB| + |aB - bB| \leqslant |a||A - B| + |a - b||B|. \tag{A.3}$$

Then, the smallness of the second factor in (A.2) is proved as suggested above, whereas the first one is typically bounded by Theorem 2.12, where, once again, we must make sure that the domain is compact.

An example of this common factor approach is estimating various integrals, which essentially boils down to

$$\left|\int_a^b f(x)dx\right| \leqslant \int_a^b |f(x)|\,dx \leqslant (b-a)\max_{a\leqslant x\leqslant b}|f(x)| \tag{A.4}$$

and its higher dimensional counterparts for double, triple, *etc*. integrals; similar estimates for line integrals eventually reduce to (A.4) (*cf.* §A.2).

These principles are illustrated by two proofs below. Considering that, in fact, our tools are quite limited, we seldom have much choice for each next step. Still, it should be kept in mind that there is no "algorithm" solving all mathematical problems (and advanced calculus is not a calculus in the literal meaning of the word); hence, the only way to learn how to prove theorems is to prove them!

A.1 Proof of Theorem 1.19

The implication Definition 1.15 ⇒ Definition 1.18 is a simple implementation of the guidelines above. Given a sequence $x_n \to a$, we need to prove the inequality

$$|f(x_n) - b| < \varepsilon \quad \text{(for any } \varepsilon > 0), \tag{A.5}$$

see Definition 1.10. By the hypothesis $f(x) \to b$, this is given by Definition 1.15 *provided that* $|x_n - a| < \delta$ (for some $\delta > 0$), and the latter inequality for $n \gg 0$ is given by Definition 1.10 in view of $x_n \to a$.

The converse implication Definition 1.18 ⇒ Definition 1.15 is an example of *reductio ad absurdum*, where the key is a thorough understanding of the *negation* of what we are trying to prove. We need (1.16):

$$\forall\varepsilon > 0\;\exists\delta > 0\;\forall x\colon\; 0 < |x - a| < \delta \Rightarrow |f(x) - b| < \varepsilon,$$

where the implicit quantifier $\forall x$ has been added explicitly. Following the rules of logic (most notably, changing each quantifier to the opposite), we need to assume that

$$\exists\varepsilon > 0\;\forall\delta > 0\;\exists x\colon\; \bigl(0 < |x - a| < \delta\bigr)\;\&\;\bigl(|f(x) - b| \geqslant \varepsilon\bigr). \tag{A.6}$$

Thus, we fix one offending $\varepsilon > 0$ ($\exists\varepsilon$) and, for *each* $\delta > 0$ ($\forall\delta$) find one "bad" point x ($\exists x$) such that … In practice, in calculus, it suffices to consider a sequence $\delta_n \searrow 0$ (here, we can just take $\delta_n := 1/n$), giving rise to a sequence x_n, one term for each δ_n. (Recall that finding one "bad" sequence is precisely what we need to arrive at a contradiction to the hypothesis.) Now, indeed, x_n is "bad": using for both assertions (1.12), we have

- $|x_n - a| < \delta_n \searrow 0$, hence $x_n \to a$, but
- $|f(x) - b| \geqslant \varepsilon > 0$, hence $f(x_n) \not\to b$. $\square$

Remark A.7. I would like to emphasize that this proof is *literally* straightforward (and so are most others), in the sense that, at every single step, we essentially have no choice. In the first part, we start with stating the inequality (A.5) that we need to proof. Then, step-by-step, we search for a definition/assumption that would guarantee the current inequality in question, possibly modulo some other inequality/assumption. This is repeated until the very last inequality ($n \gg 0$ in our case) is or can be assumed known.

In the second part, once (A.6) has been written down, we merely decipher this statement step-by-step and analyze the result.

A.2 Continuity of line integrals

As an example of a more "advanced" series of estimates, we prove a theorem that can be used, *e.g.*, to approximate line integrals by those parametrized by polynomials (see Theorem 5.10).

Theorem A.8. *Let $P, Q \colon U \to \mathbb{R}$ be a pair of continuous functions on an open subset $U \subset \mathbb{R}^2$, and let $\gamma_n \colon [0,1] \to U$ be a sequence of smooth* (i.e.*, continuously differentiable*) *curves such that $\gamma_n \rightrightarrows \gamma \colon [0,1] \to U$ and $\gamma_n' \rightrightarrows \gamma'$. Then,*

$$\int_{\gamma_n} Pdx + Qdy \xrightarrow[n\to\infty]{} \int_\gamma Pdx + Qdy.$$

Proof. We start with a few technicalities. First, by Theorem 5.6, γ is also a smooth curve and all integrals in the statement are well defined, so that it only remains to estimate their difference. Second, we can replace U with a compact subset $K \subset U$. Indeed, $\operatorname{Im}\gamma$ is compact by Theorem 2.13 and, by Lemma 2.11, this image has a compact neighborhood containing (due to the uniform convergence) the images of all γ_n, $n \gg 0$.

