Qiuzhen Lectures on Analysis

BIN GUI

December 6, 2023

1 Basic set theory and numbers

In this chapter, we discuss informally some of the basic notions in set theory and basic properties about numbers. A more thorough treatment can be found in [Mun, Ch. 1] (for set theory) and [Rud-P, Ch. 1] (for numbers).

Let me first list the notations and conventions that will be used throughout the notes. We use frequently the abbreviations:

$$\text{iff=if and only if}$$
$$\text{LHS=left hand side} \qquad \text{RHS=reft hand side}$$
$$\exists\text{=there exists} \qquad \forall\text{=for all}$$
$$\text{i.e.=id est=that is=namely} \qquad \text{e.g.=for example}$$
$$\text{cf.=compare/check/see/you are referred to}$$
$$\text{resp.=respectively} \qquad \text{WLOG=without loss of generality}$$

If P, Q are properties, then

$$P \wedge Q = P \text{ and } Q \qquad P \vee Q = P \text{ or } Q \qquad \neg P = \text{ Not } P$$

When we write $A := B$ or $A \overset{\text{def}}{=\!=} B$, we mean that A is defined by the expression B. When we write $A \equiv B$, we mean that A are B are different symbols of the same object.

If $\mathbb{F}$ is any field (e.g. $\mathbb{Q}, \mathbb{R}, \mathbb{C}$), we let $\mathbb{F}^{\times} = \mathbb{F}\backslash\{0\}$. If α is a complex number and $n \in \mathbb{N}$, we define the **binomial coefficient**

$$\binom{\alpha}{n} = \begin{cases} \dfrac{\alpha \cdot (\alpha - 1) \cdots (\alpha - n + 1)}{n!} & \text{if } n \geqslant 1 \\ 1 & \text{if } n = 0 \end{cases} \tag{1.1}$$

where $n! = n(n-1)(n-2) \cdots 2 \cdot 1$ and $0! = 1$. The bold letter $\mathbf{i}$ means

$$\mathbf{i} = \sqrt{-1}$$

If $z = a + b\mathbf{i}$ where $a, b \in \mathbb{R}$, we let

$$\mathrm{Re}(z) = a \qquad \mathrm{Im}(z) = b$$

If x, y are two elements, we let

$$\delta_{x,y} = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases} \tag{1.2}$$

Topics marked with $\star\star$ are technical and/or their methods are rarely used in later studies. Topics marked with $\star$ are interesting, but not necessarily technical or difficult. They are not essential for understanding the rest of the notes. You can skim or skip the starred topics on first reading. When a chapter/section/subsection is starred, it means that all of the material in that chapter/section/subsection is starred.

1.1 Basic operations and axioms

Intuitively, a set denotes a collection of elements. For instance:

$$\mathbb{Z} = \{\text{all integers}\} \qquad \mathbb{N} = \mathbb{Z}_{\geqslant 0} = \{n \in \mathbb{Z} : n \geqslant 0\} \qquad \mathbb{Z}_+ = \{n \in \mathbb{Z} : n > 0\}$$

have infinitely many elements. (In this course, we will not be concerned with the rigorous construction of natural numbers and integers from Peano axioms.) We also let

$$\mathbb{Q} = \{\text{all rational numbers}\} \qquad \mathbb{R} = \{\text{all real numbers}\}$$

if we that rational and real numbers exist and satisfy the properties we are familiar with in high school mathematics. (We will construct $\mathbb{Q}$ and $\mathbb{R}$ rigorously, by the way.)

Set theory is the foundation of modern mathematics. It consists of several Axioms telling us what we can do about the sets. For example, the following way of describing sets

$$\{x : x \text{ satisfies property...}\} \tag{1.3}$$

is illegal, since it gives **Russell's paradox**: Consider

$$S = \{A : A \text{ is a set and } A \notin A\} \tag{1.4}$$

If S were a set, then $S \in S \Rightarrow S \notin S$ and $S \notin S \Rightarrow S \in S$. This is something every mathematician doesn't want to happen.

Instead, the following way of defining sets is legitimate:

$$\{x \in X : x \text{ satisfies property} \dots\} \tag{1.5}$$

where X *is a given set*. For instance, we can define the **difference** of two sets:

$$A \backslash B = A - B = \{x \in A : x \notin B\}$$

So let us figure out the legal way of defining unions and intersections of sets. The crucial point is that we assume the following axiom:

Axiom. If $\mathscr{A}$ is a set of sets, then there exists a set X such that $A \subset X$ for all $A \in \mathscr{A}$.

Thus, if $\mathscr{A}$ is a set of sets, let X satisfy $A \subset X$ for all $A \in \mathscr{A}$, then we can define the **union** and the **intersection**

$$\bigcup_{A \in \mathscr{A}} A = \{x \in X : \text{there exists } A \in \mathscr{A} \text{ such that } x \in A\} \tag{1.6a}$$

$$\bigcap_{A \in \mathscr{A}} A = \{x \in X : \text{for all } A \in \mathscr{A} \text{ we have } x \in A\} \tag{1.6b}$$

It is clear that this definition does not rely on the particular choice of X.

Remark 1.1. In many textbooks, it is not uncommon that sets are defined as in (1.3). You should interpret such definition as (1.5), where the set X is omitted because it is clear from the context. For instance, if the context is clear, the set $\{x \in \mathbb{R} : x \geqslant 0\}$ could be simply written as $\{x : x \geqslant 0\}$ or even $\{x \geqslant 0\}$. By the same token, the phrase "$\in X$" in (1.6) could be omitted. So we can also write

$$A \cup B = \{x : x \in A \text{ or } x \in B\} \qquad A \cap B = \{x : x \in A \text{ and } x \in B\}$$

which are special cases of (1.6).

Remark 1.2. In the same spirit, when discussing subsets of a given "large" set X, and if X is clear from the context, we shall write $X \backslash A$ (where $A \subset X$) as A^c and call it the **complement** of A.

Example 1.3. We have

$$\bigcup_{x \in (1, +\infty)} [0, x) = [0, +\infty) \qquad \bigcap_{n \in \mathbb{Z}_+} (0, 1 + 1/n) = (0, 1] \qquad \bigcup_{n \in \mathbb{N}} (0, 1 - 1/n] = (0, 1)$$

The readers may notice that these examples are not exactly in the form (1.6). They are actually unions and intersections of indexed families of sets. (See Def. 1.10.) We need some preparation before discussing this notion.

Axiom. If $A_1, \ldots, A_n$ are sets, their **Cartesian product** exists:

$$A_1 \times \cdots \times A_n = \{(a_1, \ldots, a_n) : a_i \in A_i \text{ for all } 1 \leqslant i \leqslant n\}$$

where two elements $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$ of the Cartesian product are regarded equal iff $a_1 = b_1, \ldots, a_n = b_n$. We also write

$$(a_1, \ldots, a_n) = a_1 \times \cdots \times a_n$$

especially when a, b are real numbers and (a, b) can mean an open interval. We understand $A_1 \times \cdots \times A_n$ as $\varnothing$ if some A_i is $\varnothing$.

If $A_1 = \cdots = A_n = A$, we write the Cartesian product as A^n. $\qquad\square$

Example 1.4. Assume that the set of real numbers $\mathbb{R}$ exists. Then the set of complex numbers $\mathbb{C}$ is defined to be $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ as a set. We write $(a, b) \in \mathbb{C}$ as $a + b\mathbf{i}$ where $a, b \in \mathbb{R}$. Define

$$(a + b\mathbf{i}) + (c + d\mathbf{i}) = (a + c) + (b + d)\mathbf{i}$$
$$(a + b\mathbf{i}) \cdot (c + d\mathbf{i}) = (ac - bd) + (ad + bc)\mathbf{i}$$

Define the zero element 0 of $\mathbb{C}$ to be $0 + 0\mathbf{i}$. More generally, we consider $\mathbb{R}$ as a subset of $\mathbb{C}$ by viewing $a \in \mathbb{R}$ as $a + 0\mathbf{i} \in \mathbb{C}$. This defines the usual arithmetic of complex numbers.

If $z = a + b\mathbf{i}$, we define its **absolute value** $|z| = \sqrt{a^2 + b^2}$. Then $z = 0$ iff $|z| = 0$. We define the **(complex) conjugate** of z to be $\overline{z} = a - b\mathbf{i}$. Then $|z|^2 = z\overline{z}$.

If $z \neq 0$, then there clearly exists a unique $z^{-1} \in \mathbb{C}$ such that $zz^{-1} = z^{-1}z = 1$: $z^{-1} = |z|^{-2} \cdot \overline{z}$. Thus, using the language of modern algebra, $\mathbb{C}$ is a **field**.[1]   □

The axiom of Cartesian product allows us to define relations and functions:

Definition 1.5. If A, B are sets, a subset R of $A \times B$ is called a **relation**. For $(a, b) \in A \times B$, we write aRb iff $(x, y) \in R$. We understand "aRb" as "a is related to b through the relation R".

Definition 1.6. A relation f of A, B is called a **function** or a **map** (or a **mapping**), if for every $a \in A$ there is a unique $b \in B$ such that afb. In this case, we write $b = f(a)$.

When we write $f : A \to B$, we always mean that A, B are sets and f is a function from A to B. A and B are called respectively the **domain** and the **codomain** of f. (Sometimes people also use the words "source" and "target" to denote A and B.)

If $E \subset A$ and $F \subset B$, we define the **image under** f of E and the **preimage under** f of F to be

$$f(E) = \{b \in B : \exists a \in E \text{ such that } b = f(a)\}$$
$$f^{-1}(F) = \{a \in A : f(a) \in F\}.$$

$f(A)$ is simply called the **image** of f, or the **range** of f. If $b \in B$, $f^{-1}(\{b\})$ is often abbreviated to $f^{-1}(b)$. The function

$$f|_E : E \to B \qquad x \mapsto f(x)$$

is called the **restriction** of f to E.   □

The intuition behind the definition of functions is clear: we understand functions as the same as their graphs. So a subset f of the "coordinate plane" $A \times B$ is the graph of a function iff it "intersects every vertical line precisely once".

Remark 1.7. According to our definition, $\varnothing$ (as a subset of $\varnothing \times B$) is the only function from $\varnothing$ to B. (A false assumption implies any statement.) If $A \neq \varnothing$, there are no functions $A \to \varnothing$.

Definition 1.8. A function $x : \mathbb{Z}_+ \to A$ is called a **sequence in** A. We write $x(n)$ as x_n, and write this sequence as $(x_n)_{n \in \mathbb{Z}_+}$ (or simply $(x_n)_n$ or (x_n)).

---

[1]See Rem. 6.7 or [Rud-P, Def. 1.12] for the definition of fields. Rather than memorizing the full definition of fields, it is more important to keep in mind some typical (counter)examples: $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are fields. $\mathbb{Z}$ is not a field, because not every non-zero element of $\mathbb{Z}$ has an inverse. The set of quaternions $\{a+b\mathbf{i}+c\mathbf{j}+d\mathbf{k} : a, b, c, d \in \mathbb{R}\}$ is not a field because it is not commutative ($\mathbf{ij} = -\mathbf{ji} = \mathbf{k}$). The set of rational functions $P(x)/Q(x)$, where P, Q are polynomials with coefficients in $\mathbb{R}$ and $Q \neq 0$, is a field.

Many people write such a sequence as $\{x_n\}_{n \in \mathbb{Z}_+}$. We do not use this notation, since it can be confused with $\{x_n : n \in \mathbb{Z}_+\}$ (the range of the function x).

Axiom. If X is a set, then the **power set** 2^X exists, where

$$2^X = \{\text{Subsets of } X\}$$

Example 1.9. The set $2^{\{1,2,3\}}$ has 8 elements: $\varnothing$, $\{1\}$, $\{2\}$, $\{3\}$, $\{1,2\}$, $\{1,3\}$, $\{2,3\}$, $\{1,2,3\}$. Surprisingly, $8 = 2^3$. As we shall see in Exp. 1.54 and Cor. 1.57, this relationship holds more generally, which explains the terminology 2^X.

Now we are ready to define indexed families of sets.

Definition 1.10. An **indexed family of sets** $(S_i)_{i \in I}$ is defined to be a function $S : I \to 2^X$ for some sets I, X. We write $S(i)$ as S_i. (So S_i is a subset of X.) I is called the **index set**. Define

$$\bigcup_{i \in I} S_i = \bigcup_{T \in S(I)} T \qquad \bigcap_{i \in I} S_i = \bigcap_{T \in S(I)} T$$

Note that $S(I)$ is the image of the function S.

Example 1.11. In the union $\bigcup_{x \in (1,+\infty)} [0, x)$, the index set is $I = (1, +\infty)$, and X can be the set of real numbers $\mathbb{R}$. Then $S : I \to 2^X$ is defined to be $S_i = S(i) = [0, i)$.

Exercise 1.12. Let $f : A \to B$ be a function. We say that f is **injective** if for all $a_1, a_2 \in A$ satisfying $a_1 \neq a_2$ we have $f(a_1) \neq f(a_2)$. We say that f is **surjective** if for each $b \in B$ we have $f^{-1}(b) \neq \varnothing$. f is called **bijective** if it is both surjective and bijective. Define the **identity maps** $\mathrm{id}_A : A \to A, a \mapsto a$ and id_B in a similar way. Prove that

$$f \text{ is injective} \iff \text{there is } g : B \to A \text{ such that } g \circ f = \mathrm{id}_A \qquad (1.7a)$$
$$f \text{ is surjective} \iff \text{there is } g : B \to A \text{ such that } f \circ g = \mathrm{id}_B \qquad (1.7b)$$
$$f \text{ is bijective} \iff \text{there is } g : B \to A \text{ such that } g \circ f = \mathrm{id}_A \text{ and } f \circ g = \mathrm{id}_B \quad (1.7c)$$

Show that the g in (1.7a) (resp. (1.7b), (1.7c)) is surjective (resp. injective, bijective).

The equivalence (1.7b) is subtler, since its proof requires Axiom of Choice.

Axiom. Let $(S_i)_{i \in I}$ be an indexed family of sets. The **Axiom of Choice** asserts that if $S_i \neq \varnothing$ for all $i \in I$, then there exists a function (the **choice function**)

$$f : I \to \bigcup_{i \in I} S_i$$

such that $f(i) \in S_i$ for each $i \in I$.

Intuitively, the axiom of choice says that for each $i \in I$ we can choose an element $f(i)$ of S_i. And such choice gives a function f.

Example 1.13. Let $f : A \to B$ be surjective. Then each member of the family $(f^{-1}(b))_{b \in B}$ is nonempty. Thus, by axiom of choice, there is a choice function g defined on the index set B such that $g(b) \in f^{-1}(b)$ for each b. Clearly $f \circ g = \mathrm{id}_B$.

Remark 1.14. Suppose that each member S_i of the family $(S_i)_{i \in I}$ has exactly one element. Then the existence of a choice function does not require Axiom of Choice: Let $X = \bigcup_{i \in I} S_i$ and define relation

$$f = \{(i, x) \in I \times X : x \in S_i\}$$

Then one checks easily that this relation between I and X is a function, and that it is the (necessarily unique) choice function of $(S_i)_{i \in I}$.

According to the above remark, one does not need Axiom of Choice to prove (1.7a) and (1.7c). Can you see why?

1.2   Partial and total orders, equivalence relations

Definition 1.15. Let A be a set. A **partial order** (or simply an **order**) $\leqslant$ on A is a relation on $A \times A$ satisfying for all $a, b, c \in A$ that:

- (Reflexivity) $a \leqslant a$.

- (Antisymmetry) If $a \leqslant b$ and $b \leqslant a$ then $a = b$.

- (Transitivity) If $a \leqslant b$ and $b \leqslant c$ then $a \leqslant c$.

We write $b \geqslant a$ iff $a \leqslant b$. Write $a > b$ iff $a \geqslant b$ and $a \neq b$. Write $a < b$ iff $b > a$. So $\geqslant$ is also an order on A. The pair $(A, \leqslant)$ is called a **partially ordered set**, or simply a **poset**. A partial order $\leqslant$ on A is called a **total order**, if for every $a, b \in A$ we have either $a \leqslant b$ or $b \leqslant a$.

Example 1.16. The following are examples of orders.

- Assume that $\mathbb{R}$ exists. Then $\mathbb{R}$ has the canonical total order, which restricts to the total order of $\mathbb{Z}$. This is the total order that everyone is familiar with.

- Let X be a set. Then $(2^X, \subseteq)$ is a poset.

- $\mathbb{R}^2$ is a poset, if we define $(a, b) \leqslant (c, d)$ to be $a \leqslant c$ and $b \leqslant d$.

Definition 1.17. A relation $\sim$ on a set A is called an **equivalence relation**, if for all $a, b, c \in A$ we have

- (Reflexivity) $a \sim a$.

- (Symmetry) $a \sim b$ iff $b \sim a$.

- (Transitivity) If $a \sim b$ and $b \sim c$ then $a \sim c$.

Later, we will use the notions of partial orders and equivalence relation not just for a set, but for a collection of objects "larger" than a set. See Sec. 1.4.

Definition 1.18. Let A be a set, together with an equivalence relation $\sim$. Define a new set

$$A/\sim = \{[a] : a \in A\}$$

where the notion $[a]$ can be understood in the following two equivalent ways (we leave it to the readers to check the equivalence):

(1) $[a]$ is a new symbol. We understand $[a]$ and $[b]$ as equal iff $a \sim b$.

(2) $[a] = \{x \in A : x \sim a\}$

We call $[a]$ the **equivalence class** (or the **residue class**) of a, and call $A/\sim$ the **quotient set** of A under $\sim$. The surjective map $\pi : a \in A \mapsto [a] \in A/\sim$ is called the **quotient map**.

Exercise 1.19. Prove that every surjective map is equivalent to a quotient map. More precisely, prove that for every surjection $f : A \to B$, there is an equivalence relation $\sim$ on A and a bijective map $\Phi : A/\sim \to B$ such that the following diagram commutes (i.e. $f = \Phi \circ \pi$):

$$
\begin{array}{ccc}
 & A & \\
\pi \swarrow & & \searrow f \\
A/\sim & \xrightarrow{\Phi} & B
\end{array}
\tag{1.8}
$$

This is the first time we see commutative diagrams. Commutative diagrams are very useful for indicating that certain maps between sets are "equivalent" or are satisfying some more general relations. For example, (1.8) shows that the maps f and π are equivalent, and that this equivalence is implemented by the bijection Φ. The formal definition of commutative diagrams is the following:

Definition 1.20. A diagram (i.e. some sets denoted by symbols, and some maps denoted by arrows) is called a **commutative diagram**, if all directed paths in the diagram with the same start and endpoints lead to the same result.

Here is an example of commutative diagram in linear algebra. This example assumes some familiarity with the basic properties of vector spaces and linear maps.[2]

Example 1.21. Let V, W be vector spaces over a field $\mathbb{F}$ with finite dimensions m, n respectively. Let $e_1, \ldots, e_m$ be a basis of V, and let $\varepsilon_1, \ldots, \varepsilon_n$ be a basis of W. We know that there are unique linear isomorphisms $\Phi : \mathbb{F}^m \xrightarrow{\simeq} V$ and $\Psi : \mathbb{F}^n \xrightarrow{\simeq} W$ such that

$$\Phi(a_1, \ldots, a_m) = a_1 e_1 + \cdots + a_m e_m \qquad \Psi(b_1, \ldots, b_n) = b_1 \varepsilon_1 + \cdots + b_n \varepsilon_n$$

Let $T : V \to W$ be a **linear map**, i.e., a map satisfying $T(a\xi + b\eta) = aT\xi + bT\eta$ for all $a, b \in \mathbb{F}, \xi, \eta \in V$. Then there is a unique $n \times m$ matrix $A \in \mathbb{F}^{n \times m}$ (viewed as a linear map $\mathbb{F}^m \to \mathbb{F}^n$ defined by matrix multiplication) such that the following diagram commutes:

$$\begin{array}{ccc} \mathbb{F}^m & \xrightarrow[\simeq]{\Phi} & V \\ {\scriptstyle A}\downarrow & & \downarrow{\scriptstyle T} \\ \mathbb{F}^n & \xrightarrow[\simeq]{\Psi} & W \end{array} \qquad (1.9)$$

namely, $T\Phi = \Psi A$. This commutative diagram tells us that T is equivalent to its **matrix representation** A under the bases $e_\bullet, \varepsilon_\star$, and that this equivalence is implemented by the linear isomorphisms Φ (on the sources) and Ψ (on the targets).

Commutative diagrams are ubiquitous in mathematics. You should learn how to read commutative diagrams and understand their intuitive meanings. We will see more examples in the future of this course.

1.3 $\mathbb{Q}, \mathbb{R}$, and $\overline{\mathbb{R}} = [-\infty, +\infty]$

Using equivalence classes, one can construct rational numbers from integers, and real numbers from rationals. We leave the latter construction to the future, and discuss the construction of rationals here.

Example 1.22 (Construction of $\mathbb{Q}$ from $\mathbb{Z}$). Define a relation on $\mathbb{Z} \times \mathbb{Z}^\times$ (where $\mathbb{Z}^\times = \mathbb{Z} \backslash \{0\}$) as follows. If $(a, b), (a', b') \in \mathbb{Z} \times \mathbb{Z}^\times$, we say $(a, b) \sim (a', b')$ iff $ab' = a'b$. It is a routine check that $\sim$ is an equivalence relation. Let $\mathbb{Q} = (\mathbb{Z} \times \mathbb{Z}^\times)/\sim$, and write the equivalence class of (a, b) as a/b or $\frac{a}{b}$. Define additions and multiplications in $\mathbb{Q}$ to be

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd} \qquad \frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}$$

---

[2]Again, we refer the readers to Internet or any Linear Algebra textbook (e.g. [Axl]) for the definition of vector spaces and linear maps.

We leave it to the readers to check that this definition is **well-defined**: If $(a, b) \sim (a', b')$ and $(c, d) \sim (c', d')$ then $(ad + bc, bd) \sim (a'd' + b'c', b'd')$ and $(ac, bd) \sim (a'c', b'd')$.

We regard $\mathbb{Z}$ as a subset of $\mathbb{Q}$ by identifying $n \in \mathbb{Z}$ with $\frac{n}{1}$. (This is possible since the map $n \in \mathbb{Z} \mapsto \frac{n}{1} \in \mathbb{Q}$ is injective.) Each $a/b \in \mathbb{Q}$ has additive inverse $\frac{-a}{b}$. If $a/b \in \mathbb{Q}$ is not zero (i.e. $(a, b) \not\sim (0, 1)$), then a/b has multiplicative inverse b/a. This makes $\mathbb{Q}$ a field: the field of **rational numbers**.

If $a/b \in \mathbb{Q}$, we say $a/b \geqslant 0$ if $ab \geqslant 0$. Check that this is well-defined (i.e., if $(a, b) \sim (a', b')$, then $ab \geqslant 0$ iff $a'b' \geqslant 0$). More generally, if $a/b, c/d \in \mathbb{Q}$, we say $\frac{a}{b} \geqslant \frac{c}{d}$ if $\frac{a}{b} - \frac{c}{d} \geqslant 0$. Check that $\geqslant$ is a total order on $\mathbb{Q}$. Check that $\mathbb{Q}$ is an Archimedean ordered field, defined below. □

Definition 1.23. A field $\mathbb{F}$, together with a total order $\leqslant$, is called an **ordered field**, if for every $a, b, c \in \mathbb{F}$ we have

- (Addition preserves $\leqslant$) If $a \leqslant b$ then $a + c \leqslant b + c$.

- (Multiplication by $\mathbb{F}_{\geqslant 0}$ preserves $\geqslant 0$) If $a, b \geqslant 0$ then $ab \geqslant 0$.

These two properties relate $\leqslant$ to $+$ and $\cdot$ respectively.

Remark 1.24. Many familiar properties about inequalities in $\mathbb{Q}$ hold for an ordered field. For instance:

$$a \geqslant b \ \wedge \ c \geqslant d \qquad \Longrightarrow \qquad a + c \geqslant b + d$$
$$a \geqslant 0 \qquad \Longleftrightarrow \qquad -a \leqslant 0$$
$$a \geqslant 0 \ \wedge \ b \geqslant c \qquad \Longleftrightarrow \qquad ab \geqslant ac$$
$$a \leqslant 0 \ \wedge \ b \geqslant c \qquad \Longleftrightarrow \qquad ab \leqslant a$$
$$a^2 \geqslant 0$$
$$0 < a \leqslant b \qquad \Longrightarrow \qquad 0 < b^{-1} \leqslant a^{-1}$$

Check them yourself, or see [Rud-P, Prop. 1.18].

Definition 1.25. We say that an ordered field $\mathbb{F}$ satisfies **Archimedean property** if for each $a, b \in \mathbb{F}$ we have

$$a > 0 \qquad \Longrightarrow \qquad \exists n \in \mathbb{N} \text{ such that } na > b$$

where na means $\underbrace{a + \cdots + a}_{n}$.

Example 1.26. $\mathbb{Q}$ satisfies Archimedean property. Indeed, let $a, b \in \mathbb{Q}$ and $a > 0$. Then $a = p/q$ and $b = r/s$ where $p, q, s \in \mathbb{Z}_+$ and $r \in \mathbb{Z}$. So $na > b$ where $n = q|r| + q$.

Prop. 1.29 gives an important application of Archimedian property. We will use this in the construction of $\mathbb{R}$ from $\mathbb{Q}$, and in the proof that $\mathbb{Q}$ is dense in $\mathbb{R}$.

Definition 1.27. Let $\mathbb{F}$ be a field. A subset $\mathbb{E} \subset \mathbb{F}$ is called a **subfield** of $\mathbb{F}$, if $\mathbb{E}$ contains the 1 of $\mathbb{F}$, and if $\mathbb{E}$ is closed under the operations of addition, multiplication, taking negative, and taking inverse in $\mathbb{F}$ (i.e. if $a, b \in \mathbb{E}$ then $a + b, ab, -a \in \mathbb{E}$, and $a^{-1} \in \mathbb{E}$ whenever $a \neq 0$). We also call $\mathbb{F}$ a **field extension** of $\mathbb{E}$, since $\mathbb{E}$ is clearly a field.

Note that if $\mathbb{E}$ is a subfield of $\mathbb{F}$, the 0 of $\mathbb{F}$ is in $\mathbb{E}$ since $0 = 1 + (-1) \in \mathbb{E}$.

Definition 1.28. Let $\mathbb{E}$ be an ordered field. A field extension $\mathbb{F}$ of $\mathbb{E}$ is called an **ordered field extension**, if $\mathbb{F}$ is equipped with a total order $\leqslant$ such that $\mathbb{F}$ is an ordered field, and if the order $\leqslant$ of $\mathbb{F}$ restricts to that of $\mathbb{E}$. We also call $\mathbb{E}$ an **ordered subfield** of $\mathbb{F}$.

Our typical example of ordered field extension will be $\mathbb{Q} \subset \mathbb{R}$.

Proposition 1.29. *Let $\mathbb{F}$ be an ordered field extension of $\mathbb{Q}$. Assume that $\mathbb{F}$ is Archimedean. Then for every $x, y \in \mathbb{F}$ satisfying $x < y$, there exists $p \in \mathbb{Q}$ such that $x < p < y$.*

Proof. Assume $x, y \in \mathbb{F}$ and $x < y$. Then $y - x > 0$ (since $y - x \neq 0$ and $-x + x \leqslant -x + y$). By Archimedean property, there exists $n \in \mathbb{Z}_+$ such that $n(y - x) > 1$. So $y - x > \dfrac{1}{n}$ and hence $x + \dfrac{1}{n} < y$.

Let us prove that the subset

$$A = \Big\{k \in \mathbb{Z} : \frac{k}{n} \leqslant x\Big\}$$

is nonempty and bounded from above in $\mathbb{Z}$. By Archimedean property, there is $m \in \mathbb{Z}_+$ such that $m > nx$, i.e. $\dfrac{m}{n} > x$. So for each $k \in \mathbb{Z}_+$ satisfying $k \geqslant m$, we have $\dfrac{k}{n} = \dfrac{m}{n} + \dfrac{k - m}{n} > x$. Therefore, for each $k \in A$ we have $k < m$. So A is bounded above. By Archimedean property again, there is $l \in \mathbb{Z}_+$ such that $\dfrac{l}{n} > -x$. So $-\dfrac{l}{n} < x$, and hence A is nonempty.

We now use the fact that *every nonempty subset of $\mathbb{Z}$ bounded above has a maximal element*. Let $k = \max A$. Since $k + 1 \notin A$, we have $x < \dfrac{k + 1}{n}$. Since $\dfrac{k}{n} \leqslant x$, we have

$$\frac{k + 1}{n} = \frac{k}{n} + \frac{1}{n} \leqslant x + \frac{1}{n} < y$$

This proves $x < p < y$ with $p = \dfrac{k + 1}{n}$. $\qquad\square$

To introduce $\mathbb{R}$ formally, we need more definitions:

Definition 1.30. Let $(X, \leqslant)$ be a poset and $E \subset X$. An **upper bound of E in X** is an element $x \in X$ satisfying $e \leqslant x$ for all $e \in E$. An upper bound $x \in X$ of E is called a **least upper bound** or a **supremum** if $x \leqslant y$ for every upper bound $y \in Y$ of E. In this case, we write the supremum as $\sup E$. It is not hard to check that supremums are unique if they exist.

We leave it to the readers to define **lower bounds** and the **greatest lower bound** (if exists) of E, also called the **infimum** and is denoted by $\inf E$. $\qquad\square$

Definition 1.31. Let $(X, \leqslant)$ be a poset. We say that X satisfies the **least-upper-bound property**, if every every nonempty subset $E \subset X$ which is bounded above (i.e. E has an upper bound) has a supremum in X. The **greatest-lower-bound property** is defined in the opposite way.

Example 1.32. $\mathbb{Z}$ satisfies the least-upper-bound and the greatest-lower-bound property: Let $A \subset \mathbb{Z}$. If A is bounded above (resp. below), then the maximum $\max A$ (resp. minimum $\min A$) exists and is the supremum (resp. infimum) of A.

Example 1.33. Let X be a set. Then $(2^X, \subset)$ satisfies the least-upper-bound and the greatest-lower-bound property: Let $\mathscr{A} \subset 2^X$, i.e., $\mathscr{A}$ is a set of subsets of X. Then $\mathscr{A}$ is bounded from above by X, and is bounded from below by $\varnothing$. Moreover,

$$\sup \mathscr{A} = \bigcup_{A \in \mathscr{A}} A \qquad \inf \mathscr{A} = \bigcap_{A \in \mathscr{A}} A$$

Theorem 1.34. *There is an ordered field extension of $\mathbb{Q}$ which is Archimedian and satisfies the least-upper-bound property. This field is denoted by $\mathbb{R}$. Its elements are called* ***real numbers****.*

Thm. 1.34 will be proved in Ch. 6. Note that by taking negative, we see that $\mathbb{R}$ also satisfies the greatest-lower-bound property.

Remark 1.35. The ordered field extensions satisfying the conditions in Thm. 1.34 are unique "up to isomorphisms". (The words "**isomorphism**" and "equivalence" are often interchangeable, though "isomorphism" is more often used in the algebraic setting, whereas "equivalence" can be used in almost every context. For example, in point-set topology, "equivalence" means "homeomorphism".) We leave it to the readers to give the precise statement. We will not use this uniqueness in this course.

Note that to compare two extensions $\mathbb{F}, \mathbb{R}$ of $\mathbb{Q}$, it is very confusing to regard $\mathbb{Q}$ as a subset of both $\mathbb{F}$ and $\mathbb{R}$. You'd better consider two different injective maps $\tau : \mathbb{Q} \to \mathbb{F}$ and $\iota : \mathbb{Q} \to \mathbb{R}$ preserving the algebraic operations and the order of $\mathbb{Q}$, and use a commutative diagram to indicate that τ and ι are equivalent. (Thus, what's happening here is that we have an equivalence of maps, not just an equivalence of the fields $\mathbb{F}$ and $\mathbb{R}$.) $\qquad\square$

Definition 1.36. Let $-\infty, +\infty$ be two different symbols, and extend the total order $\leqslant$ of $\mathbb{R}$ to the **extended real line**

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$$

by letting $-\infty < x < +\infty$ for all $x \in \mathbb{R}$. Define for each $x \in \mathbb{R}$ that

$$x \pm \infty = \pm\infty + x = \pm\infty \qquad +\infty - (-\infty) = +\infty$$

$$x \cdot (\pm\infty) = \pm\infty \cdot x = \begin{cases} \pm\infty & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ \mp\infty & \text{if } x < 0 \end{cases}$$

$$\frac{x}{\pm\infty} = 0$$

$$\frac{\pm\infty}{x} = x^{-1} \cdot (\pm\infty) \qquad \text{if } x \neq 0$$

If $a, b \in \overline{\mathbb{R}}$ and $a \leqslant b$, we define **intervals** with endpoints a, b:

$$[a, b] = \{x \in \overline{\mathbb{R}} : a \leqslant x \leqslant b\} \qquad (a, b) = \{x \in \overline{\mathbb{R}} : a < x < b\}$$
$$(a, b] = \{x \in \overline{\mathbb{R}} : a < x \leqslant b\} \qquad [a, b) = \{x \in \overline{\mathbb{R}} : a \leqslant x < b\} \qquad (1.10)$$

So $\mathbb{R} = (-\infty, +\infty)$ and $\overline{\mathbb{R}} = [-\infty, +\infty]$. If a, b are in $\mathbb{R}$, we say that the corresponding interval is **bounded**.

In this course, unless otherwise stated, an interval always means one of the four sets in (1.10). The first two intervals are called respectively a **closed interval** and an **open interval**.

Remark 1.37. Clearly, every subset E of $\overline{\mathbb{R}}$ is bounded and has a supremum and an infimum. We have that $\sup E = +\infty$ iff E is not bounded above in $\mathbb{R}$, and that $\inf E = -\infty$ iff E is not bounded below in $\mathbb{R}$.

1.4 Cardinalities, countable sets, and product spaces Y^X

Definition 1.38. Let A and B be sets. We say that A and B have the same **cardinality** and write $\mathrm{card}(A) = \mathrm{card}(B)$ (or simply $A \approx B$), if there is a bijection $f : A \to B$. We write $\mathrm{card}(A) \leqslant \mathrm{card}(B)$ (or simply $A \precsim B$) if A and a subset of B have the same cardinality.

Exercise 1.39. Show that $\mathrm{card}(A) \leqslant \mathrm{card}(B)$ iff there is an injection $f : A \to B$, iff there is a surjection $g : B \to A$. (You need either Axiom of Choice or its consequence (1.7b) to prove the last equivalence.)

It is clear that $\approx$ is an equivalence relation on the collection of sets. It is also true that $\precsim$ is a partial order: Reflexivity and transitivity are easy to show. The proof of antisymmetry is more involved:

Theorem 1.40 (Schröder-Bernstein). *Let A, B be two sets. If $A \lesssim B$ and $B \lesssim A$, then $A \approx B$.*

$\star\star$ *Proof.* Assume WLOG that $A \subset B$. Let $f : B \to A$ be an injection. Let $A_n = f^n(A)$ defined inductively by $f^0(A) = A$, $f^n(A) = f(f^{n-1}(A))$. Let $B_n = f^n(B)$. Then

$$B_0 \supset A_0 \supset \cdots \supset B_n \supset A_n \supset B_{n+1} \supset \cdots$$

In particular, $C = \bigcap_{n \in \mathbb{N}} A_n$ equals $\bigcap_{n \in \mathbb{N}} B_n$. Note that f gives a bijection $B_n \backslash A_n \to B_{n+1} \backslash A_{n+1}$ (since f gives bijections $B_n \to B_{n+1}$ and $A_n \to A_{n+1}$). Therefore, we have a bijection $g : B \to A$ defined by

$$g(x) = \begin{cases} f(x) & \text{if } x \in B_n \backslash A_n \text{ for some } n \in \mathbb{N} \\ x & \text{otherwise} \end{cases}$$

where "otherwise" means either $x \in C$ or $x \in A_n \backslash B_{n+1}$ for some n. $\qquad\square$

Intuition about the above proof: View B as an onion. The layers of B are $B_n \backslash A_n$ (the odd layers) and $A_n \backslash B_{n+1}$ (the even layers). The bijection g maps each odder layer to the subsequent odd one, and fixes the even layers and the core C.

Example 1.41. If $-\infty < a < b < +\infty$, then $(0, 1) \approx (a, b)$.

Proof. $f : (0, 1) \to (a, b)$ sending x to $(b - a)x + a$ is a bijection. $\qquad\square$

Example 1.42. If $-\infty < a < b < +\infty$, then $\mathbb{R} \approx (a, b)$

Proof. By the previous example, it suffices to prove $\mathbb{R} \approx (-1, 1)$. The function

$$f : \mathbb{R} \to (-1, 1) \qquad f(x) = \begin{cases} \frac{x}{1+x} & \text{if } x \geqslant 0 \\ -f(-x) & \text{if } x < 0 \end{cases} \tag{1.11}$$

is bijective. $\qquad\square$

Alternatively, one may use the tangent function to give a bijection between $(-\pi/2, \pi/2)$ and $\mathbb{R}$. I have avoided this method, since (1.11) is more elementary than trigonometric functions. The mathematically rigorous definition of trigonometric functions and the verification of their well-known properties are far from easy tasks.

Proposition 1.43. *Let I be an interval with endpoints $a < b$ in $\overline{\mathbb{R}}$. Then $I \approx \mathbb{R}$.*

Proof. Let $A = (0, 1) \cup \{-\infty, +\infty\}$. By Exp. 1.42, we have

$$(a, b) \subset I \lesssim \overline{\mathbb{R}} \approx A \approx [0, 1] \subset (-2, 2) \approx (a, b)$$

So $I \approx \overline{\mathbb{R}}$ by Schröder-Bernstein Thm. 1.40. In particular, $\mathbb{R} = (-\infty, +\infty) \approx \overline{\mathbb{R}}$. $\quad\square$

Definition 1.44. A set A is called **finite** if $A \precsim \{1, \ldots, n\}$ for some $n \in \mathbb{Z}_+$. A is called **countable** if $A \precsim \mathbb{N}$.

Clearly, a set A is finite iff either $A \approx \varnothing$ or $A \approx \{1, \ldots, n\}$ for some $n \in \mathbb{Z}_+$.

Remark 1.45. Let $A \subset \mathbb{N}$. If A is bounded above, then $A \subset \{0, \ldots, n\}$ and hence A is finite. If A is not bounded above, then we can construct a strictly increasing sequence $(x_n)_{n \in \mathbb{N}}$ in A. (Pick any $x_0 \in A$. Suppose we have $x_n \in A$. Since x_n is not an upper bounded of A, there is $x_{n+1} \in A$ larger than x_n. So $(x_n)_{n \in \mathbb{N}}$ can be constructed inductively.) This gives an injection $\mathbb{N} \to A$. Therefore $A \succsim \mathbb{N}$, and hence $A \approx \mathbb{N}$ by Schröder-Bernstein.

It follows that if $B \precsim \mathbb{N}$, then either B is a finite set, or $B \approx \mathbb{N}$. Therefore, "a set B is **countably infinite**" means the same as "$B \approx \mathbb{N}$". □

Theorem 1.46. *A countable union of countable sets is countable. In particular,* $\mathbb{N} \times \mathbb{N} \approx \mathbb{N}$.

Proof. Recall Exe. 1.39. Let $A_1, A_2, \ldots$ be countable sets. Since each A_i is countable, there is a surjection $f_i : \mathbb{N} \to A_i$. Thus, the map $f : \mathbb{N} \times \mathbb{N} \to \bigcup_i A_i$ defined by $f(i, j) = f_i(j)$ is surjective. Therefore, it suffices to show that there is a surjection $\mathbb{N} \to \mathbb{N} \times \mathbb{N}$. This is true, since we have a bijection $g : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ where $g(0), g(1), g(2), \ldots$ are $(0, 0)$, $(1, 0)$, $(0, 1)$, $(2, 0)$, $(1, 1)$, $(0, 2)$, $(3, 0)$, $(2, 1)$, $(1, 2)$, $(0, 3)$, etc., as shown by the figure

□

As an application, we prove the extremely important fact that $\mathbb{Q}$ is countable.

Corollary 1.47. *We have* $\mathbb{N} \approx \mathbb{Z}_+ \approx \mathbb{Z} \approx \mathbb{Q}$.

Proof. Clearly $\mathbb{Z}_{<0} \approx \mathbb{N} \approx \mathbb{Z}_+$. By Thm. 1.46, $\mathbb{Z} = \mathbb{Z}_{<0} \cup \mathbb{N}$ is countably infinite, and hence $\mathbb{Z} \approx \mathbb{N}$. It remains to prove $\mathbb{Z} \approx \mathbb{Q}$. By Schröder-Bernstein, it suffices to prove $\mathbb{Q} \precsim \mathbb{Z}$. By Thm. 1.46 again, $\mathbb{Z} \times \mathbb{Z} \approx \mathbb{Z}$. By Exp. 1.22, there is a surjection from a subset of $\mathbb{Z} \times \mathbb{Z}$ to $\mathbb{Q}$. So $\mathbb{Q} \precsim \mathbb{Z} \times \mathbb{Z} \approx \mathbb{Z}$. □

Later, when we have learned Zorn's Lemma (an equivalent form of Axiom of Choice), we will be able to prove the following generalization of $\mathbb{N} \times \mathbb{N} \approx \mathbb{N}$. So we defer the proof of the following theorem to a later section.

Theorem 1.48. *Let X be a infinite set. Then* $X \times \mathbb{N} \approx X$.

Proof. See Thm. 16.7. □

Our next goal is to prove an exponential law $a^{b+c} = a^b \cdot a^c$ for cardinalities. For that purpose, we first need to define the set-theoretic operations that correspond to the summation $b + c$ and the exponential a^b.

Definition 1.49. We write $X = \bigsqcup_{\alpha \in \mathscr{A}} A_\alpha$ and call X the **disjoint union** of $(A_\alpha)_{\alpha \in \mathscr{A}}$, if $X = \bigcup_{\alpha \in \mathscr{A}} A_\alpha$ and $(A_\alpha)_{\alpha \in \mathscr{A}}$ is a family of pairwise disjoint sets (i.e. $A_\alpha \cap A_\beta = \varnothing$ if $\alpha \neq \beta$). If moreover $\mathscr{A} = \{1, \dots, n\}$, we write $X = A_1 \sqcup \cdots \sqcup A_n$.

Definition 1.50. Let X, Y be sets. Then

$$Y^X = \{\text{functions } f : X \to Y\} \tag{1.12}$$

A more precise definition of Y^X (in the spirit of (1.5)) is $\{f \in X \times Y \mid f : X \to Y \text{ is a function}\}$. Note that by Rem. 1.7,

$$Y^\varnothing = \{\varnothing\} \tag{1.13}$$

This new notation is compatible with the old one $Y^n = Y \times \cdots \times Y$:

Example 1.51. Let $n \in \mathbb{Z}_+$. We have $Y^{\{1,\dots,n\}} \approx Y^n$ due to the bijection

$$Y^{\{1,\dots,n\}} \to Y^n \qquad f \mapsto (f(1), \dots, f(n))$$

Remark 1.52. The above example suggests that in the general case that X is not necessarily finite, we can view each function $f : X \to Y$ as $(f(x))_{x \in X}$, an **indexed family of elements** of Y with index set X. Thus, intuitively and hence not quite rigorously,

$$Y^X = \underbrace{Y \times Y \times \cdots}_{\text{card}(X) \text{ pieces}} \tag{1.14}$$

This generalizes the intuition in Def. 1.8 that a function $f : \mathbb{Z}_+ \to Y$ is equivalently a sequence $(f(1), f(2), f(3), \dots)$.

The viewpoint that Y^X is a **product space** with index set X is very important and will be adopted frequently in this course. More generally, we can define: □

Definition 1.53. Let $(X_i)_{i \in I}$ be a family of sets with index set I. Their **product space** is defined by

$$\prod_{i \in I} X_i = \{f \in \mathfrak{X}^I : f(i) \in X_i \text{ for all } i \in I\}$$

where $\mathfrak{X} = \bigcup_{i \in I} X_i$. If each X_i is nonempty, then $\prod_{i \in I} X_i$ is nonempty by Axiom of Choice. An element of $\prod_{i \in I} X_i$ is also written as $(f_i)_{i \in I}$ when the i-th component of it is $f_i \in X_i$.

In particular, if all X_i are equal to X, then $X^I = \prod_{i \in I} X$.

Example 1.54. Let X be a set. For each $A \subset X$, define the **characteristic function** $\chi_A : X \to \{0, 1\}$ to be

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

Then we have

$$2^X \approx \{0, 1\}^X$$

since the following map is bijective:

$$2^X \to \{0, 1\}^X \qquad A \mapsto \chi_A$$

Its inverse is $f \in \{0, 1\}^X \mapsto f^{-1}(1) \in 2^X$.

Proposition 1.55 (Exponential Law). *Suppose that $X = A_1 \sqcup \cdots \sqcup A_n$. Then*

$$Y^X \approx Y^{A_1} \times \cdots \times Y^{A_n}$$

Proof. We have a bijection

$$\begin{aligned} \Phi : Y^X &\to Y^{A_1} \times \cdots \times Y^{A_n} \\ f &\mapsto (f|_{A_1}, \ldots, f|_{A_n}) \end{aligned} \tag{1.15}$$

where we recall that $f|_{A_i}$ is the restriction of f to A_i. $\qquad\square$

Exercise 1.56. Assume that $A_1, \ldots, A_n$ are subsets of X. Define Φ by (1.15). Prove that Φ is injective iff $X = A_1 \cup \cdots \cup A_n$. Prove that Φ is surjective iff $A_1, \ldots, A_n$ are pairwise disjoint.