Now, we spell out the integrals (the Pdx terms only) and estimate their difference. Denote the components of γ_n and γ by (x_n, y_n) and (x, y), respectively. The $'$ stands for the derivative in t. Then, the first inequality in (A.4),

$$\left|\int_0^1 P(\gamma_n)x_n'dt - \int_0^1 P(\gamma)x'dt\right| \leqslant \int_0^1 |P(\gamma_n)x_n' - P(\gamma)x'|\,dt$$

by the triangle inequality, *cf.* (A.3), and (A.4):

$$\leqslant \max |P(\gamma_n)x'_n - P(\gamma)x'_n| + \max |P(\gamma)x'_n - P(\gamma)x'|$$

and, taking out common factors:

$$\leqslant \max |P(\gamma_n) - P(\gamma)| \cdot \max|x'_n| + \max |P(\gamma)| \cdot \max|x'_n - x'|, \quad \text{(A.9)}$$

where $\max$ is the maximum over $t \in [0,1]$, which exists due to Theorem 2.12, and we omit the length $(1-0)$ of the segment.

Now, each of the four quantities in the right hand side is estimated separately (insertion of the quantifiers is left to the reader; $\forall t \in [0,1]$ is always assumed):

- by the uniform convergence $\gamma_n \rightrightarrows \gamma$ and the uniform continuity of P on the compact K (Theorem 2.14), we have
$$n \gg 0 \Rightarrow |\gamma_n - \gamma| < \delta \Rightarrow |P(\gamma_n) - P(\gamma)| < \varepsilon;$$
- by the uniform convergence $\gamma'_n \rightrightarrows \gamma'$,
$$n \gg 0 \Rightarrow |x'_n - x'| < \varepsilon;$$
- by the continuity of $P \circ \gamma$ and Theorem 2.12,
$$|P(\gamma)| \leqslant M < \infty;$$
- by the continuity of x' (Theorem 5.6) and uniform convergence $x'_n \rightrightarrows x$,
$$n \gg 0 \Rightarrow |x'_n| \leqslant |x'| + |x'_n - x'| \leqslant N + 1 < \infty.$$

  (Warning: just the continuity of x'_n would not suffice as we need a bound independent of n. Thus, we also used Definition 5.3 with $\varepsilon = 1$.)

Combining everything, we conclude that, for any $\varepsilon > 0$, the difference of the two integrals is bounded by $(M + N + 1)\varepsilon$ for $n \gg 0$, which is good enough. Here, M, N are two constants that we do not know, but they are not important. □

The argument leading to the last bound $|x'_n| \leqslant N + 1$ may deserve a separate statement, which can as well be derived from the fact that f_n is a convergent sequence in the function space $C(K, Y)$, see §5.3.

Theorem A.10. *On a compact space K, a uniformly convergent sequence of continuous maps $f_n\colon K \to Y$ is uniformly bounded.*

B Proof of Theorem 1.8

This theorem depends on quite a few interdependent subsequent results culminating in Theorem 2.5, and we start with observing that all these results do hold for one favorite norm, *e.g.*, L^1 (see Example 1.5). Furthermore, as the name suggests, equivalence of norms is an equivalence relation and, therefore, it suffices to show that **a** norm $\|\cdot\|$ is equivalent to the L^1-norm $|\cdot|$. We reserve this notation $\|\cdot\|$ and $|\cdot|$ till the rest of this section.

The next lemma holds in any normed space.

Lemma B.1. *For any norm $\|\cdot\|$ on X and any pair $x, y \in X$ we have*

$$\big|\|x\| - \|y\|\big| \leqslant \|x - y\|.$$

Proof. By the triangle inequality and symmetry we have

$$\begin{aligned} \|x\| &= \|y + (x - y)\| \leqslant \|y\| + \|x - y\|, \\ \|y\| &= \|x + (y - x)\| \leqslant \|x\| + \|x - y\|. \end{aligned}$$

Combining the two, we get $-\|x - y\| \leqslant \|x\| - \|y\| \leqslant \|x - y\|$. □

Lemma B.2. *Any norm $\|\cdot\|$ on $\mathbb{R}^n$ is continuous with respect to $|\cdot|$.*