Corollary 1.57. *Let X, Y be finite sets with cardinalities $m, n \in \mathbb{N}$ respectively. Assume that $Y \neq \varnothing$. Then Y^X is a finite set with cardinality n^m.*

Proof. The special case that $m = 0$ (i.e. $X = \varnothing$, cf. (1.13)) and $m = 1$ is clear. When $m > 1$, assume WLOG that $X = \{1, \ldots, m\}$. Then $X = \{1\} \sqcup \cdots \sqcup \{m\}$. Apply Prop. 1.55 to this disjoint union. We see that $Y^X \simeq Y \times \cdots \times Y \simeq \{1, \ldots, n\}^m$ has n^m elements. $\qquad\square$

We end this section with some (in)equalities about the cardinalities of product spaces. To begin with, we write $X \precnsim Y$ (or $\mathrm{card}(X) < \mathrm{card}(Y)$) if $X \precsim Y$ and $X \not\approx Y$.

Proposition 1.58. *Let X, Y be sets with $\mathrm{card}(Y) \geqslant 2$ (i.e. Y has at least two elements). Then $X \precnsim Y^X$. In particular, $X \precnsim 2^X$.*

Proof. The case $X = \varnothing$ is obvious since $0 < 1$. So we assume $Y \neq \varnothing$. Clearly $2^X \simeq \{0, 1\}^X$ is $\precsim Y^X$. So it suffices to prove $X \precsim 2^X$. Since the map $X \to 2^X$ sending x to $\{x\}$ is injective, $X \precsim 2^X$. Let us prove $X \not\approx 2^X$.

Assume that $X \approx 2^X$. So there is a bijection $\Phi : X \to 2^X$ sending each $x \in X$ to a subset $\Phi(x)$ of X. Motivated by Russell's Paradox (1.4), we define

$$S = \{x \in X : x \notin \Phi(x)\}$$

Since Φ is surjective, there exists $y \in X$ such that $S = \Phi(y)$. If $y \in \Phi(y)$, then $y \in S$, and hence $y \notin \Phi(y)$ by the definition of S. If $y \notin \Phi(y)$, then $y \notin S$, and hence $y \in \Phi(y)$ by the definition of S. This gives a contradiction. $\qquad\square$

Remark 1.59. Write $\{1, \ldots, n\}^X$ as n^X for short. Assuming that real numbers have decimal, binary, or (more generally) base-n representations where $n \in \mathbb{Z}_{\geqslant 2}$, then $\mathbb{R} \approx n^{\mathbb{N}}$. So by Prop. 1.58, $\mathbb{N} \precnsim \mathbb{R}$, i.e. $\mathbb{R}$ *is uncountable*. The base-n representations of real numbers suggest that $\operatorname{card}(n^{\mathbb{N}})$ is independent of n. This fact can be proved by elementary methods without resorting to the analysis of real numbers:

Theorem 1.60. *Let X be an infinite set. Then*

$$2^X \approx 3^X \approx 4^X \approx \cdots \approx \mathbb{N}^X$$

Proof. First, we assume that $X = \mathbb{N}$. Clearly, for each $n \in \mathbb{Z}_{\geqslant 2}$ we have $2^X \precsim n^X \precsim \mathbb{N}^X$. Since elements of $\mathbb{N}^X$ are subsets of $X \times \mathbb{N}$ (i.e. elements of $2^{X \times \mathbb{N}}$), we have

$$\mathbb{N}^X \subset 2^{X \times \mathbb{N}} \simeq 2^X$$

since $X \times \mathbb{N} \approx X$ by Thm. 1.46. So $2^X \approx n^X \approx \mathbb{N}^X$ by Schröder-Bernstein.

As pointed out earlier (cf. Thm. 1.48), it can be proved by Zorn's Lemma that $X \times \mathbb{N} \approx X$ for every infinite set X. So the same conclusion holds for such X. $\quad\square$

2 Metric spaces

We first give an informal introduction to metric spaces, hoping to motivate the readers from a (relatively) historical perspective. It is okay if you do not understand all of the concepts mentioned in the introduction on the first read. Simply return to this section when you feel unmotivated while formally studying these concepts in later sections. (The same suggestion applies to all the introductory sections and historical comments in our notes.)

2.1 Introduction: what is point-set topology?

> The method which has been used with success by Volterra and Hilbert consists in observing that a function (for instance a continuous one) can be replaced by a countable infinity of parameters. One treats the problem first as though one had only a finite number of parameters and then one goes to the limit... We believe that this method has played a useful role because it followed intuition, but that its time has passed... The most fruitful method in functional analysis seems to us to treat the element of which the functional depends directly as a variable and in the form in which it presents itself naturally.
>
> —- Fréchet, 1925   (cf. [Jah, Sec. 13.8])

In this chapter, we begin the study of point-set topology by learning one of its most important notions: metric spaces. Similar to [Rud-P], we prefer to introduce metric spaces and basic point-set topology at the early stage of our study. An obvious reason for doing so is that metric spaces provide a uniform language for the study of basic analysis problems in $\mathbb{R}, \mathbb{R}^n, \mathbb{C}^n$, and more generally in function spaces such as the space of continuous functions $C([a, b])$ on the interval $[a, b] \subset \mathbb{R}$. With the help of such a language, for example, many useful criteria for the convergence of series in $\mathbb{R}$ and $\mathbb{C}$ (e.g. root test, ratio test) are generalized straightforwardly to criteria for the *uniform* convergence of series of functions in $C([a, b])$.

Point-set topology was born in 1906 when Fréchet defined metric spaces, motivated mainly by the study of function spaces in analysis (i.e. *functional analysis*). Indeed, point-set topology and functional analysis are the two faces of the same coin: they both originated from the study of **functionals**, i.e., functions of functions. See for example (2.1). The core ideas of point-set topology are as follows:

(1) Take X to be a set of functions defined on a "classical space" (e.g. the set of all continuous functions $f : [a, b] \to \mathbb{C}$). Then a functional is a function $S : X \to \mathbb{C}$. This is a generalization of functions on $\mathbb{R}, \mathbb{C}, \mathbb{R}^n, \mathbb{C}^n$ or on their subsets.

(2) Unlike $\mathbb{R}^n$, a function space X is usually "infinite dimensional". Thus, one may think that a functional S is a function with infinite variables. In point-set topology, this viewpoint is abandoned; the philosophy of point-set topology is diametrically opposed to that of multivariable calculus.[1] Instead, one should view a functional S as a function with one variable x, where x denotes a general point of the function space X.

(3) Rather than looking at each variable/component and doing explicit mutivariable calculations, one uses geometric intuitions to study the analytic properties of functionals.[2] These geometric intuitions (e.g. distances, open balls, convergence) are borrowed from $\mathbb{R}$ and $\mathbb{R}^n$ and are mostly irrelevant to dimensions or numbers of variables.

(Sequential) **compactness**, **completeness**, and **separability** are prominent geometric properties that are useful in the study of the analytic properties of functionals. The importance of these three notions was already recognized by Fréchet by the time he defined metric spaces. The study of these three properties will be a main theme of our course.

Consider sequential compactness for example. The application of compactness to function spaces originated from the problems in calculus of variations. For instance, let $L(x, y, z)$ be a polynomial or (more generally) a continuous function in 3 variables. We want to find a "good" (e.g. differentiable) function $f : [0, 1] \to \mathbb{R}$ minimizing or maximizing the expression

$$S(f) = \int_0^1 L(t, f(t), f'(t))dt \tag{2.1a}$$

This is the general setting of Lagrangian mechanics. In the theory of integral equations, one considers the extreme values and points of the functional

$$S(f) = \int_0^1 \int_0^1 f(x)K(x, y)\overline{f(y)}dxdy \tag{2.1b}$$

where $K : [0, 1]^2 \to \mathbb{R}$ is continuous and $f : [0, 1] \to \mathbb{C}$ is subject to the condition $\int_0^1 |f(x)|^2 dx \leqslant 1$. Any f maximizing (resp. minimizing) $S(f)$ is an eigenvector of the linear operator $g \mapsto \int_0^1 K(x, y)g(y)dy$ with maximal (resp. minimal) eigenvalue.

As we shall learn, (sequential) compactness is closely related to the problem of finding (or proving the existence of) maximal/minimal values of a continuous

---

[1]Very often, the formula of $S(f)$ involves an integral. See e.g. (2.1). Mathematicians (e.g. Volterra, Lévy, Fredholm, and early Hilbert) used to study $S(f)$ by discretizing $S(f)$, i.e., by approximating integrals by finite sums. Thus, S is approximated by a sequence of functions with n variables where $n \to \infty$. This viewpoint is abandoned in point-set topology.

[2]This is similar to linear algebra where one prefers vectors, linear subspaces, and linear operators to n-tuples, sets of solutions, and matrices.

function and the points at which the function attains its maximum/minimum. So, in 19th century, when people were already familiar with sequential compactness in $\mathbb{R}^n$ (e.g. Bolzano-Weierstrass theorem, Heine-Borel theorem), they applied compactness to function spaces and functionals. The idea is simple: Suppose we are given X, a set of functions (say continuous and differentiable) from $[a, b]$ to $\mathbb{R}$. We want to find $f \in X$ maximizing $S(f)$. Here is an explicit process (see also the proof of Lem. 3.2):

(A) Find a sequence $(f_n)_{n \in \mathbb{Z}_+}$ in X such that $S(f_n)$ increases to $M = \sup S(X)$.

(B) Define convergence in X in a suitable way, and verify that $S : X \to \mathbb{R}$ is continuous (i.e. if f_n converges to f in the way we define, then $S(f_n) \to S(f)$).

(C) Suppose we can find a subsequence $(f_{n_k})_{k \in \mathbb{Z}_+}$ converging to some $f \in X$, then S attains its maximum at f. In particular, $S(f) = M$ and hence $M < +\infty$.

To carry out step (B), we need to define suitable geometric structures for a function space X so that the convergence of sequences in X and the continuity of functions $S : X \to \mathbb{R}$ can be defined and studied in a similar pattern as that for $\mathbb{R}^n$. **Metric** (of a metric space) and **topology** (of a topological space) are such geometric structures. As we shall learn, the topology of a metric space is uniquely determined by the convergence of sequences in this space.

Step (C) can be carried out if every sequence in X has a convergent subsequent, i.e., if X is sequentially compact. Thus, we need a good criterion for sequential compactness for subsets of a function space. Arzelà-Ascoli theorem, the $C([a, b])$-version of Heine-Borel theorem, is such a criterion. This famous theorem was proved in late 19th century (and hence before the birth of point-set topology), and it gave an important motivation for Fréchet to consider metric spaces in general. We will learn this theorem at the end of the first semester.

To summarize: Metric spaces are defined not just for fun. We introduce such geometric objects because we want to study the convergence of sequences and the analytic properties of continuous functions using geometric intuitions. And moreover, the examples we are interested in are not just subsets of $\mathbb{R}^n$, but also subsets of function spaces. With this in mind, we now begin our journey into point-set topology.

2.2 Basic definitions and examples

Definition 2.1. Let X be a set. A function $d : X \times X \to \mathbb{R}_{\geqslant 0}$ is called a **metric** if for all $x, y, z \in X$ we have

(1) $d(x, y) = d(y, x)$.

(2) $d(x, y) = 0$ iff $x = y$.

(3) (Triangle inequality) $d(x, z) \leqslant d(x, y) + d(y, z)$.

The pair (X, d), or simply X, is called a **metric space**. If $x \in X$ and $r \in (0, +\infty]$,[3] the set

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

often abbreviated to $B(x, r)$, is called the **open ball** with center x and radius r. If $r \in [0, +\infty)$,

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

also abbreviated to $\overline{B}(x, r)$, is called the **closed ball** with center x and radius r.

We make some comments on this definition.

Remark 2.2. That "$d(x, y) = 0$ iff $x = y$" is very useful. Think about X as a set of functions $[0, 1] \to \mathbb{R}$ and d is a metric on X. To show that $f, g \in X$ are equal, instead of checking that infinitely many values are equal (i.e. $f(t) = g(t)$ for all $t \in \mathbb{R}$), it suffices to check that one value (i.e. $d(f, g)$) is zero.

Remark 2.3. Triangle inequality clearly implies "polygon inequality":

$$d(x_0, x_n) \leqslant \sum_{j=1}^{n} d(x_{j-1}, x_j) \tag{2.2}$$

Remark 2.4. Choose distinct points $x, y \in X$. Then x, y are separated by two open balls centered at them: there exists $r, \rho > 0$ such that $B(x, r) \cap B(y, \rho) = \varnothing$. This is called the **Hausdorff property**.

To see this fact, note that $d(x, y) > 0$. Choose r, ρ such that $r + \rho \leqslant d(x, y)$. If $z \in B(x, r) \cap B(y, \rho)$, then $d(x, z) + d(y, z) < r + \rho \leqslant d(x, y)$, contradicting triangle inequality.

We will see (cf. Prop. 2.19) that Hausdorff property guarantees that any sequence in a metric space cannot converge to two different points. Intuition: one cannot find a point which is very close to x and y at the same time. $\square$

We give some examples, and leave it to the readers to check that they satisfy the definition of metric spaces. We assume that square roots of positive real numbers can be defined. (We will rigorously define square roots after we define e^x using the series $\sum_{n \in \mathbb{N}} x^n / n!$.)

Example 2.5. $\mathbb{R}$ is a metric space if we define $d(x, y) = |x - y|$

---

[3]We want open and closed balls to be nonempty. So we assume $r \neq 0$ only for open balls.

Example 2.6. On $\mathbb{R}^n$, we can define **Euclidean metric**

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

if $x_\bullet, y_\bullet$ are the components of x, y. The following are also metrics:

$$d_1(x, y) = |x_1 - y_1| + \cdot + |x_n - y_n|$$
$$d_\infty(x, y) = \max\{|x_1 - y_1|, \ldots, |x_n - y_n|\}$$

Example 2.7. The **Euclidean metric** on $\mathbb{C}^n$ is

$$d(z, w) = \sqrt{|z_1 - w_1|^2 + \cdots + |z_n - w_n|^2}$$

which agrees with the Euclidean metric on $\mathbb{R}^{2n}$. The following are also metrics:

$$d_1(z, w) = |z_1 - w_1| + \cdot + |z_n - w_n|$$
$$d_\infty(z, w) = \max\{|z_1 - w_1|, \ldots, |z_n - w_n|\}$$

Convention 2.8. Unless otherwise stated, the metrics on $\mathbb{R}^n$ and $\mathbb{C}^n$ (and their subsets) are assumed to be the Euclidean metrics.

Remark 2.9. One may wonder what the subscripts $1, \infty$ mean. This notation is actually due to the general fact that

$$d_p(z, w) = \sqrt[p]{|z_1 - w_1|^p + \cdots + |z_n - w_n|^p}$$

is a metric where $1 \leqslant p < +\infty$, and $d_\infty = \lim_{q \to +\infty} d_q$. It is not easy to prove that d_p satisfies triangle inequality: one needs Minkowski inequality. For now, we will not use such general d_p. And we will discuss Minkowski inequality in later sections.

Example 2.10. Let $X = X_1 \times \cdots \times X_N$ where each X_i is a metric space with metric d_i. Write $x = (x_1, \ldots, x_N) \in X$ and $y = (y_1, \ldots, y_N) \in Y$. Then the following are metrics on X:

$$d(x, y) = d_1(x_1, y_1) + \cdots + d_N(x_N, y_N)$$
$$\delta(x, y) = \max\{d_1(x_1, y_1), \ldots, d_N(x_N, y_N)\}$$
$$\rho(x, y) = \sqrt{d_1(x_1, y_1)^2 + \cdots + d_N(x_N, y_N)^2}$$

With respect to the metric δ, the open balls of X are "polydisks"

$$B_X(x, r) = B_{X_1}(x_1, r) \times \cdots \times B_{X_N}(x_N, r)$$

There is no standard choice of metric on the product of metric spaces. d, δ, ρ are all good, and they are equivalent in the following sense:

Definition 2.11. We say that two metrics d_1, d_2 on a set X are **equivalent** and write $d_1 \approx d_2$, if there exist $\alpha, \beta > 0$ such that for any $x, y \in X$ we have

$$d_1(x,y) \leqslant \alpha d_2(x,y) \qquad d_2(x,y) \leqslant \beta d_1(x,y)$$

This is an equivalence relation. More generally, we may write $d_1 \lesssim d_2$ if $d_1 \leqslant \alpha d_2$ for some $\alpha > 0$. Then $d_1 \approx d_2$ iff $d_1 \lesssim d_2$ and $d_2 \lesssim d_1$.

Example 2.12. In Exp. 2.10, we have $\delta \leqslant \rho \leqslant d \leqslant N\delta$. So $\delta \approx \rho \approx d$.

Convention 2.13. Given finitely many metric spaces $X_1, \ldots, X_N$, the metric on their product space $X = X_1 \times \cdots \times X_N$ is chosen to be any one that is equivalent to the ones defined in Exp. 2.10. In the case that each X_i is a subset of $\mathbb{R}$ or $\mathbb{C}$, we follow Convention 2.8 and choose the metric on X to be the Euclidean metric (unless otherwise stated).

Definition 2.14. Let (X, d) be a metric space. Then a **metric subspace** denotes an object $(Y, d|_Y)$ where $Y \subset X$ and $d|_Y$ is the restriction of d to Y, namely, for all $y_1, y_2 \in Y$ we set

$$d|_Y(y_1, y_2) = d(y_1, y_2)$$

Convention 2.15. Suppose Y is a subset of a given metric space (X, d). Unless otherwise stated, the metric of Y is chosen to be $d|_Y$ whenever Y is viewed as a metric space. For example, the metric of any subset of $\mathbb{R}^n$ is assumed to be the Euclidean metric, unless otherwise stated.

2.3 Convergence of sequences

Definition 2.16. Let $(x_n)_{n \in \mathbb{Z}_+}$ be a sequence in a metric space X. Let $x \in X$. We say that x is the **limit** of x_n and write $\lim_{n \to \infty} x_n = x$ (or $x_n \to x$), if:

- For every real number $\varepsilon > 0$ there exists $N \in \mathbb{Z}_+$ such that for every $n \geqslant N$ we have $d(x_n, x) < \varepsilon$.

Equivalently, this means that every (nonempty) open ball centered at x contains **all but finitely many** x_n.[4]

Remark 2.17. The negation of $x_n \to x$ is clear:

- There exists $\varepsilon > 0$ such that for all $N \in \mathbb{Z}_+$ there exists $n \geqslant N$ such that $d(x_n, x) \geqslant \varepsilon$.

---

[4]"All but finitely many x_n satisfies..." means "for all but finitely many n, x_n satisfies...". It does NOT mean that "all but finitely many elements of the set $\{x_n : n \in \mathbb{Z}_+\}$ satisfies...".

Namely, one changes each "for all" to "there exists", changes each "there exists" to "for all", and negate the last sentence.

Exercise 2.18. Show that in the above definition of limits, it suffices to consider rational numbers $\varepsilon > 0$. (Note: You need to use Prop. 1.29.)

This exercise implies that the definition of limits does not require the existence of real numbers, i.e., does not assume Thm. 1.34. Indeed, we will use limits of sequences (and "double sequences") to prove Thm. 1.34.

In many textbooks and research papers, you will see phrases such as

$$x_n \text{ satisfies property } P \text{ for } \textbf{sufficiently large } n \tag{2.3}$$

This means that "there exists $N \in \mathbb{Z}_+$ such that P holds for all $n \geqslant N$". (We also say that x_n **eventually** satisfies P.) Then $\lim_{n\to\infty} x_n = x$ means that "for every $\varepsilon > 0$, we have $d(x_n, x) < \varepsilon$ for sufficiently large n".

Proposition 2.19. *Any sequence $(x_n)_{n\in\mathbb{Z}_+}$ in a metric space X has at most one limit.*

Proof. Suppose $(x_n)_{n\in\mathbb{Z}_+}$ converges to $x, y \in X$ where $x \neq y$. By Hausdorff property (Rem. 2.4), there exist $r, \rho > 0$ such that $B(x, r) \cap B(y, \rho) = \varnothing$. By the definition of $x_n \to x$, there exists $N_1 \in \mathbb{Z}_+$ such that $x_n \in B(x, r)$ for all $n \geqslant N_1$. Similarly, $x_n \to y$ means that there is $N_2 \in \mathbb{Z}_+$ such that $x_n \in B(y, \rho)$ for all $n \geqslant N_2$. Choose any $n \geqslant N_1, N_2$ (e.g. $n = \max\{N_1, N_2\}$). Then $x_n \in B(x, r) \cap B(y, \rho) = \varnothing$, impossible. $\square$

2.3.1 Methods for proving convergence and computing limits

Example 2.20. $\lim_{n\to\infty} \dfrac{1}{n} = 0$.

Proof. Choose any $\varepsilon \in \mathbb{Q}_{>0}$. By Archimedean property, there exists $N \in \mathbb{Z}_+$ such that $N\varepsilon > 1$, i.e. $1/N < \varepsilon$. Thus, for all $n \geqslant N$ we have $1/n < \varepsilon$. $\square$

Proposition 2.21. *Let $\mathbb{F} \in \{\mathbb{Q}, \mathbb{R}\}$ and $(x_n), (y_n)$ be sequences in $\mathbb{F}$ converging to $x, y \in \mathbb{R}$. If $x_n \leqslant y_n$ for all n, then $x \leqslant y$.*

Proof. If $y < x$, let $\varepsilon = x - y$. Then all but finitely many members of (x_n) are in $(x - \varepsilon/2, x + \varepsilon/2)$, and all but finitely many members of (y_n) are in $(y - \varepsilon/2, y + \varepsilon/2)$. Since $y + \varepsilon/2 < x - \varepsilon/2$, there must exist n such that $y_n < x_n$. $\square$

Definition 2.22. If A and B are posets (or more generally, preordered sets, see Def. 5.1), we say a function $f : A \to B$ is **increasing** (resp. **strictly increasing**), if for each $x, y \in A$ we have

$$x \leqslant y \qquad \Longrightarrow \qquad f(x) \leqslant f(y)$$

$$\text{resp.}$$
$$x < y \qquad \Longrightarrow \qquad f(x) < f(y)$$

We leave the definitions of **decreasing** and **strictly decreasing** to the readers. We say that f is **monotonic** (resp. **strictly monotonic**), if f is either increasing or decreasing (resp. either strictly increasing or strictly decreasing).

Proposition 2.23. *Let $(x_n)_{n \in \mathbb{Z}}$ be a sequence in $[a, b] \subset \mathbb{R}$. If (x_n) is increasing (resp. decreasing), then $\lim_{n \to \infty} x_n$ equals $\sup\{x_n : n \in \mathbb{Z}_+\}$ (resp. $\inf\{x_n : n \in \mathbb{Z}_+\}$).*

Proof. Assume (x_n) increases. (The case of decreasing is similar and hence its proof is omitted.) Let $A = \sup\{x_n : n \in \mathbb{Z}_+\} < +\infty$. Then for each $\varepsilon > 0$ there is N such that $x_N > A - \varepsilon$ (since $A - \varepsilon$ is not an upper bound). Since (x_n) is increasing, for all $n \in \mathbb{N}$ we have $A - \varepsilon < x_n \leqslant A$ and so $|x_n - A| < \varepsilon$. $\qquad\square$

Example 2.24. Let $(x_n)_{n \in \mathbb{Z}_+}$ be a sequence in a metric space X, and let $x \in X$. It is easy to see that

$$\lim_{n \to \infty} x_n = x \qquad \Longleftrightarrow \qquad \lim_{n \to \infty} d(x_n, x) = 0$$

Example 2.25. Suppose that (a_n) and (b_n) are sequences in $\mathbb{R}_{\geqslant 0}$, that $a_n \leqslant b_n$ for all n, and that $b_n \to 0$. Then $a_n \to 0$.

Proof. For each $\varepsilon > 0$, $[0, \varepsilon)$ contains all but finitely many b_n, and hence all but finitely many a_n. $\qquad\square$

More generally, we have:

Proposition 2.26 (Squeeze theorem). *Suppose that (x_n) is a sequence in a metric space X. Let $x \in X$. Suppose that there is a sequence (a_n) in $\mathbb{R}_{\geqslant 0}$ such that $\lim_{n \to \infty} a_n = 0$ and that $d(x_n, x) \leqslant a_n$ for all n. Then $\lim_{n \to \infty} x_n = x$.*

Proof. This follows immediately from Exp. 2.24 and 2.25. $\qquad\square$

The above proposition explains why many people say that "analysis is the art of inequalities": It transforms the problem of convergence to the problem of finding a sequence $(a_n) \in \mathbb{R}_{\geqslant 0}$ converging to 0 such that the inequality $d(x_n, x) \leqslant a_n$ holds. And very often, a good (hard) analyst is one who knows how to find such good sequences!

Proposition 2.27. *Let $X = X_1 \times \cdots \times X_N$ be a product of metric spaces (X_i, d_i). Let d be any of the three metrics of X in Exp. 2.10. Let $\mathbf{x}_n = (x_{1,n}, \ldots, x_{N,n})$ be a sequence in X. Let $\mathbf{y} = (y_1, \ldots, y_N)$. Then*

$$\lim_{n \to \infty} \mathbf{x}_n = \mathbf{y} \qquad \Longleftrightarrow \qquad \lim_{n \to \infty} x_{i,n} = y_i \quad (\forall 1 \leqslant i \leqslant N)$$

Proof. We let d be the metric δ in Exp. 2.10, i.e. defined by $\max_j d_j(x_j, y_j)$. Now choose a sequence $(\mathbf{x}_n)$ and an element $\mathbf{y}$ in X. Then

$$\mathbf{x}_n \to \mathbf{y} \iff d_X(\mathbf{x}_n, \mathbf{y}) \to 0 \iff \max_{1 \leqslant j \leqslant N} d_j(x_{j,n}, y_j) \to 0 \tag{2.4}$$

Suppose that the RHS of (2.4) is true. Fix any $1 \leqslant i \leqslant N$. Then $d_i(x_{i,n}, y_i) \leqslant \max_j d_j(x_{j,n}, y_j)$. So $x_{i,n} \to y_i$ by Prop. 2.26.

Conversely, assume that for every i we have $x_{i,n} \to y_i$. Then for every $\varepsilon > 0$ there is $K_i \in \mathbb{Z}_+$ such that $d_i(x_{i,n}, y_i) < \varepsilon$ for every $n \geqslant K_i$. Let $K = \max\{K_1, \ldots, K_N\}$. Then for all $n \geqslant K$ we have $\max_j d_j(x_{j,n}, y_j) < \varepsilon$. This proves the RHS of (2.4).

If d is one of the other two metrics in Exp. 2.10, one can either use a similar argument, or use the following important (but easy) fact. $\square$

Proposition 2.28. *Let d, δ be two equivalent metrics on a set X. Let $(x_n)_{n \in \mathbb{Z}_+}$ and x be in X. Then (x_n) converges to x under the metric d iff (x_n) converges to x under δ. Namely, $d(x_n, x) \to 0$ iff $\delta(x_n, x) \to 0$.*

Proof. Prove it yourself. (Or see Prop. 2.62.) $\square$

More useful formulas about limit will be given in Exp. 4.24.

2.3.2 Criteria for divergence

Definition 2.29. A subset E of a metric space (X, d) is called **bounded** if either $E = \varnothing$ or there exist $p \in X$ and $R \in \mathbb{R}_{>0}$ such that $E \subset B_X(p, R)$. If X is bounded, we also say that d is a **bounded metric**.

Remark 2.30. Note that if E is bounded, then for *any* $q \in X$ there exists $\widetilde{R} \in \mathbb{R}_{>0}$ such that $E \subset B_X(q, \widetilde{R})$. (Indeed, choose $\widetilde{R} = R + d(p, q)$, then by triangle inequality, $B(p, R) \subset B(q, \widetilde{R})$.)

Some easy examples are as follows.

Example 2.31. In a metric space X, if $x \in X$ and $0 < r < +\infty$, then $B(x, r)$ is bounded. Hence $\overline{B}(x, r)$ is bounded (since it is inside $B(x, 2r)$).

Also, it is easy to see:

Example 2.32. A finite union of bounded subsets is bounded.

Proposition 2.33. *Let $(x_n)_{n \in \mathbb{Z}_+}$ be a convergent sequence in a metric space X. Then $(x_n)_{n \in \mathbb{Z}_+}$ is bounded.*

By saying that a sequence $(x_n)_{n \in \mathbb{Z}_+}$ in X is **bounded**, we mean that its range in X (namely $\{x_n : n \in \mathbb{Z}_+\}$) is bounded.

Proof. Suppose that $x_n \to x$. Then for each $\varepsilon > 0$, say $\varepsilon = 1$, all but finitely many elements of x_n (say $x_1, \dots, x_N$) are in $B(x, 1)$. So this whole sequence is in $A = \{x_1\} \cup \cdots \{x_N\} \cup B(x, 1)$. Since each $\{x_i\}$ is bounded, and since $B(x, 1)$ is bounded, A is bounded by Exp. 2.32. $\square$

Remark 2.34. Prop. 2.33 gives our first criterion on divergence: If a sequence is unbounded (e.g. $x_n = n^2$), then it does not converge. But there are many bounded and divergent sequences. (See Exp. 2.37.) In this case, we need the second criterion: If a sequence has two subsequences converging to two different points, then this sequence diverge. (See Prop. 2.36)

Definition 2.35. Let $(x_n)_{n \in \mathbb{Z}_+}$ be a sequence in a set X. If $(n_k)_{k \in \mathbb{Z}_+}$ is a strictly increasing sequence in $\mathbb{Z}_+$, we say that $(x_{n_k})_{k \in \mathbb{Z}_+}$ is a **subsequence** of $(x_n)_{n \in \mathbb{Z}_+}$.

Thus, a subsequence of $(x_n)_{n \in \mathbb{Z}_+}$ is equivalently the restriction of the function $x : \mathbb{Z}_+ \to X$ to an infinite subset of $\mathbb{Z}_+$.

Proposition 2.36. *Let $(x_n)_{n \in \mathbb{Z}_+}$ be a sequence in a metric space X converging to $x \in X$. Then every subsequence $(x_{n_k})_{k \in \mathbb{Z}_+}$ converges to x.*

Proof. For every $\varepsilon > 0$, $B(x, \varepsilon)$ contains all but finitely many $\{x_n : n \in \mathbb{Z}_+\}$, and hence all but finitely many $\{x_{n_k} : k \in \mathbb{Z}_+\}$. $\square$

Example 2.37. The sequence $x_n = (-1)^n$ in $\mathbb{R}$ is divergent, because the subsequence $(x_{2k})_{k \in \mathbb{Z}_+}$ converges to 1, whereas $(x_{2k-1})_{k \in \mathbb{Z}_+}$ converges to -1.

One may wonder if the two criteria in Rem. 2.34 are complete in order to determine whether a sequence diverges. This is true for sequences in $\mathbb{R}^n$. We will discuss this topic later. (See Cor. 3.17.)

2.4 Continuous maps of metric spaces

Continuous maps are a powerful tool for showing that a sequence converges.

Definition 2.38. Let $f : X \to Y$ be a map of metric spaces. Let $x \in X$. We say that f is **continuous at** x if one of the following equivalent conditions hold:

(1) For every sequence $(x_n)_{n \in \mathbb{Z}_+}$ in X, we have

$$\lim_{n \to \infty} x_n = x \qquad \Longrightarrow \qquad \lim_{n \to \infty} f(x_n) = f(x)$$

(2) For every $\varepsilon > 0$, there exists $\delta > 0$ such that for every $p \in X$ satisfying $d(p, x) < \delta$, we have $d(f(p), f(x)) < \varepsilon$.

(2') For every $\varepsilon > 0$, there exists $\delta > 0$ such that $B_X(x, \delta) \subset f^{-1}(B_Y(f(x), \varepsilon))$.

We say that f is a **continuous map/function**, if f is continuous at every point of X.

Proof of the equivalence. (2)⇔(2′): Obvious.

(2)⇒(1): Assume (2). Assume $x_n \to x$. For every $\varepsilon > 0$, let $\delta > 0$ be as in (2). Then since $x_n \to x$, there is $N \in \mathbb{Z}_+$ such that for all $n \geqslant N$ we have $d(x_n, x) < \delta$. By (2), we have $d(f(x_n), f(x)) < \varepsilon$ for all $n \geqslant N$. This proves $f(x_n) \to f(x)$.

¬(2)⇒ ¬(1): Assume that (2) is not true. Then there exists $\varepsilon > 0$ such that for every $\delta > 0$, there exists $p \in X$ such that $d(p, x) < \delta$ and $d(f(p), f(x)) \geqslant \varepsilon$. Thus, for every $n \in \mathbb{Z}_+$, by taking $\delta = 1/n$, we see that there exists $x_n \in X$ such that $d(x_n, x) < 1/n$ and $d(f(x_n), f(x)) \geqslant \varepsilon$. By Squeeze theorem (Prop. 2.26), $x_n \to x$. But $f(x_n) \nrightarrow f(x)$ (i.e. $d(f(x_n), f(x)) \nrightarrow 0$). So (1) is not true. $\qquad \square$

Remark 2.39. One can compare Def. 2.38-(1) to the definition of linear maps. A map is continuous iff it *preserves the convergence of sequences*, i.e., iff it maps convergent sequences to convergent ones. A map (between vector spaces) is linear iff it perserves the addition and the scalar multiplication of vectors. In general, a good map between two sets with "structures" is a map which preserves the structures. (Thus, linear combinations encode the linear structures of vector spaces. Similarly, the convergence of sequences remembers the "topological" structures of metric spaces.) As another example, we will define an isometry of metric spaces to be one that preserves the metrics (the structures of metric spaces), see Exe. 2.63.

Remark 2.40. In this section, we mainly use Def. 2.38-(1) to study continuity. But in later sections we will also use Def. 2.38-(2′). An advantage of (2′) is that it is more geometric. Indeed, if X is a metric space and $E \subset X$, we say that $x \in E$ is an **interior point of E in X** if there exists $\delta > 0$ such that $B_X(x, \delta) \subset E$. For example, a point $z \in \mathbb{C}$ is an interior point of the closed unit disk $\overline{B}_{\mathbb{C}}(0, 1) = \{w \in \mathbb{C} : |w| \leqslant 1\}$ iff $|z| < 1$.

Thus, Def. 2.38-(2′) says that for any map of metric spaces $f : X \to Y$ and $x \in X$, the following are equivalent:

(a) f is continuous at x.

(b) For each $\varepsilon > 0$, every $x \in X$ is an interior point of $f^{-1}\big(B_Y(f(x), \varepsilon)\big)$.

We say that a subset $U \subset X$ is **open** if each point of U is an interior point. For example, by triangle inequality, every open ball in a metric space is an open set. Thus, we have:

- A map of metric spaces $f : X \to Y$ is continuous iff the preimage under f of every open ball of Y is an open subset of X.

In the study of point-set topology, we will see that many properties can be studied in two approaches: using sequences (or using nets, the natural generalizations of sequences) and their convergence, and using open sets. The first example

of such property is continuity, as we have seen in Def. 2.38. Another prominent example is sequential compactness vs. compactness. These two approaches represent two (very) different intuitions: one is dynamic, while the other is static and more geometric. (So it is surprising that these two very things are actually equivalent!) Sometimes both approaches work for a problem, but sometimes only one of them works, or one of them is much simpler. If you are a beginner in analysis and point-set topology, I suggest that whenever you see one approach applied to a problem, try to think about whether the other approach also works and which one is better. □

2.4.1 Methods for proving continuity

Lemma 2.41. *Let $f : X \to Y$ be a map of metric spaces. Let $(B_i)_{i \in I}$ be a collection of open balls in X such that $X = \bigcup_{i \in I} B_i$. Suppose that for each i, the restriction $f|_{B_i} : B_i \to Y$ is continuous. Then f is continuous.*

This lemma shows that if we can prove that f is "locally" continuous, then f is globally continuous.

Proof. Choose (x_n) in X converging to $x \in X$. We shall show $f(x_n) \to f(x)$. Choose i such that $x \in B_i$. Then one can find $\delta > 0$ such that $B(x, \delta) \subset B_i$. (In the language of point-set topology: x is an interior point of B_i.) To see this, write $B_i = B(y, r)$. Since $x \in B(y, r)$, we have $r - d(x, y) > 0$. Choose $0 < \delta \leqslant r - d(x, y)$. Then triangle inequality implies $B(x, \delta) \subset B(y, r)$.

Since $x_n \to x$, there is $N \in \mathbb{Z}_+$ such that $x_n \in B(x, \delta)$ for all $n \geqslant N$. Thus, $(x_{k+N})_{k \in \mathbb{Z}_+}$ converges in B_i to x. Since $f|_{B_i}$ is continuous, $\lim_{k \to \infty} f(x_{k+N}) = f(x)$. So $f(x_n) \to f(x)$. □

Definition 2.42. A map of metric spaces $f : X \to Y$ is called **Lipschitz continuous** if there exists $L \in \mathbb{R}_{>0}$ (called **Lipschitz constant**) such that for all $x_1, x_2 \in X$,

$$d_Y\big(f(x_1), f(x_2)\big) \leqslant L \cdot d_X(x_1, x_2) \tag{2.5}$$

Lemma 2.43. *Lipschitz continuous maps are continuous.*

Proof. Suppose that $f : X \to Y$ is Lipschitz continuous with Lipschitz constant L. Suppose $x_n \to x$ in X. Then $L \cdot d(x_n, x) \to 0$. By (2.5) and Squeeze theorem (Prop. 2.26), $f(x_n) \to f(x)$. (You can also use Def. 2.38-(2) to prove this lemma.) □

Example 2.44. Let $\mathbb{F} \in \{\mathbb{Q}, \mathbb{R}, \mathbb{C}\}$. The map $z \in \mathbb{F} \backslash \{0\} \mapsto z^{-1} \in \mathbb{F}$ is continuous.

Proof. Let this map be f. Since $\mathbb{F}$ is covered by open balls $B(z, \delta)$ where $z \in \mathbb{F} \backslash \{0\}$ and $0 < \delta < |z|$, by Lem. 2.41, it suffices to prove that f is continuous when restricted to every such $B(z, \delta)$. Let $\varepsilon = |z| - \delta > 0$. Choose $x, y \in B(z, \delta)$. Then $|x| = |x - z + z| \geqslant |z| - |z - x| > \varepsilon$ by triangle inequality. Similarly, $|y| > \varepsilon$. So

$$|f(x) - f(y)| = |x^{-1} - y^{-1}| = |x - y|/|xy| \leqslant \varepsilon^{-2}|x - y|$$

So $f|_{B(z,\delta)}$ has Lipschitz constant ε^{-2}, and hence is continuous. □

(Question: in the above proof, is the map $f : \mathbb{F}\backslash\{0\} \to \mathbb{F}$ Lipschitz continuous?)

We have given a fancy way of proving that if (z_n) is a sequence in $\mathbb{F}\backslash\{0\}$ converging to $z \in \mathbb{F}\backslash\{0\}$, then $z_n^{-1} \to z^{-1}$. You should think about how to prove this fact directly using $\varepsilon - N$ language, and compare your proof with the above proof to find the similarities!

Proposition 2.45. *Let $\mathbb{F} \in \{\mathbb{Q}, \mathbb{R}, \mathbb{C}\}$. Then the following maps are continuous:*

$$
\begin{aligned}
+ &: \mathbb{F} \times \mathbb{F} \to \mathbb{F} & (x, y) &\mapsto x + y \\
- &: \mathbb{F} \times \mathbb{F} \to \mathbb{F} & (x, y) &\mapsto x - y \\
\times &: \mathbb{F} \times \mathbb{F} \to \mathbb{F} & (x, y) &\mapsto xy \\
\div &: \mathbb{F} \times \mathbb{F}^\times \to \mathbb{F} & (x, y) &\mapsto x/y
\end{aligned}
$$

Recall our Convention 2.13 on the metrics of finite product spaces.

Proof. We only prove that the last two are continuous: the first two can be treated in a similar (and easier) way.

Denote the multiplication map by μ. We choose the metric on $\mathbb{F}^2$ to be $d(\mathbf{x}, \mathbf{x}') = \max\{|x_1 - x_1'|, |x_2 - x_2'|\}$. Since $\mathbb{F} \times \mathbb{F}$ is covered by open balls of the form $B(0, r) = \{(x, y) \in \mathbb{F}^2 : |x| < r, |y| < r\}$ where $0 < r < +\infty$, by Lem. 2.41 and 2.43, it suffices to show that $\mu|_{B(0,r)}$ is Lipschitz continuous. This is true, since for each $(x, y), (x', y') \in B(0, r)$, we have

$$
\begin{aligned}
|\mu(x, y) - \mu(x', y')| &= |xy - x'y'| \leqslant |(x - x')y| + |x'(y - y')| \\
&< 2r \cdot \max\{|x - x'|, |y - y'|\} = 2r \cdot d((x, y), (x', y'))
\end{aligned}
\tag{2.6}
$$

By Exp. 2.44 and Prop. 2.47, the map $(x, y) \in \mathbb{F} \times \mathbb{F}^\times \mapsto (x, y^{-1}) \in \mathbb{F} \times \mathbb{F}$ is continuous. So its composition with the continuous map μ is continuous, thanks to Prop. 2.46. So $\div$ is continuous. $\qquad\square$

Proposition 2.46. *Suppose that $f : X \to Y$ and $g : Y \to Z$ are continuous maps of metric spaces. Then $g \circ f : X \to Z$ is continuous.*

Proposition 2.47. *Suppose that $f_i : X_i \to Y_i$ is a map of metric spaces, where $1 \leqslant i \leqslant N$. Then the product map*

$$
\begin{aligned}
f_1 \times \cdots \times f_N &: X_1 \times \cdots \times X_N \to Y_1 \times \cdots \times Y_N \\
(x_1, \ldots, x_N) &\mapsto (f_1(x_1), \ldots, f_N(x_N))
\end{aligned}
$$

is continuous if and only if $f_1, \ldots, f_N$ are continuous.

Proof of Prop. 2.46 and 2.47. Immediate from Def. 2.38-(1) and Prop. 2.27. $\qquad\square$

Corollary 2.48 (Squeeze theorem). *Let $\mathbb{F} \in \{\mathbb{Q}, \mathbb{R}\}$ and $(x_n), (y_n), (z_n)$ be sequences in $\mathbb{R}$. Assume that $x_n \leqslant y_n \leqslant z_n$ for all n. Assume that x_n and z_n both converge to $A \in \mathbb{R}$. Then $\lim_{n\to\infty} y_n = A$.*

Proof. Let $a_n = y_n - x_n$ and $b_n = z_n - x_n$. Then $0 \leqslant a_n \leqslant b_n$, and $\lim_n b_n = \lim_n z_n - \lim_n x_n = 0$ because the subtraction map is continuous (Prop. 2.45). By Exp. 2.25, $a_n \to 0$. Since $x_n \to A$, $y_n = x_n + a_n$ converges to A, since the addition map is continuous by Prop. 2.45. □

Again, this is a fancy way of proving Squeeze theorem. The readers should know how to prove it directly from the definition of limits of sequences.

We give some more examples of continuous maps.

Example 2.49. By Prop. 2.45, $f(x, y, z) = x^2 y + 5y^4 z^7 - 3xyz^2$ is a continuous function on $\mathbb{C}^3$. Clearly $z \in \mathbb{C} \mapsto \overline{z} \in \mathbb{C}$ is continuous. So $g(x, y, z) = f(\overline{x}, \overline{y}, z) + 2\overline{f(z, x^2, xy^{-9})} - 5xy^{-2}\overline{z^{-3}}$ is a continuous function on $\mathbb{C} \times \mathbb{C}^\times \times \mathbb{C}^\times$.

Example 2.50. Let $f, g : X \to \mathbb{F}$ be continuous functions where $\mathbb{F} \in \{\mathbb{Q}, \mathbb{R}, \mathbb{C}\}$. Then by Prop. 2.45 and 2.46, $f \pm g$ and fg are continuous, and f/g is continuous when $0 \notin g(X)$. Here

$$(f \pm g)(x) = f(x) \pm g(x) \qquad (fg)(x) = f(x)g(x) \qquad (f/g)(x) = f(x)/g(x) \quad (2.7)$$

Example 2.51. Let $f : X \to Y$ be a map of metric spaces. Let E, F be subsets of X, Y respectively. (Recall that the metrics of subsets are chosen as in Def. 2.14.)

- The inclusion map $\iota_E : E \to X, x \mapsto x$ is clearly continuous. Thus, if f is continuous, then $f|_E : E \to Y$ is continuous, since $f|_E = f \circ \iota_E$.

- If $f(X) \subset F$, then we can restrict the codomain of f from Y to F: let $\widetilde{f} : X \to F$ be $\widetilde{f}(x) = f(x)$. Then it is clear that f is continuous iff $\widetilde{f}$ is continuous.

Proposition 2.52. *Let $X_1, \ldots, X_N$ be metric spaces. The following projection is clearly continuous:*

$$\pi_{X_i} : X_1 \times \cdots \times X_N \to X_i \qquad (x_1, \ldots, x_N) \mapsto x_i$$

Proposition 2.53. *Let $f_i : X \to Y_i$ be maps where $X, Y_1, \ldots, Y_N$ are continuous. Then*

$$f_1 \vee \cdots \vee f_N : X \to Y_1 \times \cdots \times Y_N \qquad x \mapsto (f_1(x), \ldots, f_N(x))$$

is continuous iff $f_1, \ldots, f_N$ are continuous.

Example 2.54. Let X be a metric space. Then $d : X \times X \to \mathbb{R}, (x, y) \mapsto d(x, y)$ is Lipschiz continuous with Lipschitz constant 1 (by triangle inequality). So d is continuous.

Proposition 2.55. *Let X be a metric space and $E \subset X$ is nonempty. Define **distance function***

$$d(\cdot, E) : X \to \mathbb{R}_{\geqslant 0} \qquad d(x, E) = \inf_{e \in E} d(x, e)$$

Then $d(\cdot, E)$ is has Lipschitz constant 1. So $d(\cdot, E)$ is continuous.

Proof. Choose any $x, y \in X$. By triangle inequality, for each $e \in E$ we have $d(x, e) \leqslant d(x, y) + d(y, e)$. Since $d(x, E) \leqslant d(x, e)$, we get $d(x, E) \leqslant d(x, y) + d(y, e)$. Applying $\inf_{e \in E}$ to the RHS gives $d(x, E) \leqslant d(x, y) + d(y, E)$. Hence $d(x, E) - d(y, E) \leqslant d(x, y)$. Exchanging x and y gives

$$\left| d(x, E) - d(y, E) \right| \leqslant d(x, y) \tag{2.8}$$

This proves that $d(\cdot, E)$ has Lipschitz constant 1. $\qquad\square$

Definition 2.56. More generally, if E, F are subsets of a metric space E, we can define

$$d(E, F) = \inf_{e \in E, f \in F} d(e, f) \tag{2.9}$$

to be the **distance between E and F**.

Exercise 2.57. Let $E, F \subset X$. Prove that

$$d(E, F) = \inf_{f \in F} d(E, f) \tag{2.10}$$

Example 2.58. If X is a metric space and $p \in X$, then by Prop. 2.55 (or simply by triangle inequality), the function $d_p : x \in X \mapsto d(x, p) \in \mathbb{R}$ has Lipschitz constant 1 and hence is continuous. In particular, if $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, the function $z \in \mathbb{F} \mapsto |z|$ is continuous (since $|z| = d_{\mathbb{F}}(z, 0)$). Thus, if $f : X \to \mathbb{F}$ is continuous, then $|f| : X \to \mathbb{R}_{\geqslant 0}$ is continuous where

$$|f|(x) = |f(x)| \tag{2.11}$$

Example 2.59. Let $N \in \mathbb{Z}_+$. Then the following function is continuous:

$$\max : \mathbb{R}^N \to \mathbb{R} \qquad (x_1, \ldots, x_N) \mapsto \max\{x_1, \ldots, x_N\} \in \mathbb{R}$$

Similarly, the minimum function is continuous.

Proof. To avoid confusion, we write $\max$ as $\max_N$. The case $N = 1$ is obvious. When $N = 2$, we have

$$\max(x_1, x_2) = \frac{x_1 + x_2 + |x_1 - x_2|}{2} \tag{2.12}$$

So $\max_2$ is continuous by Exp. 2.50 and 2.58.

We use induction. Suppose we have proved that $\max_N$ is continuous. Then $\max_N \times \mathrm{id}_{\mathbb{R}} : \mathbb{R}^N \times \mathbb{R} \to \mathbb{R} \times \mathbb{R}$ is continuous. So $\max_{N+1} = \max_2 \circ (\max_N \times \mathrm{id}_{\mathbb{R}})$ is continuous. $\qquad\square$

2.5 Homeomorphisms and isometric isomorphisms; convergence in $\overline{\mathbb{R}}$

2.5.1 General theory

Definition 2.60. A bijection of metric spaces $f : X \to Y$ is called a **homeomorphism** if one of the following equivalent (cf. Def. 2.38) statements holds:

(1) $f : X \to Y$ and its inverse map $f^{-1} : Y \to X$ are continuous.

(2) For each sequence (x_n) in X and each $x \in X$, we have $\lim\limits_{n\to\infty} x_n = x$ iff $\lim\limits_{n\to\infty} f(x_n) = f(x)$.

If such f exists, we say that X, Y are **homeomorphic**.

A special case of the above definition is:

Definition 2.61. Let X be a set with metrics d, δ. We say that d and δ induce the **same topology** on X (or that d, δ are **topologically equivalent**) if one of the following clearly equivalent statements holds:

(1) The map $(X, d) \to (X, \delta), x \mapsto x$ is a homeomorphism.[5]

(2) For each sequence (x_n) in X and each $x \in X$, (x_n) converges to x under the metric d iff (x_n) converges to x under δ.

Proposition 2.62. *Suppose that d, δ are equivalent metrics on a set X. Then d, δ are topologically equivalent.*

Proof. Suppose $\delta \leqslant \alpha d$ and $d \leqslant \beta \delta$ for some $\alpha, \beta > 0$. Then the map $f : (X, d) \to (X, \delta), x \mapsto x$ and its inverse f^{-1} have Lipschitz constants α and β respectively. So f, f^{-1} are continuous. $\qquad\square$

Exercise 2.63. Let $f : X \to Y$ be a map of metric spaces. We say that $f : X \to Y$ is an **isometry** (or is **isometric**) if for all $x_1, x_2 \in X$ we have

$$d_Y(f(x_1), f(x_2)) = d_X(x_1, x_2) \tag{2.13}$$

Show that an isometry is injective and continuous.

We say that f is an **isometric isomorphism** if f is a surjective isometry. If an isometric isomorphism between two metric spaces X, Y exists, we say that X and Y are **isometric metric spaces**. Show that an isometric isomorphism is a homeomorphism. $\qquad\square$

---

[5]We prefer not to call this map the identity map, because the metrics on the source and on the target are different.

Remark 2.64. Isometric isomorphisms are important examples of homeomorphisms. That $f : X \to Y$ is an isometric isomorphism means that X and Y are equivalent as metric spaces, and that this equivalence can be implemented by the bijection f.

We now look at isometric isomorphisms in a different direction. Suppose that $f : X \to Y$ is a bijection of sets. Suppose that Y is a metric space. Then there is unique metric d_X on X such that f is an isometric isomorphism: one defines d_X using (2.13). We write such d_X as f^*d_Y, i.e.,

$$f^*d_Y(x_1, x_2) = d_Y(f(x_1), f(x_2))$$

and call f^*d_Y the **pullback metric** of d_Y by f. $\hspace{1cm} \square$

Pullback metrics are a very useful way of constructing metrics on a set. We consider some examples below.

Exercise 2.65. Two metrics inducing the same topology are not necessarily equivalent metrics. For example, let $f : [0, 1] \to [0, 1]$ be $f(x) = x^2$. Let $X = [0, 1]$, and let d_X be the Euclidean metric: $d_X(x, y) = |x - y|$. So

$$f^*d_X(x, y) = |x^2 - y^2|$$

is a metric on X. It is not hard to check that $f : (X, d_X) \to (X, d_X)$ is a homeomorphism. So d_X and f^*d_X give the same topology on $[0, 1]$ (cf. Exe. 2.66). Show that f^*d_X and d_X are not equivalent metrics.

Exercise 2.66. Let $f : X \to Y$ be a bijection of sets with metrics d_X, d_Y. Show that d_X and f^*d_Y give the same topology on X iff $f : (X, d_X) \to (Y, d_Y)$ is a homeomorphism.

In particular, if $f : X \to X$ is a bijection, and d_X is a metric on X. Then d_X and f^*d_X give the same topology on X iff $f : (X, d_X) \to (X, d_X)$ is a homeomorphism.

2.5.2   Convergence in $\overline{\mathbb{R}}$

Our second application of pullback metrics is the convergence in $\overline{\mathbb{R}}$.

Definition 2.67. We say that a sequence (x_n) in $\overline{\mathbb{R}}$ **converges to** $+\infty$ (resp. $-\infty$), if for every $A \in \mathbb{R}$ there is $N \in \mathbb{Z}_+$ such that for all $n \geqslant N$ we have $x_n > A$ (resp. $x_n < A$).

Suppose $x \in \mathbb{R}$. We say that a sequence (x_n) in $\overline{\mathbb{R}}$ **converges to** x, if there is $N \in \mathbb{Z}_+$ such that $x_n \in \mathbb{R}$ for all $n \geqslant N$, and that the subsequence $(x_{k+N})_{k \in \mathbb{Z}_+}$ converges in $\mathbb{R}$ to x. $\hspace{1cm} \square$

This notion of convergence is weird: it is not defined by a metric. So one wonders if there is a metric d on $\overline{\mathbb{R}}$ such that convergence of sequences under d agrees with that in Def. 2.67. We shall now find such a metric.

Lemma 2.68. *Let $-\infty \leqslant a < b \leqslant +\infty$ and $-\infty \leqslant c < d \leqslant +\infty$. Then there is a strictly increasing bijective map $[a, b] \to [c, d]$.*

Note that this map clearly sends a to c and b to d. So it restricts to strictly increasing bijections $(a, b) \to (c, d)$, $(a, b] \to (c, d]$, $[a, b) \to [c, d)$.

Proof. We have a strictly increasing bijection $f : \mathbb{R} \to (-1, 1)$ defined by (1.11). f can be extended to a strictly increasing bijective map $\overline{\mathbb{R}} \to [-1, 1]$ if we set $f(\pm\infty) = \pm 1$. Thus, f restricts to a strictly increasing bijection $[a, b] \to [f(a), f(b)]$. Choose a linear function $g(x) = \alpha x + \beta$ (where $\alpha > 0$) giving an increasing bijection $[f(a), f(b)] \to [0, 1]$. Then $h = g \circ f : [a, b] \to [0, 1]$ is a strictly increasing bijection. Similarly, we have a strictly increasing bijection $k : [c, d] \to [0, 1]$. Then $k^{-1} \circ h : [a, b] \to [c, d]$ is a strictly increasing bijection. $\qquad\square$

Theorem 2.69. *Let $\varphi : \overline{\mathbb{R}} \to [a, b]$ be a strictly increasing bijective map where $[a, b] \subset \mathbb{R}$ is equipped with the Euclidean metric $d_{[a,b]}$. Then a sequence (x_n) in $\overline{\mathbb{R}}$ converges to $x \in \overline{\mathbb{R}}$ in the sense of Def. 2.67 iff $\varphi(x_n)$ converges to $\varphi(x)$ under the metric $d_{[a,b]}$. In other words, the convergence in $\overline{\mathbb{R}}$ is given by the metric $\varphi^* d_{[a,b]}$.*

Proof. Let $y = \varphi(x)$ and $y_n = \varphi(x_n)$. We need to prove that $x_n \to x$ (in the sense of Def. 2.67) iff $y_n \to y$ (under the Euclidean metric). Write $\psi = \varphi^{-1}$, which is a strictly increasing map $[a, b] \to \overline{\mathbb{R}}$. Note that $\varphi(+\infty) = b$ and $\varphi(-\infty) = a$.

Case 1: $x \in \mathbb{R}$. By discarding the first several terms, we may assume that (x_n) is always in $\mathbb{R}$. If $x_n \to x$, then for every $\varepsilon > 0$, all but finitely many x_n are inside the open interval $(\psi(y-\varepsilon), \psi(y+\varepsilon))$. So all but finitely many y_n are inside $(y-\varepsilon, y+\varepsilon)$. So $y_n \to y$. That $y_n \to y$ implies $x_n \to x$ is proved in a similar way.

Case 2: $x = \pm\infty$. We consider $x = +\infty$ only; the other case is similar. Note that if $0 < \varepsilon < b - a$, then $B_{[a,b]}(b, \varepsilon) = (b - \varepsilon, b]$. If $x_n \to +\infty$, then for each $0 < \varepsilon < b - a$, all but finitely many x_n are $> \psi(b - \varepsilon)$. So all but finitely many y_n are inside $(b - \varepsilon, b]$. This proves $y_n \to b$. Conversely, if $y_n \to b$, then for each $A \in \mathbb{R}$, all but finitely many y_n are inside $(\varphi(A), b]$ and hence $> \varphi(A)$. So all but finitely many x_n are $> A$. $\qquad\square$

Convention 2.70. Unless otherwise stated, a metric on $\overline{\mathbb{R}}$ is one that makes Def. 2.67 true, for instance $\varphi^* d_{[a,b]}$ in Thm. 2.69. Unless otherwise stated, we do NOT view $\mathbb{R}$ (or any subset of $\mathbb{R}$) as a metric subspace of $\overline{\mathbb{R}}$. Namely, we do not follow Convention 2.15 for $\mathbb{R} \subset \overline{\mathbb{R}}$, or more generally for $\mathbb{R}^N \subset \overline{\mathbb{R}}^N$. Instead, we choose Euclidean metrics on $\mathbb{R}^N$, following Convention 2.8.

The main reason for not following Convention 2.15 here is that metrics on $\overline{\mathbb{R}}$ are all bounded (by Prop. 3.7). Thus, every subset of $\mathbb{R}$ is bounded if we view $\mathbb{R}$ as a metric subspace of $\overline{\mathbb{R}}$. However, we want a subset of $\mathbb{R}$ to be bounded precisely when it is contained in $[a, b]$ for some $-\infty < a < b < +\infty$. (Recall also Def. 1.36.)

After learning topological spaces, we shall forget about the metrics on $\overline{\mathbb{R}}$ and only care about its topology. (See Conv. 7.18.)

Remark 2.71. By Thm. 2.69, the properties of $[a, b]$ about convergence of sequences and inequalities can be transported to $\overline{\mathbb{R}}$, for example:

1. If $(x_n), (y_n)$ are sequences in $\overline{\mathbb{R}}$ converging to $A, B \in \overline{\mathbb{R}}$, and if $x_n \leqslant y_n$ for all n, then $A \leqslant B$.

2. **Squeeze theorem**: Suppose that $(x_n), (y_n), (z_n)$ are sequences in $\overline{\mathbb{R}}$, $x_n \leqslant y_n \leqslant z_n$ for all n, and x_n and z_n both converge to $A \in \overline{\mathbb{R}}$. Then $y_n \to A$.

3. Prop. 2.23 also holds for $[-\infty, +\infty] = \overline{\mathbb{R}}$: if (x_n) is an increasing resp. decreasing sequence in $\overline{\mathbb{R}}$, then $\lim_n x_n$ exists in $\overline{\mathbb{R}}$ and equals $\sup_n x_n$ resp. $\inf_n x_n$.

We will see more examples when studying $\limsup$ and $\liminf$ in the future.

We have shown that there is a metric on $\overline{\mathbb{R}}$ which defines the convergence in Def. 2.67. However, there is no standard choice of such a metric on $\overline{\mathbb{R}}$. Even worse, two possible choices of metrics might not be equivalent: Let $\varphi, \psi : \overline{\mathbb{R}} \to [0, 1]$ be a strictly increasing bijections where $\psi \circ \varphi^{-1} : [0, 1] \to [0, 1]$ is $x \mapsto x^2$. Then by Exe. 2.65, $\varphi^* d_{[0,1]}$ and $\psi^* d_{[0,1]}$ are non-equivalent but topologically equivalent metrics on $\overline{\mathbb{R}}$. This is the first example that metrics are not convenient for the description of convergence. When studying the convergence in $\overline{\mathbb{R}}$, thinking about metrics is distracting. In the future, we will see a better notion for the study of convergence: the notion of topological spaces.

We end this section with a generalization of Thm. 2.69.

Theorem 2.72. *Let φ be a strictly increasing bijection in one of the following forms*

$$[a, b] \to [c, d] \qquad (a, b) \to (c, d)$$
$$(a, b] \to (c, d] \qquad [a, b) \to [c, d)$$

where $-\infty \leqslant a \leqslant b \leqslant +\infty$ and $-\infty \leqslant c \leqslant d \leqslant +\infty$. Then φ is a homeomorphism, i.e., if (x_n) and x are in the domain, then $x_n \to x$ iff $\varphi(x_n) \to \varphi(x)$ (in the sense of Def. 2.67).

Proof. The case $a = b$ is obvious. So we consider $a < b$, and hence $c < d$. We consider the left-open-right-closed case for example. The other cases are treated in a similar way. If the theorem can be proved for $(-\infty, +\infty] \to (c, d]$, then it can also be proved $(-\infty, +\infty] \to (a, b]$. By composing the inverse of the second map with the first map, we see that the theorem holds for $(a, b] \to (c, d]$.

Let us consider $\varphi : (-\infty, +\infty] \to (c, d]$. φ can be extended to a strictly increasing bijection $\varphi : \overline{\mathbb{R}} \to [c, d]$ by letting $\varphi(-\infty) = c$. It suffices to prove that this φ is a homeomorphism. When $-\infty < c < d < +\infty$, then the theorem holds by Thm. 2.69. If one of c, d is $\pm\infty$, the same argument as in the proof of Thm. 2.69 proves that φ is a homeomorphism. We leave it to the readers to fill in the details. $\qquad\square$

2.6 Problems and supplementary material

Definition 2.73. Let A be a subset of $\mathbb{R}$ satisfying $x + y \in A$ for all $x, y \in A$. (Or more generally, let A be an abelian semigroup.) We say that a function $f : A \to \mathbb{R}$ is **subadditive** if for every $x, y \in A$ we have $f(x + y) \leqslant f(x) + f(y)$.

Problem 2.1. Consider the following increasing functions:

$$f_1 : \mathbb{R}_{\geqslant 0} \to [0, 1) \qquad f_1(x) = \frac{x}{1 + x}$$
$$f_2 : \mathbb{R}_{\geqslant 0} \to [0, 1] \qquad f_2(x) = \min\{x, 1\}$$

Prove that they are subadditive functions.

Problem 2.2. Let $f : \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$ be an increasing subadditive function satisfying the following conditions:

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

(2) For any $(x_n)_{n \in \mathbb{Z}_+}$ in $\mathbb{R}_{\geqslant 0}$ we have $x_n \to 0$ iff $f(x_n) \to 0$.

Let (X, d) be a metric space. Define

$$\delta : X \times X \to [0, A) \qquad \delta(x, y) = f \circ d(x, y)$$

Prove that δ is a metric, and that δ and d are topologically equivalent.

Proposition 2.74. *Let (X, d) be a metric space. Then there is a bounded metric δ on X such that d and δ are topologically equivalent.*

Proof. Let f be either f_1 or f_2 defined in Pb. 2.1. Then f satisfies the assumptions in Pb. 2.2. So $\delta = f \circ d$ is a desired metric due to Pb. 2.2. We write down the formulas explicitly:

$$\delta_1(x, y) = \frac{d(x, y)}{1 + d(x, y)} \qquad \delta_2(x, y) = \min\{d(x, y), 1\}$$

$\square$

Problem 2.3. Let $(X_i, d_i)_{i \in \mathbb{Z}_+}$ be a sequence of metric spaces. Assume that $d_i \leqslant 1$ for each i. Let $S = \prod_{i \in \mathbb{Z}_+} X_i$. For each elements $f = (f(i))_{i \in \mathbb{Z}_+}$ and $g = (g(i))_{i \in \mathbb{Z}_+}$ of S, define

$$d(f, g) = \sup_{i \in \mathbb{Z}_+} \frac{d_i(f(i), g(i))}{i} \tag{2.14}$$

Prove that d is a metric on S. Let $f_n = (f_n(i))_{i \in \mathbb{Z}_+}$ be a sequence in S. Let $g \in S$. Prove that the following are equivalent:

(a) $\lim\limits_{n \to \infty} f_n = g$ under the metric d.

(b) f_n **converges pointwise** to g, namely, $\lim\limits_{n \to \infty} f_n(i) = g(i)$ for every $i \in \mathbb{Z}_+$.

Remark 2.75. The above problem gives our first non-trivial example of function spaces as metric spaces, where the domain of functions is a countable set. After learning series, the readers can check that

$$\delta(f, g) = \sum_{i \in \mathbb{Z}_+} 2^{-i} d_i(f(i), g(i)) \tag{2.15}$$

also defines a metric, and that (a) (with d replaced by δ) and (b) are equivalent. So δ and d (defined by (2.14)) induce the same topology on X, called the **pointwise convergence topology** or simply **product topology**. Unfortunately, if the index set $\mathbb{Z}_+$ is replaced by an uncountable set, there is in general no metric inducing the product topology. We will prove this in Pb. 7.9.

★ **Problem 2.4.** Let $X = \bigsqcup_{\alpha \in \mathscr{A}} X_\alpha$ be a disjoint union of metric spaces (X_α, d_α). Assume that $d_\alpha \leqslant 1$ for all i. For each $x, y \in X$, define

$$d(x, y) = \begin{cases} d_\alpha(x, y) & \text{if } x, y \in X_\alpha \text{ for some } \alpha \in \mathscr{A} \\ \frac{1}{2} & \text{otherwise} \end{cases}$$

1. Prove that d defines a metric on X.

2. Choose $(x_n)_{n \in \mathbb{Z}_+}$ in X and $x \in X$. What does $\lim\limits_{n \to \infty} x_n = x$ mean in terms of the convergence in each X_α?

Think about the question: Let X be a set. For each $x, y \in X$ define $d(x, y) = 0$ if $x = y$, and $d(x, y) = 1$ if $x \neq y$. What does convergence in (X, d) mean?

3 Sequential compactness and completeness

3.1 Sequential compactness

3.1.1 Basic properties of sequentially compact spaces

Definition 3.1. Let X be a metric space. We say that X is **sequentially compact** if every sequence in X has a subsequence converging to some point of X.

The notion of sequential compactness is extremely useful for finding solutions in an analysis problem. In general, suppose we want to find a point $x \in X$ which makes a property $P(x)$ to be true. Suppose that we can find an "approximate solution", i.e. an $y \in X$ such that $P(y)$ is close to being true. Thus, we can find a sequence (x_n) in X such that $P(x_n)$ is closer and closer to being true when $n \to \infty$. Now, if X is sequentially compact, then (x_n) has a subsequence (x_{n_k}) converging to $x \in X$. Then $P(x)$ is true, and hence x is a solution for the problem. (See also Sec. 2.1.) Let us see an explicit example:

Lemma 3.2 (**Extreme value theorem**)**.** *Let X be a sequentially compact metric space. Let $f : X \to \mathbb{R}$ be a continuous function. Then f attains its maximum and minimum at points of X. In particular, $f(X)$ is a bounded subset of $\mathbb{R}$.*

This extremely important result is the main reason for introducing sequentially compact spaces. We call this a lemma, since we will substantially generalize this result later. (See Exe. 3.6.)

Note that the **boundedness** of subsets of $\mathbb{R}$ (or more generally, of $\mathbb{R}^N$) is always understood under the Euclidean metric of $\mathbb{R}$, not under any metric of $\overline{\mathbb{R}}$ or $\overline{\mathbb{R}}^N$. (Recall Convention 2.70.)

Proof. We show that f attains its maximum on X. The proof for minimum is similar. Let $A = \sup f(X)$. Then $A \in (-\infty, +\infty]$. If $A < +\infty$, then for each $n \in \mathbb{Z}_+$ there is $x_n \in X$ such that $A - 1/n < f(x_n) \leqslant A$ (since $A - 1/n$ is not an upper bound of $f(X)$). If $A = +\infty$, then for each n there is $x_n \in X$ such that $f(x_n) > n$. In either case, we have a sequence (x_n) in X such that $f(x_n) \to A$ in $\overline{\mathbb{R}}$.

Since X is sequentially compact, (x_n) has a subsequence $(x_{n_k})_{k \in \mathbb{Z}_+}$ converging to some $x \in X$. Now, consider f as a map $f : X \to \overline{\mathbb{R}}$, which is continuous (cf. Exp. 2.51). Since $f(x_n) \to A$, its subsequence $f(x_{n_k})$ also converges to A. But since $x_{n_k} \to x$ and f is continuous at x, we have $A = f(x)$. So f attains its maximum at x. Since $f(X) \subset \mathbb{R}$, we have $A \in \mathbb{R}$. $\qquad\square$

The following are some elementary examples of sequential compactness:

Exercise 3.3. Show that finite unions of sequentially compact spaces is sequentially compact. (In particular, a finite set is sequentially compact.)

More precisely, let X be a metric space. Assume $X = A_1 \cup \cdots \cup A_N$ where each metric subspace A_i is sequentially compact. Show that X is sequentially compact.

Proposition 3.4. *Let $X_1, \ldots, X_N$ be sequentially compact metric spaces. Then $X = X_1 \times \cdots \times X_N$ is sequentially compact.*

Proof. Since $X = (X_1 \times \cdots \times X_{N-1}) \times X_N$, by induction, it suffices to assume $N = 2$. So we write $X = A \times B$ where A, B are sequentially compact. Let (a_n, b_n) be a sequence in X. Since A is sequentially compact, (a_n) has a convergent subsequence (a_{n_k}). Since B is sequentially compact, (b_{n_k}) has a convergent subsequence $(b_{n_{k_l}})$. So $(a_{n_{k_l}}, b_{n_{k_l}})$ is a convergent subsequence of (a_n, b_n). $\qquad\square$

Proposition 3.5. *Let $f : X \to Y$ be a continuous map of metric spaces. Assume that X is sequentially compact. Then $f(X)$, as a metric subspace of Y, is sequentially compact.*

Proof. Choose any sequence (y_n) in $f(X)$. We can write $y_n = f(x_n)$ where $x_n \in X$. Since X is sequentially compact, (x_n) has a subsequence (x_{n_k}) converging to some $x \in X$. Since f is continuous, $y_{n_k} = f(x_{n_k})$ converges to $f(x)$. $\qquad\square$

Exercise 3.6. Prove that if Y is a sequentially compact subset of $\mathbb{R}$, then $\sup Y \in Y$ and $\inf Y \in Y$. Therefore, Prop. 3.5 generalizes Lem. 3.2.

Proposition 3.7. *Let X be a sequentially compact metric space. Then X is bounded under its metric d.*

Proof. Choose any $p \in X$. The function $d_p : x \in X \mapsto d(x, p) \in \mathbb{R}_{\geqslant 0}$ is continuous by Exp. 2.58. So, by Lem. 3.2, d_p is bounded by some $0 < R < +\infty$. So $X = \overline{B}_X(p, R) \subset B_X(p, 2R)$. $\qquad\square$

3.1.2 Limits inferior and superior, and Bolzano-Weierstrass

The goal of this subsection is to prove that closed intervals are sequentially compact.

Definition 3.8. Let (x_n) be a sequence in a metric space X. We say that $x \in X$ is a **cluster point** of (x_n), if (x_n) has a subsequence (x_{n_k}) converging to x.

Warning: In a general topological space, the cluster points of a sequence will be defined in a different way. (See Pb. 7.2 and Rem. 7.117.)

Definition 3.9. Let (x_n) be a sequence in $\overline{\mathbb{R}}$. Define

$$\alpha_n = \inf\{x_k : k \geqslant n\} \qquad \beta_n = \sup\{x_k : k \geqslant n\} \tag{3.1}$$

It is clear that $\alpha_n \leqslant x_n \leqslant \beta_n$, that (α_n) is increasing and (β_n) is decreasing. Define

$$\liminf_{n \to \infty} x_n = \sup\{\alpha_n : n \in \mathbb{Z}_+\} = \lim_{n \to \infty} \alpha_n \tag{3.2a}$$

$$\limsup_{n \to \infty} x_n = \inf\{\beta_n : n \in \mathbb{Z}_+\} = \lim_{n \to \infty} \beta_n \tag{3.2b}$$

(cf. Rem. 2.71), called respectively the **limit inferior** and the **limit superior** of (x_n).

Remark 3.10. Let $(x_n), (y_n)$ be sequences in $\overline{\mathbb{R}}$. Suppose that $x_n \leqslant y_n$ for every n. It is clear that

$$\liminf_{n\to\infty} x_n \leqslant \limsup_{n\to\infty} x_n \qquad \liminf_{n\to\infty} x_n \leqslant \liminf_{n\to\infty} y_n \qquad \limsup_{n\to\infty} x_n \leqslant \limsup_{n\to\infty} y_n$$

Theorem 3.11. *Let (x_n) be a sequence in $\overline{\mathbb{R}}$, and let S be the set of cluster points of (x_n). Then $\liminf_{n\to\infty} x_n$ and $\limsup_{n\to\infty} x_n$ belong to S. They are respectively the minimum and the maximum of S.*

In particular, every sequence in $\overline{\mathbb{R}}$ has at least one cluster point.

Proof. We use the notations in Def. 3.9. Let $A = $ (3.2a) and $B = $ (3.2b). If $x \in S$, pick a subsequence (x_{n_k}) converging to x. Since $\alpha_{n_k} \leqslant x_{n_k} \leqslant \beta_{n_k}$, we have $A \leqslant x \leqslant B$ by Rem. 2.71. It remains to show that $A, B \in S$. We prove $B \in S$ by constructing a subsequence (x_{n_k}) converging to B; the proof of $A \in S$ is similar.

Consider first of all the special case that (x_n) is bounded, i.e., is inside $[a, b] \subset \mathbb{R}$. Choose an arbitrary $n_1 \in \mathbb{Z}_+$. Suppose $n_1 < \cdots < n_k$ have been constructed. By the definition of β_{1+n_k}, there is $n_{k+1} \geqslant 1 + n_k$ such that $x_{n_{k+1}}$ is close to β_{1+n_k}, say

$$\beta_{1+n_k} - \frac{1}{k} < x_{n_{k+1}} \leqslant \beta_{1+n_k} \tag{3.3}$$

Since the left most and the right most of (3.3) both converge to B as $k \to \infty$, by Squeeze theorem (Cor. 2.48) we conclude $\lim_k x_{n_k} = B$.

In general, by Lem. 2.68 and Thm. 2.72, there is an increasing (i.e. order-preserving) homeomorphism (i.e. topopgy-preserving map) $\varphi : \overline{\mathbb{R}} \to [0, 1]$. Then $\varphi(\beta_n) = \sup\{\varphi(x_k) : k \geqslant n\}$ (cf. Exe. 3.13) and $\varphi(B) = \lim_n \varphi(\beta_n)$. So $\varphi(B) = \limsup_n \varphi(x_n)$. By the above special case, $(\varphi(x_n))$ has a subsequence $(\varphi(x_{n_k}))$ converging to $\varphi(B)$. So (x_{n_k}) converges to B. $\qquad\square$

Remark 3.12. One can also prove the above general case directly using a similar idea as in the special case. And you are encouraged to do so! (Pay attention to the case $B = \pm\infty$.)

The proof given above belongs to a classical proof pattern: To prove that a space X satisfies some property, one first prove it in a convenient case. Then, in the general case, one finds an "isomorphism" (i.e. "equivalence" in a suitable sense) $\varphi : X \to Y$ where Y is in the convenient case. Then the result on Y can be translated via φ^{-1} to X, finishing the proof.

For example, to solve a linear algebra problem about linear maps between finite-dimensional vector spaces V, W, one first proves it in the special case that $V = \mathbb{F}^m$ and $W = \mathbb{F}^n$. Then, the general case can be translated to the special case via an equivalence as in Exp. 1.21. $\qquad\square$

Exercise 3.13. Let X, Y be posets. Let $\varphi : X \to Y$ be an increasing bijection whose inverse is also increasing. (Namely, φ induces an equivalence of posets). Suppose $E \subset X$ has supremum $\sup E$. Explain why $\varphi(E)$ has supremum $\varphi(\sup E)$.

It is now fairly easy to prove the famous

Theorem 3.14 (Bolzano-Weierstrass)**.** *Let* $[a_1, b_1], \dots, [a_N, b_N]$ *be closed intervals in* $\overline{\mathbb{R}}$. *Then* $[a_1, b_1] \times \cdots \times [a_N, b_N]$ *is sequentially compact.*

Proof. By Prop. 3.4, it suffices to assume $N = 1$. Write $a_1 = a, b_1 = b$. Let (x_n) be a sequence in $[a, b]$. By Thm. 3.11, (x_n) has a subsequence (x_{n_k}) converging to some $x \in \overline{\mathbb{R}}$. (E.g. $x = \limsup_n x_n$.) Since $a \leqslant x_{n_k} \leqslant b$, we have $a \leqslant x \leqslant b$ by Rem. 2.71. $\square$

Bolzano-Weierstrass theorem illustrates why we sometimes prefer to work with $\overline{\mathbb{R}}$ instead of $\mathbb{R}$: $\overline{\mathbb{R}}$ is sequentially compact, while $\mathbb{R}$ is not. That every sequence has limits superior and inferior in $\overline{\mathbb{R}}$ but not necessarily in $\mathbb{R}$ is closely related to this fact. In the language of point-set topology, $\overline{\mathbb{R}}$ is a **compactification** of $\mathbb{R}$.

Bolzano-Weierstrass theorem (restricted to $\mathbb{R}^N$) will be generalized to **Heine-Borel theorem**, which says that a subset of $\mathbb{R}^N$ is sequentially compact iff it is bounded and closed (cf. Def. 3.26 for the definition of closed subsets). See Thm. 3.55.

3.1.3 A criterion for convergence in sequentially compact spaces

At the end of Sec. 2.3, we have raised the following question: Suppose that (x_n) is a bounded sequence in a metric space X such that any two convergent subsequences converge to the same point. Does (x_n) converge?

When X is sequentially compact, (x_n) is automatically bounded due to Prop. 3.7. The answer to the above question is yes:

Theorem 3.15. *Let* X *be a sequentially compact metric space. Let* (x_n) *be a sequence in* X. *Then the following are equivalent.*

(1) *The sequence* (x_n) *converges in* X.

(2) *Any two convergent subsequences of* (x_n) *converge to the same point. In other words,* (x_n) *has only one cluster point.*

Proof. (1)$\Rightarrow$(2): By Prop. 2.36.

(2)$\Rightarrow$(1): Assume that (x_n) has at most one cluster point. Since X is sequentially compact, (x_n) has at least one cluster point $x \in X$. We want to prove $\lim_{n \to \infty} x_n = x$. Suppose not. Then there exists $\varepsilon > 0$ such that for every $N \in \mathbb{Z}_+$ there is $n \geqslant N$ such that $d(x_n, x) \geqslant \varepsilon$. Thus, one can inductively construct a subsequence (x_{n_k}) of (x_n) such that $d(x_{n_k}, x) \geqslant \varepsilon$ for all k. Since X is sequentially compact, (x_{n_k}) has a subsequence x'_n converging to $x' \in X$. So $d(x'_n, x) \geqslant \varepsilon$ for all n. Since the function $y \in X \mapsto d(y, x)$ is continuous (Exp. 2.58), we have $\lim_{n \to \infty} d(x'_n, x) = d(x', x)$. This proves that $d(x', x) \geqslant \varepsilon > 0$. However, x', x are both cluster points of (x_n), and so $x = x'$. This gives a contradiction. $\square$

Remark 3.16. Thm. 3.15 can be used in the following way. Suppose that we want to prove that a given sequence (x_n) in a sequentially compact space X converges to x. Then it suffices to prove that if (x'_n) is a subsequence of (x_n) converging to some $y \in X$, then $y = x$. This is sometimes easier to prove than directly proving the convergence of (x_n). We will use this strategy in the proof of L'Hôpital's rule, for example. (See Subsec. 12.1.2.)

Corollary 3.17. *Let (x_n) be a sequence in $\mathbb{R}^N$. The following are equivalent.*

(1) The sequence (x_n) converges in $\mathbb{R}^N$.

(2) The sequence (x_n) is bounded. Moreover, any two convergent subsequences of (x_n) converge to the same point of $\mathbb{R}^N$.

Proof. (1)$\Rightarrow$(2): By Prop. 2.33 and 2.36.

(2)$\Rightarrow$(1): Assume (2). Since (x_n) is bounded, it can be contained in $X = I_1 \times \cdots \times I_N$ where each I_i is a closed interval in $\mathbb{R}$. Clearly, any two cluster points of (x_n) are inside X, and are equal by (2). By Bolzano-Weierstrass, X is sequentially compact. Thus, by Thm. 3.15, (x_n) converges in X and hence in $\mathbb{R}^N$. $\qquad\square$

Corollary 3.18. *The following are true.*

1. *Let (x_n) be a sequence in $\overline{\mathbb{R}}$. Then (x_n) converges in $\overline{\mathbb{R}}$ iff $\limsup_{n\to\infty} x_n$ equals $\liminf_{n\to\infty} x_n$.*

2. *Let (x_n) be a sequence in $\mathbb{R}$. Then (x_n) converges in $\mathbb{R}$ iff $\limsup_{n\to\infty} x_n$ equals $\liminf_{n\to\infty} x_n$ and (x_n) is bounded.*

Note that if (x_n) converges in $\overline{\mathbb{R}}$, we must have $\lim x_n = \limsup x_n = \liminf x_n$ by Thm. 3.11.

Proof. 1. Let $A = \liminf x_n$ and $B = \limsup x_n$. Let S be the set of cluster points of (x_n). By Thm. 3.11, $A = \min S, B = \max S$. So $A = B$ iff S has only one element. This is equivalent to the convergence of (x_n) in $\overline{\mathbb{R}}$ due to Thm. 3.15 (since $\overline{\mathbb{R}}$ is sequentially compact by Bolzano-Weierstrass.)

2. If (x_n) converges, then $A = B$ by part 1. And (x_n) is bounded due to Prop. 2.33. Conversely, if $A = B$ and if (x_n) is bounded, say $\alpha \leqslant x_n \leqslant \beta$ for all n where $-\infty < \alpha < \beta < +\infty$. Then $\alpha \leqslant A \leqslant B \leqslant \beta$. So $A, B \in \mathbb{R}$. By part 1, (x_n) converges to $A \in \mathbb{R}$. $\qquad\square$

3.2 Outlook: sequentially compact function spaces

In Sec. 2.1, we mentioned that metric spaces and (more generally) point-set topology were introduced by mathematicians in order to study (typically infinite dimensional) function spaces with the help of the geometric intuition of $\mathbb{R}^N$. Now we have learned a couple of important results about sequentially compact spaces. But we have not met any example arising from function spaces. So let me show one example to the curious readers: The product space $[0,1]^{\mathbb{Z}_+}$, equipped with the metric defined in Pb. 2.3, is sequentially compact. We will prove this result at the end of this chapter. (Indeed, we will prove a slightly more general version. See Thm. 3.54.) This is a famous result, not only because it has many important applications (some of which will be hinted at in this section), but also because its proof uses the clever "diagonal method".

Moreover, we will later prove an even more surprising fact: every sequentially compact metric space is homeomorphic to a closed subset of $[0,1]^{\mathbb{Z}_+}$. (See Thm. 8.45.) Thus, all sequentially compact metric spaces can be constructed explicitly, in some sense.

The readers may still complain that functions on $\mathbb{Z}_+$ are very different from those we often see and use in analysis and (especially) in differential equations: We are ultimately interested in functions on $\mathbb{R}$ or on $[a,b]$, but not on countable sets. This is correct. But $[0,1]^{\mathbb{Z}_+}$ (and its closed subsets) are in fact very helpful for the study of spaces of functions on $\mathbb{R}$ and on $[a,b]$. In this course, we shall learn two major examples that the sequential compactness of $[0,1]^{\mathbb{Z}_+}$ helps with:

1. $A = \mathbb{Q} \cap [a,b]$ is a countable dense subset of $[a,b]$. Thus, if we let $C([a,b])$ denote the set of continuous $\mathbb{R}$-functions on $[a,b]$, then the restriction map $f \in C([a,b]) \mapsto f|_A \in \mathbb{R}^A$ is injective. In many applications, we are interested in a subset $\mathcal{X} \subset C([a,b])$ of uniformly bounded functions, say all $f \in \mathcal{X}$ take values in $[-1,1]$. Then we have an injective map

$$\Phi : \mathcal{X} \to [-1,1]^A \qquad f \mapsto f|_A$$

   If $\mathcal{X}$ satisfies a condition called "**equicontinuous**", then a sequence f_n in $\mathcal{X}$ converges *uniformly* to $f \in C([a,b])$ iff $f_n|_A$ converges *pointwise* to $f|_A$. (See Rem. 3.62.) Thus, from the sequential compactness of $[-1,1]^A$ under pointwise convergence topology, one concludes that every sequence in $\mathcal{X}$ has a subsequence converging uniformly in $C([a,b])$. This remarkable sequential compactness result on (the closure of) $\mathcal{X}$ is called **Arzelà-Ascoli theorem**, and will be used to prove the fundamental Peano existence theorem in ordinary differential equations. We also see that the fact that $[a,b]$ has a countable dense subset A plays a crucial role. This property of metric spaces is called "**separable**" and will be studied later.

2. **Fourier series** are powerful for the study of partial differential equations. A continuous function $f : [-\pi, \pi] \to \mathbb{C}$ satisfying $f(-\pi) = f(\pi)$ has Fourier

series expansion $f(x) = \sum_{n \in \mathbb{Z}} a_n e^{\mathrm{i}nx}$ where $a_n \in \mathbb{C}$. However, for the sake of studying differential equations, one needs to consider series $\sum_{n \in \mathbb{Z}} a_n e^{\mathrm{i}nx}$ converging to a function much worse than a continuous function. For example, in the study of integral equations (which are closely related to certain partial differential equations), Hilbert and Schmidt discovered that one has to consider all $f(x) = \sum_{n \in \mathbb{Z}} a_n e^{\mathrm{i}nx}$ satisfying $\sum_n |a_n|^2 \leqslant 1$. Therefore, one lets $\overline{B} = \{z \in \mathbb{C} : |z| \leqslant 1\}$ and considers $\widehat{f} : n \in \mathbb{Z} \mapsto a_n \in \mathbb{C}$ as an element of $\overline{B}^{\mathbb{Z}}$. The sequential compactness of $\overline{B}^{\mathbb{Z}}$ helps one find the $\widehat{f}$ such that the corresponding $f(x) = \sum_n \widehat{f}(n) \cdot e^{\mathrm{i}nx}$ is a desired solution of the integral equation.

3.3 Complete metric spaces and Banach spaces

In this section, we let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, and assume that all vector spaces are over $\mathbb{F}$.

3.3.1 Cauchy sequences and complete metric spaces

Definition 3.19. A sequence (x_n) in a metric space X is called a **Cauchy sequence**, if:

- For every $\varepsilon > 0$ there exists $N \in \mathbb{Z}_+$ such that for all $m, n \geqslant N$ we have $d(x_m, x_n) < \varepsilon$.

Here, "$\varepsilon > 0$" can mean either "$\varepsilon \in \mathbb{R}_{>0}$" or "$\varepsilon \in \mathbb{Q}_{>0}$". The choice of this meaning does not affect the definition. The above definition can be abbreviated to "for every $\varepsilon > 0$, we have $d(x_m, x_n) < \varepsilon$ for sufficiently large m, n".

Remark 3.20. It is an easy consequence of triangle inequality that (x_n) is a Cauchy sequence iff

- For every $\varepsilon > 0$ there exists $N \in \mathbb{Z}_+$ such that for all $n \geqslant N$ we have $d(x_n, x_N) < \varepsilon$.

Also, it is clear that every Cauchy sequence is bounded.

Proposition 3.21. *Every convergent sequence in a metric space X is a Cauchy sequence.*

Proof. Assume (x_n) converges to x in X. Then for every $\varepsilon > 0$ there is $N \in \mathbb{Z}_+$ such that $d(x, x_n) < \varepsilon/2$ for all $n \geqslant N$. Since this is true for every $m \geqslant N$, we have $d(x_m, x_n) \leqslant d(x, x_n) + d(x, x_m) < \varepsilon/2 + \varepsilon/2 = \varepsilon$. $\qquad\square$

Definition 3.22. A metric space X is called **complete** if every Cauchy sequence in X converges.

We have many examples of complete metric spaces:

Theorem 3.23. *If (x_n) is a Cauchy sequence in a metric space X with at least one cluster point, then (x_n) converges in X. Consequently, every sequentially compact metric space is complete.*

Proof. Let (x_n) be a Cauchy sequence in X with subsequence (x_{n_k}) converging to $x \in X$. Let us show that $x_n \to x$.

Since (x_n) is Cauchy, for every $\varepsilon > 0$ there is $N \in \mathbb{Z}_+$ such that $d(x_n, x_m) < \varepsilon/2$ for all $m, n \geqslant N$. Since $x_{n_k} \to x$, there is $k \geqslant N$ such that $d(x_{n_k}, x) < \varepsilon/2$. Since n_k is strictly increasing over k, we have $n_k \geqslant k$. So $n_k \geqslant N$. So we can let $m = n_k$. This gives $d(x_n, x_{n_k}) < \varepsilon/2$. Therefore $d(x_n, x) \leqslant d(x_n, x_{n_k}) + d(x_{n_k}, x) < \varepsilon$ for all $n \geqslant N$. $\square$

Example 3.24. Let $X = [a_1, b_1] \times \cdots \times [a_N, b_N]$ where each $[a_i, b_i]$ is a closed interval in $\mathbb{R}$, then X is sequentially compact by Bolzano-Weierstrass. Thus, by Thm. 3.23. X is complete.

Corollary 3.25. $\mathbb{R}^N$ and $\mathbb{C}^N$ *are complete (under the Euclidean metrics).*

Proof. Since $\mathbb{C}^N$ is isometrically isomorphic to $\mathbb{R}^{2N}$, it suffices to prove that $\mathbb{R}^N$ is complete. Choose a Cauchy sequence (x_n) in $\mathbb{R}^N$. Since (x_n) is bounded, (x_n) is contained inside $X = I_1 \times \cdots \times I_N$ where each $I_i = [a, b]$ is in $\mathbb{R}$. By Exp. 3.24, X is complete. So (x_n) converges to some $x \in X$. $\square$

Definition 3.26. We say that a subset A of a metric space X is **closed** if the following condition is true: For every sequence (x_n) in A converging to a point $x \in X$, we have $x \in A$.

Thus, the word "closed" here means "closed under taking limits".

Proposition 3.27. *Let A be a metric subspace of a metric space X. Recall that the metric of A inherits from that of X (cf. Def. 2.14). Consider the statements:*

(1) A is complete.

(2) A is a closed subset of X.

Then (1)$\Rightarrow$(2). If X is complete, then (2)$\Rightarrow$(1).

Proof. First, assume that X is complete and (2) is true. Let (x_n) be a Cauchy sequence in A. Then it is a Cauchy sequence in X. So $x_n \to x \in X$ because X is complete. So $x \in A$ by the definition of closedness. This proves (1).