Proof. Pick a basis $e_1, \ldots, e_n$ for $\mathbb{R}^n$ and let $x = \sum \alpha_k e_k$ and $y = \sum \beta_k e_k$. Then, by the triangle inequality and homogeneity,

$$\|x - y\| = \Big\|\sum(\alpha_k - \beta_k)e_k\Big\| \leqslant \sum |\alpha_k - \beta_k|\|e_k\| \leqslant M|x - y|,$$

where $M := \max_k \|e_k\|$. This implies uniform continuity of $\|\cdot\|$. □

Corollary B.3 (of Lemma B.2, Theorem 2.5, and Theorem 2.12)**.** *On the L^1-unit sphere $S^{n-1} := \big\{|x| = 1\big\}$ any norm $\|\cdot\|$ takes its maximal value C and minimal value $c > 0$. (Here, $c > 0$ since it is the* value *of $\|\cdot\|$ on a nonzero vector.)*

Finally, for any $0 \neq x \in \mathbb{R}^n$ we have $x/|x| \in S^{n-1}$. Hence, by Corollary B.3,

$$c \leqslant \left\|\frac{x}{|x|}\right\| \leqslant C,$$

and it remains to apply the homogeneity of $\|\cdot\|$. □

C Proof of Theorem 2.4

The "learned" proof of the theorem goes as follows:

- any metric space is *first countable* (hence, Definition 1.26 works);
- a first countable compact space is sequentially compact

for the implication (compact) $\Rightarrow$ (sequentially compact), and

- a sequentially compact metric space is *separable*;
- a separable metric space is *second countable*;
- a second countable space is *Lindelöf*;
- a Lindelöf sequentially compact space is compact

for the opposite implication. Now, I will try to outline these steps without using the scary words.

Assume that X is compact, pick a sequence x_n, and consider $A := \{x_n\}$ as a subset of X. If x_n has no converging subsequences, then A is closed (as there is nothing to be checked in Definition 1.26), and the same applies to any subset of A. Hence, A is discrete and, by Corollary 2.7, A is finite; thus, x_n has constant subsequences, which obviously converge.

Warning C.1. Do not be mislead: this proof does use the fact that X is metric, as otherwise Definition 1.26 does not apply and closed sets may not be detectable in terms of limits of sequences.

Lemma C.2. *If X is a sequentially compact metric space, then, for each $\varepsilon > 0$, there is a* **finite** *ε-grid,* i.e., *finite subset $G_\varepsilon \subset X$ such that, for each $x \in X$, there is $g \in G_\varepsilon$ with $|x - g| < \varepsilon$.*

Proof. Pick any $g_1 \in X$. If $\{g_1\}$ is an ε-grid, we stop; otherwise, there is $g_2 \in X$ with $|g_2 - g_1| \geqslant \varepsilon$. Continuing in this manner, we construct points $g_1, g_2, \ldots \in X$ such that $|g_n - g_m| \geqslant \varepsilon$ whenever $n \neq m$. This process must terminate, as otherwise we would obtain a sequence g_n without converging subsequences. $\square$

Corollary C.3. *Any sequentially compact metric space X has a* **countable** *dense set,* i.e., *subset $G \subset X$ that intersects any nonempty open set.*

Proof. Just take $G := \bigcup_{n \geqslant 1} G_{1/n}$ as in Lemma C.2: this set intersects any open ball of positive radius, hence, by Definition 1.21, any nonempty open set. $\square$

Pick a set $G \subset X$ as in Corollary C.3 and consider the collection

$$\mathcal{G} := \{B_r(g) \mid g \in G,\ r \in \mathbb{Q}_+\}.$$

This **countable** collection has the property that

$$\text{any open set } U \subset X \text{ is the union of some of the sets } V \in \mathcal{G}. \tag{C.4}$$

Indeed, if U is open and $a \in U$, there is an open ball $B_\varepsilon(a) \subset U$. Pick an integer $n > 1/2\varepsilon$. By the assumption, there is a common point $g \in G \cap B_{1/n}(a)$; then, clearly, $a \in B_{1/n}(g) \subset B_\varepsilon(a) \subset U$. Thus, for each $a \in U$, there is $V_a \in \mathcal{G}$ such that $a \in V_a \subset U$. Then,

$$U = \bigcup_{a \in U} V_a. \tag{C.5}$$

Remark C.6. The last equation is a very common trick to represent a set U as a union of subsets with certain desired properties: do it for a single point $a \in U$ first, so that $a \in V_a \subset U$, and then apply (C.5). It is this trick that is also used in the proof of Theorem 5.11. It is worth remembering.