Next, we assume (1). Choose a sequence (x_n) in A converging to a point $x \in X$. By Prop. 3.21, (x_n) is a Cauchy sequence in X, and hence a Cauchy sequence in A. Since A is complete, there is $a \in A$ such that $x_n \to a$. So we must have $x = a$ because any sequence has at most one limit in a metric space. This proves $x \in A$. So (2) is proved. $\square$

A similar result holds for sequential compactness. See Pb. 3.4.

Example 3.28. Let $-\infty < a < b < +\infty$. Then (a, b) is not complete (under the Euclidean metric), because (a, b) is not closed in the metric space $\mathbb{R}$. (For sufficiently large n, $b - 1/n$ is in (a, b), but $\lim_{n\to\infty}(b - 1/n) = b$ is not in b.)

Example 3.29. By Prop. 1.29, for each $x \in \mathbb{R}\backslash\mathbb{Q}$, we can choose an increasing sequence in $\mathbb{Q}$ converging to x. So $\mathbb{Q}$ is not closed in $\mathbb{R}$. So $\mathbb{Q}$ is not complete under the Euclidean topology.

Example 3.30. Let X be a metric space. Let $p \in X$ and $0 \leqslant R < +\infty$. Then $\overline{B}_X(x, R)$ is a closed subset of X. Therefore, if X is complete, then $\overline{B}_X(p, R)$ is complete by Prop. 3.27.

Proof of closedness. Let (x_n) be a sequence in $\overline{B}(p, R)$ converging to $x \in X$. Then $d(p, x_n) \leqslant R$. Since the function $y \in X \mapsto d(p, y) \in \mathbb{R}$ is continuous (Exp. 2.58), we have $d(p, x) = \lim_{n\to\infty} d(p, x_n) \leqslant R$. So $x \in \overline{B}(p, R)$. $\qquad\square$

Exercise 3.31. Let d, δ be two equivalent metrics on a set X. Show that a sequence (x_n) in X is Cauchy under d iff (x_n) is Cauchy under δ.

Note that if, instead of assuming d, δ are equivalent, we only assume that d, δ are topologically equivalent. Then the above conclusion is not necessarily true:

Exercise 3.32. Find a non-complete metric δ on $\mathbb{R}$ topologically equivalent to the Euclidean metric.

3.3.2 Normed vector spaces and Banach spaces

A major application of complete metric spaces is to show that many series converge without knowing to what exact values these series converge. A typical example is the convergence of $\sum_{n\in\mathbb{Z}_+} \sin(\sqrt{2}n)/n^2$ in $\mathbb{R}$. We are also interested in the convergence of series in function spaces, for instance: the uniform convergence of $f(x) = \sum_{n\in\mathbb{Z}_+} \sin(\sqrt{2}nx^3)/n^2$ on $\mathbb{R}$; a suitable convergence of the Fourier series $\sum_{n\in\mathbb{Z}} a_n e^{inx}$. But we cannot take sum in a general metric space since it has no vector space structures. Therefore, we need a notion which combines complete metric spaces with vector spaces. Banach spaces are such a notion.

Definition 3.33. Let V be a vector space over $\mathbb{F}$ with zero vector 0_V. A function $\|\cdot\| : V \to \mathbb{R}_{\geqslant 0}$ is called a **norm** if for every $u, v \in V$ and $\lambda \in \mathbb{F}$, the following hold:

- (Subadditivity) $\|u + v\| \leqslant \|u\| + \|v\|$.

- (Absolute homogeneity) $\|\lambda v\| = |\lambda| \cdot \|v\|$. In particular, (by taking $\lambda = 0$) we have $\|0_V\| = 0$.

- If $\|v\| = 0$ then $v = 0_V$.

We call $(V, \|\cdot\|)$ (often abbreviated to V) a **normed vector space**.

Remark 3.34. To check the absolute homogeneity, it suffices to check

$$\|\lambda v\| \leqslant |\lambda| \cdot \|v\|$$

for all λ and v. Then clearly $\|\lambda v\| = |\lambda| \cdot \|v\|$ when $\lambda = 0$. Suppose $\lambda \neq 0$. Then

$$\|v\| = \|\lambda^{-1} \lambda v\| \leqslant |\lambda|^{-1} \|\lambda v\|$$

which implies $\|\lambda v\| = |\lambda| \cdot \|v\|$.

Remark 3.35. Let V be a vector space. If V is a normed vector space, then

$$d(u, v) = \|u - v\| \tag{3.4}$$

clearly defines a metric. (Note that triangle inequality follows from subadditivity.) Unless otherwise stated, we always assume that the metric of a normed vector space is defined by (3.4).

Definition 3.36. Let V be a normed vector space. We say that V is a **Banach space** if V is a complete metric space where the metric is the canonical one (3.4). If V is over the field $\mathbb{C}$ (resp. $\mathbb{R}$), we call V a **complex** (resp. **real**) **Banach space**.

Example 3.37. We always assume that the norm on $\mathbb{F}^N$ is the **Euclidean norm**

$$\|(a_1, \ldots, a_N)\| = \sqrt{|a_1|^2 + \cdots + |a_N|^2} \tag{3.5}$$

The canonical metric it gives is the Euclidean metric. Thus, by Cor. 3.25, $\mathbb{F}^N$ is a Banach space.

If (λ_n) is a sequence in $\mathbb{F}$ converging to λ, and if (x_n) is a sequence in $\mathbb{F}^N$ converging to x, then one can show that $\lambda_n x_n$ converges to λx by checking that each component of $\lambda_n x_n$ converges to the corresponding component of λx. This is due to Prop. 2.27. However, if (x_n) is in general a sequence in a normed vector space, this method fails. So we need a different argument:

Proposition 3.38. *Let V be a normed vector space. The following maps are continuous*

$$
\begin{aligned}
+ &: V \times V \to V & (u, v) &\mapsto u + v \\
- &: V \times V \to V & (u, v) &\mapsto u - v \\
\times_{\mathbb{F}} &: \mathbb{F} \times V \to V & (\lambda, v) &\mapsto \lambda v \\
\|\cdot\| &: V \to \mathbb{R}_{\geqslant 0} & v &\mapsto \|v\|
\end{aligned}
$$

We didn't mention the continuity of the division map $(\lambda, v) \in \mathbb{F}^\times \times V \mapsto \lambda^{-1} v$ since it follows from that of $\times_{\mathbb{F}}$ and of the inversion map $\lambda \mapsto \lambda^{-1}$ by Exp. 2.44.

Proof. One can check that the addition map, the subtraction map, and the last map $\|\cdot\|$ are Lipschitz continuous.

Define metric $d((\lambda, v), (\lambda', v')) = \max\{|\lambda - \lambda'|, \|v - v'\|\}$ on $\mathbb{F} \times V$. Then $\mathbb{F} \times V$ is covered by open balls of the form $B(0, r) = \{(\lambda, v) \in \mathbb{F} \times V : |\lambda| < r, \|v\| < r\}$. Similar to the argument in (2.6), one uses subadditivity (i.e. triangle inequality) and absolute homogeneity to show that $\times_{\mathbb{F}}$ has Lipschitz constant $2r$ on $B(0, r)$. So $\times_{\mathbb{F}}$ is continuous by Lem. 2.41 and 2.43. $\qquad\square$

3.4   The Banach spaces $l^\infty(X, V)$ and $C(X, V)$

In this section, we let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and assume that the vector spaces V are over $\mathbb{F}$. As the title suggests, in this section we shall introduce two important examples of Banach spaces: the space of bounded functions $l^\infty(X, V)$ and its subspace of continuous functions $C(X, V)$ (when X is a sequentially compact metric space). In order for these two spaces to be Banach spaces, we must assume that V is also Banach.

In application, the main examples are $V = \mathbb{R}, \mathbb{C}, \mathbb{R}^N, \mathbb{C}^N$. Indeed, $C([a, b], \mathbb{R}^N)$ is one of the main examples of function spaces considered by Fréchet when he defined metric spaces. Therefore, the readers can assume that V is one of such spaces if they want to make life easier. Just keep in mind that we sometimes also consider the case where V itself is a function space.

Definition 3.39. Let X be a set and let V be a vector space. The set V^X is a vector space if we define for each $f, g \in V^X$ and $\lambda \in \mathbb{F}$:

$$f + g : X \to V \qquad (f + g)(x) = f(x) + g(x)$$
$$\lambda f : X \to V \qquad (\lambda f)(x) = \lambda f(x)$$

We also define the **absolute value function**

$$|f| : X \to \mathbb{R}_{\geqslant 0} \qquad x \in X \mapsto \|f(x)\| \tag{3.6}$$

The symbol $|f|$ is sometimes also written as $\|f\|$ when it will not be confused with $\|f\|_\infty$ or other norms of f.

Definition 3.40. Let X be a set and let V be a normed vector space. For each $f \in V^X$, define the l^∞**-norm**

$$\|f\|_{l^\infty(X,V)} \equiv \|f\|_{l^\infty} \equiv \|f\|_\infty = \sup_{x \in X} \|f(x)\| \tag{3.7}$$

where $\|f(x)\|$ is defined by the norm of V. Define the l^∞**-space**

$$l^\infty(X, V) = \{f \in V^X : \|f\|_\infty < +\infty\} \tag{3.8}$$

which is a vector subspace of V^X. Then $l^\infty(X, V)$ is a normed vector space under the l^∞-norm. A function $f : X \to V$ is called **bounded** if $f \in l^\infty(X, V)$.

Exercise 3.41. Prove that for every $f, g \in V^X$ and $\lambda \in \mathbb{F}$ we have

$$\begin{aligned}
\|f + g\|_\infty &\leqslant \|f\|_\infty + \|g\|_\infty \\
\|\lambda f\|_\infty &= |\lambda| \cdot \|f\|_\infty
\end{aligned} \tag{3.9}$$

(Note that clearly we have that $\|f\|_\infty = 0$ implies $f = 0$.) Here, we understand $0 \cdot (+\infty) = 0$. Use these relations to verify that $l^\infty(X, V)$ is a linear subspace of V^X (i.e. it is closed under addition and scalar multiplication) and that $\|\cdot\|_\infty$ is a norm on $l^\infty(X, V)$. □

Definition 3.42. Let V be a normed vector space. We say that a sequence (f_n) in V^X **converges uniformly** to $f \in V^X$ if $\lim_{n\to\infty} \|f - f_n\|_\infty = 0$. In this case, we write $f_n \rightrightarrows f$.

We say that (f_n) **converges pointwise** to $f \in V^X$ if for every $x \in X$ we have $\lim_{n\to\infty} f_n(x) = f(x)$, i.e. $\lim_{n\to\infty} \|f_n(x) - f(x)\| = 0$.

The same definition will be applied to nets $(f_\alpha)_{\alpha \in I}$ in V^X after learning net convergence in Sec. 5.2. □

In more details, the uniform convergence of f_n to f means that "for every $\varepsilon > 0$ there is $N \in \mathbb{Z}_+$ such that for all $n \geqslant N$ and *for all $x \in X$*, we have $\|f_n(x) - f(x)\| < \varepsilon$". If we place the words "for all $x \in X$" at the very beginning of the sentence, we get pointwise convergence.

Uniform convergence implies pointwise convergence: If $\|f - f_n\|_\infty \to 0$, then for each $x \in X$ we have $\|f_n(x) - f(x)\| \to 0$ since $\|f(x) - f_n(x)\| \leqslant \|f - f_n\|_\infty$.

Example 3.43. Let $f_n : (0, 1) \to \mathbb{R}$ be $f_n(x) = x^n$. Then f_n converges pointwise to 0 (cf. Exp. 4.10). But $\sup_{x \in (0,1)} |x^n - 0| = 1$ does not converge to 0. So f_n does not converge uniformly to 0.

Remark 3.44. The uniform convergence of sequences in $l^\infty(X, V)$ is induced by the l^∞-norm, and hence is induced by the metric $d(f, g) = \|f - g\|_\infty$. However, this formula cannot be extended to a metric on V^X, since for arbitrary $f, g \in V^X$, $\|f - g\|_\infty$ is possibly $+\infty$.

In fact, it is true that the uniform convergence of sequences in V^X is induced by a metric, see Pb. 3.6. When X is countable, we have seen in Pb. 2.3 that the pointwise convergence in V^X is also given by a metric. □

Theorem 3.45. *Let X be a set, and let V be a Banach space (over $\mathbb{F}$). Then $l^\infty(X, V)$ is a Banach space (over $\mathbb{F}$).*

Proof. Let (f_n) be a Cauchy sequence in $l^\infty(X, V)$. Then for every $\varepsilon > 0$ there is $N \in \mathbb{Z}_+$ such that for all $m, n \geqslant N$ we have that $\sup_{x \in X} \|f_n(x) - f_m(x)\| < \varepsilon$, and hence $\|f_n(x) - f_m(x)\| < \varepsilon$ for each $x \in X$. This shows that for each $x \in X$, $(f_n(x))$ is a Cauchy sequence in V, which converges to some element $f(x) \in V$ because V is complete.

We come back to the statement that for each $\varepsilon > 0$, there exists $N \in \mathbb{Z}_+$ such that for all $n \geqslant N$ and all x,

$$\|f_n(x) - f_m(x)\| < \varepsilon$$

for every $m \geqslant N$. Let $m \to \infty$. Then by the continuity of subtraction and taking norm (cf. Prop. 3.38.), we obtain $\|f_n(x) - f(x)\| \leqslant \varepsilon$ for all $n \geqslant N$ and $x \in X$. In other words, $\|f_n - f\|_\infty \leqslant \varepsilon$ for all $n \geqslant N$. In particular, $\|f\|_\infty \leqslant \|f_N\|_\infty + \|f_N - f\|_\infty < +\infty$ by (3.9). This proves $f \in l^\infty(X, V)$ and $f_n \rightrightarrows f$. $\qquad \square$

Mathematicians used to believe that "if a sequence of continuous functions $f_n : [0, 1] \to \mathbb{R}$ converges pointwise to a function $f : [0, 1] \to \mathbb{R}$, then f is continuous". Cauchy, one of the main figures in 19th century working on putting analysis on a rigorous ground, has given a problematic proof of this wrong statement. Counterexamples were later found in the study of Fourier series: Let $f : \mathbb{R} \to \mathbb{R}$ be a function with period 2π such that $f(x) = x$ when $-\pi < x < \pi$, and $f(x) = 0$ when $x = \pm\pi$. Then the Fourier series of this noncontinuous function f converges pointwise to f, yet the partial sums of this series are clearly continuous functions. Later, it was realized that uniform convergence is needed to show the continuity of the limit function. (See Thm. 3.48.) This was the first time the importance of uniform convergence was realized.

The following discussions about (resp. sequentially compact) metric spaces also apply to general (resp. compact) topological spaces. The reader can come back and check the proofs for these more general spaces after studying them in the future.

Definition 3.46. Let X, Y be metric spaces (resp. topological spaces). Then $C(X, Y)$ denotes the set of continuous functions from X to Y.

Lemma 3.47. *Let X be a metric space (resp. a topological space), and let V be a normed vector space. Then $C(X, V)$ is a linear subspace of V^X. If X is sequentially compact (resp. compact), then $C(X, V)$ is a linear subspace of $l^\infty(X, V)$.*

Proof. Using Prop. 3.38, one checks easily that $C(X, V)$ is a linear subspace of V^X. For any $f \in C(X, V)$, the absolute value function $|f| : x \in X \mapsto \|f(x)\|$ is continuous. Thus, assuming that X is sequentially compact, then by Lem. 3.2, $|f|$ is bounded on X. This proves that $\|f\|_\infty < +\infty$. Thus $C(X, V)$ is a subset (and hence a linear subspace) of $l^\infty(X, V)$. $\qquad \square$

Theorem 3.48. *Let X be a metric space (resp. a topological space), and let V be a normed vector space. Then $C(X, V) \cap l^\infty(X, V)$ is a closed linear subspace of $l^\infty(X, V)$. In particular, if X is sequentially compact (resp. compact), then $C(X, V)$ is a closed linear subspace of $l^\infty(X, V)$.*

Proof. Choose a sequence (f_n) in $C(X, V) \cap l^\infty(X, V)$ converging in $l^\infty(X, V)$ to f. Namely, $f_n \rightrightarrows f$. We want to prove that f is continuous. We check that f satisfies Def. 2.38-(2'). (One can also use Def. 2.38-(1). The proofs using these two definitions are not substantially different.)

Fix $p \in X$. Choose any $\varepsilon > 0$. Since $f_n \rightrightarrows f$, there exists $N \in \mathbb{Z}_+$ such that for all $n \geqslant N$ and we have $\|f - f_n\|_\infty < \varepsilon$. Since f_N is continuous, there exists $r > 0$ such that for each $x \in B_X(p, r)$ we have $\|f_N(x) - f_N(p)\| < \varepsilon$. Thus, for each $x \in B_X(p, r)$ we have

$$\|f(x) - f(p)\| \leqslant \|f(x) - f_N(x)\| + \|f_N(x) - f_N(p)\| + \|f_N(p) - f(p)\| < 3\varepsilon$$

This finishes the proof. $\qquad\square$

Convention 3.49. Unless otherwise stated, if X is sequentially compact metric space (or more generally, a compact topological space to be defined latter), and if V is a normed vector space, the norm on $C(X, V)$ is chosen to be the l^∞-norm.

Corollary 3.50. *Let X be a metric space (resp. a topological space), and let V be a Banach space. Then $C(X, V) \cap l^\infty(X, V)$ is a Banach space under the l^∞-norm. In particular, if X is sequentially compact (resp. compact), then $C(X, V)$ is a Banach space.*

Proof. This follows immediately from Prop. 3.27, Thm. 3.48, and the fact that $l^\infty(X, V)$ is complete (Thm. 3.45). $\qquad\square$

3.5 Problems and supplementary material

Problem 3.1. Let (x_n) be a sequence in a metric space X. Let $x \in X$. Prove that the following are equivalent.

(1) x is a cluster point of (x_n), i.e., the limit of a convergent subsequence of (x_n).

(2) For each $\varepsilon > 0$ and each $N \in \mathbb{Z}_+$, there exists $n \geqslant N$ such that $d(x_n, x) < \varepsilon$.

(Note: in a general topological space, these two statements are not equivalent.)

Remark 3.51. Condition (2) is often abbreviated to "for each $\varepsilon > 0$, the sequence (x_n) is frequently in $B(x, \varepsilon)$". In general, we say "(x_n) **frequently** satisfies P" if for each $N \in \mathbb{Z}_+$ there is $n \geqslant N$ such that x_n satisfies P. We say that "(x_n) **eventually** satisfies P" if there exists $N \in \mathbb{Z}_+$ such that for every $n \geqslant N$, x_n satisfies P.

Thus "(x_n) eventually satisfies P" means the same as "all but finitely many x_n satisfies P". Its negation is "(x_n) frequently satisfies $\neg$P". $\qquad\square$

Remark 3.52. Condition (2) of Pb. 3.1 is sometimes easier to use than (1). For example, compared to the original definition of cluster points, it is much easier to find an explicit negation of (2) by using the rule suggested in Rem. 2.17: There exist $\varepsilon > 0$ and $N \in \mathbb{Z}_+$ such that $d(x_n, x) \geqslant \varepsilon$ for all $n \geqslant N$. (Or simply: there exists $\varepsilon > 0$ such that x_n is eventually not in $B(x, \varepsilon)$.)

Problem 3.2. Use Pb. 3.1-(2) to prove that if (x_n) is a sequence in $\overline{\mathbb{R}}$, then $\limsup\limits_{n\to\infty} x_n$ is a cluster point of (x_n).

Remark 3.53. You will notice that your proof of Pb. 3.2 is slightly simpler than the proof we gave for Thm. 3.11. This is because our construction of subsequence as in (3.3) has been incorporated into your proof of (2)$\Rightarrow$(1) in Pb. 3.1.

Problem 3.3. Let $f : X \to Y$ be a continuous map of metric spaces. Assume that f is bijective and X is sequentially compact. Prove that f is a homeomorphism using the following hint.

Hint. You need to prove that if (y_n) is a sequence in Y converging to $y \in Y$, then $x_n = f^{-1}(y_n)$ converges to $x = f^{-1}(y)$. Prove that (x_n) has only one cluster point, and hence converges to some point $x' \in X$ (why?). Then prove $x' = x$. (In the future, we will use the language of open sets and closed sets to prove this result again. Do not use this language in your solution.) $\qquad\square$

Theorem 3.54 (Tychonoff theorem, countable version). *Let $(X_n)_{n\in\mathbb{Z}_+}$ be a sequence of sequentially compact metric spaces. Then the product space $S = \prod\limits_{n\in\mathbb{Z}_+} X_n$ is sequentially compact under the metric defined as in Pb. 2.3.*

The method of choosing subsequence in the following proof is the reknowned **diagonal method**. A different method will be given in Pb. 8.7.

Proof. Let $(x_m)_{m\in\mathbb{Z}_+}$ be a sequence in S. Since $(x_m(1))_{m\in\mathbb{Z}_+}$ is a sequence in the sequentially compact space X_1, $(x_m)_{m\in\mathbb{Z}_+}$ has a subsequence $x_{1,1}, x_{1,2}, x_{1,3} \ldots$ whose value at $n = 1$ converges in X_1. Since X_2 is sequentially compact, we can choose a subsequence $x_{2,1}, x_{2,2}, x_{2,3}, \ldots$ of the previous subsequence such that its values at $n = 2$ converge in X_2. Then pick a subsequence from the previous one whose values at 3 converge in X_3.

By repeating this process, we get an $\infty \times \infty$ matrix $(x_{i,j})_{i,j\in\mathbb{Z}_+}$:

$$
\begin{matrix}
x_{1,1} & x_{1,2} & x_{1,3} & \cdots \\
x_{2,1} & x_{2,2} & x_{2,3} & \cdots \\
x_{3,1} & x_{3,2} & x_{3,3} & \cdots \\
\vdots & \vdots & \vdots & \ddots
\end{matrix}
\tag{3.10}
$$

such that the following hold:

- The 1-st line is a subsequence of the original sequence $(x_m)_{m\in\mathbb{Z}_+}$.

- The $(i + 1)$-th line is a subsequence of the i-th line.

- For each n, $\lim_{j\to\infty} x_{n,j}(n)$ converges in X_n.

Then the diagonal line $(x_{i,i})_{i \in \mathbb{Z}_+}$ is a subsequence of the original sequence $(x_m)_{m \in \mathbb{Z}_+}$. Moreover, for each n, $(x_{i,i})_{i \geqslant n}$ is a subsequence of the n-th line, whose value at n therefore converges in X_n. Thus $\lim_{i \to \infty} x_{i,i}(n)$ converges in X_n. Thus, by Pb. 2.3, $(x_{i,i})_{i \in \mathbb{Z}_+}$ converges under any metric inducing the product topology. □

Problem 3.4. Let X be a sequentially compact metric space. Let $A \subset X$ be a metric subspace. Consider the statements:

(1) A is sequentially compact.

(2) A is a closed subset of X.

Prove that (1)⇒(2). Prove that if X is sequentially compact, then (2)⇒(1).

The above problem implies immediately:

Theorem 3.55 (Heine-Borel theorem). *Let A be a subset of $\mathbb{R}^N$. Then A is sequentially compact iff A is a bounded closed subset of $\mathbb{R}^N$.*

Proof. Suppose that A is sequentially compact. Then A is bounded under the Euclidean metric by Prop. 3.7. By Pb. 3.4, A is a closed subset of $\mathbb{R}^N$.

Conversely, assume that A is a bounded and closed subset of $\mathbb{R}^N$. Then $A \subset B$ where B is the product of N pieces of closed intervals in $\mathbb{R}$. Then B is sequentially compact by Bolzano-Weierstrass. Since A is closed in $\mathbb{R}^N$, it is not hard to check that A is closed in B.[1] Thus A is sequentially compact by Pb. 3.4. □

Example 3.56. Choose any $p \in \mathbb{R}^N$ and $0 \leqslant R < +\infty$. Then $\overline{B}_{\mathbb{R}^N}(p, R)$ is a bounded closed subset of $\mathbb{R}^N$ (Exp. 3.30), and hence is sequentially compact by Heine-Borel.

Remark 3.57. Think about the question: Equip $\mathbb{R}^{\mathbb{Z}_+}$ with metric

$$d(x, y) = \sup_{n \in \mathbb{Z}_+} \frac{\min\{|x(n), y(n)|, 1\}}{n}$$

What are the sequentially compact subsets of $\mathbb{R}^{\mathbb{Z}_+}$? (Namely, think about how to generalize Heine-Borel theorem to $\mathbb{R}^{\mathbb{Z}_+}$.)

Problem 3.5. Do Exercise 3.41.

Problem 3.6. Let V be a normed vector space. For every $f, g \in V^X$ define

$$d(f, g) = \min\{1, \|f - g\|_\infty\} \tag{3.11}$$

1. Show that d defines a metric on V^X.

---

[1]If Z is a metric space, if $X \subset Y \subset Z$, and if X is closed in Z, then it is easy to check that X is closed in Y.

2. Show that for every sequence (f_n) in V^X and every $g \in V^X$, we have $f_n \to g$ under the metric d iff $f_n \rightrightarrows g$.

Definition 3.58. Let X be a set, and let V be a normed vector space. A metric on V^X is called a **uniform convergence metric** if it is equivalent to (3.11). Thus, by Def. 2.61, a uniform convergence metric is one such that a sequence (f_n) in V^X converges to f under this metric iff $f_n \rightrightarrows f$.

Problem 3.7. Let X, Y be metric spaces, and assume that Y is sequentially compact. Let V be a normed vector space. Choose $f \in C(X \times Y, V)$, i.e., $f : X \times Y \to V$ is continuous. For each $x \in X$, let

$$f_x : Y \to V \qquad y \mapsto f(x, y)$$

Namely $f_x(y) = f(x, y)$. It is easy to check that $f_x \in C(Y, V)$. Define a new function

$$\Phi(f) : X \to C(Y, V) \qquad x \mapsto f_x \tag{3.12}$$

Recall that $C(Y, V)$ is equipped with the l^∞-norm.

1. Prove that $\Phi(f)$ is continuous. In other words, prove that if (x_n) is a sequence in X converging to $x \in X$, then $f_{x_n} \rightrightarrows f_x$ on Y, i.e.

$$\lim_{n \to \infty} \|f_{x_n} - f_x\|_{l^\infty(Y,V)} = 0$$

⋆ 2. Give an example of $f \in C(X \times Y, \mathbb{R})$ where Y is not sequentially compact, (x_n) converges to x in X, and f_{x_n} does not converge uniformly to f_x. (Note: you may consider $X = Y = \mathbb{R}$.)

Hint. In part 1, to prove that $\Phi(f)$ is continuous, one can prove the equivalent fact that for every fixed $x \in X$ the following is true:

- For every $\varepsilon > 0$ there exists $\delta > 0$ such that for all $p \in B_X(x, \delta)$, we have $\sup_{y \in Y} \|f(p, y) - f(x, y)\| < \varepsilon$.

(Cf. Def. 2.38.) Prove this by contradiction and by using the sequential compactness of Y appropriately. □

Remark 3.59. Let $X = \mathbb{Z}_+ \cup \{\infty\}$, equipped with the metric

$$d(m, n) = |m^{-1} - n^{-1}|$$

In other words, the metric on X is $\tau^* d_\mathbb{R}$ where $d_\mathbb{R}$ is the Euclidean metric on $\mathbb{R}$, and $\tau : X \to \mathbb{R}, n \mapsto n^{-1}$. It is not hard to show that X is sequentially compact: either prove it directly, or apply Heine-Borel to $\tau(X)$.

Let Y be a metric space. Let $(y_n)_{n \in \mathbb{Z}_+}$ be a sequence in Y, and let $y_\infty \in Y$. It is not hard to see that the following two statements are equivalent:

(1) The function $F : X \to Y, n \mapsto y_n$ is continuous.

(2) The sequence $(y_n)_{n \in \mathbb{Z}_+}$ converges to y_∞.

The following problem is a generalization of this equivalence. □

$\star$ **Problem 3.8.** Let V be a normed vector space. Let Y be a metric space. Let $X = \mathbb{Z}_+ \cup \{\infty\}$ with metric defined as in Rem. 3.59. Let $(f_n)_{n \in \mathbb{Z}_+}$ be a sequence in $C(Y, V)$. Let $f_\infty \in V^Y$. Prove that the following are equivalent:

(1) The following function is continuous:

$$F : X \times Y \to V \qquad (n, y) \mapsto f_n(y) \tag{3.13}$$

In particular, by restricting F to $\infty \times Y$, we see that $f_\infty \in C(Y, V)$.

(2) $(f_n)_{n \in \mathbb{Z}_+}$ converges pointwise to f_∞. Moreover, $(f_n)_{n \in \mathbb{Z}_+}$ is **pointwise equicontinuous**, which means the following:

- For every $y \in Y$ and every $\varepsilon > 0$, there exists $\delta > 0$ such that for all $p \in B_Y(y, \delta)$ we have

$$\sup_{n \in \mathbb{Z}_+} \| f_n(p) - f_n(y) \| < \varepsilon$$

Note. In part (1), the only nontrivial thing to prove is that F is continuous at (∞, y) for every $y \in Y$.

Remark 3.60. There is a concise way to define pointwise equicontinuity: a sequence $(f_n)_{n \in \mathbb{Z}_+}$ in V^Y is pointwise equicontinuous iff the function

$$Y \mapsto V^{\mathbb{Z}_+} \qquad y \mapsto (f_1(y), f_2(y), \dots) \tag{3.14}$$

is continuous, where $V^{\mathbb{Z}_+}$ is equipped with any uniform convergence metric (cf. Def. 3.58).

$\star$ **Remark 3.61.** In Pb. 3.8, there is a quick and tricky way to conclude (1)$\Rightarrow$(2): Use Pb. 3.7 and the sequential compactness of X. (Do not use this method in your solution. Prove (1)$\Rightarrow$(2) directly; it is a good exercise and is not difficult.)

$\star$ **Remark 3.62.** Pb. 3.7 and 3.8, together with Thm. 3.48, imply the following fact (can you see why?):

- Let Y be a sequentially compact metric space. Let V be a normed vector space. Let $(f_n)_{n \in \mathbb{Z}_+}$ be a pointwise equicontinuous sequence of functions $Y \to V$ converging pointwise to some $f : Y \to V$. Then $f_n \rightrightarrows f$ on Y.

You can also try to give a straightforward proof of this fact without using Pb. 3.7 and 3.8.

4 Series

In this chapter, we assume that vector spaces are over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ unless otherwise stated.

4.1 Definitions and basic properties

Definition 4.1. Let V be a Banach space (over $\mathbb{F}$). A **series** in V is an expression of the form

$$\sum_{i=1}^{\infty} v_i \tag{4.1}$$

where $(v_i)_{i \in \mathbb{Z}_+}$ is a sequence in V. If $s \in V$, we say that the series (4.1) **converges to** s if

$$s = \lim_{n \to \infty} \sum_{i=1}^{n} v_i$$

namely, $s_n \to s$ where s_n is the **partial sum** $s_n = \sum_{i=1}^{n} v_i$. In this case, we write

$$s = \sum_{i=1}^{\infty} v_i$$

Remark 4.2. Since V is complete, the series (4.1) converges iff the sequence of partial sum (s_n) is a Cauchy sequence: for every $\varepsilon > 0$ there exists $N \in \mathbb{Z}_+$ such that for all $n > m \geqslant N$ we have $\|s_n - s_m\| < \varepsilon$, i.e.,

$$\left\| \sum_{i=m+1}^{n} v_i \right\| < \varepsilon \tag{4.2}$$

Proposition 4.3. *Suppose that $\sum_{i=1}^{\infty} v_i$ is a convergent series in a Banach space V. Then $\lim_{n \to \infty} v_n = 0$.*

Proof. Let $s_n = v_1 + \cdots + v_n$, which converges to $s \in V$. Then $\lim_{n \to \infty} s_{n+1} = s$. So $v_n = s_{n+1} - s_n \to s - s = 0$ since subtraction in continuous (Prop. 3.38). $\qquad \square$

Thus, for example, $\sum_{n=1}^{\infty} (-1)^n$ diverges in the Banach space $\mathbb{R}$ since $\lim_{n \to \infty} (-1)^n$ does not converge to 0.

Definition 4.4. Consider a **series** in $\overline{\mathbb{R}}_{\geqslant 0}$:

$$\sum_{i=1}^{\infty} a_i \tag{4.3}$$

namely, each a_i is in $\overline{\mathbb{R}}_{\geqslant 0}$. Note that the partial sum $s_n = \sum_{i=1}^n a_i$ is increasing. We say that $\lim_{n \to \infty} s_n$ (which exists in $\overline{\mathbb{R}}_{\geqslant 0}$ and equals $\sup\{s_n : n \in \mathbb{Z}_+\}$, cf. Rem. 2.71) is the value of the series (4.3) and write

$$\sum_{i=1}^\infty a_i = \lim_{n \to \infty} s_n$$

Definition 4.5. We say that a series $\sum_{i=1}^\infty a_i$ in $\mathbb{R}_{\geqslant 0}$ **converges** if it converges in $\mathbb{R}$ (but not just converges in $\overline{\mathbb{R}}_{\geqslant 0}$, which is always true). Clearly, $\sum_{i=1}^\infty a_i$ converges iff

$$\sum_{i=1}^\infty a_i < +\infty$$

More generally, we say that a series $\sum_{i=1}^\infty v_i$ in a Banach space V **converges absolutely**, if

$$\sum_{i=1}^\infty \|v_i\| < +\infty$$

Remark 4.6. By the Cauchy condition of convergence, $\sum_{i=1}^\infty v_i$ converges absolutely iff for every $\varepsilon > 0$ there exists $N \in \mathbb{Z}_+$ such that for all $n > m \geqslant N$ we have

$$\sum_{i=m+1}^n \|v_i\| < \varepsilon \tag{4.4}$$

By comparing (4.4) with (4.2) and using the subadditivity of the norm (recall Def. 3.33), we immediately see:

Proposition 4.7. *Let $\sum_{i=1}^\infty v_i$ be a series in a Banach space. The following are true.*

1. *If $\sum_{i=1}^\infty v_i$ converges absolutely, then it converges.*

2. *For each i we choose $a_i \in \mathbb{R}_{\geqslant 0}$ satisfying $\|v_i\| \leqslant a_i$. Suppose that $\sum_{i=1}^\infty a_i < +\infty$. Then $\sum_{i=1}^\infty v_i$ converges absolutely.*

Proof. Part 1 has been explained above. In part 2, we have $\sum \|v_i\| \leqslant \sum a_i < +\infty$. So $\sum v_i$ converges absolutely. $\qquad\square$

Exercise 4.8. Suppose that $\sum_{i=1}^\infty u_i$ and $\sum_{i=1}^\infty v_i$ are convergent (resp. absolutely convergent) series in a Banach space V. Let $\lambda \in \mathbb{F}$. Show that the LHS of the following equations converges (resp. converges absolutely) in V, and that the following equations hold:

$$\sum_{i=1}^\infty (u_i + v_i) = \sum_{i=1}^\infty u_i + \sum_{i=1}^\infty v_i$$

$$\sum_{i=1}^\infty \lambda v_i = \lambda \cdot \sum_{i=1}^\infty v_i$$

Remark 4.9. We have seen that absolute convergence implies convergence. In fact, at least when $V = \mathbb{F}^N$, absolute convergence is in many ways more natural than convergence. For example, we will learn that if a series $\sum_i v_i$ in $\mathbb{F}^N$ converges absolutely, then the value of $\sum_i v_i$ is invariant under rearrangement of the series: for every bijection $\varphi : \mathbb{Z}_+ \to \mathbb{Z}_+$ we have $\sum_i v_i = \sum_i v_{\varphi(i)}$. In the next semester, we shall learn Lebesgue integral theory and, more generally, measure theory. When applying measure theory to infinite sums over the countable set $\mathbb{Z}_+$, many good results (e.g. dominated convergence theorem, Fubini's theorem) hold only for absolute convergence series, but not for arbitrary convergent series in general. In fact, there is no analog of convergent (but not absolutely convergent) series in measure theory at all!

When V is not necessarily finite-dimensional, the situation is subtler: there is a version of convergence which lies between the usual convergence and absolute convergence, and which coincides with absolute convergence when $V = \mathbb{F}^N$. This version of convergence is defined using nets instead of sequences. Moreover, many good properties (as mentioned above) hold for this convergence, and these properties can be proved in a very conceptual way (rather than using brute-force computation). We will learn this convergence in the next chapter. $\qquad\square$

4.2 Basic examples

Let us study the **geometric series** $\sum_{n=0}^{\infty} z^n$ where $z \in \mathbb{C}$. We first note the famous **binomial formula**: for each $z, w \in \mathbb{C}$ and $n \in \mathbb{N}$,

$$(z + w)^n = \sum_{k0}^{n} \binom{n}{k} z^k w^{n-k} \qquad (4.5)$$

In particular,

$$(1 + z)^n = 1 + nz + \frac{n(n-1)}{2} z^2 + \frac{n(n-1)(n-2)}{6} z^3 + \cdots + nz^{n-1} + z^n \qquad (4.6)$$

Example 4.10. Assume $z \in \mathbb{C}$ and $|z| < 1$. Then $\lim_{n\to\infty} z^n = 0$.

Proof. If $z = 0$ then it is obvious. Assume that $0 < |z| < 1$. Choose $\delta > 0$ such that $|z| = 1/(1 + \delta)$. By (4.6), $(1 + \delta)^n \geqslant 1 + n\delta$. So

$$0 \leqslant |z^n| \leqslant (1 + n\delta)^{-1}$$

Since $\lim_{n\to\infty} (1+n\delta)^{-1} = 0$, we have $|z^n| \to 0$ by squeeze theorem. Hence $z^n \to 0$. $\quad\square$

Example 4.11. Let $z \in \mathbb{C}$. If $|z| < 1$, then $\displaystyle\sum_{n=0}^{\infty} z^n$ converges absolutely, and

$$\sum_{n=0}^{\infty} z^n = \frac{1}{1 - z} \qquad (4.7)$$

where 0^0 is understood as 1. If $|z| \geqslant 1$, then $\sum_{n=0}^{\infty} z_n$ diverges in $\mathbb{C}$.

Proof. The partial sum $s_n = 1 + z + z^2 + \cdots + z^n$ equals $(1 - z^{n+1})/(1 - z)$ when $z \neq 1$. Therefore, when $|z| < 1$, $s_n \to 1/(1 - z)$. When $|z| \geqslant 1$, we have $|z^n| \geqslant 1$ and hence $z^n \not\to 0$. So $\sum_{n=0}^{\infty} z^n$ diverges by Prop. 4.3. $\square$

Example 4.12. The **harmonic series** $\sum_{n=1}^{\infty} \dfrac{1}{n}$ diverges (in $\mathbb{R}$).

Proof. We want to show that the Cauchy condition (cf. Rem. 4.2) does not hold. Thus, we want to prove that there exists $\varepsilon > 0$ such that for every $N \in \mathbb{Z}_+$ there exist $n > m \geqslant N$ such that $|(m + 1)^{-1} + (m + 2)^{-1} + \cdots + n^{-1}| \geqslant \varepsilon$.

To see this, for each N we choose $m = 2^N$ and $n = 2^{N+1}$. Then $n > m > N$, and

$$\left| \sum_{i=m+1}^{n} i^{-1} \right| = \left| \frac{1}{2^N + 1} + \frac{1}{2^N + 2} + \cdots + \frac{1}{2^N + 2^N} \right|$$

$$\geqslant \underbrace{\left| \frac{1}{2^{N+1}} + \frac{1}{2^{N+1}} + \cdots + \frac{1}{2^{N+1}} \right|}_{2^N \text{ terms}} = \varepsilon$$

where $\varepsilon = \frac{1}{2}$. $\square$

Exercise 4.13. Choose any $p \in \mathbb{Z}$. Prove that $\sum_{n=1}^{\infty} n^{-p}$ converges iff $p \geqslant 2$.

Hint. Use Prop. 4.7 and Exp. 4.12 to reduce the problem to the case $p = 2$. Prove this case by proving $\sum_{n=1}^{\infty} 1/n(n + 1) = 1 < +\infty$. $\square$

Definition 4.14. Let V be a Banach space, let X be a set, and let (f_n) be a sequence in $l^\infty(X, V)$, and let $g \in l^\infty(X, V)$. We say that the series of functions $\sum_{i=1}^{\infty} f_i$ **converges uniformly to** g (on X) if it converges to g as a series in the Banach space $l^\infty(X, V)$ and under the l^∞-norm. Equivalently, this means that the partial sum function $s_n = f_1 + \cdots + f_n$ converges uniformly to g as $n \to \infty$.

Example 4.15. The series of functions $\sum_{n=1}^{\infty} \dfrac{\sin |nz^3|}{n^2}$ converges uniformly on $\mathbb{C}$ to a continuous function $g : \mathbb{C} \to \mathbb{R}$ which is bounded (i.e. $\sup_{z \in \mathbb{C}} |g(z)| < +\infty$).

Proof. Let $f_n(z) = \sin |nz^3|/n^2$. Then each f_n is in $\mathfrak{X} = C(\mathbb{C}, \mathbb{R}) \cap l^\infty(\mathbb{C}, \mathbb{R})$ where $\mathfrak{X}$ is a real Banach space under the l^∞-norm by Cor. 3.50. Note that $\|f_n\|_\infty \leqslant n^{-2}$. By Exe. 4.13, $\sum_{n=1}^{\infty} n^{-2} < +\infty$. Therefore, by Prop. 4.7, the series $\sum_n f_n$ converges in $\mathfrak{X}$, i.e., it converges uniformly to an element $g \in \mathfrak{X}$. (In particular, $\sum_n f_n(z) = g(z)$ for all $z \in \mathbb{C}$.) $\square$

4.3 Root test and ratio test; power series; construction of e^z

Root test and ratio test are useful criteria for proving the convergence or divergence of series, especially power series. In addition, the method of power series provides a unified and elegant proof for many useful formulas about limit (see Prop. 4.23 and Exp. 4.24). We begin our discussion with the following easy observation:

Remark 4.16. Let (x_n) be a sequence in $\overline{\mathbb{R}}$, and let $A \in \overline{\mathbb{R}}$. The following are true.

1. If $\limsup\limits_{n\to\infty} x_n < A$, then $x_n < A$ is eventually true.

2. If $\limsup\limits_{n\to\infty} x_n > A$, then $x_n > A$ is frequently true.

By taking negative, we obtain similar statements for $\liminf$.

Proof. Recall that $\limsup x_n = \inf_{n\in\mathbb{Z}_+} \alpha_n$ where where $\alpha_n = \sup\{x_n, x_{n+1}, \dots\}$.

Assume that $\inf_{n\in\mathbb{Z}_+} \alpha_n < A$. Then A is not a lower bound of $\{\alpha_n : n \in \mathbb{Z}_+\}$. Thus, there exists $N \in \mathbb{Z}_+$ such that $\alpha_N < A$. Then $x_n < A$ for all $n \geqslant N$.

Assume that $\inf_{n\in\mathbb{Z}_+} \alpha_n > A$. Then for each $N \in \mathbb{Z}_+$ we have $\alpha_N > A$. So A is not an upper bound of $\{x_n, x_{n+1}, \dots\}$. So there is $n \geqslant N$ such that $x_n > A$. $\square$

We will heavily use $\sqrt[n]{x}$ (where $x \geqslant 0$ and $n \in \mathbb{Z}_+$) in the following discussions. $\sqrt[n]{x}$ will be rigorously constructed in Exp. 7.108, whose proof does not rely on the results of this section.

Proposition 4.17 (Root test). *Let* $\sum\limits_{n=1}^{\infty} v_n$ *be a series in a Banach space* V. *Let* $\beta = \limsup\limits_{n\to\infty} \sqrt[n]{\|v_n\|}$. *Then:*

1. *If* $\beta < 1$, *then* $\sum v_n$ *converges absolutely, and hence converges in* V.

2. *If* $\beta > 1$, *then* $\sum v_n$ *diverges in* V.

Proof. Suppose $\beta < 1$. Then we can choose γ such that $\beta < \gamma < 1$. So $\limsup \sqrt[n]{\|v_n\|} < \gamma$. By Rem. 3.8, there exists $N \in \mathbb{Z}_+$ such that for all $n \geqslant N$, we have $\sqrt[n]{\|v_n\|} < \gamma$, and hence $\|v_n\| < \gamma^n$. Since $\sum_{n=0}^{\infty} \gamma^n = (1 - \gamma)^{-1} < +\infty$ (Exp. 4.11), the series $\sum_{n=N}^{\infty} v_n$ converges absolutely by Prop. 4.7. So the original series converges absolutely.

Assume that $\beta > 1$. Then by Rem. 4.16, for each N there is $n \geqslant N$ such that $\sqrt[n]{\|v_n\|} > 1$ and hence $\|v_n - 0\| > 1$. So $v_n \nrightarrow 0$. So $\sum v_n$ diverges by Prop. 4.3. $\square$

Example 4.18. Let $V = \mathbb{R}$ and $v_n = 1/n$ resp. $v_n = 1/n^2$. Then $\beta = 1$, and $\sum v_n$ diverges resp. converges absolutely due to Exe. 4.13. So Root test gives no information on the convergence of series when $\beta = 1$. The same can be said about ratio test.

Proposition 4.19 (**Ratio test**)**.** *Let* $\sum\limits_{n=1}^{\infty} v_n$ *be a series in a Banach space* V *such that* $v_n \neq 0$ *for all n. Let* $\alpha = \liminf\limits_{n\to\infty} \dfrac{\|v_{n+1}\|}{\|v_n\|}$ *and* $\beta = \limsup\limits_{n\to\infty} \dfrac{\|v_{n+1}\|}{\|v_n\|}$*. Then:*

1. *If* $\beta < 1$*, then* $\sum v_n$ *converges absolutely, and hence converges in* V*.*

2. *If* $\alpha > 1$*, then* $\sum v_n$ *diverges in* V*.*

Proof. Suppose $\beta < 1$. Choose γ such that $\beta < \gamma < 1$. Then by Rem. 4.16, there is N such that for all $n \geqslant N$ we have $\|v_{n+1}\|/\|v_n\| < \gamma$. So $\|v_n\| < \gamma^{n-N}\|v_N\|$. So $\sum_{n\geqslant N}\|v_n\| \leqslant \|v_N\| \cdot \sum_{n\geqslant N}\gamma^{n-N} = \|v_N\| \cdot (1-\gamma)^{-1} < +\infty$. So $\sum v_n$ converges absolutely.

Suppose $\alpha > 1$. Then by Rem. 4.16, there is N such that for all $n \geqslant N$ we have $\|v_{n+1}\|/\|v_n\| > 1$. So $\|v_n\| \geqslant \|v_N\| > 0$ for all $n \geqslant N$. So $v_n \nrightarrow 0$ and hence $\sum v_n$ diverges, as in the proof of root test. $\square$

Definition 4.20. A **power series** in a complex Banach space V is an expression of the form $\sum\limits_{n=0}^{\infty} v_n z^n$ where the **coefficients** $v_0, v_1, v_2, \ldots$ are elements of V, and z is a **complex variable**, i.e., a symbol which can take arbitrary values in $\mathbb{C}$. If the power series $\sum v_n z^n$ converges at $z_0 \in \mathbb{C}$, we often let $\sum v_n z_0^n$ denote this limit.

Proposition 4.21. *Let* $\sum v_n z^n$ *be a power series in a complex Banach space* V*. Then there is a unique* $0 \leqslant R \leqslant +\infty$ *satisfying the following properties:*

(a) *If* $z \in \mathbb{C}$ *and* $|z| < R$*, then* $\sum v_n z^n$ *converges absolutely in* V*.*

(b) *If* $z \in \mathbb{C}$ *and* $|z| > R$*, then* $\sum v_n z^n$ *diverges in* V*.*

Such R *is called the* ***radius of convergence*** *of* $\sum v_n z^n$*. Moreover, we have*

$$R = \frac{1}{\limsup\limits_{n\to\infty} \sqrt[n]{\|v_n\|}} = \liminf\limits_{n\to\infty} \frac{1}{\sqrt[n]{\|v_n\|}} \tag{4.8}$$

Proof. Clearly, there are at most one R satisfying (a) and (b). Let us define R using (4.8) (note that the second and the third terms of (4.8) are clearly equal), and prove that R satisfies (a) and (b). Let

$$\beta(z) = \limsup\limits_{n\to\infty} \sqrt[n]{\|v_n z^n\|}$$

Then $\beta(z) = |z|/R$. So (a) and (b) follow immediately from root test. $\square$

Remark 4.22. Note that if one can find $0 \leqslant r \leqslant R$ such that $\sum v_n z^n$ converges whenever $|z| < r$, then $r \leqslant R$ where R is the radius of convergence: otherwise, the series diverges for any positive z satisfying $R < z < r$, impossible.

It follows that if $\sum v_n z^n$ converges for all $|z| < r$, then $\sum v_n z^n$ converges *absolutely* for all $|z| < r$. $\square$

Prop. 4.21 provides a useful method for computing limits of a positive sequence:

Proposition 4.23. *Let (λ_n) be a sequence in $\mathbb{R}_{>0}$. Then*

$$\liminf_{n\to\infty} \frac{\lambda_{n+1}}{\lambda_n} \leqslant \liminf_{n\to\infty} \sqrt[n]{\lambda_n} \leqslant \limsup_{n\to\infty} \sqrt[n]{\lambda_n} \leqslant \limsup_{n\to\infty} \frac{\lambda_{n+1}}{\lambda_n} \tag{4.9}$$

In particular, (by Cor. 3.18) we have

$$\lim_{n\to\infty} \sqrt[n]{\lambda_n} = \lim_{n\to\infty} \frac{\lambda_{n+1}}{\lambda_n} \tag{4.10}$$

provided that the limit on the RHS of (4.10) exists in $\overline{\mathbb{R}}$.

The four numbers in (4.9) can be completely different. See [Rud-P, Exp. 3.35].

Proof. Let R be the radius of convergence of $\sum \lambda_n z^n$. Then $R = 1/\limsup \sqrt[n]{\lambda_n}$ by (4.10). Thus, by Prop. 4.21, if $|z| > R$ then $\sum \lambda_n z^n$ diverges, and hence $\limsup |\lambda_{n+1} z^{n+1}|/|\lambda_n z^n| \geqslant 1$ by ratio test. Therefore,

$$|z| > \frac{1}{\limsup \sqrt[n]{\lambda_n}} \qquad \Longrightarrow \qquad |z| \cdot \limsup \frac{\lambda_{n+1}}{\lambda_n} \geqslant 1$$

This proves

$$\limsup \sqrt[n]{\lambda_n} \leqslant \limsup \frac{\lambda_{n+1}}{\lambda_n}$$

Replacing λ_n by λ_n^{-1}, we get

$$\frac{1}{\liminf \sqrt[n]{\lambda_n}} = \limsup \sqrt[n]{\lambda_n^{-1}} \leqslant \limsup \frac{\lambda_n}{\lambda_{n+1}} = \frac{1}{\liminf \dfrac{\lambda_{n+1}}{\lambda_n}}$$

This proves (4.9). $\qquad\qquad\square$

Example 4.24. Let $a \in \mathbb{R}_{>0}$ and $p \in \mathbb{Z}$. The following formulas follow immediately from Prop. 4.23 (especially, from (4.10)):

$$\lim_{n\to\infty} \sqrt[n]{a} = 1 \tag{4.11a}$$

$$\lim_{n\to\infty} \sqrt[n]{n!} = +\infty \tag{4.11b}$$

$$\lim_{n\to\infty} \sqrt[n]{n^p} = 1 \tag{4.11c}$$

Note that (4.11c) follows from

$$\lim_{n\to\infty} \left(\frac{n}{n+1}\right)^p = 1 \tag{4.11d}$$

(This is clearly true when $p = \pm 1$, and hence is true for any p by induction.) By (4.11c), the radius of convergence of $\sum_n n^p z^n$ is 1. Therefore, by Prop. 4.21, $\sum n^p A^{-n}$ converges absolutely when $A > 1$. Thus, by Prop. 4.3,

$$\lim_{n\to\infty} \frac{n^p}{A^n} = 0 \qquad (\text{if } A > 1) \tag{4.11e}$$

This means that "polynomials grow slower than exponentials".

The same conclusions hold for arbitrary $p \in \mathbb{R}$ once we know how to define x^p and prove the continuity of $x \in \mathbb{R}_{>0} \mapsto x^p$. (See Sec. 7.8.) $\qquad\square$

$\star$ **Exercise 4.25.** Prove (4.11a) directly. Then use (4.11a) to give a direct proof of Prop. 4.23. Do not use root test, ratio test, or any results about power series.

Definition 4.26. By (4.11b), the power series

$$\exp(z) \equiv e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}$$

has radius of convergence $+\infty$, and hence converges absolutely on $\mathbb{C}$. (In particular, $\lim_{n\to\infty} z^n/n! = 0$ for all $z \in \mathbb{C}$.) This gives a function $\exp : \mathbb{C} \to \mathbb{C}$, called the **exponential function**.

Part (a) of Prop. 4.21 can be strengthened in the following way.

Theorem 4.27. *Let $\sum v_n z^n$ be a power series with coefficients in a complex Banach space V. Let R be its radius of convergence, and assume that $0 < R \leqslant +\infty$. For each $z \in B_{\mathbb{C}}(0, R)$, let $f(z)$ denote the value of this series at z (which is an element of V). Then $f : B_{\mathbb{C}}(0, R) \to V$ is continuous. Moreover, for each $0 < \rho < R$, the series of functions $\sum v_n z^n$ converges uniformly on $\overline{B}_{\mathbb{C}}(0, \rho)$ to f.*

Note that by calling $\sum v_n z^n$ a series of functions, we understand each term $v_n z^n$ as a function $\mathbb{C} \to V$.

Proof. For each $0 < \rho < R$, let $X_\rho = \overline{B}_{\mathbb{C}}(0, \rho)$. Then X_ρ is clearly a bounded closed subset of $\mathbb{C}$, and hence is sequentially compact by Heine-Borel Thm. 3.55. Let $g_n = v_n z^n$, which is a continuous function $X_\rho \to V$. We view g_n as an element of the Banach space (cf. Cor. 3.50) $C(X_\rho, V)$. Then $\|g_n\|_\infty = \rho^n \|v_n\|$. Thus

$$\limsup_{n\to\infty} \sqrt[n]{\|g_n\|_\infty} = \limsup_{n\to\infty} \rho \sqrt[n]{\|v_n\|} = \rho/R < 1$$

Therefore, by root test, the series $\sum g_n$ converges in the Banach space $C(X_\rho, V)$ to some $f_\rho \in C(X_\rho, V)$.

We have proved that for each $0 < \rho < R$, the series of functions $\sum v_n z^n$ converges uniformly on X_ρ to a continuous function f_ρ. Let $f : B_{\mathbb{C}}(0, R) \to V$ whose value at each z is the value of the original series at z. Thus, if $|z| \leqslant \rho$, then

$f_\rho(z) = f(z)$. Namely, $f|_{X_\rho} = f_\rho$. This shows that $\sum v_n z^n$ converges uniformly on X_ρ to f. It also shows that $f|_{B_{\mathbb{C}}(0,\rho)}$ is continuous (because f_ρ is continuous). Therefore, since $B_{\mathbb{C}}(0, R)$ is covered by all open disks $B_{\mathbb{C}}(0, \rho)$ (where $0 < \rho < R$), we conclude from Lem. 2.41 that f is continuous on $B_{\mathbb{C}}(0, R)$. $\qquad\square$

Example 4.28. By Thm. 4.27, the exponential function $\exp : \mathbb{C} \to \mathbb{C}$ is continuous; moreover, $\sum_{n=0}^{\infty} z^n/n!$ converges uniformly to e^z on $\overline{B}_{\mathbb{C}}(0, R)$ for every $0 < R < +\infty$, and hence on every bounded subset of $\mathbb{C}$.

4.4   Problems and supplementary material

Let V be a Banach space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$.

Problem 4.1. Use the following discrete version of "integration by part"

$$\sum_{k=m+1}^{n} f_k g_k = F_n g_n - F_m g_m + \sum_{k=m}^{n-1} F_k(g_k - g_{k+1}) \tag{4.12}$$

(where $F_n = \sum_{j=1}^{n} f_j$) to prove the following Dirichlet's test.

Theorem 4.29 (Dirichlet's test). *Let X be a set. Assume that (f_n) is a sequence in $l^\infty(X, V)$ such that $\sup_{n \in \mathbb{Z}_+} \|F_n\|_{l^\infty} < +\infty$ (where $F_n = \sum_{j=1}^{n} f_j$). Assume that (g_n) is a decreasing sequence (i.e. $g_1 \geqslant g_2 \geqslant g_3 \geqslant \cdots$) in $l^\infty(X, \mathbb{R}_{\geqslant 0})$ converging uniformly to 0. Then $\sum_{n=1}^{\infty} f_n g_n$ converges uniformly on X.*

Problem 4.2. Let $e_n : \mathbb{R} \to \mathbb{C}$ be $e_n(x) = e^{\mathbf{i}nx} = \cos(nx) + \mathbf{i}\sin(nx)$. Show that for each $x \in \mathbb{R}$, $\sum_{n=1}^{\infty} e_n(x)/n$ does not converge absolutely. Use Dirichlet's test to show that the series of functions $\sum_{n=1}^{\infty} e_n/n$ converges pointwise on $\mathbb{R} \backslash \{2k\pi : k \in \mathbb{Z}\}$, and uniformly on $[\delta, 2\pi - \delta]$ for every $0 < \delta < \pi$.

5 Nets and discrete integrals

5.1 Introduction: why do we need nets?

Nets were introduced by Moore and Smith in 1922 as a generalization of sequences. The most well-known motivation for introducing nets is that sequences are not enough for the study of non-metrizable topological spaces (i.e. topological spaces whose topologies are not induced by metrics). Here are two examples:

- In a general topological space, the definition of continuous maps using sequential convergence (as in Def. 2.38-(1)) is weaker than the definition using interior points and open sets (as in Def. 2.38-(2'), see also Rem. 4.23). Therefore, the dynamic intuition of sequences is not equivalent to the static intuition of open sets.

- Some important topological spaces are compact (i.e. every open cover has a finite subcover) but not sequentially compact. $[0,1]^I$ (where I is uncountable), equipped with the "product topology" (i.e. "pointwise convergence topology"), is such an example.

As we shall see, nets provide a remedy for these issues: For a general topological space, the definition of continuity using net convergence is equivalent to that using open sets; compactness is equivalent to "net-compactness", where the latter means that every net has a convergent subset. Thus, by generalizing sequences to nets, the dynamic intuition and the static and geometric intuition are unified again.

Nevertheless, the most common topological spaces appearing in analysis are metrizable. This raises the question: Why should we care about nets, given that our primary interest is in metrizable topological spaces? Here is my answer: Even though we are mainly interested in metrizable spaces, we can still find nets helpful in the following aspects.

First of all, many convergence processes cannot be described by sequential convergence, but can be described by net convergence. For example, the following limits can be formulated and understood in the language of net convergence:

(1) The limit of a function $\lim_{x \to x_0} f(x)$ where $f : X \to Y$ is a map of metric spaces and $x_0 \in X$.

(2) The limit $\lim_{m,n \to \infty} a_{m,n}$ where $(a_{m,n})_{m,n \in \mathbb{Z}_+}$ is a **double sequence** in a metric space X. Note that this is not the same as (but is more natural than) the **iterated limit** $\lim_{n \to \infty} \lim_{m \to \infty} a_{m,n}$. Moreover, the limit $\lim_{m,n \to \infty} a_{m,n}$ is the key to understanding the problem of commutativity of iterated integrals:

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

(3) The Riemann integral $\int_a^b f(x)dx$. This is the limit of the Riemann sum $\lim \sum f(\xi_i)(a_i - a_{i-1})$ as the partition of the interval $[a, b]$ is getting finer and finer.

Moreover, as for (3), we shall see that the net version of Cor. 3.18 provides a quick and conceptual proof of the following fact: If the upper and lower Darboux integrals are equal, then the Riemann integral exists and are equal to the two Darboux integrals. Indeed, the upper and lower Darboux integrals are respectively the $\lim \sup$ and $\lim \inf$ of a net in $\mathbb{R}$.

Second, nets provide a conceptual solution to many problems about **double series**. Let $(a_{m,n})_{m,n\in\mathbb{Z}_+}$ be a double sequence in $\mathbb{R}$. Think about the following questions, which arise naturally when one is trying to prove $e^z e^w = e^{z+w}$.

(a) When is it true that $\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} a_{m,n} = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} a_{m,n}$?

(b) Since $\text{card}(\mathbb{Z}_+ \times \mathbb{Z}_+) = \text{card}(\mathbb{Z}_+)$, why not use an ordinary series to study a double series? So let us **parametrize** $\mathbb{Z}_+ \times \mathbb{Z}_+$ by $\mathbb{Z}_+$: choose a bijection $\varphi : \mathbb{Z}_+ \to \mathbb{Z}_+ \times \mathbb{Z}_+$. When is it true that $\sum_{k=1}^{\infty} a_{\varphi(k)} = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} a_{m,n}$?

(c) Choose another parametrization (i.e. bijection) $\psi : \mathbb{Z}_+ \to \mathbb{Z}_+ \times \mathbb{Z}_+$. When is it true that $\sum_{k=1}^{\infty} a_{\varphi(k)} = \sum_{k=1}^{\infty} a_{\psi(k)}$?

(d) More generally, let X be a countably infinite set, and let $f : X \to \mathbb{R}$. Intuitively, we can take an infinite sum $\sum_{x\in X} f(x)$. How to define it rigorously? One may think about choosing a parametrization, i.e., a bijection $\varphi : \mathbb{Z}_+ \to X$. Then one defines the infinite sum by $\sum_{k=1}^{\infty} f(\varphi(k))$. Is this definition independent of the choice of parametrization?

(e) As a special case of (d), when is a series invariant under **rearrangement**? Namely, choose a bijection $\varphi : \mathbb{Z}_+ \to \mathbb{Z}_+$, and choose a sequence (a_n) in $\mathbb{R}$, when is it true that $\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} a_{\varphi(n)}$?

Modern differential geometry (whose "intrinsic" spirit stems from Gauss's Theorema Egregium) teaches us that in order to answer these questions, one should first define the infinite sum $\sum_{x\in X} f(x)$ in a parametrization-independent way. (The reason we call a bijection $\varphi : \mathbb{Z}_+ \to X$ a parametrization is that we want the readers to compare it with the parametrizations of curves, surfaces, and more generally manifolds.) We will call this sum a **discrete integral**. Then, one tries to answer when this definition agrees with those that depend on parametrizations

(such as the sums in (a)-(e) above). These goals can be achieved with the help of nets.

5.2 Nets

5.2.1 Directed sets and nets

Definition 5.1. A relation $\leqslant$ on a set I is called a **preorder** if for all $\alpha, \beta, \gamma \in I$, the following are satisfied:

- (Reflexivity) $\alpha \leqslant \alpha$.

- (Transitivity) If $\alpha \leqslant \beta$ and $\beta \leqslant \gamma$ then $a \leqslant \gamma$.

The pair $(I, \leqslant)$ (or simply I) is called a **preordered set**.

Therefore, a partial order is a preorder satisfying antisymmetry: $(\alpha \leqslant \beta) \wedge (\beta \leqslant \alpha) \Rightarrow (\alpha = \beta)$.

Definition 5.2. A preordered set $(I, \leqslant)$ is called a **directed set** if

$$\forall \alpha, \beta \in I \quad \exists \gamma \in I \quad \text{such that } \alpha \leqslant \gamma, \beta \leqslant \gamma \tag{5.1}$$

If I is a directed set and X is a set, then a function $x : I \to X$ is called a **net** with directed set/index set I. We often write $x(\alpha)$ as x_α if $\alpha \in I$, and write x as $(x_\alpha)_{\alpha \in I}$.

Example 5.3. $(\mathbb{Z}_+, \leqslant)$ is a directed set. A net with index set $\mathbb{Z}_+$ in a set X is precisely a sequence in X.

Definition 5.4. Suppose that $(I, \leqslant_I)$ and $(J, \leqslant_J)$ are preordered set (resp. directed set), then the **product** $(I \times J, \leqslant)$ is a preordered set (resp. directed set) if for every $\alpha, \alpha' \in I, \beta, \beta' \in J$ we define

$$(\alpha, \beta) \leqslant (\alpha', \beta') \qquad \Longleftrightarrow \qquad \alpha \leqslant_I \alpha' \text{ and } \beta \leqslant_J \beta' \tag{5.2}$$

Unless otherwise stated, the preorder on $I \times J$ is assumed to be defined by (5.2).

Example 5.5. $\mathbb{Z}_+ \times \mathbb{Z}_+$ (or similarly, $\mathbb{N} \times \mathbb{N}$) is naturally a directed set whose preorder is defined by (5.2). A net $(x_{m,n})_{(m,n) \in \mathbb{Z}_+ \times \mathbb{Z}_+}$ with index set $\mathbb{Z}_+ \times \mathbb{Z}_+$ is called a **double sequence** and is written as $(x_{m,n})_{m,n \in \mathbb{Z}_+}$ or simply $(x_{m,n})$. (We will even write it as (x_{mn}) when no confusion arises.)

More generally, we call $(x_{\alpha,\beta})_{(\alpha,\beta) \in I \times J} = (x_{\alpha,\beta})_{\alpha \in I, \beta \in J}$ a **double net** if its index set is $I \times J$ for some directed sets I, J. $\qquad\square$

Example 5.6. If X is a set, then $(2^X, \subset)$ and $(\text{fin}(2^X), \subset)$ are directed sets where

$$\text{fin}(2^X) = \{A \subset X : A \text{ is a finite set}\} \tag{5.3}$$

We will use nets with index set $\text{fin}(2^X)$ to study infinite sums.

Example 5.7. Let X be a metric space and $x \in X$. Then $X_x = (X, \leqslant)$ is a directed set if for each $p_1, p_2 \in X$ we define

$$p_1 \leqslant p_2 \text{ in } X_x \qquad \Longleftrightarrow \qquad d(p_1, x) \geqslant d(p_2, x) \tag{5.4}$$

(Namely, a larger element of X_x is one closer to x.) Nets with this directed set can be used to study the limits of functions (cf. Rem. 7.83). Note that X_x is our first example of directed set which is not a poset! ($d(p_1, x) = d(p_2, x)$ does not imply $p_1 = p_2$.)

5.2.2 Limits of nets

If I is a preordered set and $\beta \in I$, we write

$$I_{\geqslant \beta} = \{\alpha \in I : \alpha \geqslant \beta\} \tag{5.5}$$

Definition 5.8. Let P be a property about elements of a set X, i.e., P is a function $X \to \{\text{true, false}\}$. Let $(x_\alpha)_{\alpha \in I}$ be a net in X.

We say that x_α **eventually** satisfies P (equivalently, we say that x_α satisfies P for **sufficiently large** α) if:

- There exists $\beta \in I$ such that for every $\alpha \in I_{\geqslant \beta}$, the element x_α satisfies P.

"Sufficiently large" is also called "**large enough**".

We say that x_α **frequently** satisfies P if:

- For every $\beta \in I$ there exists $\alpha \in I_{\geqslant \beta}$ such that x_α satisfies P.

$\square$

Remark 5.9. Note that unlike sequences, for a general net, "x_α eventually satisfies P" does not imply "all but finitely many x_α satisfy P" because the complement of $I_{\geqslant \beta}$ is not necessarily a finite set.

Remark 5.10. Let P and Q be two properties about elements of X. Then

$$\neg(x_\alpha \text{ eventually satisfies } P) \quad = \quad (x_\alpha \text{ frequently satisfies } \neg P) \tag{5.6a}$$

By the crucial condition (5.1) for directed sets, we have

$$(x_\alpha \text{ eventually satisfies } P) \wedge (x_\alpha \text{ eventually satisfies } Q)$$
$$\Downarrow \tag{5.6b}$$
$$x_\alpha \text{ eventually satisfies } P \wedge Q$$

By taking contraposition and replacing P, Q by $\neg P, \neg Q$, we have

$$x_\alpha \text{ frequently satisfies } P \vee Q$$
$$\Downarrow \tag{5.6c}$$
$$(x_\alpha \text{ frequently satisfies } P) \vee (x_\alpha \text{ frequently satisfies } Q)$$

Definition 5.11. Let $(x_\alpha)_{\alpha \in I}$ be a net in a metric space X. Let $x \in X$. We say that (x_α) **converges to** x and write

$$\lim_{\alpha \in I} x_\alpha \equiv \lim_\alpha x_\alpha = x$$

or simply $x_\alpha \to x$ if the following statement holds:

- For every $\varepsilon > 0$, x_α is eventually in $B_X(x, \varepsilon)$.

Clearly, $x_\alpha \to x$ iff $d(x_\alpha, x) \to 0$.

Definition 5.12. Let $(x_{m,n})_{m,n \in \mathbb{Z}_+}$ be a double sequence in a metric space. Then we write

$$\lim_{(m,n) \in \mathbb{Z}_+ \times \mathbb{Z}_+} x_{m,n} \equiv \lim_{m,n \to \infty} x_{m,n} \tag{5.7}$$

and call it the **(double) limit** of $(x_{m,n})$.

Remark 5.13. Let us spell out the meaning of $\lim_{m,n \to \infty} x_{m,n} = x$: For each $\varepsilon > 0$ there exists $M, N \in \mathbb{Z}_+$ such that $d(x_{m,n}, x) < \varepsilon$ for all $m \geqslant M$ and $n \geqslant N$. Clearly, this is equivalent to the statement:

- For each $\varepsilon > 0$ there exists $N \in \mathbb{Z}_+$ such that $d(x_{m,n}, x) < \varepsilon$ for all $m, n \geqslant N$.

Therefore, if (x_n) is a sequence in X, then

$$(x_n) \text{ is a Cauchy sequence} \qquad \Longleftrightarrow \qquad \lim_{m,n \to \infty} d(x_m, x_n) = 0 \tag{5.8}$$

Thus, the Cauchyness of sequences can be studied in terms of double limits, and hence in terms of nets.

Proposition 5.14. *Let $(x_\alpha)_{\alpha \in I}$ be a net in a metric space X converging to x, y. Then $x = y$.*

Proof. Suppose that $x \neq y$. Then there are $r, \rho > 0$ such that $B(x, r) \cap B(y, \rho) = \varnothing$, say $r = \rho = d(x, y)/2$. Since $x_\alpha \to x$, the point x_α is eventually in $B(x, r)$. Since $x_\alpha \to y$, the point x_α is eventually in $B(y, \rho)$. Therefore, by the logic (5.6b), x_α is eventually in $B(x, r) \cap B(y, \rho)$, impossible. $\square$

Theorem 5.15. *Let $f : X \to Y$ be map of metric spaces continuous at $x \in X$. Let $(x_\alpha)_{\alpha \in I}$ be a net in X converging to x. Then $\lim_\alpha f(x_\alpha) = f(x)$.*

Proof. Choose any $\varepsilon > 0$. By Def. 2.38-(2) and the continuity of f at x, there exists $\delta > 0$ such that for all $p \in B(x, \delta)$ we have $f(p) \in B(f(x), \varepsilon)$. Since $x_\alpha \to x$, x_α is eventually in $B(x, \delta)$. Therefore $f(x_\alpha)$ is eventually in $B(f(x), \varepsilon)$. $\square$

This theorem implies, for example, that if (v_α) is a net in a complex normed vector space converging to v, and if (λ_α) is a net in $\mathbb{C}$ converging to λ, then $\lambda_\alpha v_\alpha$ converges to λv because the scalar multiplication map is continuous (Prop. 3.38).

Exercise 5.16. Prove the generalization of Rem. 2.71:

1. If $(x_\alpha)_{\alpha \in I}, (y_\alpha)_{\alpha \in I}$ are nets in $\overline{\mathbb{R}}$ converging to $A, B \in \overline{\mathbb{R}}$, and if $x_\alpha \leqslant y_\alpha$ for all α, then $A \leqslant B$.

2. **Squeeze theorem**: Suppose that $(x_\alpha)_{\alpha \in I}, (y_\alpha)_{\alpha \in I}, (z_\alpha)_{\alpha \in I}$ are nets in $\overline{\mathbb{R}}$, $x_\alpha \leqslant y_\alpha \leqslant z_\alpha$ for all α, and x_α and z_α both converge to $A \in \overline{\mathbb{R}}$. Then $y_\alpha \to A$.

3. If (x_α) is an increasing resp. decreasing net in $\overline{\mathbb{R}}$, then $\lim_\alpha x_\alpha$ exists in $\overline{\mathbb{R}}$ and equals $\sup_\alpha x_\alpha$ resp. $\inf_\alpha x_\alpha$.

5.2.3 Subnets (in the sense of Willard)

Definition 5.17. A subset E of a directed set I is called **cofinal** if:

$$\forall \alpha \in I \quad \exists \beta \in E \quad \text{such that } \alpha \leqslant \beta$$

By the transitivity in Def. 5.1 and property (5.1), we clearly have

$$\forall \alpha_1, \ldots, \alpha_n \in I \quad \exists \beta \in E \quad \text{such that } \alpha_1 \leqslant \beta, \ldots, \alpha_n \leqslant \beta$$

Definition 5.18. Let $(x_\alpha)_{\alpha \in I}$ be a net in a set X. A **subnet** of $(x_\alpha)_{\alpha \in I}$ is, by definition, of the form $(x_{\alpha_s})_{s \in S}$ where S is a directed set, and

$$(\alpha_s)_{s \in S} : S \to I \qquad s \mapsto \alpha_s$$

is an increasing function whose range $\{\alpha_s : s \in S\}$ is cofinal in I.

Remark 5.19. There are several different definitions of subnets that are equivalent for proving the main results in point-set topology. Unfortunately, there is no common agreement on the standard definition of subnets. The definition we gave is due to Willard [Wil], and is also the one given in the famous textbook of Munkres [Mun]. Some famous analysis and topology textbooks (e.g. [Fol, Kel, RS]) use a weaker definition, which does not assume that the map $S \to I$ is increasing.

Example 5.20. A subsequence of a sequence is a subnet of that sequence.

Example 5.21. Let $(x_{m,n})_{m,n \in \mathbb{Z}_+}$ be a net with index set $\mathbb{Z}_+ \times \mathbb{Z}_+$. Then $(x_{k,k})_{k \in \mathbb{Z}_+}$ and $(x_{2k,k})_{k \in \mathbb{Z}_+}$ are subnets. $(x_{k,1})_{k \in \mathbb{Z}_+}$ is not a subnet, because the cofinal condition is not satisfied. More generally, it is not hard to show that for every function $\varphi, \psi : \mathbb{Z}_+ \to \mathbb{Z}_+$, $(x_{\varphi(k),\psi(k)})_{k \in \mathbb{Z}_+}$ is a subnet iff φ, ψ are increasing and $\lim_{k \to \infty} \varphi(k) = \lim_{k \to \infty} \psi(k) = +\infty$.

Exercise 5.22. Prove the following facts:

- The cofinal subset of a cofinal subset of a directed set I is a cofinal subset of I.

- The subnet of a subnet of a net (x_α) is a subnet of (x_α).

Note that in your proof you need to use the transitivity in Def. 5.1.

The biggest difference between subnets and subsequences is that the index set of a subnet is not necessarily a subset of the index set of the original net. Indeed, subnets are defined in this way mainly because we want to have a net version of Pb. 3.1 in any topological space. (This will be achieved in Pb. 7.2.) Let us see an elementary example of subnet whose index set is larger than that of the original net. Its importance is justified by the proofs of Exp. 5.28 and Prop. 5.34.

Example 5.23. Let J be a directed set. Then every net $(x_\alpha)_{\alpha \in I}$ has subnet $(x_\alpha)_{(\alpha,\beta) \in I \times J}$. The corresponding increasing map of directed sets is the projection $I \times J \to I$ onto the first component.

To appreciate the importance of cofinalness (as well as transitivity), we prove the following generalization of Prop. 2.36. This result has a wide range of surprising applications that are unavailable when one only considers sequences. (We will see them soon in this chapter. For instance, this result explains why the values of absolutely convergent series are invariant under rearrangement.) So I call this result a theorem, even though its proof is simple.

Theorem 5.24. *Let $(x_\alpha)_{\alpha \in I}$ be a net in a metric space (or more generally, a topological space) X converging to $x \in X$. Then every subnet $(x_{\alpha_s})_{s \in S}$ converges to x.*

The following proof for metric spaces can be generalized straightforwardly to topological spaces. The readers can come back and check the details after learning topological spaces.

Proof. Choose any $\varepsilon > 0$. Since $x_\alpha \to x$, there exists $\beta \in I$ such that for all $\alpha \geqslant \beta$ we have $d(x_\alpha, x) < \varepsilon$. By the cofinalness, there exists $t \in S$ such that $\alpha_t \geqslant \beta$. Thus, since $s \in S \mapsto \alpha_s \in I$ is increasing, for every $s \geqslant t$, we have $\alpha_s \geqslant \alpha_t \geqslant \beta$ and hence $\alpha_s \geqslant \beta$ by the transitivity in Def. 5.1. So $d(x_{\alpha_s}, x) < \varepsilon$ for all $s \geqslant t$. This finishes the proof. $\square$

This proposition does not hold if one does not assume cofinalness in the definition of subnets:

Example 5.25. Let (x_n) be a sequence in $\mathbb{R}$ converging to $x \in \mathbb{R}$. Since (x_n) is a Cauchy sequence, we know that $\lim_{m,n \to \infty} x_m - x_n = 0$. We have seen in Exp. 5.21 that $(x_{2k} - x_k)_{k \in \mathbb{Z}_+}$ is a subnet of $(x_{m,n})$. Therefore, $\lim_{k \to \infty} x_{2k} - x_k = 0$. But

$(x_k - x_1)_{k \in \mathbb{Z}_+}$ is not a subnet since the cofinal condition is not satisfied. And if $x \neq x_1$, then $\lim_k (x_k - x_1) = x - x_1 \neq 0$, i.e.,

$$\lim_{k \to \infty} (x_k - x_1) \neq \lim_{m,n \to \infty} (x_m - x_n)$$

In Subsec. 2.3.2, we have seen two criteria for the divergence of sequence: a sequence diverges if it is unbounded, or if it has two subsequences converging to different points. By Thm. 5.24, the second criterion can be generalized to nets. However, the following example shows that the first criterion does not has its net version:

Example 5.26. A convergent net $(x_\alpha)_{\alpha \in I}$ in a metric space X is not necessarily **bounded**. Namely, it is not necessarily true that $\{x_\alpha : \alpha \in I\}$ is a bounded subset of X. Let $f : \mathbb{R}_{>0} \to \mathbb{R}$ be $f(x) = 1/x$. Then f is net in $\mathbb{R}$ with directed set $(\mathbb{R}_{>0}, \leqslant)$. This net is not bounded, although $\lim f(x) = 0$.

Example 5.27. The double sequence $x_{m,n} = n/(m+n)$ in $\mathbb{R}$ has subnets $x_{n,n} = n/(n+n) = 1/2$ and $x_{2n,n} = 1/3$. Since these two subnets converge to different values, Thm. 5.24 implies that $\lim_{m,n} x_{m,n}$ does not exist. However, the **iterated limits** exist and take different values:

$$\lim_{m \to \infty} \lim_{n \to \infty} \frac{n}{m+n} = 1 \qquad \lim_{n \to \infty} \lim_{m \to \infty} \frac{n}{m+n} = 0$$

As we shall see, this gives another criterion for the divergence of double series: If the two iterated limits exist and are different, then the double series diverge.

Finally, we do an example of convergent double sequence:

Example 5.28. Let $x_{m,n} = (m^{-2} - n^{-1}) \sin \dfrac{\pi(m + \sqrt{n})}{4}$. Then $\lim_{m,n \to \infty} x_{m,n} = 0$.

Proof. The sequence $(m^{-2})_{m \in \mathbb{Z}_+}$ converges to 0. By Exp. 5.23, the double sequence $(m^{-2})_{m,n \in \mathbb{Z}_+}$ is its subnet, and hence converges to 0 by Thm. 5.24. Similarly, the double sequence $(n^{-1})_{m,n \in \mathbb{Z}_+}$ converges to 0. Therefore, $m^{-2} + n^{-1}$ converges to 0 due to Thm. 5.15 and the continuity of the addition map $(x, y) \in \mathbb{R}^, \to x + y \in \mathbb{R}$ (Prop. 2.45). Since $0 \leqslant |x_{m,n}| \leqslant m^{-2} + n^{-1}$, we conclude $|x_{m,n}| \to 0$ (and hence $x_{m,n} \to 0$) by squeeze theorem (Exe. 5.16). $\qquad \square$

5.2.4 Double limits and iterated limits

Theorem 5.29. *Let* $(x_{\alpha,\beta})_{\alpha \in I, \beta \in J}$ *be a double net in a metric space* X. *Assume that the following are true:*

(1) The limit $\lim_{(\alpha,\beta) \in I \times J} x_{\alpha,\beta}$ *exists in* X.

(2) For each $\alpha \in I$, the limit $\lim\limits_{\beta \in J} x_{\alpha,\beta}$ exists in X.

Then the LHS limit in the following equation exists and equals the RHS:

$$\lim_{\alpha \in I} \lim_{\beta \in J} x_{\alpha,\beta} = \lim_{(\alpha,\beta) \in I \times J} x_{\alpha,\beta} \qquad (5.9)$$

In particular, suppose that the following is also true:

(3) For each $\beta \in J$, the limit $\lim\limits_{\alpha \in J} x_{\alpha,\beta}$ exists in X.

Then the following limits exist and are equal:

$$\lim_{\alpha \in I} \lim_{\beta \in J} x_{\alpha,\beta} = \lim_{\beta \in J} \lim_{\alpha \in I} x_{\alpha,\beta} \qquad (5.10)$$

Proof. Let $x_\alpha = \lim\limits_{\beta} x_{\alpha,\beta}$ and $x = \lim\limits_{\alpha,\beta} x_{\alpha,\beta}$. We want to show that $\lim\limits_{\alpha} x_\alpha = x$. Choose any $\varepsilon > 0$. Then there exist $A \in I$, $B \in J$ such that for every $\alpha \geqslant A$ and $\beta \geqslant B$ we have $d(x_{\alpha,\beta}, x) < \varepsilon/3$. In particular, $d(x_{\alpha,\beta}, x) \leqslant \varepsilon/2$. Using Thm. 5.15 and the fact that $p \in X \mapsto d(p, x) \in \mathbb{R}$ is continuous (Exp. 2.58), we see that for every $\alpha \geqslant A$ we have $d(x_\alpha, x) = \lim\limits_{\beta \in J_{\geqslant B}} d(x_{\alpha,\beta}, x) \leqslant \varepsilon/2 < \varepsilon$. $\qquad \square$

The readers may skip the next remark and proof and come back to them when they have learned about topological spaces.

$\star$ **Remark 5.30.** Thm. 5.29 can be generalized to the case that X is a regular topological space. By saying that the topological space X is **regular**, we mean that for every $x \in X$ and every open set U containing x, there is a smaller open set V containing x such that the closure $\overline{V}$ (cf. Def. 7.28) is contained in U.

$\star$ *Proof.* Let $x_\alpha = \lim\limits_{\beta} x_{\alpha,\beta}$ and $x = \lim\limits_{\alpha,\beta} x_{\alpha,\beta}$. Choose any open set U containing x. We want to prove that x_α is eventually in U. Choose an open set V containing x such that $\overline{V} \subset U$. Then there are $A \in I$, $B \in J$ such that for all $\alpha \geqslant A$ and $\beta \geqslant B$ we have $x_{\alpha,\beta} \in V$. Thus, for each $\alpha \geqslant A$, since $x_{\alpha,\beta}$ approaches x_α, we have $x_\alpha \in \overline{V}$ and hence $x_\alpha \in U$. $\qquad \square$

Corollary 5.31. *Let $(x_{\alpha,\beta})_{\alpha \in I, \beta \in J}$ be a double net in $\overline{\mathbb{R}}$. Assume that $x_{\bullet,\bullet}$ is increasing, i.e., $x_{\alpha,\beta} \leqslant x_{\alpha',\beta'}$ if $\alpha \leqslant \alpha'$ and $\beta \leqslant \beta'$. Then the following equation (5.11) hold, where all the limits (5.11) exist in $\overline{\mathbb{R}}$:*

$$\lim_{\alpha \in I} \lim_{\beta \in J} x_{\alpha,\beta} = \lim_{\beta \in J} \lim_{\alpha \in I} x_{\alpha,\beta} = \lim_{(\alpha,\beta) \in I \times J} x_{\alpha,\beta} = \sup\{x_{\alpha,\beta} : \alpha \in I, \beta \in J\} \qquad (5.11)$$

Clearly, a similar result holds for decreasing double nets in $\overline{\mathbb{R}}$.

Proof. By Exe. 5.16, the three limits $\lim\limits_{\alpha} x_{\alpha,\beta}$, $\lim\limits_{\beta} x_{\alpha,\beta}$, and $\lim\limits_{\alpha,\beta} x_{\alpha,\beta}$ exist in $\overline{\mathbb{R}}$. Therefore, by Thm. 5.29, the three limits in (5.11) exist and are equal. The last equality in (5.11) is also due to Exe. 5.16. $\qquad \square$

5.2.5 Cauchy nets

Definition 5.32. A net $(x_\alpha)_{\alpha \in I}$ in a metric space X is called a Cauchy net if

$$\lim_{\alpha,\beta \in I} d(x_\alpha, x_\beta) = 0$$

Equivalently, this means that

$$\forall \varepsilon > 0 \quad \exists \gamma \in I \quad \text{such that } \forall \alpha, \beta \geqslant \gamma \quad \text{we have } d(x_\alpha, x_\beta) < \varepsilon \qquad (5.12)$$

Exercise 5.33. Show that the subnet of a Cauchy net is Cauchy.

Proposition 5.34. *A convergent net in a metric space is a Cauchy net.*

Proof. Let $(x_\alpha)_{\alpha \in I}$ converge to x in a metric space X. Then $\lim_\alpha d(x_\alpha, x) = 0$. Since $(d(x_\alpha, x))_{\alpha,\beta \in I}$ is a subnet (cf. Exp. 5.23), we have $\lim_{\alpha,\beta} d(x_\alpha, x) = 0$ by Thm. 5.24. Similarly, we have $\lim_{\alpha,\beta} d(x, x_\beta) = 0$. Since $0 \leqslant d(x_\alpha, x_\beta) \leqslant d(x_\alpha, x) + d(x, x_\beta)$, by Squeeze theorem (Exe. 5.16) we have $\lim_{\alpha,\beta} d(x_\alpha, x_\beta) = 0$. $\square$

Proposition 5.35. *Let $(x_\alpha)_{\alpha \in I}$ be a Cauchy net in a metric space X. Suppose that $(x_\alpha)_{\alpha \in I}$ has a convergent subnet $(x_{\alpha_s})_{s \in S}$ converging to $x \in X$. Then $(x_\alpha)_{\alpha \in I}$ converges to x.*

Proof. Choose any $\varepsilon > 0$. Since (x_α) is a Cauchy net, there exists $\gamma \in I$ such that $d(x_\alpha, x_\beta) \leqslant \varepsilon$ for all $\alpha, \beta \geqslant \gamma$. Since $(\alpha_s)_{s \in S}$ has cofinal range, $\alpha_{s_0} \geqslant \gamma$ for some $s_0 \in S$. Thus $\alpha_s \geqslant \gamma$ for all $s \geqslant s_0$ because $(\alpha_s)_{s \in S}$ is increasing and because of the transitivity in Def. 5.1. Thus, for every $\beta \geqslant \gamma$, $d(x_{\alpha_s}, x_\beta) \leqslant \varepsilon$ for sufficiently large s. By taking limit over s and using the continuity of $y \in X \mapsto d(y, x_\beta)$ as well as Thm. 5.15, we get $d(x, x_\beta) \leqslant \varepsilon$ for all $\beta \geqslant \gamma$. $\square$

Definition 5.36. Two nets $(x_\alpha)_{\alpha \in I}$ and $(y_\alpha)_{\alpha \in I}$ in a metric space X are called **Cauchy-equivalent** if

$$\lim_{\alpha \in I} d(x_\alpha, y_\alpha) = 0$$

Two Cauchy nets are simply called **equivalent** if they are Cauchy-equivalent. It is not hard to see that Cauchy-equivalence is an equivalence relation (recall Def. 1.17) on X^I.

Exercise 5.37. Let $(x_\alpha)_{\alpha \in I}$ and $(y_\alpha)_{\alpha \in I}$ be nets in a metric space X.

1. Assume that $(x_\alpha)_{\alpha \in I}$ and $(y_\alpha)_{\alpha \in I}$ are Cauchy-equivalent. Prove that (x_α) is a Cauchy net iff (y_α) is a Cauchy net.

2. Assume that $(x_\alpha)_{\alpha \in I}$ converges to x. Prove that $(y_\alpha)_{\alpha \in I}$ converges to x iff $(x_\alpha)_{\alpha \in I}$ and $(y_\alpha)_{\alpha \in I}$ are Cauchy-equivalent.

Theorem 5.38. *Every Cauchy net $(x_\alpha)_{\alpha \in I}$ in a complete metric space X is convergent.*

We give a hint of the proof and leave the details to the readers as an exercise.

Hint. Construct an increasing sequence $(\alpha_n)_{n \in \mathbb{Z}_+}$ in I such that for every $\beta, \gamma \geqslant \alpha_n$ we have $d(x_\beta, x_\gamma) < 1/n$. Prove that $(x_{\alpha_n})_{n \in \mathbb{Z}_+}$ is a Cauchy sequence, and hence converges to some $x \in X$. Prove that $(x_\alpha)_{\alpha \in I}$ converges to x. (Warning: $(x_{\alpha_n})_{n \in \mathbb{Z}_+}$ is not necessarily a subnet of $(x_\alpha)_{\alpha \in I}$.) □

5.3 Discrete integrals $\sum_{x \in X} f(x)$

In this section, we fix V to be a Banach space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. We fix a (non-necessarily countable) set X. Note that if $f : X \to V$ is a function and X is finite, then $\sum_{x \in X} f(x)$ can be understood in its most obvious way.

Definition 5.39. Let $f : X \to V$ be a map. The expression

$$\sum_{x \in X} f(x)$$

(or simply $\sum_X f$) is called a **discrete integral**. If $v \in V$, we say that $\sum_{x \in X} f(x)$ equals (or converges to) v, if

$$\lim_{A \in \text{fin}(2^X)} \sum_{x \in A} f(x) = v \tag{5.13}$$

In this case, we write

$$\sum_{x \in X} f(x) = v \tag{5.14}$$

Remark 5.40. Recall from Exp. 5.6 that $\text{fin}(2^X)$ is the directed set of finite subsets of 2^X. Its preorder is "$\subset$". So (5.14) means more precisely that:

- For every $\varepsilon > 0$, there exists a finite set $B \subset X$ such that for every finite set A satisfying $B \subset A \subset X$, we have $\|v - \sum_{x \in A} f(x)\| < \varepsilon$.