Lemma C.7. *If a space X admits a countable collection $\mathcal{G}$ as in* (C.4)*, then any open covering of X has a* **countable** *subcovering.*

Proof. Consider an open covering $\mathcal{U} := \{U_\alpha \mid \alpha \in S\}$. Each set U_α is a union of some $V_{\alpha\beta} \in \mathcal{G}$, and clearly all these $V_{\alpha\beta}$ involved also cover X. On the other hand, since $\mathcal{G}$ itself is countable, in the collection $\{V_{\alpha\beta}\}$ there are but countably many distinct sets. Picking one $U \in \mathcal{U}$ for each $V_{\alpha\beta}$ involved, $U \supset V_{\alpha\beta}$, we obtain a countable subcollection of sets $U \in \mathcal{U}$ that cover X. □

Now, assume that a metric space X is sequentially compact, consider an open covering of X, and apply the statements above to find a countable subcovering $U_1 \cup U_2 \cup \ldots = X$. Assuming that no finite union of these sets U_n equals X, for each $n \geqslant 1$ we can find a point $x_n \in X \smallsetminus (U_1 \cup \ldots \cup U_n)$. We assert that the sequence x_n thus obtained has no converging subsequence, contradicting to the assumption. Indeed, each point $a \in X$ (the prospective limit of a subsequence) lies in at least one of U_n. If a subsequence of x_n converged to a, this neighborhood $U_n \ni a$ would have to contain infinitely many members of the sequence, see (1.22). On the other hand, $x_k \notin U_n$ for all $k \geqslant n$. □

D Proof of Theorem 6.12

Trivial as it seems, the statement is quite involved. Given a_n and b_n, we consider the *double series*

$$\sum_{m,n=0}^{\infty} a_m b_n :$$

$$\begin{array}{ccccc}
a_0b_0 & a_1b_0 & a_2b_0 & a_3b_0 & \dots \\
a_0b_1 & a_1b_1 & a_2b_1 & a_3b_1 & \dots \\
a_0b_2 & a_1b_2 & a_2b_2 & a_3b_2 & \dots \\
a_0b_3 & a_1b_3 & a_2b_3 & a_3b_3 & \dots \\
\vdots & \vdots & \vdots & \vdots &
\end{array}$$

To give this gadget a meaning we need to fix the *order of summation*. Definition (6.11) suggests the summation along the *diagonals* $\{m + n = N\}$,

$$(a_0b_0) + (a_1b_0 + a_0b_1) + (a_2b_0 + a_1b_1 + a_0b_2) + \dots, \tag{D.1}$$

whereas what we can try to control is the summation along the *principal minors* (squares with one vertex at a_0b_0) $\{0 \leqslant m, n \leqslant N\}$:

$$(a_0b_0) + (a_1b_0 + a_1b_1 + a_0b_1) + (a_2b_0 + a_1b_1 + a_2b_2 + a_1b_2 + a_0b_2) + \dots \tag{D.2}$$

Denote the partial sums of $\sum a_n$, $\sum b_n$, and (D.2) (without grouping of the terms) by A_n, B_n, and P_n, respectively, and let A'_n, B'_n, P'_n be the partial sums of the respective series (6.5) of norms. For the *subsequence* P'_{N^2}, we have

$$P'_{N^2} = \sum_{m,n=0}^{N} |a_m b_n| \leqslant \sum_{m,n=0}^{N} |a_n||b_n| = A'_N B'_N \to A'B' < \infty.$$

In view of Theorem 6.7, in this particular case (an increasing sequence of real numbers), the convergence of a subsequence implies the convergence of the whole sequence P'_n. Thus, (D.2) is absolutely convergent, and to compute its sum we can use the same trick, confining ourselves to the subsequence

$$P_{N^2} = \sum_{m,n=0}^{N} a_m b_n = \sum_{m=0}^{N} a_m \sum_{n=0}^{N} b_n = A_N B_N \to AB.$$

Finally, due to Corollary 6.9, we can reshuffle (and then group) the terms and conclude that (D.1) also converges to the same sum AB. $\square$

Remark D.3. The assumption that both series should converge absolutely is not a mere technicality used in the proof. Consider the alternating p-series

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

by Leibniz's test, it is conditionally convergent. The n-th term of the product of this series by itself is

$$c_n = (-1)^n \sum_{p+q=n} \frac{1}{\sqrt{pq}}.$$

If $n = 2k$ is even, the smallest term in this sum is $1/\sqrt{k^2} = 1/k$ in the middle. Since the number of terms is $2k - 1$,

$$|c_{2k}| \geqslant \frac{2k-1}{k} \underset{k\to\infty}{\longrightarrow} 2 \neq 0$$

and the series cannot converge in view of Theorem 6.3.