Remark 5.41. Discrete integrals are one of the most important and representative examples in Moore and Smith's original paper on nets (cf. [MS22]), explaining why nets are called nets: Imagine an infinitely large *fishing net* whose vertices form the set $X = \mathbb{Z}^2$. You grab the net with your hands and pull it up. As you pull it up, the lifted part $A \in \text{fin}(2^X)$ becomes larger and larger.

Remark 5.42. One of the advantages of discrete integrals over series is that discrete integrals are clearly invariant under rearrangement: For every bijection $\varphi : X \to X$, if one side of the following equation converges in V, then the other side converges, and the equation holds true:

$$\sum_{x \in X} f(x) = \sum_{x \in X} f(\varphi(x)) \tag{5.15}$$

or simply $\sum_X f = \sum_X f \circ \varphi$.

Remark 5.43. Let us spell out what Cauchyness means for the net $(\sum_A f)_{A \in \mathrm{fin}(2^X)}$:

(1) For every $\varepsilon > 0$, there exists a finite set $B \subset X$ such that for any finite sets A_1, A_2 satisfying $B \subset A_1 \subset X, B \subset A_2 \subset X$, we have

$$\left\| \sum_{A_1 \setminus A_2} f - \sum_{A_2 \setminus A_1} f \right\| < \varepsilon$$

Note that the term inside the norm is $\sum_{A_1} f - \sum_{A_2} f$. This is also equivalent to:

(2) For every $\varepsilon > 0$, there exists a finite set $B \subset X$ such that for any finite set $E \subset X \setminus B$, we have

$$\left\| \sum_{E} f \right\| < \varepsilon$$

We shall mainly use (2) as the Cauchy criterion for the convergence of $\sum_X f$.

Proof of the equivalence. (2) follows from (1) by taking $A_1 = B$ and $A_2 = B \cup E$. (1) follows from (2) by taking $E_1 = A_1 \setminus A_2$ and $E_2 = A_2 \setminus A_1$ and then concluding $\| \sum_{E_1} f - \sum_{E_2} f \| < 2\varepsilon$. $\square$

Definition 5.44. Let $g : X \to \overline{\mathbb{R}}_{\geqslant 0}$ be a map. Note that the net $(\sum_A g)_{A \in \mathrm{fin}(2^X)}$ is increasing. Hence, its limit exists in $\overline{\mathbb{R}}$ and equals $\sup_{A \in \mathrm{fin}(2^X)} \sum_A g$ (by Exe. 5.16). We write this as $\sum_X g$, or more precisely:

$$\sum_X g \equiv \sum_{x \in X} g(x) \overset{\text{def}}{=\!=} \lim_{A \in \mathrm{fin}(2^X)} \sum_A g = \sup_{A \in \mathrm{fin}(2^X)} \sum_A g \tag{5.16}$$

We say that $\sum_X g$ **converges** or **converges absolutely**, if $\sum_X g < +\infty$.

It is clear that $\sum_X g < +\infty$ iff there exists $C \in \mathbb{R}_{\geqslant 0}$ such that $\sum_A g < C$ for all $A \in \mathrm{fin}(2^X)$.

Remark 5.45. Note that when $g : X \to \mathbb{R}_{\geqslant 0}$, the convergence in Def. 5.44 agrees with that in Def. 5.39. Therefore, Rem. 5.43 still gives a Cauchy criterion for convergence.

Definition 5.46. Let $f : X \to V$. We say that $\sum_X f$ **converges absolutely** if

$$\sum_{x \in X} \| f(x) \| < +\infty$$

Proposition 5.47. *Let $f : X \to V$. If $\sum_X f$ converges absolutely, then it converges, and*

$$\left\| \sum_{x \in X} f(x) \right\| \leqslant \sum_{x \in X} \| f(x) \| \tag{5.17}$$

We write this simply as $\| \sum_X f \| \leqslant \sum_X |f|$. (Recall Def. 3.39.)

Proof. (5.17) clearly holds when X is finite. In the general case, assume that $\sum_X f$ converges absolutely. Then by Cauchy criterion (Rem. 5.43-(2)), for every $\varepsilon > 0$ there is $A \in \text{fin}(2^X)$ such that for each finite $E \subset X\backslash A$ we have $\sum_E |f| < \varepsilon$, and hence $\|\sum_E f\| < \varepsilon$. Therefore $\sum_X f$ converges by Cauchy criterion again.

By the continuity of the norm function $v \in V \mapsto \|v\| \in \mathbb{R}_{\geqslant 0}$, and by Thm. 5.15, we have

$$\Big\| \sum_X f \Big\| = \Big\| \lim_A \sum_A f \Big\| = \lim_A \Big\| \sum_A f \Big\|$$

Since $\|\sum_A f\| \leqslant \sum_A |f|$, by Exe. 5.16, the above expression is no less than

$$\lim_A \sum_A |f| = \sum_X |f|$$

$\square$

The following proposition gives another demonstration that discrete integrals are more natural than series. We leave the proof to the readers.

Proposition 5.48. *Let $f : X \to \mathbb{R}^N$ where $N \in \mathbb{Z}_+$. Then*

$$\sum_{x \in X} f(x) \text{ converges} \qquad \Longleftrightarrow \qquad \sum_{x \in X} f(x) \text{ converges absolutely}$$

Hint. Reduce to the case $N = 1$. Consider $A = \{x \in X : f(x) \geqslant 0\}$ and $B = X\backslash A$. $\square$

When $\mathbb{R}^N$ is replaced by an infinite-dimensional Banach space, the convergence of a discrete integral may not imply absolute convergence. See Pb. 5.6.

5.4 Fubini's theorem for discrete integrals

Fix a Banach space V over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. Let X, Y be sets.

Theorem 5.49 (Fubini's theorem-A). *Let $f : X \times Y \to V$. Assume that $\sum_{X \times Y} f$ converges. Then $\sum_Y f(x, \cdot)$ converges for each $x \in X$, and $\sum_X f(\cdot, y)$ converges for each $y \in Y$, and*

$$\sum_{x \in X} \sum_{y \in Y} f(x, y) = \sum_{y \in Y} \sum_{x \in X} f(x, y) = \sum_{(x,y) \in X \times Y} f(x, y) \tag{5.18}$$

where all discrete integrals converge in V.

We abbreviate (5.18) to $\sum_X \sum_Y f = \sum_Y \sum_X f = \sum_{X \times Y} f$.

Proof. For each $x \in X$, let $f_x(y) = f(x, y)$. Let us prove that $\sum_Y f_x$ converges. Choose any $\varepsilon > 0$. Since $\sum_{X \times Y} f$ converges, by Cauchy criterion (Rem. 5.43-(2)), there exists a finite $S \subset X \times Y$ such that the sum of f over any finite subset outside S has norm $< \varepsilon$. The projection $X \times Y \to Y$ maps S to a finite set $B \subset Y$. Thus, for each finite $E \subset Y \backslash B$, we have $\| \sum_E f_x \| < \varepsilon$ since $x \times E$ is outside S. Therefore $\sum_Y f_x$ converges. By the same reasoning, $\sum_X f(\cdot, y)$ converges for all y.

Recall that $\sum_{X \times Y} f$ is the limit of the net $(\sum_S f)_{S \in \text{fin}(2^{X \times Y})}$. This net has subnet

$$\left(\sum_{(x,y) \in A \times B} f(x, y) \right)_{A \in \mathcal{I}, B \in \mathcal{J}} \qquad \text{where } \mathcal{I} = \text{fin}(2^X) \ \mathcal{J} = \text{fin}(2^Y)$$

(Its index set is $\mathcal{I} \times \mathcal{J}$.) Thus, by Thm. 5.24,

$$\sum_{(x,y) \in X \times Y} f(x, y) = \lim_{A \in \mathcal{I}, B \in \mathcal{J}} \sum_{(x,y) \in A \times B} f(x, y) \tag{5.19}$$

We are now going to use Thm. 5.29 to show that

$$\lim_{A \in \mathcal{I}, B \in \mathcal{J}} \sum_{(x,y) \in A \times B} f(x, y) = \lim_{A \in \mathcal{I}} \lim_{B \in \mathcal{J}} \sum_{(x,y) \in A \times B} f(x, y) \tag{5.20}$$

where the RHS limit exists. For that purpose, we need to check for each $A \in \mathcal{I}$ the convergence of $\lim_B \sum_{(x,y) \in A \times B} f(x, y)$. Since $\sum_Y f_x$ converges, we have

$$\lim_{B \in \mathcal{J}} \sum_{(x,y) \in A \times B} f(x, y) = \lim_{B \in \mathcal{J}} \sum_{x \in A} \sum_{y \in B} f(x, y)$$
$$= \sum_{x \in A} \lim_{B \in \mathcal{J}} \sum_{y \in B} f(x, y) = \sum_{x \in A} \sum_{y \in Y} f(x, y) \tag{5.21}$$

(Note that $\sum_A$ is a finite sum and hence commutes with $\lim_B$.) Thus, the assumption in Thm. 5.29 ensuring (5.20) has now been proved true. So (5.20) is true. Moreover, combining (5.19), (5.20), (5.21) together, we get

$$\sum_{(x,y) \in X \times Y} f(x, y) = \lim_{A \in \mathcal{I}} \sum_{x \in A} \sum_{y \in Y} f(x, y) = \sum_{x \in X} \sum_{y \in Y} f(x, y)$$

where the second and the third limits exist. This proves a half of (5.18). The other half can be proved in the same way. $\qquad\square$

Theorem 5.50 (Fubini's theorem-B). *Let* $g : X \times Y \to \overline{\mathbb{R}}_{\geqslant 0}$*. Then the five discrete integrals in (5.22) exist in* $\overline{\mathbb{R}}_{\geqslant 0}$*, and equations (5.22) hold in* $\overline{\mathbb{R}}_{\geqslant 0}$*:*

$$\sum_{x \in X} \sum_{y \in Y} f(x, y) = \sum_{y \in Y} \sum_{x \in X} f(x, y) = \sum_{(x,y) \in X \times Y} f(x, y) \tag{5.22}$$

Proof. The existence in $\overline{\mathbb{R}}_{\geqslant 0}$ of the five discrete integrals is clear. (Recall Def. 5.44.) Formula (5.22) can be proved in the same way as (5.18). Note that when applying Thm. 5.29 to prove (5.22), the assumption in Thm. 5.29 on the existence of limits is satisfied because all nets involved are increasing in $\overline{\mathbb{R}}$. (Recall Exe. 5.16.) □

Corollary 5.51 (Fubini's theorem-C). *Let* $f : X \times Y \to V$*. Then the following are equivalent.*

(1) $\sum_{X \times Y} f$ *converges absolutely.*

(2) $\sum_{x \in X} \sum_{y \in Y} \|f(x,y)\| < +\infty$*.*

(3) $\sum_{y \in Y} \sum_{x \in X} \|f(x,y)\| < +\infty$*.*

Proof. Immediate from Thm. 5.50. It is also not hard to prove it directly. □

5.5 Parametrization theorem for discrete integrals

We fix a Banach space V over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. In the following sections, we shall apply the results about discrete integrals to the study of series and double series. For the convenience of applications (e.g. the proof of $e^z e^w = e^{z+w}$), we enlarge the concept of series a little:

5.5.1 Series over $\mathbb{Z}$

Definition 5.52. A **series over** $\mathbb{Z}$ is an expression

$$\sum_{n=-\infty}^{+\infty} f(n)$$

where f is a function from $\mathbb{Z}$ to either V or $\overline{\mathbb{R}}_{\geqslant 0}$. We say that this series **converges to** (or equals) $\mu \in V$ resp. equals $\mu \in \overline{\mathbb{R}}_{\geqslant 0}$, if

$$\lim_{m,n \to +\infty} \sum_{i=-m}^{n} f(i) = \mu \tag{5.23}$$

In this case, we write

$$\sum_{n=-\infty}^{+\infty} f(n) = \mu$$

We say that $\sum_{n=-\infty}^{+\infty} f(n)$ **converges absolutely** if $\sum_{n=-\infty}^{+\infty} \|f(n)\| < +\infty$.

Remark 5.53. Note that in the case of $\overline{\mathbb{R}}_{\geqslant 0}$, the limit on the LHS of (5.23) must exist in $\overline{\mathbb{R}}_{\geqslant 0}$. Again, this is due to the fact that the involved net is increasing, and so one can use Exe. 5.16.

Exercise 5.54. Let $\sum_{n=-\infty}^{+\infty} f(n)$ be a series in either V or $\overline{\mathbb{R}}_{\geq 0}$.

1. Fix $k \in \mathbb{Z}$. Prove that $\sum_{n=-\infty}^{+\infty} f(n)$ converges iff the following limits converge:

$$\sum_{n=k}^{+\infty} f(n) = \lim_{n \to +\infty} \sum_{i=k}^{n} f(i) \tag{5.24a}$$

$$\sum_{n=-\infty}^{k-1} f(n) = \lim_{m \to +\infty} \sum_{i=-m}^{k-1} f(i) \tag{5.24b}$$

Moreover, if these limits converge, then

$$\sum_{n=-\infty}^{+\infty} f(n) = \sum_{n=k}^{+\infty} f(n) + \sum_{n=-\infty}^{k-1} f(n) \tag{5.25}$$

2. In the case that f has codomain V, prove that $\sum_{n=-\infty}^{+\infty} f(n)$ converges if it converges absolutely.

3. Prove that if f is zero outside $\mathbb{Z}_+$, then

$$\sum_{n=-\infty}^{+\infty} f(n) = \sum_{n=1}^{+\infty} f(n) \tag{5.26}$$

Thus, by (5.26), our following results about series over $\mathbb{Z}$ can be directly applied to series over $\mathbb{Z}_+$ (or over $\mathbb{Z}_{\geq k}$ where $k \in \mathbb{Z}$).

5.5.2 Parametrization theorem

The following theorem relates series and discrete integrals. The structure of this theorem is similar to that of Fubini's theorem-A,B,C in Sec. 5.4.

Theorem 5.55 (Parametrization theorem). *Let X be an infinite countable set. Let $\varphi : \mathbb{Z} \to X$ be a bijection (called a **parametrization** of X). The following are true.*

1. *Let $f : X \to V$. If the RHS of (5.27) converges in V, then the LHS converges, and (5.27) holds:*

$$\sum_{n=-\infty}^{+\infty} f \circ \varphi(n) = \sum_{x \in X} f(x) \tag{5.27}$$

2. *Let $f : X \to \overline{\mathbb{R}}_{\geq 0}$. Then (5.27) holds in $\overline{\mathbb{R}}_{\geq 0}$.*

3. Let $f : X \to V$. Then the discrete integral $\sum_{x \in X} f(x)$ converges absolutely iff the series $\sum_{n=-\infty}^{+\infty} f \circ \varphi(n)$ converges absolutely.

The same conclusions hold if we assume that $\varphi : \mathbb{Z}_+ \to X$ is a bijection.

Proof. We prove the case $\varphi : \mathbb{Z} \to X$; the other case is similar.

Assume that $\sum_X f$ converges, which means that the limit of the net $(\sum_A f)_{A \in \text{fin}(2^X)}$ converges to some $v \in V$. Therefore, by Thm. 5.24, the subnet

$$\left(\sum_{x \in A_{m,n}} f(x) \right)_{m,n \in \mathbb{Z}_+} = \left(\sum_{i=-m}^{n} f \circ \varphi(i) \right)_{m,n \in \mathbb{Z}_+}$$

converges to v, where $A_{m,n} = \{\varphi(i) : i \in \mathbb{Z}, -m \leqslant i \leqslant n\}$. This proves part 1. The same method proves part 2. Part 3 follows directly from part 2. $\qquad\square$

5.6 Application to (double) series and power series; $e^z e^w = e^{z+w}$

Fix a Banach space V over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$.

5.6.1 General results about series and double series

Corollary 5.56. *Let $f : \mathbb{Z} \to V$, and let $\psi : \mathbb{Z} \to \mathbb{Z}$ be a bijection. Suppose that*

$$\sum_{n=-\infty}^{+\infty} \|f(n)\| < +\infty \tag{5.28}$$

Then (5.29) holds true, where the RHS of (5.29) converges absolutely:

$$\sum_{n=-\infty}^{+\infty} f(n) = \sum_{n=-\infty}^{+\infty} f \circ \psi(n) \tag{5.29}$$

The same conclusion clearly holds if $\mathbb{Z}$ is replaced by $\mathbb{Z}_+$.

Proof. By (5.28) and Thm. 5.55-3, the discrete integral $\sum_{\mathbb{Z}} f$ converges absolutely, and hence converges. By Thm. 5.55-1, the LHS resp. RHS of (5.29) converges to the value of $\sum_{\mathbb{Z}} f$ if we choose the parametrization to be $\text{id}_{\mathbb{Z}}$ resp. ψ. This proves (5.29) and the convergence of the RHS of (5.29). Applying the same conclusion to $\|f(\cdot)\|$ proves the absolute convergence of the RHS of (5.29). $\qquad\square$

Corollary 5.57. *Let $f : \mathbb{Z}^2 \to V$. Let $\Phi : \mathbb{Z}^2 \to \mathbb{Z}^2$ be a bijection. Suppose that*

$$\sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \|f(m,n)\| < +\infty \tag{5.30}$$

Then (5.31) *holds true, where the six series involved in* (5.31) *converge absolutely.*

$$\sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} f(m,n) = \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} f(m,n) = \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} f \circ \Phi(k,l) \tag{5.31}$$

Similar results clearly hold if $\mathbb{Z}^2$ is replaced by $\mathbb{Z}_+^2$: One extends the domain of f and the domain and codomain of Φ from $\mathbb{Z}_+^2$ to $\mathbb{Z}^2$. Then one apply Cor. 5.57.

Also, note that the second term of (5.31) is redundant: it follows from the equality of the first and the third terms of (5.31) if we choose $\Phi(k,l) = (l,k)$.

Proof. By Thm. 5.55-2 and Thm. 5.50, we have

$$\sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \|f(m,n)\| = \sum_{x\in\mathbb{Z}} \sum_{y\in\mathbb{Z}} \|f(x,y)\| = \sum_{(x,y)\in\mathbb{Z}^2} \|f(x,y)\| \tag{5.32}$$

where all the limits exist in $\overline{\mathbb{R}}_{\geq 0}$. Therefore, by (5.30), the discrete integral $\sum_{\mathbb{Z}^2} f$ is absolutely convergent and hence convergent.

Similar to the argument for (5.32), Thm. 5.55-1 and Thm. 5.49 imply that the first two terms of (5.31) exist and are both equal to the discrete integral $\sum_{\mathbb{Z}^2} f$. Since $\sum_{\mathbb{Z}^2} f = \sum_{\mathbb{Z}^2} f \circ \Phi$ (recall Rem. 5.42), by Thm. 5.55-1 and Thm. 5.49 again, the last term of (5.31) converges to $\sum_{\mathbb{Z}^2} f \circ \Phi$.

We have proved that the six series in (5.31) converge, and (5.31) holds. Replacing $f(\cdot, \cdot)$ with $\|f(\cdot, \cdot)\|$ and applying a similar argument, we see that the six series in (5.31) converge absolutely. $\qquad\square$

Remark 5.58. Using the same method as in the above proof, one can easily prove a more general version of Cor. 5.57: Let $N \in \mathbb{Z}_+$. Let $f : \mathbb{Z}^2 \to V$ such that (5.30) holds true. Let $\Psi : \mathbb{Z}^N \to \mathbb{Z}^2$ be a bijection. Then the N series involved in the expression of (5.33) (from innermost to outermost) converge absolutely:

$$\sum_{n_1=-\infty}^{+\infty} \cdots \sum_{n_N=-\infty}^{+\infty} f \circ \Psi(n_1, \ldots, n_N) \tag{5.33}$$

Moverover, the outermost series of (5.33) converges to (5.31). And of course, a similar result holds if $\mathbb{Z}^2$ is replaced by $\mathbb{Z}^M$ for every $M \in \mathbb{Z}_+$. We leave it to the readers to fill in the details.

Corollary 5.59. *Assume that*

$$A = \sum_{n=-\infty}^{+\infty} a_n \qquad B = \sum_{n=-\infty}^{+\infty} b_n$$

are absolutely convergent series in $\mathbb{C}$. *Then for each* $k \in \mathbb{Z}$, *the series*

$$c_k = \sum_{l=-\infty}^{+\infty} a_{k-l} b_l$$

converges absolutely. Moreover, the LHS of the (5.34) *converges absolutely to the RHS:*

$$\sum_{k=-\infty}^{+\infty} c_k = AB \tag{5.34}$$

Proof. Apply Cor. 5.57 to the case that $f(m, n) = a_m b_n$ and $\Phi(k, l) = (k - l, l)$. □

5.6.2 Application to power series

Corollary 5.60. *Let* $f(z) = \sum_{n=0}^{+\infty} a_n z^n$ *and* $g(z) = \sum_{n=0}^{+\infty} b_n z^n$ *be power series in* $\mathbb{C}$ *with radii of convergence* R_1, R_2 *respectively. Let* $R = \min\{R_1, R_2\}$. *For each* $k \in \mathbb{Z}_+$, *let*

$$c_k = \sum_{l=0}^{k} a_{k-l} b_l$$

Then the power series $h(z) = \sum_{k=0}^{+\infty} c_k z^k$ *has radius of convergence* $\geqslant R$. *Moreover, for each* $z \in \mathbb{C}$ *satisfying* $0 \leqslant |z| < R$, *we have*

$$h(z) = f(z) \cdot g(z)$$

Proof. For each $0 \leqslant |z| < R$, apply Cor. 5.59 by replacing the a_n, b_n, c_k of Cor. 5.59 with $a_n z^n, b_n z^n, c_k z^k$. This shows that $h(z)$ converges absolutely to $f(z) \cdot g(z)$. Since this is true for all $|z| < R$, $h(z)$ must have radius of convergence at least R by Rem. 4.21. □

The above result also holds more generally for Laurent series. See Exe. 5.66.

Corollary 5.61. *For each* $z, w \in \mathbb{C}$ *we have*

$$e^z e^w = e^{z+w}$$

Proof. Apply Cor. 5.59 to the case $a_n = z^n/n!$ and $b_n = w^n/n!$. (We set $a_n = b_n = 0$ if $n < 0$.) Then

$$c_k = \sum_{l=0}^{k} \frac{z^{k-l}}{(k-l)!} \cdot \frac{w^l}{l!} = \sum_{l=0}^{k} \binom{k}{l} \frac{z^{k-l} w^l}{k!} = \frac{(z+w)^k}{k!}$$

by (4.5). □

5.7 Summary

The following are some fundamental questions about series and double series:

(a) Are they invariant under rearrangement? (Cf. (5.29).)

(b) Does the value of an interated double series remain unchanged if the order of the two infinite sums is changed? (Cf. the first equality in (5.31).)

(c) A mixture of the above two questions. (Cf. the last term of (5.31).)

We address these questions by relating them to discrete integral, a version of infinite sums which is parametrization-independent. The following are some key features of this theory.

1. (General principle) A discrete integral is to a series (defined by parametrization) as a net to a subnet.

2. (Net $\Rightarrow$ Subnet) All subnets of a convergent net converge to the same value: the limit of the original net.

3. (Discrete integral $\Rightarrow$ Series) Therefore, different series converge to the same value if they are different parametrizations of the same convergent discrete integral.

4. (Discrete integrals $\Rightarrow$ Series) Fubini-type theorems (any theorems about exchanging the orders of iterated sums/integrals) hold for convergent double discrete integrals. Therefore, they hold when passing to subnets, in particular, when passing to double series.

5. (Subnet $\Rightarrow$ Net) Every increasing net in $\overline{\mathbb{R}}_{\geqslant 0}$ has a limit in $\overline{\mathbb{R}}_{\geqslant 0}$. Therefore, if an increasing net in $\mathbb{R}_{\geqslant 0}$ has a subnet converging to a number $< +\infty$, then the original net converges in $\mathbb{R}_{\geqslant 0}$ (to a finite number).

6. (General principle) The discrete integral $\sum_{x \in X} \|f(x)\|$ is defined by the limit of an increasing net in $\overline{\mathbb{R}}_{\geqslant 0}$.

7. (Series $\Rightarrow$ Discrete integral) Therefore, if any series or double series corresponds in a reasonable way to a discrete integral, then the absolute convergence of this (double) series (more specifically: (5.28) or (5.30)) implies the absolute convergence (and hence convergence) of the original discrete integral. This implies the absolute convergence of any other (double) series arising from that discrete integral.

8. (Conclusion) Thus, when a (double) series converges absolutely (in the form of (5.28) or (5.30)), the three problems (a), (b), (c) have satisfying answers. The reason absolutely convergent (double) series are so good is because increasing nets in $\overline{\mathbb{R}}_{\geqslant 0}$ are very good!

9. (Counterexamples) Non-absolutely convergent series in $\mathbb{R}$ have rearrangements converging to different values. This is because non-convergent nets may have two subnets converging to different values, cf. Pb. 5.5. (Recall from Prop. 5.48 that for discrete integrals in $\mathbb{R}$, absolute convergence is equivalent to convergence.)

5.8   Problems and supplementary material

Let X be a set, and let V be a Banach space over $\mathbb{R} \in \{\mathbb{R}, \mathbb{C}\}$.

Problem 5.1. Compute $\lim_{p,q \to +\infty} a_{p,q}$ where $(a_{p,q})_{p,q \in \mathbb{Z}_+}$ are given below. Or explain why the limit does not exist.

$$a_{p,q} = \frac{(-1)^p \cdot p}{p+q} \qquad a_{p,q} = \frac{(-1)^p}{p} \qquad a_{p,q} = \frac{\cos(p\pi/4)}{p+q}$$

Problem 5.2. Prove Thm. 5.38. (Every Cauchy net in a complete metric space converges.)

Problem 5.3. Let $f : X \to V$. Define the **support** of f to be

$$\text{Supp}(f) = \{x \in X : f(x) \neq 0\} \tag{5.35}$$

Prove that if $\sum_X f$ converges absolutely, then $\text{Supp}(f)$ is a countable set.

Hint. Consider $\{x \in X : |f(x)| \geqslant \varepsilon\}$ where $\varepsilon > 0$. ☐

Problem 5.4. Prove Prop. 5.48.

⋆ **Problem 5.5.** Prove **Riemann rearrangement theorem**, which says the following: Let $\sum_{n=1}^{+\infty} x_n$ be a series in $\mathbb{R}$ which converges and which does not converge absolutely. Choose any $A \in \overline{\mathbb{R}}$. Then $\sum_{n=1}^{+\infty} x_n$ has a rearrangement converging to A (i.e., there is bijection $\varphi : \mathbb{Z}_+ \to \mathbb{Z}_+$ such that $\sum_{n=1}^{+\infty} f \circ \varphi(n) = A$).

Remark 5.62. By Riemann rearrangement theorem, it is clear that every convergent series in $\mathbb{R}^N$ which is not absolutely convergent must have two rearrangements converging to two different points. However, when $\mathbb{R}^N$ is replaced by an infinite dimensional Banach space, one may find a series $\sum_{n=1}^{+\infty} v_n$ which does not converge absolutely but converge to some v, and every rearrangement of $\sum_{n=1}^{+\infty} v_n$ converges to v. See Pb. 5.6.

Problem 5.6. Consider the case that V is the real Banach space $V = l^\infty(\mathbb{Z}_+, \mathbb{R})$. For each $n \in \mathbb{Z}_+$, let $e_n \in V$ be the characteristic function $\chi_{\{n\}}$. Namey, e_n takes value 1 at n, and takes 0 at the other points. Prove that the discrete integral

$$\sum_{n \in \mathbb{Z}_+} \frac{1}{n} e_n \tag{5.36}$$

converges in V, and find the limit. Prove that (5.36) does not converge absolutely.

Remark 5.63. A more important example that will be considered later is $V = l^2(\mathbb{Z}_+, \mathbb{C})$, the set of all functions $f : \mathbb{Z}_+ \to \mathbb{C}$ satisfying that the **l^2-norm** $\|f\|_{l^2} = \sqrt{\sum_{n \in \mathbb{Z}_+} |f(n)|^2}$ is finite. Then V is in fact a Banach space. (Actually, it is a so-called **Hilbert space**.) Again, let $e_n = \chi_{\{n\}}$. (These e_n will be called an **orthonormal basis** of V.) Then for each $f \in V$, the discrete integral $\sum_{n \in \mathbb{Z}_+} f(n) \cdot e_n$ converges to f. But it does not converge absolutely if $\sum_{n \in \mathbb{Z}_+} |f(n)| = +\infty$. Take for example $f(n) = n^{-1}$. We will study these objects in the second semester.

Problem 5.7. Define $(x_{j,k})_{(j,k) \in \mathbb{Z}_+ \times \mathbb{Z}_+}$ to be

$$x_{j,k} = \begin{cases} \dfrac{1}{j^2 - k^2} & \text{if } j \neq k \\ 0 & \text{if } j = k \end{cases}$$

Prove that the discrete integral $\displaystyle\sum_{(j,k) \in \mathbb{Z}_+^2} x_{j,k}$ does not converge in $\mathbb{R}$.

Hint. Consider $(x_{j,k})$ as a net over $\mathbb{Z}^2$. Find a good bijection $\Phi : \mathbb{Z}^2 \to \mathbb{Z}^2$. $\qquad\square$

Definition 5.64. For each $f \in V^X$, define the **l^1-norm**

$$\|f\|_{l^1(X,V)} \equiv \|f\|_{l^1} \equiv \|f\|_1 = \sum_{x \in X} \|f(x)\|$$

Define the l^1-space

$$l^1(X,V) = \{f \in V^X : \|f\|_{l^1} < +\infty\}$$

Namely, $l^1(X,V)$ is the set of all $f \in V^X$ where $\sum_X f$ converges absolutely. In particular, $\sum_X f$ converges for such f.

Exercise 5.65. Prove that for each $f, g \in V^X$ and $\lambda \in \mathbb{F}$, we have

$$\|f + g\|_1 \leqslant \|f\|_1 + \|g\|_1 \qquad \|\lambda f\|_1 = |\lambda| \cdot \|f\|_1 \tag{5.37}$$

Show that $l^1(X,V)$ is a linear subspace of $l^\infty(X,V)$, and that $\|\cdot\|_{l^1}$ is a norm on $l^1(X,V)$

Problem 5.8. Prove that $l^1(X,V)$ is a Banach space. Namely, prove that the metric on $l^1(X,V)$ defined by the l^1-norm is complete.

Problem 5.9. Prove the **dominated convergence theorem** for discrete integrals: Let $(f_\alpha)_{\alpha \in I}$ be a net in V^X satisfying the following conditions:

(1) There exists $g \in l^1(X, \mathbb{R})$ satisfying $g \geqslant 0$ (i.e. $g(x) \geqslant 0$ for all $x \in X$) such that for every $\alpha \in I, x \in X$ we have

$$\|f_\alpha(x)\| \leqslant g(x)$$

We simply write the above condition as $|f_\alpha| \leqslant g$.

(2) $(f_\alpha)_{\alpha \in I}$ converges pointwise some $f \in V^X$. Namely, $\lim_\alpha f_\alpha(x) = f(x)$ for every $x \in X$.

Prove that $f \in l^1(X, V)$. Prove that the LHS of (5.38) exists and equals the RHS:

$$\lim_{\alpha \in I} \sum_{x \in X} f_\alpha(x) = \sum_{x \in X} f(x) \tag{5.38}$$

⋆ **Problem 5.10.** Assume $v_n \in V$ for each n. Let z be a complex variable. Then the expression

$$f(z) = \sum_{n=-\infty}^{+\infty} v_n z^n$$

is called a **Laurent series** in V.

Prove that there exist unique $r, R \in \overline{\mathbb{R}}_{\geqslant 0}$ such that $f(z)$ converges absolutely when $|r| < z < |R|$, and that $f(z)$ diverges when $|z| < r$ or $|z| > R$. Prove that

$$r = \limsup_{n \to +\infty} \sqrt[n]{\|v_{-n}\|} \qquad R = \frac{1}{\limsup\limits_{n \to +\infty} \sqrt[n]{\|v_n\|}} \tag{5.39}$$

(Recall that by Exe. 5.54, $f(z)$ diverges iff either $\sum_{n=0}^{\infty} v_n z^n$ or $\sum_{n=-\infty}^{-1} v_n z^n$ diverges.) We call r and R the **radii of convergence** of $f(z)$. □

⋆ **Exercise 5.66.** Consider Laurent series $f(z) = \sum_{n=-\infty}^{+\infty} a_n z^n$ (with radii of convergence $r_1 < R_1$) and $g(z) = \sum_{n=-\infty}^{+\infty} b_n z^n$ (with radii of convergence $r_2 < R_2$) in $\mathbb{C}$. Let

$$r = \max\{r_1, r_2\} \qquad R = \min\{R_1, R_2\}$$

Assume that $r < R$. Prove that for each $k \in \mathbb{Z}$, the series

$$c_k = \sum_{l=-\infty}^{+\infty} a_{k-l} b_l$$

converges absolutely. Prove that for each $z \in \mathbb{C}$ satisfying $r < |z| < R$, the LHS of the following equation converges absolutely to the RHS:

$$\sum_{k=-\infty}^{+\infty} c_k z^k = f(z)g(z) \tag{5.40}$$

6 ⋆ Construction of $\mathbb{R}$ from $\mathbb{Q}$

The goal of this chapter is to construct real numbers from rationals. More precisely, our goal is to prove Thm. 1.34. We use the method of Cantor to construct real numbers using equivalence classes of Cauchy sequences in $\mathbb{Q}$. The idea is quite simple: If we admit the existence of $\mathbb{R}$ satisfying Thm. 1.34, then by Prop. 1.29, each $x \in \mathbb{R}$ is the limit of a sequence (x_n) in $\mathbb{Q}$, which must be a Cauchy sequence. Moreover, if (y_n) is a sequence in $\mathbb{Q}$ converging to $y \in \mathbb{R}$, then by Exe. 5.37 we have $x = y$ iff (x_n) and (y_n) are Cauchy equivalent. Motivated by this, we now do not assume the existence of $\mathbb{R}$, and make the following definition:

Definition 6.1. We let $\mathscr{R}$ be the set of Cauchy sequences in $\mathbb{Q}$, [1] namely, the set of $(x_n)_{n \in \mathbb{Z}_+} \in \mathbb{Q}^{\mathbb{Z}_+}$ satisfying

$$\lim_{m,n \to +\infty} (x_m - x_n) = 0$$

We say that two elements $(x_n), (y_n)$ of $\mathscr{R}$ are Cauchy-equivalent and write $(x_n) \sim (y_n)$ if $\lim_{n \to \infty}(x_n - y_n) = 0$.

Note that the above definition does not rely on the existence of $\mathbb{R}$, because the limit of nets in $\mathbb{Q}$ can be defined using only rational numbers: a net $(\xi_\alpha)_{\alpha \in I}$ converges to ξ iff for every $\varepsilon \in \mathbb{Q}_{>0}$, ξ_α is eventually satisfies $|\xi_\alpha - \xi| < \varepsilon$. The readers can check that all the properties about limit used in this chapter does not rely on the existence of $\mathbb{R}$.

Cauchy-equivalence is clearly an equivalence condition on $\mathscr{R}$: For example, if $\lim(x_n - y_n) = \lim(y_n - z_n) = 0$ then $|x_n - z_n| \leqslant |x_n - y_n| + |y_n - z_n| \to 0$. So $|x_n - z_n| \to 0$. This proves the transitivity. The other two conditions are obvious. Therefore, we can define:

Definition 6.2. We let $\mathbb{R} = \mathscr{R}/\sim$ where $\sim$ is the Cauchy-equivalence relation. (Recall Def. 1.18). The equivalence class of $(x_n)_{n \in \mathbb{Z}_+}$ is denoted by $[x_n]_{n \in \mathbb{Z}_+} = [x_1, x_2, \dots]$, simply written as $[x_n]$. The zero element 0 of $\mathbb{R}$ is defined to be $[0, 0, \dots]$.

Exercise 6.3. Choose $[x_n] \in \mathbb{R}$ (i.e. $(x_n) \in \mathscr{R}$). The following are equivalent:

(1) $[x_n] \neq 0$. Namely, $\lim_n x_n$ does not converge to 0.

(2) There exists $\varepsilon \in \mathbb{Q}_{>0}$ such that either $x_n > \varepsilon$ eventually, or $x_n < -\varepsilon$ eventually. In particular, $x_n \neq 0$ eventually.

Consequently, the map $a \in \mathbb{Q} \mapsto [a, a, \dots] \in \mathbb{R}$ is injective. With the help of this injective map, $\mathbb{Q}$ can be viewed as a subset of $\mathbb{R}$.

---

[1] Only in this chapter do we use $\mathscr{R}$ for this meaning.

Exercise 6.4. Let $[x_n] \in \mathbb{R}$ and $a \in \mathbb{Q}$. Suppose that $x_n \geqslant a$ (resp. $x_n \leqslant a$) frequently. Then for every $\varepsilon \in \mathbb{Q}_{>0}$, we have that $x_n \geqslant a - \varepsilon$ (resp. $x_n \leqslant a + \varepsilon$) eventually.

Definition 6.5. If $\xi, \eta \in \mathbb{R}$, write $\xi = [x_n]$ and $\eta = [y_n]$. In the case that $\eta \neq 0$, we assume $y_n \neq 0$ for all n, which is possible by Exe. 6.3. Define

$$[x_n] + [y_n] = [x_n + y_n]$$
$$-[x_n] = [-x_n]$$
$$[x_n] \cdot [y_n] = [x_n y_n]$$
$$1/[y_n] = [1/y_n] \qquad (\text{if } [y_n] \neq 0)$$

Exercise 6.6. Prove that the above formulas are well-defined: For example, if $(x_n) \sim (x'_n)$ in $\mathscr{R}$, then $(x_n y_n) \sim (x'_n y_n)$. (You may need the easy fact that every Cauchy sequence is bounded.)

Remark 6.7. It is clear that Def. 6.5 makes $\mathbb{R}$ a **field**, which means that for every $\alpha, \beta, \gamma \in \mathbb{R}$, the following are satisfied:

$$\alpha + \beta = \beta + \alpha \qquad (\alpha + \beta) + \gamma = \alpha + (\beta + \gamma) \qquad 0 + \alpha = \alpha \qquad \alpha + (-\alpha) = 0$$
$$\alpha\beta = \beta\alpha \qquad (\alpha\beta)\gamma = \alpha(\beta\gamma) \qquad 1 \cdot \alpha = \alpha \qquad (\alpha + \beta)\gamma = \alpha\gamma + \beta\gamma$$
$$\alpha \cdot \frac{1}{\alpha} = 1 \qquad (\text{if } \alpha \neq 0)$$

Moreover, $\mathbb{Q}$ is a subfield of $\mathbb{R}$ where the addition, taking negative, multiplication, and inverse of $\mathbb{R}$ restrict to those of $\mathbb{Q}$.

Definition 6.8. Let $[x_n], [y_n] \in \mathbb{R}$. We write $[x_n] < [y_n]$ if one of the following equivalent (due to Exe. 6.4) statements hold:

- There exists $\varepsilon \in \mathbb{Q}_{>0}$ such that $y_n - x_n > \varepsilon$ eventually.

- There exists $\varepsilon \in \mathbb{Q}_{>0}$ such that $y_n - x_n > \varepsilon$ frequently.

It is not hard show that " $<$ " is well-defined, and that (by Exe. 6.3) if $[x_n] < [y_n]$ then $[x_n] \neq [y_n]$. We write $[x_n] \leqslant [y_n]$ if $[x_n] < [y_n]$ or $x_n = y_n$.

Lemma 6.9. $(\mathbb{R}, \leqslant)$ *is a totally ordered set.*

Proof. Choose $[x_n], [y_n], [z_n] \in \mathbb{R}$. If $[x_n] < [y_n]$ and $[y_n] < [z_n]$, then clearly $[x_n] < [z_n]$. This proves that $\leqslant$ is a preorder.

Suppose $[x_n] \leqslant [y_n]$ and $[y_n] \leqslant [x_n]$. Let us prove $[x_n] = [y_n]$. Suppose not. Then $[x_n] < [y_n]$ and $[y_n] < [x_n]$ by the definition of "$\leqslant$". So there is $\varepsilon > 0$ such that $y_n - x_n > \varepsilon$ eventually, and $x_n - y_n > \varepsilon$ eventually. Impossible. So $\leqslant$ is a partial order.

Suppose $[x_n] \neq [y_n]$. Then $(x_n - y_n) \nsim 0$. So Exe. 6.3 implies that either $[x_n] < [y_n]$ or $[y_n] < [x_n]$. So $\leqslant$ is a total order. □

94

Lemma 6.10. *Let $[x_n], [y_n] \in \mathbb{R}$. Then the following are equivalent.*

- $[x_n] \geqslant [y_n]$.

- *For every $\varepsilon \in \mathbb{Q}_{>0}$, $x_n - y_n \geqslant -\varepsilon$ frequently.*

- *For every $\varepsilon \in \mathbb{Q}_{>0}$, $x_n - y_n \geqslant -\varepsilon$ eventually.*

Proof. Since $\leqslant$ is a total order, the negation of $>$ is $\leqslant$. So the statements follow immediately by negating Def. 6.8. $\qquad\square$

Lemma 6.11. $\mathbb{R}$ *is an ordered field extension of $\mathbb{Q}$, and $\mathbb{R}$ is Archimedean.*

Proof. The order of $\mathbb{R}$ clearly restricts to that of $\mathbb{Q}$. We want to prove that $\mathbb{R}$ is an ordered field. (Recall Def. 1.23). Clearly, if $[x_n] < [y_n]$ and $[z_n] \in \mathbb{R}$, then $[x_n] + [z_n] = [x_n + z_n] < [y_n + z_n] = [y_n] + [z_n]$. If $[x_n] > 0$ and $[y_n] > 0$, then there are $\varepsilon > 0$ such that $x_n > \varepsilon$ eventually and $y_n > \varepsilon$ eventually. So $x_n y_n > \varepsilon^2$ eventually. So $[x_n][y_n] > 0$. This proves that $\mathbb{R}$ is an ordered field.

Now let $[x_n] > 0$ and $[y_n] \in \mathbb{R}$. So there exist $\varepsilon \in \mathbb{Q}_{>0}$ such that $x_n > \varepsilon$ eventually. Since (y_n) is Cauchy, one checks easily that (y_n) is bounded. So there is $M \in \mathbb{Q}_{>0}$ such that $|y_n| \leqslant M$ for all n. Since $\mathbb{Q}$ is Archimedean, there exists $k \in \mathbb{Z}_+$ such that $k\varepsilon > M + 1$. So $kx_n > M + 1$ eventually. So $k[x_n] > M$. This proves that $\mathbb{R}$ is Archimedean. $\qquad\square$

To finish the proof of Thm. 1.34, it remains to prove that $\mathbb{R}$ satisfies the least-upper-bound property.

Lemma 6.12. *Thm. 1.34 holds if every bounded increasing sequence in $\mathbb{R}$ converges.*

Proof. Suppose that every bounded increasing sequence in $\mathbb{R}$ converges. Choose any nonempty $E \subset \mathbb{R}$ bounded from above. We shall show that E has a least upper bound.

Let F be the set of upper bounds of E. Namely, $F = \{\eta \in \mathbb{R} : \eta \geqslant \xi, \forall \xi \in E\}$. So $F \neq \varnothing$. We construct an increasing sequence (ξ_k) in E and an decreasing sequence (η_k) in F as follows. Since E, F are nonempty, we choose arbitrary $\xi_1 \in E$ and $\eta_1 \in F$. Then $\xi_1 \leqslant \eta_1$. Suppose $\xi_1 \leqslant \cdots \leqslant \xi_k \in E$ and $\eta_1 \geqslant \cdots \geqslant \eta_k \in F$ have been constructed. Let $\psi_k = (\xi_k + \eta_k)/2$. Let

$$\begin{cases} \xi_{k+1} = \psi_k, \eta_{k+1} = \eta_k & \text{if } \psi_k \in E \\ \xi_{k+1} = \xi_k, \eta_{k+1} = \psi_k & \text{if } \psi_k \in F \end{cases}$$

Then the sequences we have constructed satisfy $\lim_{k\to\infty}(\eta_k - \psi_k) = 0$.

By assumption, $\alpha = \lim_{k\to\infty} \xi_k$ exists, and it equals $\lim_k \eta_k$. So α is an upper bound of E. (If $\lambda \in E$, then $\lambda \leqslant \eta_k$ for all k since $\eta_k \in F$. So $\lambda \leqslant \lim_k \eta_k = \alpha$.) We now show that α is the least upper bound. Let $\varepsilon > 0$. Since $\xi_k \to \alpha$, there is k such that $\alpha - \xi_k < \varepsilon$. So $\xi_k > \alpha - \varepsilon$, and hence $\alpha - \varepsilon$ is not an upper bound of E. $\qquad\square$

Lemma 6.13. *Thm. 1.34 holds if every bounded increasing sequence in $\mathbb{Q}$ converges to an element of $\mathbb{R}$.*

Proof. Suppose that every increasing sequence in $\mathbb{Q}$ converges in $\mathbb{R}$. By Lem. 6.12, it suffices to prove that every increasing sequence (ξ_k) in $\mathbb{R}$ converges. If $\{\xi_k : k \in \mathbb{Z}_+\}$ is a finite subset of $\mathbb{R}$, then (ξ_k) clearly converges. If $\{\xi_k : k \in \mathbb{Z}_+\}$ is infinite, then (ξ_k) clearly has a strictly increasing subsequence (ξ_{k_l}). If we can prove that (ξ_{k_l}) converges to some $\psi \in \mathbb{R}$, then (ξ_k) converges to ψ. (Choose any $\varepsilon > 0$. Choose $L \in \mathbb{Z}_+$ such that $|\psi - \xi_{k_L}| < \varepsilon$ and hence $0 \leqslant \psi - \xi_{k_L} < \varepsilon$. Then for all $k \geqslant k_L$ we have $0 \leqslant \psi - \xi_k < \varepsilon$.)

Thus, it remains to prove that every strictly increasing sequence (η_k) in $\mathbb{R}$ converges. Since we have proved that $\mathbb{R}$ is an Archimedean ordered field extension of $\mathbb{Q}$, by Prop. 1.29, for each k, there exists $a_k \in \mathbb{Q}$ such that $\xi_k < a_k < \xi_{k+1}$. By assumption, (a_k) converges to some $\alpha \in \mathbb{R}$. Since $a_{k-1} < \xi_k < a_k$, by squeeze theorem, (ξ_k) converges to α. $\qquad\square$

Proof of Thm. 1.34. By Lem. 6.13, it suffices to show that every bounded increasing sequence (a_k) in $\mathbb{Q}$ converges in $\mathbb{R}$. Let $M \in \mathbb{Q}$ such that $a_k \leqslant M$ for all k.

We first prove that (a_k) is a Cauchy sequence. If not, then there exists $\varepsilon \in \mathbb{Q}_{>0}$ such that for every $K \in \mathbb{Z}_+$ there is $k > K$ such that $|a_k - a_K| > \varepsilon$, and hence $a_k - a_K > \varepsilon$. Thus, we can find a subsequence (a_{k_l}) such that $a_{k_{l+1}} - a_{k_l} > \varepsilon$. By the Archimedean property for $\mathbb{Q}$, there is $l \in \mathbb{Z}_+$ such that $a_{k_1} + l \cdot \varepsilon > M$. So $a_{k_{l+1}} > M$, impossible.

Note that each a_k is identified with $\xi_k = [a_k, a_k, \dots]$. Let $\psi = [a_1, a_2, a_3, \dots]$, which is an element of $\mathbb{R}$ since we just proved that $(a_n) \in \mathscr{R}$. Then for each k, $\psi - \xi_k = [a_1 - a_k, a_2 - a_k, \dots]$, where the terms are eventually $\geqslant 0$. So $\xi_k \leqslant \psi$ by Lem. 6.10. We have proved that ψ is an upper bound for the sequence (ξ_k).

Let us prove that $\lim_k \xi_k = \psi$. Choose any $\varepsilon \in \mathbb{Q}_{>0}$. Let us prove that there exists k such that $\psi - \varepsilon < \xi_k$. Then for every $k' \geqslant k$ we have $\psi - \varepsilon < \xi_{k'} \leqslant \psi$, finishing the proof of $\lim_k \xi_k = \psi$.

We have proved that $a_1, a_2, \dots$ is a Cauchy sequence in $\mathbb{Q}$. So there exists k such that $a_l - a_k < \varepsilon/2$ for all $l \geqslant k$. Thus, for all $l \geqslant k$ we have $a_k - (a_l - \varepsilon) > \varepsilon/2$. Thus, the l-th term of $\xi_k = [a_k, a_k, \dots]$ minus that of $\psi - \varepsilon = [a_1 - \varepsilon, a_2 - \varepsilon, \dots]$ is $> \varepsilon/2$ for sufficiently large l. By Def. 6.8, we have that $\psi - \varepsilon < \xi_k$. $\qquad\square$

7 Topological spaces

7.1 The topologies of metric spaces

In this chapter, we begin our study of topological spaces, which were introduced by Hausdorff in 1914 [Hau14] as a generalization of metric spaces. As we have seen, focusing on metrics in order to study convergence and continuity is often distracting. For example, in $\overline{\mathbb{R}}$, we only care about how the convergence of sequences look like, but not about the particular metrics. The same is true about the countable product of metric spaces $S = \prod_{i \in \mathbb{Z}_+} X_i$: the metrics (2.14) and (2.15) give the same topology, although they look very different. Moreover, the shapes of the open balls defined by these two metrics are not very simple. This makes it more difficult to study the continuity of functions on S by using (2) or (2′) of Def. 2.38.

Topological spaces generalize metric spaces by giving a set of axioms satisfied by the open sets of the spaces.

Definition 7.1. Let X be a metric space, and let $E \subset X$. A point $x \in E$ is called an **interior point** of E if $B_X(x, r) \subset E$ for some $r > 0$. We say that E is an **open (sub)set** of X, if every point of E is an interior point.

Definition 7.2. Let $\mathcal{T}$ be the set of open sets of X. We call $\mathcal{T}$ the **topology of the metric space** X.

Example 7.3. By triangle inequality, every open ball of a metric space X is open. $\varnothing$ and X are open subsets of X. If $p, q \in \mathbb{R}^N$ and $d(p, x) = r$ (where $0 \leqslant r < +\infty$), then p is not an interior point of $\overline{B}_{\mathbb{R}^N}(x, r)$. So the closed balls of $\mathbb{R}^N$ are not open sets. In particular, $[a, b]$ are not open subsets of $\mathbb{R}$ since a, b are not interior points.

Example 7.4. It is not hard to see that a finite intersection of open sets is open.

In topological spaces, open sets play the role of open balls in metric spaces due to the following facts:

Exercise 7.5. Let (x_n) be a sequence in a metric space X. Let $x \in X$. Show that the following are equivalent:

(1) (x_n) converges to x.

(2) For every **neighborhood** U **of** x (i.e. every open set containing x) there is $N \in \mathbb{Z}_+$ such that for every $n \geqslant N$ we have $x_n \in U$.

Exercise 7.6. Let $f : X \to Y$ be a map of metric spaces. Let $x \in X$ and $y = f(x)$. Prove that the following are equivalent.

(1) f is continuous at x.

(2) For every neighborhood V of y there is a neighborhood U of x such that $f(U) \subset V$ (equivalently, $U \subset f^{-1}(V)$.)

But there is an important difference between the intuitions of open sets and open balls: We want the open balls at a point x to be small so that they can be used to describe the approximation to x. However, an arbitrary open set can be very large. For example, when studying convergence and continuity in $\mathbb{R}$, we really want a neighborhood of 1 to be $(1 - \varepsilon, 1 + \varepsilon)$ but not the more complicated and bigger one $(-\infty, -2) \cup (0, 100 - \varepsilon)$. Indeed, open sets can be very big:

Lemma 7.7. *Let X be a metric space. If $(U_\alpha)_{\alpha \in I}$ is a family of open subsets of X, then $W = \bigcup_{\alpha \in I} U_\alpha$ is open in X.*

Proof. Choose $x \in W$. Then $x \in U_\alpha$ for some α. So $B_X(x, r) \subset U_\alpha$ for some $r > 0$. So x is an interior point of W. $\qquad \square$

Thus, people very often choose a class $\mathcal{B}$ of smaller open sets (such as the set of open balls) to study the analytic properties of a topological space.

Definition 7.8. Let $\mathcal{B}$ be a set of open sets of a metric space (or more generally, a topological space) X. We say that $\mathcal{B}$ is a **basis for the topology $\mathcal{T}$** of X if one of the following (clearly) equivalent statements holds:

- For every point $x \in X$ and every neighborhood W of x there exists $U \in \mathcal{B}$ such that $x \in U$ and $U \subset W$.

- Every open subset of X is a union of some members of $\mathcal{B}$.

Thus, according to Def. 7.1, the set of open balls of a metric space X form a basis for the topology of X. Nevertheless, even in the case of metric spaces, we sometimes consider more convenient bases than the set of open balls. We will see this when we study the topologies of infinite product spaces.

7.2 Topological spaces

7.2.1 Definitions and basic examples

Definition 7.9. We say that a pair $(X, \mathcal{T})$ (or simply X) is a **topological space** if X is a set, and if $\mathcal{T}$ (called the **topology** of X) is a set of subsets of X satisfying the following conditions

- $\varnothing \in \mathcal{T}$ and $X \in \mathcal{T}$.

- (Union property) If $(U_\alpha)_{\alpha \in I}$ is a family of elements of $\mathcal{T}$, then $\bigcup_{\alpha \in I} U_\alpha$ is an element of $\mathcal{T}$.

- (Finite intersection property) If $n \in \mathbb{Z}_+$ and $U_1, \ldots, U_n \in \mathcal{T}$, then $U_1 \cap \cdots \cap U_n$ is an element of $\mathcal{T}$.

Elements of $\mathcal{T}$ are called **open (sub)sets** of X.

Definition 7.10. Let X be a topological space, and $x \in X$. A subset $U \subset X$ is called a **neighborhood** of x, if U is an open subset of X containing x.[1] We define $(\mathrm{Nbh}_X(x), \leqslant)$, the **directed set of neighborhoods of** x, to be

$$\mathrm{Nbh}_X(x) = \{\text{neighborhoods of } x \text{ in } X\}$$
$$U \leqslant U' \qquad \Longleftrightarrow \qquad U \supset U' \tag{7.1}$$

(Note that one needs the finite intersection property to show that $\mathrm{Nbh}_X(x)$ is a directed set.) We abbreviate this set to $\mathrm{Nbh}_X(x)$ or simply $\mathrm{Nbh}(x)$.

Example 7.11. In Subsec. 7.1, we have proved that the topology of a metric space satisfies the above axioms of a topological space.

In particular, if X is a normed vector space, the topology induced by the metric $d(x, x') = \|x - x'\|$ is called the **norm topology**. If X is a subset of $\mathbb{R}^N$ or $\mathbb{C}^N$, the topology on X induced by the Euclidean metric is called the **Euclidean topology**. $\qquad\square$

Definition 7.12. A topological space $(X, \mathcal{T})$ is called **metrizable**, if there is a metric on X inducing the topology $\mathcal{T}$.

We have seen that the open balls of a metric space generate a topology. In general, one may ask what possible subsets of 2^X generate a topology on a set X. Here is a description, whose proof is left to the readers as an exercise.

Proposition 7.13. *Let X be a set, and let $\mathcal{B} \subset 2^X$. Define*

$$\mathcal{T} = \{\text{Unions of elements of } \mathcal{B}\} \tag{7.2}$$

The following are equivalent.

(1) $(X, \mathcal{T})$ is a topological space.

(2) The following are satisfied:

(2-a) $X = \bigcup_{U \in \mathcal{B}} U$.

(2-b) If $U_1, U_2 \in \mathcal{B}$, then $U_1 \cap U_2 \in \mathcal{T}$ (i.e., for each $x \in U_1 \cap U_2$ there exists $V \in \mathcal{B}$ such that $x \in V$ and $V \subset U_1 \cap U_2$).

---

[1]We are following the convention in [Mun, Rud-R]. But many people refer to the word "neighborhood" with slightly different meaning: a subset A is called a neighborhood of x if there is an open set U such that $x \in U \subset A$. And our neighborhoods are called "open neighborhoods" by them.

When (1) or (2) holds, we call $\mathcal{T}$ the **topology generated by** $\mathcal{B}$. Clearly, $\mathcal{B}$ is a basis for $\mathcal{T}$ (cf. Def. 7.8).

Exercise 7.14. Let X be a set. Let $\mathcal{B}, \mathcal{B}'$ be subsets of 2^X generating topologies $\mathcal{T}, \mathcal{T}'$ respectively. Prove that the following are equivalent.

(1) $\mathcal{T} = \mathcal{T}'$.

(2) Each $U \in \mathcal{B}$ is a union of elements of $\mathcal{B}'$. Each $U' \in \mathcal{B}'$ is a union of elements of $\mathcal{B}$.

(3) For each $U \in \mathcal{B}$ and $x \in U$, there exists $U' \in \mathcal{B}'$ such that $x \in U' \subset U$. For each $U' \in \mathcal{B}$ and $x \in U'$, there exists $U \in \mathcal{B}$ such that $x \in U \subset U'$.

Example 7.15. If $(X, \mathcal{T})$ is a topological space, then $\mathcal{T}$ is a basis for $\mathcal{T}$.

Example 7.16. Let (X, d) be a metric space with topology $\mathcal{T}$. Let $\mathcal{B} = \{B_X(x, r) : x \in X, 0 < r < +\infty\}$. Then $\mathcal{B}$ is a basis for $\mathcal{T}$. For each $\varepsilon > 0$, the set $\mathcal{B}' = \{B_X(x, r) : x \in X, 0 < r < \varepsilon\}$ is also a basis for $\mathcal{T}$.

Example 7.17. Let $\mathcal{B} \subset 2^{\overline{\mathbb{R}}}$ be defined by

$$\mathcal{B} = \{(a, b), (c, +\infty], [-\infty, d) : a, b, c, d \in \mathbb{R}\} \tag{7.3}$$

Using Prop. 7.13, one easily checks that $\mathcal{B}$ is a basis for a topology $\mathcal{T}$. We call this the **standard topology** of $\overline{\mathbb{R}}$.

Let $\varphi : \overline{\mathbb{R}} \to [u, v]$ be a strictly increasing bijection where $-\infty < u < v < +\infty$. Let $d_{[u,v]}$ be the Euclidean metric, and let $\mathcal{T}'$ be the topology on $\overline{\mathbb{R}}$ defined by $d_{\overline{\mathbb{R}}} = \varphi^* d_{[u,v]}$. Then the set of open balls under $\mathcal{T}'$ is

$$\mathcal{B}' = \{(\varphi^{-1}(y - \varepsilon), \varphi^{-1}(y + \varepsilon)), (\varphi^{-1}(v - \varepsilon'), +\infty], [-\infty, \varphi^{-1}(u + \varepsilon'')) :$$
$$y \in (u, v) \text{ and } \varepsilon, \varepsilon', \varepsilon'' \in \mathbb{R}_{>0}\}$$

(Note that the three types of intervals in the definition of $\mathcal{B}'$ are open balls centered at $\varphi^{-1}(y), +\infty, -\infty$ respectively.) Using Exe. 7.14, one easily checks $\mathcal{T} = \mathcal{T}'$. $\qquad\square$

Convention 7.18. Unless otherwise stated, the topology on $\overline{\mathbb{R}}$ is defined to be the standard one, i.e., the one generated by (7.3). We shall forget about the metric on $\overline{\mathbb{R}}$, and view $\overline{\mathbb{R}}$ only as a (metrizable) topological space.

Definition 7.19. Let A be a subset of a topological space $(X, \mathcal{T}_X)$. Then

$$\mathcal{T}_A = \{U \cap A : U \in \mathcal{T}_X\}$$

is clearly a topology on A, called the **subspace topology**. Unless otherwise stated, when viewing a subset as a topological subspace, we always choose the subspace topology for the subset.

Exercise 7.20. Let (X, d_X) be a metric space, inducing a topology $\mathcal{T}_X$. Let A be a metric subspace of X. (So $A \subset X$, and d_X restricts to d_A.) Prove that the topology on A induced by d_A is the subspace topology.

According to the above exercise, if X is a metric space, then viewing a subset A as a topological subspace is compatible with viewing A as a metric subspace.

Exercise 7.21. Let A be a subset of a topological space X. Let $\mathcal{B}$ be a basis for the topology of X. Show that $\{U \cap A : U \in \mathcal{B}\}$ is a basis for the subspace topology of A.

7.2.2 Convergence of nets

Definition 7.22. Let $(x_\alpha)_{\alpha \in I}$ be a net in a topological space X. Let $x \in X$. We say that (x_α) **converges to** x and write

$$\lim_{\alpha \in I} x_\alpha \equiv \lim_\alpha x_\alpha = x$$

or simply write $x_\alpha \to x$, if the following statement holds:

- For every $U \in \mathrm{Nbh}_X(x)$, we have that x_α is eventually in U.

Clearly, if $\mathcal{B}$ is a basis for the topology, then $x_\alpha \to x$ iff:

- For every $U \in \mathcal{B}$ containing x, we have that x_α is eventually in U.

In the case that X is a metric space (and the topology of X is induced by the metric), the definition here agrees with Def. 5.11.

Exercise 7.23. Let (x_α) be a net in X converging to $x \in X$. Prove that every subnet of (x_α) converges to x.

Exercise 7.24. Let A be a subset of a topological space X, equipped with the subspace topology. Let (x_α) be a net in A, and let $x \in A$. Show that $x_\alpha \to x$ in A iff $x_\alpha \to x$ in X.

Example 7.25. Let X be a set. Let $\mathcal{T} = \{\varnothing, X\}$. Then every net in X converges to every point of X. Thus, if X has at least two elements, then the limit of a net in X is not unique. Therefore, a general topological space might be very pathological. To avoid this uniqueness issue, we introduce the following notion:

Definition 7.26. Let X be a topological space with a basis for the topology $\mathcal{B}$. We say that X is a **Hausdorff space** if the following equivalent conditions are satisfied:

(1) (Hausdorff condition) If $x, y \in X$ and $x \neq y$, then there exist neighborhoods U of x and V of y such that $U \cap V = \varnothing$.

(1′) If $x, y \in X$ and $x \neq y$, then there exist $U \in \mathcal{B}$ containing x and $V \in \mathcal{B}$ containing y such that $U \cap V = \varnothing$.

(2) If $(x_\alpha)_{\alpha \in I}$ is a net in X converging to both x and y, then $x = y$.

Proof of the equivalence. (1)⟺(1′): Obvious.

(1)⟹(2): Suppose that (x_α) converges to x and y. Suppose $x \neq y$. By (1), we have disjoint neighborhoods $U \ni x$ and $V \ni y$. Since $x_\alpha \to x$, x_α is eventually in U. Similarly, x_α is eventually in V. Therefore, by the logic (5.6b), x_α is eventually in $U \cap V = \varnothing$, impossible.

¬(1)⟹ ¬(2): Suppose that (1) is not true. Then there exist $x \neq y$ such that every neighborhood of x intersects every neighborhood of y. Let $I = \mathrm{Nbh}_X(x) \times \mathrm{Nbh}_X(y)$. For each $\alpha = (U, V) \in I$, by assumption, there exists $x_\alpha \in U \cap V$. Then $(x_\alpha)_{\alpha \in I}$ is a net in X. We leave it to the readers to check that $x_\alpha \to x$ and $x_\alpha \to y$. □

Remark 7.27. In Hausdorff's 1914 paper introducing topological spaces, the Hausdorff condition is one of the axioms of topological spaces. Non-Hausdorff topological spaces were studied much later. The reason that Hausdorff spaces appeared first may be as follows: The original motivation for topological spaces lies in the study of analysis (especially functional analysis). But in analysis, most spaces are Hausdorff, because we want the limits of sequences or nets to be unique.

In differential geometry and in topology[2], people are also mainly concerned with topological spaces that are Hausdorff. This is related to the fact that in these areas people often use tools from analysis. But in algebraic geometry, the main examples of topological spaces (e.g. varieties and schemes, whose topologies are called **Zariski topology**) are not Hausdorff. As a related fact, sequences and nets are not effective tools in the study of algebraic geometry. □

7.3 Closures, interiors, and closed sets

In this section, we fix a topological space X.

7.3.1 Closure points; dense subsets

Definition 7.28. Let A be a subset of X. We say that $x \in X$ is a **closure point** of A, if the following equivalent conditions hold:

(1) There is a net $(x_\alpha)_{\alpha \in I}$ in A converging to x.

(2) Each $U \in \mathrm{Nbh}_X(x)$ intersects A.

---

[2]Here, I mean genuine topology, such as algebraic topology, differential topology, geometric topology, etc., but not point-set topology, which is analysis under the guise of topology.

The **closure** of A is defined to be

$$\overline{A} \equiv \mathrm{Cl}(A) \equiv \mathrm{Cl}_X(A) = \{\text{closure points of } A\}$$

Clearly $A \subset \overline{A}$. Clearly, if $A \subset B \subset X$, then $\overline{A} \subset \overline{B}$.

Unless otherwise stated, if several subsets are involved, we always understand $\overline{A}$ as $\mathrm{Cl}_X(A)$ where X is the ambient topological space.

Proof of equivalence. (1)$\Rightarrow$(2): Assume (1). Choose any $U \in \mathrm{Nbh}_X(x)$. Since $x_\alpha \to x$, we have that x_α is eventually in U. So U must contain some x_α. But $x_\alpha \in A$. So $U \cap A \neq \varnothing$.

(2)$\Rightarrow$(1): By (2), for each $U \in \mathrm{Nbh}_X(x)$ we can choose $x_U \in U \cap A$. Then $(x_U)_{U \in \mathrm{Nbh}_X(x)}$ is a net in A converging to x. $\qquad\square$

Exercise 7.29. Let $\mathcal{B}$ be a basis for the topology of X. Show that $x \in X$ is a closure point of A iff every $U \in \mathcal{B}$ containing x must intersect A.

Exercise 7.30. Let A be a subset of a metric space. Show that $x \in X$ is a closure point of A iff there is a sequence $(x_n)_{n \in \mathbb{Z}_+}$ in A converging to x.

Exercise 7.31. Recall that if X is a metric space, then $\overline{B}_X(x, r) = \{y \in X : d(x, y) \leqslant r\}$. Show that

$$\overline{B_X(x, r)} \subset \overline{B}_X(x, r) \tag{7.4}$$

and that these two sets are not necessarily equal.

Remark 7.32. Our proof of (2)$\Rightarrow$(1) in Def. 7.28 is an indirect proof, because it uses axiom of choice. (Given $U \in \mathrm{Nbh}_X(x)$, the choice of $x_U \in U \cap A$ is highly arbitrary.) Here is a direct proof: Assume (2). Define a direct set $(I, \leqslant)$ where

$$I = \{(p, U) : U \in \mathrm{Nbh}_X(x), p \in U \cap A\}$$
$$(p, U) \leqslant (p', U') \qquad \Longleftrightarrow \qquad U \supset U'$$

The fact that I is a directed set is due to (2). Then $(p)_{(p,U) \in I}$ is a net in A converging to x.

We will often prove results about nets in topological spaces using axiom of choice, not only because it is simpler than direct proofs (as above), but also because it is parallel to our use of sequences in metric spaces. (For example, see the proof of (1)$\Rightarrow$(2) in Def. 2.38.) However, it is important to know how to give a direct proof. This is because the studies of topological spaces using nets and using open sets are often equivalent, and direct proofs using nets can be more easily translated into proofs using open sets and vice versa. $\qquad\square$

Exercise 7.33. Prove $\neg$(1)$\Rightarrow \neg$(2) of Def. 7.26 without using axiom of choice.

Remark 7.34. There is a notion closely related to closure points, called accumulation points. Let A be a subset of X. A point $x \in X$ is called a **accumulation point** (or **limit point** or **cluster point**) of A, if x is a closure point of $A\backslash\{x\}$.

We will not use the notion of accumulation points, although this concept is widely used in many textbooks on analysis or point-set topology. We use closure points instead. (But note that if $x \notin A$, then x is a closure point iff x is an accumulation point.) On the other hand, the following opposite notion of accumulation points is important and has a clear geometric picture: $\qquad\qquad\square$

Definition 7.35. We say that $x \in X$ is an **isolated point** of X, if the following (clearly) equivalent conditions hold:

(1) $x \notin \overline{X\backslash\{x\}}$.

(2) There is no net in $X\backslash\{x\}$ converging to x.

(3) There is a neighborhood of x disjoint from $X\backslash\{x\}$.

If X is a metric space, then x is an isolated point iff there is no sequence in $X\backslash\{x\}$ converging to x.

We return to the study of closures.

Proposition 7.36. *Let A be a subset of X. Then $\overline{\overline{A}} = \overline{A}$.*

Proof. Choose any $x \in \overline{\overline{A}}$. To prove $x \in \overline{A}$, we choose any $U \in \mathrm{Nbh}_X(x)$, and try to prove $U \cap A \neq \varnothing$. Since x is a closure point of $\overline{A}$, U intersects $\overline{A}$. Pick $y \in U \cap \overline{A}$. Then y is a closure point of A, and $U \in \mathrm{Nbh}_X(y)$. So U intersects A. $\qquad\square$

One should think of $\overline{\overline{A}} = \overline{A}$ not only as a "geometric" fact about closures. Instead, one should also understand its analytic content: A closure point of A is a point which can be approximated by elements of A. Thus, $\overline{\overline{A}} = \overline{A}$ says that "approximation is transitive": If x can be approximated by some elements which can be approximated by elements of A, then x can be approximated by elements of A. Alternatively, one can use the language of density:

Definition 7.37. A subset A of X is called **dense** (in X) if $\overline{A} = X$.

Exercise 7.38. Show that A is dense in X iff every nonempty open subset of X intersects A.

Remark 7.39. Let $A \subset B \subset X$. From Def. 7.28-(1), it is clear that

$$\mathrm{Cl}_B(A) = \mathrm{Cl}_X(A) \cap B \tag{7.5}$$

Thus, A is dense in B iff $B \subset \mathrm{Cl}_X(A)$.

Thus, the following property has the same meaning as $\overline{\overline{A}} = A$.

Corollary 7.40. *Let $A \subset B \subset X$. Assume that A is dense in B, and B is dense in X, then A is dense in X.*

Proof. Choose any $x \in X$. Then $x \in \mathrm{Cl}_X(B)$ since B is dense in X. Since A is dense in B, we have $B \subset \mathrm{Cl}_X(A)$. Therefore $x \in \mathrm{Cl}_X(\mathrm{Cl}_X(A))$, and hence $x \in \mathrm{Cl}_X(A)$ by Prop. 7.36. $\qquad\square$

Example 7.41. Let $X = C([0,1], \mathbb{R})$, equipped with the l^∞-norm. Let B be the set of polynomials with real coefficients, regarded as continuous functions on $[0,1]$. By Weierstrass approximation theorem (which will be studied in the future), B is a dense subset of X. Then the set A of polynomials with rational coefficients is clearly a dense subset of B under the l^∞-norm. (Proof: Let $f(x) = a_0 + a_1 x + \cdots + a_k x^k$. For each $0 \leqslant i \leqslant k$, choose a sequence $(a_{i,n})_{n \in \mathbb{Z}_+}$ in $\mathbb{Q}$ converging to a_i. Let $f_n(x) = a_{0,n} + a_{1,n} x + \cdots + a_{k,n} x^k$. Then $f_n \rightrightarrows f$ on $[0,1]$.) Therefore, A is dense in X. To summarize:

- Since each continuous function on $[0,1]$ can be uniformly approximated by polynomials with $\mathbb{R}$-coefficients, and since each polynomial can be uniformly approximated polynomials with $\mathbb{Q}$-coefficients, therefore each continuous function on $[0,1]$ can be uniformly approximated by polynomials with $\mathbb{Q}$-coefficients.

7.3.2 Interior points

Interior points are dual to closure points:

Definition 7.42. Let A be a subset of X. A point $x \in X$ is called an **interior point** of A if the following equivalent conditions hold:

(1) There exists $U \in \mathrm{Nbh}_X(x)$ such that $U \subset A$.

(2) x is not a closure point of $X \backslash A$.

The set of interior points of A is called the **interior** of A and is denoted by $\mathrm{Int}_X(A)$ or simply $\mathrm{Int}(A)$. So

$$X \backslash \mathrm{Int}(A) = \overline{X \backslash A} \qquad \text{(or simply } \mathrm{Int}(A)^c = \overline{A^c}) \tag{7.6}$$

according to (2). In particular, $\mathrm{Int}(A) \subset A$.

Proof of equivalence. A contains no neighborhoods of x with respect to X iff A^c intersects every neighborhood of x iff x is a closure point of A^c. $\qquad\square$

It is clear that if $\mathcal{B}$ is a basis for the topology, then $x \in \mathrm{Int}(A)$ iff there exists $U \in \mathcal{B}$ such that $x \in U \subset A$.

In analysis, interior points are not as commonly used as closure points. The following property is an important situation where interior points are used:

Proposition 7.43. *Let U be a subset of X. Then U is open iff every point of U is an interior point.*

In other words, U is open iff $U = \mathrm{Int}_X(U)$.

Proof. If U is open and $x \in U$, then $U \in \mathrm{Nbh}_X(x)$. So x is an interior point of U.

Conversely, suppose that each $x \in U$ is interior. Choose $V_x \in \mathrm{Nbh}_X(x)$. Then $U = \bigcup_{x \in U} V_x$. So U is open by the union property in Def. 7.9. $\qquad\square$

Note that this is the first time we seriously use the fact that a union of open sets is open.

7.3.3 Closed sets and open sets

Definition 7.44. We say that $A \subset X$ is a **closed (sub)set** of X if $\overline{A} = A$.

Exercise 7.45. Show that the above definition of closed subsets agrees with Def. 3.26 when X is a metric space.

Exercise 7.46. Show that a finite subset of a Hausdorff space is closed. Give an example of non-closed finite subset of a non-Hausdorff topological space.

Remark 7.47. The closure $\overline{A}$ is the smallest closed set containing A. (Proof: By Prop. 7.36, $\overline{A}$ is closed. If B is closed and contains A, then $\overline{A} \subset \overline{B} = B$.)

Theorem 7.48. *Let A be a subset of X. Then A is closed iff $X \backslash A$ is open.*

Proof. Let $B = X \backslash A$. Then A is closed iff every closure point of A is in A, iff every non-interior point of B is not in B, iff every point in B is an interior point of B. By Prop. 7.43, this is equivalent to that B is open. $\qquad\square$

Corollary 7.49. $\varnothing$ *and X are closed subsets of X. An intersection of closed subsets is closed. A finite union of closed subsets is closed.*

Proof. Take the complement of Def. 7.9, and apply Thm. 7.48. (Of course, they can also be proved directly using the condition $A = \overline{A}$ for closedness.) $\qquad\square$

Corollary 7.50. X *is Hausdorff iff for every distinct $x, y \in X$ there exists $U \in \mathrm{Nbh}_X(x)$ such that $y \notin \overline{U}$.*

Proof. "$\Leftarrow$": Let $x \neq y$. Choose $U \in \mathrm{Nbh}(x)$ such that $y \notin \overline{U}$. Then $X \backslash \overline{U} \in \mathrm{Nbh}(y)$ by Thm. 7.48. So x and y are separated by neighborhoods $U, X \backslash \overline{U}$.

"$\Rightarrow$": Let $x \neq y$. Choose disjoint $U \in \mathrm{Nbh}(x)$ and $V \in \mathrm{Nbh}(y)$. Then $X \backslash V$ is closed by Thm. 7.48. So $\overline{U} \subset X \backslash V$ by Rem. 7.47. So $y \notin \overline{U}$. $\qquad\square$

Corollary 7.51. *Let Y be a subset of X, and let $A \subset Y$. Then the following are equivalent.*

(1) A is a closed subset of Y.

(2) $A = B \cap Y$ for some closed subset B of X.

Note that the "open subset" version of this corollary is true due to the definition of the subspace topology of Y (cf. Def. 7.19).

First proof. A is closed in Y iff $Y \backslash A = Y \cap A^c$ is open in Y, iff $Y \cap A^c$ equals $Y \cap U$ for some open subset $U \subset X$, iff $Y \cap A$ (which is A) equals $Y \cap U^c$ for some open subset $U \subset X$. This finishes the proof, thanks to Thm. 7.48. $\qquad \square$

Second proof. Recall by (7.5) that $\overline{A} \cap Y$ is the closure of A in Y. Then A is closed in Y iff $A = \overline{A} \cap Y$. This proves (1)$\Rightarrow$(2) since $\overline{A}$ is closed by Prop. 7.36. Assume (2). Then $A = B \cap Y$ where $B = \overline{B}$. So $\overline{A} \cap Y = \overline{B \cap Y} \cap Y \subset \overline{B} \cap Y = B \cap Y = A$. This proves (1). $\qquad \square$

As an immediate consequence of Def. 7.9 and Cor. 7.51, we have:

Exercise 7.52. Let $A \subset B \subset X$.

1. Prove that if B is open in X, then A is open in B iff A is open in X.

2. Prove that if B is closed in X, then A is closed in B iff A is closed in X.

Remark 7.53. Many people define a closed set to be the complement of an open set, and then proves that a set A is closed iff $A = \overline{A}$. I went the other way because I believe that $A = \overline{A}$ is more essential for understanding of closedness from the viewpoint of analysis. In Thm. 3.48, we have already seen a classical example of closed set in analysis: $C([0,1], \mathbb{R})$ is a closed subset of $l^\infty([0,1], \mathbb{R})$, which has the clear analytic meaning that the uniform limit of a sequence/net of continuous functions $[0,1] \to \mathbb{R}$ is continuous. And we will see many more examples of this type in the future.

Remark 7.54. I defined closedness using $A = \overline{A}$, and hence using the limits of nets. This is because the intuition of closed sets is very closely related to the intuition of limits of nets/sequences. On the other hand, the intuition of open sets is very different. Let me say a few words about this.

Without a doubt, the keyword I give for the intuition of limits of nets is "approximation": Limit is not only a dynamic process, but also gives an impression of "getting smaller and smaller". When dealing with closed sets, we often do the same thing! We take an intersection of possibly infinitely many closed subsets, and the result we get is still a closed set (cf. Cor. 2.4).

The keyword I give for open sets is "local", or more precisely, "local-to-global" (as opposed to "getting smaller and smaller"!). This is not only because a union of

open sets is open, but also because open sets are very often used to prove a global result by reducing to local problems. One easy example is Exe. 7.119, which says that in order to prove that a function is continuous on the whole space X, it suffices to prove this locally. (We have already used this strategy in Sec. 2.4.) Here is a more advanced example: to define the integral for a function on a large set, one can first define it locally (i.e. on small enough open subsets), and then patch these local values together. We will see many examples in the future, for example, in the following chapter about compactness.

Remark 7.55. Very often, a theorem is an important result establishing two seemingly different (systems of) intuitions, and hence two different ways of mathematical thinking. This is why I call "closed sets are the complements of open sets" a theorem. The term "complement" implies that this theorem often manifests itself in the following way: If solving a problem using open sets is a direct proof, then solving the problem using limits of sequences/nets is a proof by contradiction/contrapositive. And vise versa.

7.4 Continuous maps and homeomorphisms

Unless otherwise stated, X and Y are topological spaces.

7.4.1 Continuous maps

Definition 7.56. Let $f : X \to Y$ be a map. Let $x \in X$. We say that f is **continuous at** x if the following equivalent conditions hold:

(1) For every net $(x_\alpha)_{\alpha \in I}$ in X converging to x, we have $\lim_{\alpha \in I} f(x_\alpha) = f(x)$.

(2) For every $V \in \mathrm{Nbh}_Y(f(x))$, there exists $U \in \mathrm{Nbh}_X(x)$ such that for every $p \in U$ we have $f(p) \in V$.

(2') For every $V \in \mathrm{Nbh}_Y(f(x))$, the point x is an interior point of $f^{-1}(V)$.

We say that f is a **continuous** function/map, if f is continuous at every point of X.

It is clear that "for every $V \in \mathrm{Nbh}_Y(f(x))$" in (2) and (2') can be replaced by "for every $V \in \mathcal{B}$ containing $f(x)$" if $\mathcal{B}$ is a basis for the topology of Y.

Note that in the case that (x_α) or $(f(x_\alpha))$ has more than one limits (which could happen when X or Y is not Hausdorff), condition (1) means that $f(x)$ is one of the limits of $(f(x_\alpha))_{\alpha \in I}$ if x is one of the limits of (x_α).

Proof of equivalence. Clearly (2) is equivalent to (2'). The proof of (2)$\Rightarrow$(1) is similar to the case of sequences in metric spaces. (See Def. 2.38.) We leave the details to the reader.

$\neg(2) \Rightarrow \neg(1)$: Assume that (2) is not true. Then there is a neighborhood V of $f(x)$ such that for every neighborhood U of x there exists $x_U \in U$ such that $f(x_U) \notin V$. Then $(x_U)_{U \in \mathrm{Nbh}_X(x)}$ is a net in X, and $\lim_U x_U = x$ since $x_U \in U$. However, for each U we have $f(x_U) \in Y \backslash V$. So $\lim_U f(x_U)$ cannot converge to $f(x)$. $\square$

Exercise 7.57. Show that when X, Y are metric spaces, Def. 7.56 agrees with Def. 2.38.

Exercise 7.58. Let $f : X \to Y$ and $g : Y \to Z$ be maps of topological spaces. Assume that f is continuous at $x \in X$, and g is continuous at $f(x)$. Prove that $g \circ f : X \to Z$ is continuous at x.

The proof of (1)$\Rightarrow$(2) in Def. 7.56 is indirect, since it uses the axiom of choice. (The merit of this proof is that it is parallel to the proof for metric spaces in Sec. 2.4.) One can also give a direct proof. Indeed, there is a particular net (x_α) converging to x such that $\lim f(x_\alpha) = f(x)$ iff (2) is true:

Exercise 7.59. Define $(\mathrm{PNbh}_X(x), \leqslant)$, the **directed set of pointed neighborhoods** of x, to be

$$\mathrm{PNbh}_X(x) = \big\{(p, U) : U \in \mathrm{Nbh}_X(x), p \in U\big\}$$
$$(p, U) \leqslant (p', U') \qquad \Longleftrightarrow \qquad U \supset U' \tag{7.7}$$

For each $\alpha = (p, U) \in \mathrm{PNbh}_X(x)$, let $x_\alpha = p$. Then $(x_\alpha)_{\alpha \in \mathrm{PNbh}_X(x)}$ is a net in X converging to x. Prove that f is continuous at x iff $\lim_\alpha f(x_\alpha) = f(x)$.

Proposition 7.60. *Let $f : X \to Y$ be a map. The following are equivalent:*

(1) f is continuous.

(2) If $V \subset Y$ is open in Y, then $f^{-1}(V)$ is open in X.

(3) If $F \subset Y$ is closed in Y, then $f^{-1}(F)$ is closed in Y.

Proof. (1)$\Leftrightarrow$(2): By Def. 7.56-(2') and Prop. 7.43. (2)$\Leftrightarrow$(3): By Thm. 7.48 and the fact that $f^{-1}(B^c) = f^{-1}(B)^c$ for every $B \subset Y$. $\square$

Remark 7.61. We first defined the continuity of f at a point, and then used this to define a continuous function f to be one continuous at every point. However, it seems that the notion of continuity at a point is used only in analysis. In geometry and in topology, only continuous maps (but not a map continuous at a point) are used, and they are defined by Prop. 7.60-(2).

One might think that continuous functions are special cases of functions which are continuous at given points. But in fact, the latter notion can also be derived from the former: $\square$

Exercise 7.62. Let $f : (X, \mathcal{T}) \to (Y, \mathcal{T}')$ be a map of topological spaces. Let $x \in X$. Define a new topological space $(X_x, \mathcal{T}_x)$ as follows. X_x equals X as a set. The topology $\mathcal{T}_x$ of X_x is generated by the basis

$$\mathcal{B}_x = \mathrm{Nbh}_X(x) \cup \big\{ \{p\} : p \neq x \big\} \tag{7.8}$$

Prove that if X is Hausdorff, then X_x is Hausdorff. Prove that the following are equivalent:

(1) $f : X \to Y$ is continuous at x.

(2) $f : X_x \to Y$ is continuous.

Continuous functions are determined by their values on a dense subset:

Proposition 7.63. *Let A be a subset of X, and let $f, g : \overline{A} \to Y$ be continuous. Then*

$$f(\overline{A}) \subset \overline{f(A)} \tag{7.9a}$$

If Y is moreover Hausdorff, then

$$f = g \qquad \Longleftrightarrow \qquad f|_A = g|_A \tag{7.9b}$$

Proof. If $y \in f(\overline{A})$, then $y = f(x)$ for some $x \in \overline{A}$. Choose a net (x_α) in A converging to x. Then $\lim_\alpha f(x_\alpha) = f(x)$, and hence $f(x) \in \overline{f(A)}$.

Assume that Y is Hausdorff. If $f = g$ then clearly $f|_A = g|_A$. Assume that $f|_A = g|_A$. For each $x \in \overline{A}$, choose a net (x_α) in A converging to x. Then $f(x) = \lim f(x_\alpha) = \lim g(x_\alpha) = g(x)$. So $f = g$. $\qquad\square$

You are encouraged to prove Prop. 7.63 using open sets instead of using nets.

7.4.2 Homeomorphisms

Definition 7.64. A map $f : X \to Y$ is called **open** (resp. **closed**) if for every open (resp. closed) subset $A \subset X$, the image $f(A)$ is open (resp. closed) in Y.

Definition 7.65. A bijection $f : X \to Y$ is called a **homeomorphism** if the following clearly equivalent conditions hold:

(1) f and f^{-1} are continuous.

(2) For every net (x_α) in X and each $x \in X$, we have that $\lim_\alpha x_\alpha = x$ iff $\lim_\alpha f(x_\alpha) = f(x)$.

(3) f is continuous and open.

(4) f is continuous and closed.

If a homeomorphism $f : X \to Y$ exists, we say that X, Y are **homeomorphic**.

Recall from Def. 2.60 that when X, Y are metric spaces, the sequential version of (2) holds.

Remark 7.66. Let $\mathcal{T}_1, \mathcal{T}_2$ be two topologies on a set X. Clearly, we have $\mathcal{T}_1 = \mathcal{T}_2$ iff

$$\varphi : (X, \mathcal{T}_1) \to (X, \mathcal{T}_2) \qquad x \mapsto x \tag{7.10}$$

is a homeomorphism. Thus, the following are equivalent:

(1) $\mathcal{T}_1 = \mathcal{T}_2$.

(2) For every net (x_α) in X and $x \in X$, we have that $\lim_\alpha x_\alpha = x$ under $\mathcal{T}_1$ iff $\lim_\alpha x_\alpha = x$ under $\mathcal{T}_2$.

This equivalence implies that

topologies are determined by net convergence

Therefore, instead of using open sets or bases of topologies to describe a topology, one can also describe a topology $\mathcal{T}$ on a set X in the following way:

$$\begin{array}{c} \mathcal{T} \text{ is the unique topology on } X \text{ such that} \\ \text{a net } (x_\alpha) \text{ in } X \text{ converges to } x \in X \text{ iff ...} \end{array} \tag{7.11}$$

Similarly, by Def. 2.60, metrizable topologies are determined by sequential convergence. Therefore, metrizable topologies can be described in the following way:

$$\begin{array}{c} \mathcal{T} \text{ is the unique metrizable topology on } X \text{ such that} \\ \text{a sequence } (x_n) \text{ in } X \text{ converges to } x \in X \text{ iff ...} \end{array} \tag{7.12}$$

7.5 Examples of topological spaces described by net convergence

Example 7.67. Let A be a subset of a topological space $(X, \mathcal{T}_X)$. Then the subspace topology $\mathcal{T}_A$ of A is the unique topology such that a net (x_α) in A converges to $x \in A$ under $\mathcal{T}_A$ iff it converges to x under $\mathcal{T}_X$.

⋆ **Example 7.68.** Let $X = \bigsqcup_{\alpha \in \mathscr{A}} X_\alpha$ be a disjoint union where each $(X_\alpha, \mathcal{T}_\alpha)$ is a topology space. Then

$$\mathcal{B} = \bigcup_{\alpha \in \mathscr{A}} \mathcal{T}_\alpha$$

is clearly a basis generating a topology $\mathcal{T}$ on X, called **disjoint union topology**. $\mathcal{T}$ is the unique topology on X such that for every net $(x_\mu)_{\mu \in I}$ in X and any $x \in X$, the following are equivalent:

(1) $(x_\mu)_{\mu \in I}$ converges to x under $\mathcal{T}$.

(2) There exists $\nu \in I$ such that for every $\mu \geqslant \nu$, the element x_μ belongs to the unique X_α containing x. Moreover, $\lim_{\mu \in I_{\geqslant \nu}} x_\mu = x$ in X_α.

We call $(X, \mathcal{T})$ the **disjoint union topological space** of $(X_\alpha)_{\alpha \in \mathscr{A}}$.

The following exercise says that "disjoint union of topological spaces" is synonymous with "disjoint union of open subsets".

$\star$ **Exercise 7.69.** Assume that $X = \bigsqcup_{\alpha \in \mathscr{A}} X_\alpha$. Assume that X has a topology $\mathcal{T}$, and equip each X_α with the subspace topology. Show that $(X, \mathcal{T})$ is the disjoint union topological space of $(X_\alpha)_{\alpha \in \mathscr{A}}$ iff each X_α is an open subset of $(X, \mathcal{T})$.

Thus, for example, $\bigcup_{n \in \mathbb{N}} [2n, 2n+1)$ (under the Euclidean topology) is the disjoint union topological space of the family $\big([2n, 2n+1)\big)_{n \in \mathbb{N}}$.

$\star$ **Exercise 7.70.** In Exp. 7.68, assume that each $(X_\alpha, \mathcal{T}_\alpha)$ is metrizable. Prove that $(X, \mathcal{T})$ is metrizable. More precisely: Choose a metric d_α inducing $\mathcal{T}_\alpha$, and assume that $d_\alpha \leqslant 1$ (cf. Prop. 2.74). Define a metric d on X as in Pb. 2.4. Solve *the net version* of part 2 of Pb. 2.4. Conclude from this that d induces the topology $\mathcal{T}$. (Warning: we cannot conclude this from the original sequential version of Pb. 2.4-2.)

Definition 7.71. Let $(X_\alpha)_{\alpha \in \mathcal{A}}$ be a family of topological spaces. Elements of the product space

$$S = \prod_{\alpha \in \mathscr{A}} X_\alpha$$

are denoted by $x = (x(\alpha))_{\alpha \in \mathscr{A}}$. One checks easily that

$$\mathcal{B} = \Big\{ \prod_{\alpha \in \mathscr{A}} U_\alpha : \text{each } U_\alpha \text{ is open in } X_\alpha, \tag{7.13}$$
$$U_\alpha = X_\alpha \text{ for all but finitely many } \alpha \Big\}$$

is a basis for a topology $\mathcal{T}$, called the **product topology** or **pointwise convergence topology** of S. We call $(S, \mathcal{T})$ the **product topological space**.

Equivalently, let

$$\pi_\alpha : S \to X_\alpha \qquad x \mapsto x(\alpha) \tag{7.14}$$

be the **projection map onto the X_α component**. Then

$$\mathcal{B} = \Big\{ \bigcap_{\alpha \in E} \pi_\alpha^{-1}(U_\alpha) : E \in \text{fin}(2^{\mathscr{A}}), \ U_\alpha \text{ is open in } X_\alpha \text{ for each } \alpha \in E \Big\} \tag{7.15}$$

Unless otherwise stated, a product of topological spaces is equipped with the product topology. $\qquad\square$

Example 7.72. Let $S = X_1 \times \cdots \times X_N$ be a finite product of topological spaces. Then the product topology has a basis

$$\mathcal{B} = \{U_1 \times \cdots \times U_N : \text{ each } U_i \text{ is open in } X_i\} \tag{7.16}$$

Theorem 7.73. *Let $S = \prod_{\alpha \in \mathscr{A}} X_\alpha$ be a product of topological spaces, equipped with the product topology. Then each projection map π_α is continuous. Moreover, for every net $(x_\mu)_{\mu \in I}$ in S and every $x \in S$, the following are equivalent.*

(1) $\lim_{\mu \in I} x_\mu = x$ *in S.*

(2) For every $\alpha \in \mathscr{A}$, we have $\lim_{\mu \in I} x_\mu(\alpha) = x(\alpha)$ in X_α.

If (x_μ) satisfies (1) or (2), we say that x_μ **converges pointwise** to x if we view (x_μ) as a net of functions with domain $\mathscr{A}$.

Proof. We leave the proof to the readers. Note that the continuity of π_α follows easily from the basis-for-topology version of Prop. 7.60-(2). And from the continuity of π_α one easily deduce (1)$\Rightarrow$(2). $\square$

Remark 7.74. In the spirit of (7.11), one says that:

- The product topology $\mathcal{T}$ on $S = \prod_{\alpha \in \mathscr{A}} X_\alpha$ is the unique topology such that a net (x_μ) converges to x under $\mathcal{T}$ iff x_μ converges pointwise to x as a net of functions on $\mathscr{A}$.

Corollary 7.75. *If each X_α is a Hausdorff space, then $S = \prod_{\alpha \in \mathscr{A}} X_\alpha$ is Hausdorff.*

Proof. Either prove this directly using the basis for the topology, or prove that any net cannot converge to two different values using Thm. 7.73. $\square$

Corollary 7.76. *Let $X_1, X_2, \ldots$ be a possibly finite sequence of metric spaces. Then $S = \prod_i X_i$ is metrizable. More precisely, for each i, choose a metric d_i on X_i topologically equivalent to the original one such that $d_i \leqslant 1$ (cf. Prop. 2.4). Then the metric d on S defined by*

$$d(f, g) = \sup_i \frac{d_i(f(i), g(i))}{i} \tag{7.17}$$

induce the product topology.

We note that the product topology is also induced by

$$\delta(f, g) = \sum_i 2^{-i} d_i(f(i), g(i)) \tag{7.18}$$

Proof. The same method for solving Pb. 2.3 also applies to its net version: One shows that a net (f_α) in S converges to f under d (or under δ) iff (f_α) converges pointwise to f. Thus, by Thm. 7.73, d and δ induce the product topology. $\square$

The next example discusses the topologies induced by uniform convergence metrics. (Recall Def. 3.58.) In this example, Y is usually a normed vector space.

Example 7.77. Let X be a set, and let (Y, d_Y) be a metric space. Then there is a unique topology $\mathcal{T}$ on Y^X such that for every net $(f_\alpha)_{\alpha \in I}$ in Y^X and every $f \in Y^X$, the following are equivalent:

(1) The net (f_α) converges to f under $\mathcal{T}$.

(2) We have $\lim\limits_{\alpha \in I} \sup\limits_{x \in X} d_Y(f_\alpha(x), f(x)) = 0$.

(If (f_α) satisfies (2), we say that f_α **converges uniformly** to f.) For example, one checks easily that $\mathcal{T}$ is induced by any **uniform convergence metric**, i.e., any metric on Y^X equivalent to d where

$$d(f, g) = \min\left\{1, \sup_{x \in X} d_Y(f(x), g(x))\right\} \tag{7.19}$$

So $\mathcal{T}$ is metrizable. We call $\mathcal{T}$ the **uniform convergence topology** on Y^X.

Theorem 7.78. *Let Y be a complete metric space. Let X be a set. Then Y^X, equipped with the metric (7.19), is complete.*

Proof. It can be proved in a similar way as Thm. 3.45. We leave the details to the readers. $\square$

Theorem 7.79. *Let V be a normed vector space over $\mathbb{R}$ or $\mathbb{C}$. Let X be a topological space. Equip V^X with the uniform convergence topology. Then $C(X, V)$ is a closed subset of V^X.*

Proof. This is similar to the proof of Thm. 3.48. Let (f_α) be a net $C(X, V)$ converging uniformly to $f : X \to V$. Choose any $x \in X$ and $\varepsilon > 0$. Then there is $\alpha \in I$ such that $\sup_{p \in X} \|f(x) - f_\alpha(x)\| < \varepsilon$. Since f_α is continuous, there is $U \in \mathrm{Nbh}_X(x)$ such that for each $p \in U$ we have $\|f_\alpha(x) - f_\alpha(p)\| < \varepsilon$. Thus, for each $p \in U$ we have

$$\|f(x) - f(p)\| \leqslant \|f(x) - f_\alpha(x)\| + \|f_\alpha(x) - f_\alpha(p)\| + \|f_\alpha(p) - f(p)\| < 3\varepsilon$$

So f is continuous. $\square$

Remark 7.80. Note that the uniform convergence topology depends on the equivalence class (not just the topological equivalence class) of d_Y. Thus, one needs metrics when talking about uniform convergence. On the other hand, the study of pointwise convergence does not require metrics.

7.6 Limits of functions

By Prop. 7.63, if $A \subset X$, and if $f : \overline{A} \to Y$ is continuous, then the value of f is uniquely determined by $f|_A$ provided that Y is Hausdorff (cf. Prop. 7.63). We now consider the opposite question of extension of continuous functions: Suppose that $f : A \to Y$ is continuous. Can we extend f to a continuous function $f : \overline{A} \to Y$? (We know that such extension must be unique if it exists.) The classical concept of the limits of functions can be understood in this light.

Definition 7.81. Let A be a subset of X. Let $f : A \to Y$ be a map. Let $x \in \overline{A} \backslash A$. Let $y \in Y$. We say that the **limit of the function** f at x is y and write

$$\lim_{\substack{p \in A \\ p \to x}} f(p) \equiv \lim_{p \to x} f(p) = y$$

if the following equivalent conditions hold:

(1) If we extend f to a function $A \cup \{x\} \to Y$ satisfying $f(x) = y$, then $f : A \cup \{x\} \to Y$ is continuous at x.

(2) For every $V \in \mathrm{Nbh}_Y(y)$, there exists $U \in \mathrm{Nbh}_X(x)$ such that for every $p \in U \cap A$, we have $f(p) \in V$.

(3) For every net $(x_\alpha)_{\alpha \in I}$ in A converging to x, we have $\lim_{\alpha \in I} f(x_\alpha) = y$.

When X, Y are metric spaces, the above three conditions and the following two are equivalent:

(2m) For every $\varepsilon > 0$, there exists $\delta > 0$ such that for every $p \in A$, if $d(p, x) < \delta$ then $d(f(p), y) < \varepsilon$.

(3m) For every sequence $(x_n)_{n \in \mathbb{Z}_+}$ in A converging to x, we have $\lim_{n \to \infty} f(x_n) = y$.

Recall that by the definition of subspace topology, we have

$$\mathrm{Nbh}_A(x) = \{U \cap A : A \in \mathrm{Nbh}_X(x)\} \tag{7.20}$$

Proof of equivalence. Extend f to $\widetilde{f} : A \cup \{x\} \to Y$ by setting $\widetilde{f}(x) = y$. Then by Def. 7.56-(2), condition (1) of Def. 7.81 means that for every $V \in \mathrm{Nbh}_Y(y)$ there is a neighborhood of x in $A \cup \{x\}$ (which, by (7.20), must be of the form $U \cap (A \cup \{x\})$ where $U \in \mathrm{Nbh}_X(x)$) such that for every $p \in U \cap (A \cup \{x\})$ we have $\widetilde{f}(p) \in V$. This is clearly equivalent to (2), since $\widetilde{f}(x) = y \in V$. The equivalence (2)⇔(3) can be proved in a similar way as the equivalence of (1) and (2) in Def. 7.56. We leave the details to the readers. When X, Y are metric spaces, (2) is clearly equivalent to (2m). The equivalence (2m)⇔(3m) can be proved in a similar way as the equivalence of (1) and (2) in Def. 2.38. □

The following remarks show that the limit of a function at a point is the limit of a single net, rather than the limit of many nets (as in Def. 7.81-(3)).

Remark 7.82. Assume the setting of Def. 7.81. In the same spirit of Exe. 7.59, define a directed set $(\mathrm{PNbh}_A(x), \leqslant)$ where

$$\mathrm{PNbh}_A(x) = \{(p, U) : U \in \mathrm{Nbh}_X(x), p \in U \cap A\}$$
$$(p, U) \leqslant (p', U') \qquad \Longleftrightarrow \qquad U \supset U' \tag{7.21}$$

(That it is a directed set is due to $x \in \overline{A}$.) We have seen this directed set in Rem. 7.32. Then $(p)_{(p,U) \in \mathrm{PNbh}_A(x)}$ is a net converging to x, and

$$\lim_{p \to x} f(p) = \lim_{(p,U) \in \mathrm{PNbh}_A(x)} f(p) \tag{7.22}$$

where the convergence of the LHS is equivalent to that of the RHS.

Remark 7.83. In the setting of Def. 7.81, assume moreover that X is a metric space, then $\lim_{p \to x} f(p)$ can be described by the limit of a simpler net. Define a directed set $(A_x, \leqslant)$ where

$$A_x = A \text{ as sets}$$
$$p \leqslant p' \qquad \Longleftrightarrow \qquad d(p', x) \geqslant d(p, x) \tag{7.23}$$

Then $(p)_{p \in A_x}$ is a net in A converging to x, and

$$\lim_{p \to x} f(p) = \lim_{p \in A_x} f(p) \tag{7.24}$$

where the convergence of the LHS is equivalent to that of the RHS.

Remark 7.84. Thanks to the above two remarks, limits of functions enjoy all the properties that limits of nets enjoy. For example, they satisfy Squeeze theorem; if $f : A \to \mathbb{F}$ and $g : A \to V$ (where V is a normed vector space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$), if $x \in \overline{A} \backslash A$, and if $\lambda = \lim_{p \to x} f(p)$ and $v = \lim_{p \to x} g(p)$ exist, then $\lim_{p \to x} f(p)g(p)$ converges to λv.

Of course, you can also conclude them by using Def. 7.81-(3) instead of Rem. 7.82. In practice, it makes no difference whether you view $\lim_{p \to x} f(p)$ as the limit of $f(x_\alpha)$ for an arbitrary net x_α in A converging to x, or whether you view $\lim_{p \to x} f(p)$ as the limit of the particular net in Rem. 7.82 or Rem. 7.83. The explicit constructions of nets in these two remarks are not important for proving results about limits of functions. □

Remark 7.85. Let $f : A \to Y$, and let $x \in \overline{A} \backslash A$. Suppose that Y is Hausdorff. Assume that there exist two nets $(x_\alpha)_{\alpha \in I}$ and $(y_\beta)_{\beta \in J}$ in A converging to x such that $(f(x_\alpha))$ and $(f(y_\beta))$ converge to two different values. Then by Def. 7.81-(2), the limit $\lim_{p \to x} f(p)$ does not exist.

The above remark gives a useful criterion for the non-convergence of limits of functions. The following proposition, on the other hand, gives a method of computing limits of functions by decomposing the domain into (non-necessarily mutually disjoint) subsets.

Proposition 7.86. *Assume the setting of Def. 7.81. Assume that $A = A_1 \cup \cdots \cup A_N$. Then the following are equivalent.*

(1) We have $\lim\limits_{p \to x} f(p) = y$.

(2) For every $1 \leqslant i \leqslant N$ such that $x \in \overline{A_i}$, we have $\lim\limits_{p \to x} f|_{A_i}(p) = y$

Proof. (1)⟹(2): Assume (1). Extend $f : A \to Y$ to a function $\widetilde{f} : A \cup \{x\} \to Y$ by setting $\widetilde{f}(x) = y$. Then $\widetilde{f}$ is continuous by (1). Thus, if $x \in \overline{A_i}$, then $\widetilde{f}|_{A_i \cup \{x\}}$ is continuous. This proves (2).

(2)⟹(1): Assume (2). Choose any $V \in \mathrm{Nbh}_Y(y)$. By (2), for each i, either $x \in \overline{A_i}$ so that there is $U_i \in \mathrm{Nbh}_X(x)$ satisfying $U_i \cap A_i \subset f^{-1}(V)$ (recall (7.20)), or that $x \notin \overline{A_i}$ so that there is $U_i \in \mathrm{Nbh}_X(x)$ disjoint from A_i. In either case, we have $U_i \cap A_i \subset f^{-1}(V)$. Let $U = U_1 \cap \cdots \cap U_N$. Then

$$U \cap A = U \cap (A_1 \cup \cdots \cup A_N) = \bigcup_i U \cap A_i \subset \bigcup_i U_i \cap A_i$$

which is therefore a subset of $f^{-1}(V)$.

Another proof of ¬(1)⟹ ¬(2): Assume (1) is not true. Then there is a net $(x_\alpha)_{\alpha \in I}$ in A converging to x such that $f(x_\alpha)$ does not converge to y. So there exists $V \in \mathrm{Nbh}_Y(y)$ such that $f(x_\alpha)$ is not eventually in V, i.e., $f(x_\alpha)$ is frequently in V^c. Then $(f(x_\alpha))$ has a subnet $(f(x_\beta))_{\beta \in J}$ which is always in V^c. For example, take

$$J = \{\beta \in I : f(x_\beta) \in V^c\}$$

Since $(x_\beta)_{\beta \in J}$ is always in A, by the logic (5.6c), there is $1 \leqslant i \leqslant N$ such that (x_β) is frequently in A_i. Thus, by the same argument as above, (x_β) has a subnet $(x_\gamma)_{\gamma \in K}$ which is always in A_i. Since $x_\alpha \to x$, we have $x_\gamma \to x$, and hence $x \in \overline{A_i}$. But $f(x_\gamma) \in V^c$. So we have found a net (x_γ) in A_i converging to x such that $(f(x_\gamma))$ does not converge to y. This disproves (2). □

Remark 7.87. In many textbooks, $\lim_{p \to x} f(x)$ is also defined more generally when x is an accumulation point of A, i.e., when $x \in \overline{A \backslash \{x\}}$. In this case, the limit of f at x simply means

$$\lim_{p \to x} f(p) \overset{\text{def}}{=\!=} \lim_{p \to x} f|_{A \backslash \{x\}}(p) \tag{7.25}$$

This more general case is important in classical analysis, but is less useful in abstract analysis. (As a matter of fact, accumulation points are less convenient than closure points.) In order not to deviate too far from the traditional analysis textbooks, let's take a look at some examples.

Definition 7.88. Let $A \subset \mathbb{R}$ and $x \in \mathbb{R}$. Let $f : A \to Y$ be a function. If x is a closure point of $A \cap \mathbb{R}_{<x}$ resp. $A \cap \mathbb{R}_{>x}$, we define the **left limit** resp. **right limit** to be

$$\lim_{t \to x^-} f(t) = \lim_{t \to x} f|_{A \cap \mathbb{R}_{<x}}(t) \tag{7.26a}$$

$$\lim_{t \to x^+} f(t) = \lim_{t \to x} f|_{A \cap \mathbb{R}_{>x}}(t) \tag{7.26b}$$

If $x \in \overline{\mathbb{R}}$ and x is a closure point of $A \backslash \{x\}$, then $\lim_{t \to x} f(t)$ is understood by (7.25).

Remark 7.89. In Def. 7.88, if $x \in \mathbb{R}$ is a closure point of $A \backslash \{x\}$, then by Prop. 7.86,

$$\lim_{p \to x} f(p) = y \qquad \Longleftrightarrow \qquad \lim_{p \to x^-} f(p) = \lim_{p \to x^+} f(p) = y \tag{7.27}$$

In particular, the existence of the limit on the LHS is equivalent to the existence and the equality of the two limits on the RHS.

Example 7.90. Let $g, h : \mathbb{R} \to \mathbb{R}$ be continuous functions. Let $c \in \mathbb{R}$. Define $f : \mathbb{R} \to \mathbb{R}$ to be

$$f(x) = \begin{cases} g(x) & \text{if } x < 0 \\ c & \text{if } x = 0 \\ h(x) & \text{if } x > 0 \end{cases}$$

Since $g|_{(-\infty,0]}$ is continuous, by Def. 7.81-(1) we have that $\lim_{x \to 0^-} f(x) = \lim_{x \to 0, t<0} g(t) = g(0)$. Similarly, $\lim_{x \to 0^+} f(x) = h(0)$. Therefore, by Rem. 7.89, $\lim_{x \to 0} f(x)$ exists iff $g(0) = h(0)$, and it converges to $g(0)$ if $g(0) = h(0)$. The value c is irrelevant to the limits.

Example 7.91. Let $f : X = \mathbb{R}^2 \backslash \{(0,0)\} \to \mathbb{R}$ be $f(x,y) = \frac{x}{x+y}$. Then $(1/n, 0)$ and $(0, 1/n)$ are sequences in X converging to 0. But $f(1/n, 0) = 1$ and $f(0, 1/n) = 0$. So $\lim_{(x,y) \to (0,0)} f(x,y)$ does not exist by Rem. 7.85.

7.7 Connected spaces

Let X be a topological space. In this section, we shall define a notion of connected space. Based on our usual geometric intuition, one might attempt to define a connected space as one satisfying that any two points can be linked by a path. Such spaces are actually called **path-connected spaces** and is stronger than connected spaces. In fact, connected spaces arise from the study of intermediate value problem.

7.7.1 Connected ⇔ IVP

Definition 7.92. We say that the topological space X is **connected** if X can not be written as the disjoint of two nonempty open sets. Namely, if $X = U \sqcup V$ where U, V are open subsets of X, then either $U = \varnothing$ (and hence $V = X$) or $V = \varnothing$ (and hence $U = X$).

Equivalently (by Thm. 7.48), X is connected iff every $U \subset X$ which is both closed and open must be $\varnothing$ or X.

Definition 7.93. We say that X satisfies the **intermediate value property** (abbreviated to **IVP**) if for every continuous function $f : X \to \mathbb{R}$ and every $x, y \in X$ we have

$$f(x) < f(y) \qquad \Longrightarrow \qquad [f(x), f(y)] \subset f(X) \tag{7.28}$$

Theorem 7.94. *X is connected iff X satisfies IVP. Moreover, if X is not connected, then there is a continuous $f : X \to \mathbb{R}$ such that $f(X) = \{0, 1\}$.*

Proof. First, assume that X does not satisfy IVP. Choose a continuous $f : X \to \mathbb{R}$ with real numbers $a < b < c$ such that $a, c \in f(X)$ but $b \notin f(X)$. So $U = f^{-1}(-\infty, b)$ and $V = f^{-1}(b, +\infty)$ are disjoint non-empty open subsets of X, and $X = U \sqcup V$. They are open, because f is continuous (cf. Prop. 7.60). So X is not connected.

Next, assume that X is not connected. Then $X = U \sqcup V$ where U, V are open subsets of X. Define $f : X \to \mathbb{R}$ to be constantly 0 on U and constantly 1 on V. It is easy to check that f is continuous. (See also Rem. 7.120.) That $f(X) = \{0, 1\}$ means that X does not satisfy IVP. $\qquad\square$

We now give a couple of elementary examples.

7.7.2 Connected subsets of $\overline{\mathbb{R}}$ are precisely intervals

Proposition 7.95. *Let A be a dense subset of X. Assume that A is connected. Then X is connected.*

Proof. If $X = \overline{A}$ is not connected, then by Thm. 7.94, there exists a continuous surjection $f : \overline{A} \to \{0, 1\}$. By Prop. 7.63, $\overline{f(A)}$ contains $f(\overline{A})$. So $f(A)$ has closure $\{0, 1\}$. So $f(A) = \{0, 1\}$. A does not satisfy IVP, and hence is not connected. $\qquad\square$

Theorem 7.96. *Let A be a nonempty subset of $\overline{\mathbb{R}}$. Then A is connected iff A is an interval.*

Proof. Step 1. Suppose that A is connected. Let $a = \inf A$ and $b = \sup B$. To show that A is one of $(a, b), (a, b], [a, b), [a, b]$, it suffices to show that every $c \in (a, b)$ belongs to A. Suppose that some $c \in (a, b)$ does not belong to A. Then A is the disjoint union of two nonempty open subsets $A \cap [-\infty, c)$ and $A \cap (c, +\infty]$, impossible.

Step 2. Every single point is clearly connected. Since every interval containing at least two points is homeomorphic to one of $[0, 1], (0, 1], [0, 1), (0, 1)$, it suffices to prove that these four intervals are connected. Since $(0, 1)$ is dense in the other three intervals, by Prop. 7.95, it suffices to prove that $(0, 1)$ is connected.

Suppose that $(0, 1)$ is not connected. Then $(0, 1) = U \sqcup V$ where U, V are disjoint open nonempty subsets. Choose $x_1 \in U$ and $y_1 \in V$, and assume WLOG that $x_1 < y_1$. In the following, we construct an increasing sequence (x_n) in U and a decreasing one (y_n) in V satisfying $x_n < y_n$ for all n by induction. Suppose x_n, y_n has been constructed. Let $z_n = (x_n + y_n)/2$.

- If $z_n \in U$, then let $x_{n+1} = z_n$ and $y_{n+1} = y_n$.

- If $z_n \in V$, then let $x_{n+1} = x_n$ and $y_{n+1} = z_n$.

Then $y_n - x_n$ converges to 0. So x_n and y_n converge to the same point $\xi \in \mathbb{R}$. We have $\xi \in (0, 1)$ since $x_1 < \xi < y_1$. Since V is open, $U = (0, 1) \backslash V$ is closed in $(0, 1)$ by Thm. 7.48. So $\xi \in \mathrm{Cl}_{(0,1)}(U) = U$. Similarly, $\xi \in \mathrm{Cl}_{(0,1)}(V) = V$. This is impossible. $\qquad \square$

7.7.3 More examples of connected spaces

Definition 7.97. Let $x, y \in X$. A **path** in X from x to y is defined to be a continuous map $\gamma : [a, b] \to X$ where $-\infty < a < b < +\infty$, such that $\gamma(a) = x$ and $\gamma(b) = y$. Unless otherwise stated, we take $[a, b]$ to be $[0, 1]$. We call x and y respectively the **initial point** and the **terminal point** of γ.

Definition 7.98. We say that X is **path-connected** if for every $x, y \in X$ there is a path in X from x to y.

Example 7.99. $\mathbb{R}^N$ is path-connected. $B_{\mathbb{R}^N}(0, R)$ and $\overline{B}_{\mathbb{R}^N}(0, R)$ (where $R < +\infty$) are path connected. $\{x \in \mathbb{R}^N : r < x < R\}$ (where $0 \leqslant r < R < +\infty$) are path connected. The region enclosed by a triangle is a connected subset of $\mathbb{R}^2$. $[0, 1]^N$ is connected.

Theorem 7.100. *Assume that X is path-connected. Then X is connected.*

Proof. If X is not connected, then $X = U \sqcup V$ where U, V are nonempty open subsets of X. Since X is path-connected, there is a path γ from a point of U to a point of V. So $[0, 1] = \gamma^{-1}(U) \sqcup \gamma^{-1}(V)$ where $\gamma^{-1}(U), \gamma^{-1}(V)$ are open (by Prop. 7.60) and nonempty. This contradicts the fact that $[0, 1]$ is connected (cf. Thm. 7.96). $\qquad \square$

Proposition 7.101. *Let $f : X \to Y$ be a continuous map of topological spaces. Suppose that X is connected. Then $f(X)$ is connected.*

Proof. By replacing $f : X \to Y$ by the restricted continuous map $f : X \to f(X)$, it suffices to assume $Y = f(X)$. If Y is not connected, then $Y = U \sqcup V$ where U, V are open and nonempty. Then $X = f^{-1}(U) \sqcup f^{-1}(V)$ are open (by Prop. 7.60) and nonempty subsets of X. So X is not connected, impossible. (One can also use Thm. 7.94 to prove that $f(X)$ is connected.) □

Remark 7.102. When $Y = \mathbb{R}$, Prop. 7.101 and Thm. 7.96 imply that $f(X)$ is an interval. So f satisfies (7.28). Therefore, Prop. 7.101 can be viewed as a generalization of IVP for connected spaces.

Corollary 7.103. *Let I be an interval, and let $f : I \to \overline{\mathbb{R}}$ be a strictly increasing continuous map. Then $J = f(I)$ is an interval, and the restriction $f : I \to J$ is a homeomorphism.*

Proof. By Thm. 7.96 and Prop. 7.101, J is connected and hence is an interval. Therefore, by Thm. 2.72, f is a homeomorphism. □

⋆ **Example 7.104.** Not all connected spaces are path-connected. Let $f : (0, 1] \to \mathbb{R}^2$ be defined by $f(x) = (x, \sin(x^{-1}))$. Then the range $f((0, 1])$ is connected by Prop. 7.101. Since $f((0, 1])$ is a dense subset of $X = f((0, 1]) \cup \{(0, 0)\}$, by Prop. 7.95, X is connected. However, it can be checked that X is not path-connected. (Prove it yourself, or see [Mun, Sec. 24].) X is called the **topologist's sine curve**.

The following proposition can be used to decompose (for example) an open subset of $\mathbb{R}^N$ into open connected subsets. (See Pb. 7.13.)

Proposition 7.105. *Assume that $X = \bigcup_{\alpha \in \mathscr{A}} X_\alpha$ where each X_α is connected. Assume that $\bigcap_{\alpha \in \mathscr{A}} X_\alpha \neq \varnothing$. Then X is connected.*

Proof. Suppose that X is not connected. By Thm. 7.94, there is a continuous surjection $f : X \to \{0, 1\}$. Let $p \in \bigcap_\alpha X_\alpha$. Then $f(p)$ is 0 or 1. Assume WLOG that $f(p) = 0$. Choose $x \in X$ such that $f(x) = 1$. Choose α such that $x \in X_\alpha$. Then $f|_{X_\alpha} : X_\alpha \to \{0, 1\}$ is a continuous surjection. So X_α does not satisfy IVP, and hence is not connected. □

Exercise 7.106. Prove a path-connected version of Prop. 7.105.

Exercise 7.107. Prove Prop. 7.95 and 7.105 directly using Def. 7.92 (but not using IVP).

7.8 Rigorous constructions of $\sqrt[n]{x}$, $\log x$, and a^x

With the help of Cor. 7.103, one can construct a lot of well-known functions rigorously.

Example 7.108. Let $f : \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$ be $f(x) = x^n$ where $n \in \mathbb{Z}_+$. Then by Cor. 7.103, $J = f(\mathbb{R}_{\geqslant 0})$ is an interval. Clearly $J \subset [0, +\infty)$. Since $0 \in J$ and $\sup J = +\infty$, we have $J = [0, +\infty)$. Therefore f is a homeomorphism. Its inverse function is a homeomorphism: the n-**th root** function

$$\sqrt[n]{} : \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0} \qquad x \mapsto \sqrt[n]{x}$$

This gives the rigorous construction of $\sqrt[n]{x}$.

A similar method gives the rigorous construction of $\log$.

Example 7.109. By Exp. 4.28, the exponential function $\exp : \mathbb{R} \to \mathbb{R}$ is continuous. We claim that $\exp$ is a strictly increasing homeomorphism from $\mathbb{R}$ to $\mathbb{R}_{>0}$. Its inverse function is called the **logarithmic function**

$$\log : \mathbb{R}_{>0} \to \mathbb{R} \qquad x \mapsto \log x$$

Proof. From $e^x = \sum_{n=0}^{\infty} x^n/n!$ we clearly have $e^0 = 1$ and $e^x > 1$ if $x > 0$. From $e^{x+y} = e^x e^y$ proved in Cor. 5.61, we have $e^x e^{-x} = e^0 = 1$, which shows that $e^x \in \mathbb{R}_{>0}$ for all $x \in \mathbb{R}$. If $x < y$, then $e^y > e^x > 0$ since $e^y = e^{y-x} e^x$ and $e^{y-x} > 1$. So $\exp$ is strictly increasing. Thus, by Cor. 7.103, $\exp$ is a homeomorphism from $\mathbb{R}$ to $J = \exp(\mathbb{R})$, and J is an interval.

When $x \geqslant 0$, we have $e^x \geqslant x$ from the definition of e^x. So $\sup_{x \geqslant 0} e^x = +\infty$. When $x \leqslant 0$, since $e^x e^{-x} = 1$, we have $\inf_{x \leqslant 0} e^x = 1/\sup_{x \geqslant 0} e^x = 0$. So $\sup J = +\infty$ and $\inf J = 0$. Since $0 \notin \exp(\mathbb{R})$ (if $e^x = 0$, then $1 = e^x e^{-x} = 0$, impossible), we have $J = \mathbb{R}_{>0}$. $\qquad\square$

Example 7.110. Let $a \in \mathbb{R}_{>0}$. For each $z \in \mathbb{C}$, define

$$a^z = e^{z \log a} \tag{7.29}$$

By Exp. 7.109, if $a > 1$ (resp. $0 < a < 1$), the map

$$\mathbb{R} \to \mathbb{R}_{>0} \qquad x \mapsto a^x \tag{7.30}$$

is an increasing (resp. decreasing) homeomorphism, since it is the composition of the increasing (resp. decreasing) homeomorphism $x \in \mathbb{R} \mapsto x \log a \in \mathbb{R}$ and the increasing one $\exp : \mathbb{R} \to \mathbb{R}_{>0}$. By the proof of Exp. 7.109, we have

$$a^0 = 1 \qquad a^x a^{-x} = 1 \qquad a^x a^y = a^{x+y} \tag{7.31}$$

And clearly

$$a^1 = a. \tag{7.32}$$

It follows that for every $n \in \mathbb{Z}_+$, $a^n = a^{1+\cdots+1} = a \cdots a$. Namely, $a^n = e^{n \log a}$ agrees with the usual understanding of a^n. Thus, since $(a^{1/n})^n$ equals $a^{1/n} \cdots a^{1/n} = a^{1/n+\cdots+1/n} = a^1 = a$, we conclude

$$a^{\frac{1}{n}} = \sqrt[n]{a}$$

Example 7.111. By Exp. 7.109, if $p > 0$ (resp. $p < 0$), then

$$\mathbb{R}_{>0} \to \mathbb{R}_{>0} \qquad x \mapsto x^p = e^{p \log x} \tag{7.33}$$

is an increasing (resp. decreasing) homeomorphism, since it is the composition of the increasing (resp. decreasing) homeomorphism $x \in \mathbb{R}_{>0} \to p \log x \in \mathbb{R}$ and the increasing homeomorphism $\exp : \mathbb{R} \to \mathbb{R}_{>0}$.

7.9 Problems and supplementary material

Let X and Y be topological spaces.

7.9.1 Open sets, closed sets, closures

Problem 7.1. Let $A, B \in X$. Let $(A_\alpha)_{\alpha \in \mathscr{A}}$ be a family of subsets of X. Prove that

$$\overline{A \cup B} = \overline{A} \cup \overline{B} \tag{7.34a}$$

$$\overline{\bigcap_{\alpha \in \mathscr{A}} A_\alpha} \subset \bigcap_{\alpha \in \mathscr{A}} \overline{A_\alpha} \tag{7.34b}$$

Problem 7.2. Let $(x_\alpha)_{\alpha \in I}$ be a net in X. Let $x \in X$. Prove that the following statements are equivalent:

(1) $(x_\alpha)_{\alpha \in I}$ has a subnet converging to x.

(2) For every neighborhood U of x, we have that x_α is frequently in U.

(3) x belongs to $\bigcap_{\alpha \in I} \overline{\{x_\beta : \beta \geqslant \alpha\}}$.

Any $x \in X$ satisfying one of these three conditions is called a **cluster point** of $(x_\alpha)_{\alpha \in I}$. (Compare Pb. 3.1.)

Hint. (2)⇔(3) is a direct translation. Assume (2). To prove (1), show that $(J, \leqslant)$ is a directed set, where

$$J = \{(\alpha, U) \in I \times \mathrm{Nbh}_X(x) : x_\alpha \in U\}$$

$$(\alpha, U) \leqslant (\alpha', U') \qquad \Longleftrightarrow \qquad \alpha \leqslant \alpha' \text{ and } U \supset U' \tag{7.35a}$$

Prove that $(x_\mu)_{\mu \in J}$ is a subnet of $(x_\alpha)_{\alpha \in I}$ if for each $(\alpha, U) \in J$ we set

$$x_{(\alpha, U)} = x_\alpha \tag{7.35b}$$

(Namely, the increasing map $J \to I$ is defined to be $(\alpha, U) \mapsto \alpha$.) Prove that $(x_\mu)_{\mu \in J}$ converges to x. You should point out where (2) is used in your proofs. □

Remark 7.112. By Pb. 7.2-(3), the set of cluster points of (x_α) is a closed subset, since intersections of closed subsets are closed (cf. Cor. 7.49, or by (7.34b)).

For the reader's convenience, we present below the sequential version of Pb. 7.2. "(1)⟺(2)" is due to Pb. 3.1. "(2)⟺(3)" is due to Pb. 7.2.

Proposition 7.113. *Assume that X is a metric space. Let $(x_n)_{n\in\mathbb{Z}_+}$ be a sequence in X. Let $x \in X$. The following statements are equivalent:*

(1) $(x_n)_{n\in\mathbb{Z}_+}$ has a subsequence converging to x.

(2) For every neighborhood U of x, we have that x_n is frequently in U.

(3) x belongs to $\bigcap_{n\in\mathbb{Z}_+} \overline{\{x_k : k \geqslant n\}}$.

*Any $x \in X$ satisfying one of these three conditions is called a **cluster point** of $(x_n)_{n\in\mathbb{Z}_+}$.*

Prop. 7.113-(3) should remind you of the definitions of $\limsup$ and $\liminf$.

⋆ **Remark 7.114.** It is not hard to show that *the (1,2,3) of Prop. 7.113 are equivalent in the more general case that X is a first countable topological space* (see below for the definition). The proof is similar to that for metric spaces, and is left to the readers as an exercise.

Definition 7.115. Let X be a topological space. A subset $\mathcal{B}_x$ of $\mathrm{Nbh}_X(x)$ is called a **neighborhood basis** of x, if for every $U \in \mathrm{Nbh}_X(x)$ there exists $V \in \mathcal{B}_x$ such that $V \subset U$. We say that X **first countable** if every point x has a neighborhood basis $\mathcal{B}_x$ which is a countable set.

Example 7.116. If X is a metric space, then X is first countable, since for every $x \in X$, $\{B_X(x, 1/n) : n \in \mathbb{Z}_+\}$ is a neighborhood basis of x.

Remark 7.117. By Pb. 7.2-(2) and Prop. 7.113-(2), if (x_n) is a sequence in a metric space X, then (x_n) has a subsequence converging to $x \in X$ iff (x_n) has a subnet converging to x. This is not necessarily true when X is a general topological space. Note that in the general case, cluster points of a sequence (x_n) mean cluster points of (x_n) as a net. Thus, they are not the limits of convergent *subsequences* of (x_n).

Problem 7.3. Assume that X is a metric space. Let $E \subset X$. Recall that $d(x, E) = d(E, x) = \inf_{e\in E} d(e, x)$. Prove that

$$\{x \in X : d(x, E) = 0\} = \overline{E} \tag{7.36}$$

Remark 7.118. If E, F are disjoint subsets of a metric space X, a continuous function $f : X \to [0, 1]$ is called an **Urysohn function** with respect to E, F, if

$$f^{-1}(1) = E \qquad f^{-1}(0) = F$$

For example, it is easy to check that

$$f : X \to [0,1] \qquad f(x) = \frac{d(x,F)}{d(x,E) + d(x,F)} \tag{7.37}$$

is an Urysohn function.

⋆ **Problem 7.4.** A topological space X is called **normal** if for every closed disjoint $E, F \subset X$, there exist disjoint open subsets $U, V \subset X$ such that $E \subset U$ and $F \subset V$.

1. Prove that X is normal iff for each $E \subset W \subset X$ where E is closed and W is open, there exists an open subset $U \subset X$ such that $E \subset U \subset \overline{U} \subset W$.

2. Prove that if X is metrizable, then X is normal.

7.9.2 Continuous maps

Exercise 7.119. Let $f : X \to Y$ be a map.

1. Suppose that F is a subset of Y containing $f(X)$. Show that $f : X \to Y$ is continuous iff $f : X \to F$ is continuous.

2. (Local to global principle) Suppose that $X = \bigcup_{\alpha \in I} U_\alpha$ where each U_α is an open subset of X, Prove that f is continuous iff $f|_{U_\alpha} : U_\alpha \to Y$ is continuous for every α.

Remark 7.120. The above local-to-global principle for continuous functions can be rephrased in the following way. Suppose that $X = \bigcup_{\alpha \in I} U_\alpha$ where each U_α is an open subset of X. Suppose that for each α we have a continuous map $f_\alpha : U_\alpha \to Y$. Assume that for each $\alpha, \beta \in I$ we have

$$f_\alpha|_{U_\alpha \cap U_\beta} = f_\beta|_{U_\alpha \cap U_\beta}$$

Then there is a (necessarily unique) continuous function $f : X \to Y$ such that $f|_{U_\alpha} = f_\alpha$ for every α.

Problem 7.5. Let $f : X \to Y$ be a map. Suppose that $X = A_1 \cup \cdots \cup A_N$ where $N \in \mathbb{Z}_+$ and $A_1, \dots, A_N$ are closed subsets of X. Suppose that $f|_{A_i} : A_i \to Y$ is continuous for each $1 \leqslant i \leqslant N$. Prove that f is continuous.

Does the conclusion remain true if $A_1, \dots, A_N$ are not assumed closed? If no, find a counterexample. ☐

Note. Do not use Prop. 7.86 in your proof. But you can think about how this problem is related to Prop. 7.86. ☐

Definition 7.121. Let $(I, \leqslant)$ be a directed set. Let ∞_I (often abbreviated to ∞) be a new symbol not in I. Then

$$I^* = I \cup \{\infty_I\}$$

is also a directed set if we extend the preorder $\leqslant$ of I to I^* by setting

$$\alpha \leqslant \infty_I \qquad (\forall \alpha \in I^*)$$

For each $\alpha \in I$, let

$$I^*_{\geqslant \alpha} = \{\beta \in I^* : \beta \geqslant \alpha\}$$

The **standard topology** on I^* is defined to be the one induced by the basis

$$\mathcal{B} = \{\{\alpha\} : \alpha \in I\} \cup \{I^*_{\geqslant \alpha} : \alpha \in I\} \tag{7.38}$$

Problem 7.6. Let $(I, \leqslant)$ be a directed set. Let I^* be as in Def. 7.121.

1. Check that $\mathcal{B}$ (defined by (7.38)) is a basis for a topology. (Therefore, $\mathcal{B}$ generates a topology $\mathcal{T}$ on I^*.)

2. Let $(x_\alpha)_{\alpha \in I}$ be a net in a topological space X. Let $x_\infty \in X$. (So we have a function $x : I^* \to X$.) Prove that

$$x \text{ is a continuous function} \qquad \Longleftrightarrow \qquad \lim_{\alpha \in I} x_\alpha = x_\infty \tag{7.39}$$

3. Is I^* Hausdorff? Prove it, or find a counterexample.

7.9.3 Product spaces

Problem 7.7. Prove Thm. 7.73.

Problem 7.8. Let $(X_\alpha)_{\alpha \in \mathscr{A}}$ and $(Y_\alpha)_{\alpha \in \mathscr{A}}$ be families of nonempty topological spaces. Let Z be a nonempty topological space. For each $\alpha \in \mathscr{A}$, choose maps $f_\alpha : X_\alpha \to Y_\alpha$ and $g_\alpha : Z \to X_\alpha$.

1. Use Thm. 7.73 to prove that

$$\prod_{\alpha \in \mathscr{A}} f_\alpha : \prod_\alpha X_\alpha \to \prod_\alpha Y_\alpha \qquad (x(\alpha))_{\alpha \in \mathscr{A}} \mapsto (f_\alpha(x(\alpha)))_{\alpha \in \mathscr{A}} \tag{7.40}$$

   is continuous iff each f_α is continuous.

2. Use Thm. 7.73 to prove that

$$\bigvee_{\alpha \in \mathscr{A}} g_\alpha : Z \to \prod_\alpha X_\alpha \qquad z \mapsto (g_\alpha(z))_{\alpha \in \mathscr{A}} \tag{7.41}$$

   is continuous iff each g_α is continuous.

Problem 7.9. Let $(X_\alpha)_{\alpha \in \mathscr{A}}$ be an uncountable family of metric spaces, where each X_α has at least two elements. Let $S = \prod_{\alpha \in \mathscr{A}} X_\alpha$ be the product space, equipped with the product topology. Prove that S is not first countable (recall Def. 7.115), and hence not metrizable.

7.9.4 Limits of functions

Problem 7.10. Prove the equivalence of (2) and (3) in Def. 7.81.

Problem 7.11. Find $\lim\limits_{(x,y)\to(0,0)} f(x,y)$, or explain why it does not exist:

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

7.9.5 Connectedness

Problem 7.12. Assume that X, Y are not empty. Prove that X and Y are connected iff $X \times Y$ is connected.

Hint. Write $X \times Y$ as a union of sets of the form $(X \times \{y\}) \cup (\{x\} \times Y)$. $\qquad\square$

Definition 7.122. A topological space X is called **locally connected** if every $x \in X$ has a neighborhood basis $\mathcal{B}_x$ (recall Def. 7.115) whose members are all connected.

Example 7.123. Every open subset of a locally connected space is clearly locally connected.

Example 7.124. $\mathbb{R}^N$ is locally connected, since open balls are path-connected and hence connected (by Thm. 7.100). Therefore, every open subset of $\mathbb{R}^N$ is locally connected.

Problem 7.13. Suppose that X is locally connected. Prove that X has a unique (disjoint) decomposition $X = \bigsqcup_{\alpha \in \mathscr{A}} X_\alpha$ where each X_α is a nonempty connected open subset of X. (Each X_α is called a **connected component** of X.)

Hint. For each $x \in X$, consider the union of all connected neighborhoods containing x. $\qquad\square$

8 Compactness

8.1 Overture: two flavors of compactness

Let X be a topological space.

Definition 8.1. An **open cover** of X means a family $\mathfrak{U} = (U_\alpha)_{\alpha \in \mathscr{A}}$ of open subsets of X such that $X = \bigcup_{\alpha \in \mathscr{A}} U_\alpha$. $\mathfrak{U}$ is called **finite** resp. **countable** if $\mathscr{A}$ is finite resp. countable. A **subcover** of $\mathfrak{U}$ means an open cover $\mathfrak{V} = (V_\beta)_{\beta \in \mathscr{B}}$ of X such that each V_β equals U_α for some $\alpha \in \mathscr{A}$.

Definition 8.2. We say that X is a **compact space** if every open cover of X has a finite subcover. We say that X is a **Lindelöf space** if every open cover of X has a countable subcover.

Remark 8.3. For the purpose of this chapter, it suffices to consider open covers $\mathfrak{U} = (U_\alpha)_{\alpha \in \mathscr{A}}$ such that $\alpha \in \mathscr{A} \mapsto U_\alpha \in 2^X$ is injective. (Namely, one can throw away repeated open sets.) In this case, we write

$$\mathfrak{U} \subset 2^X$$

and view $\mathfrak{U}$ as a subset of 2^X. A **subcover** of such $\mathfrak{U}$ is an open cover $\mathfrak{V}$ such that $\mathfrak{V} \subset \mathfrak{U}$. The readers can check that this assumption does not affect the definition of compact spaces and Lindelöf spaces.

Remark 8.4. Compactness can also be formulated in a relevant version: If A is a compact subset of X, and if $\mathfrak{U}$ is a set of open subsets of X such that $A \subset \bigcup_{U \in \mathfrak{U}} U$, then since $\{U \cap A : U \in \mathfrak{U}\}$ gives an open cover of A, we know that

$$A \subset U_1 \cup \cdots \cup U_n$$

some $U_1, \ldots, U_n \in \mathfrak{U}$. Conversely, any A satisfying such property is a compact subspace of X.

Exercise 8.5. Show that a finite union of compact spaces is compact. Show that a finite set is compact.

Exercise 8.6. Show that X is compact iff X satisfies the **finite intersection property**: The intersection of a family of non-empty closed subsets is nonempty.

A main goal of this chapter is to prove the following theorem. Its proof will be finished at the end of Sec. 8.5. In fact, the formal proof will only be in Sec. 8.3 and Sec. 8.5.

Theorem 8.7. *Let X be a metric space. Then X is sequentially compact iff X is compact.*

This is the first fundamental theorem we prove in this course. Its significance lies in the fact that it connects two seemingly very different notions of compactness, and hence two strikingly different intuitions. We hope that the reader can not only follow the logical chains of the proofs, but also understand the pictures behind the proof. More precisely, we hope that the readers can have an intuitive understanding of the following questions:

- Why are these two compactness both powerful in solving certain problems? What roles do they play in the proof? How are the roles played by these two compactness related? (This is more important than just knowing why these notions are *logically* equivalent.)

- Why is one version of compactness more powerful than the other one in solving certain problems?

Thus, I feel that it is better to look at some applications of compactness before we prove Thm. 8.7 rigorously. For pedagogical purposes, we also introduce two related notions of compactness:

Definition 8.8. We say that X is **net-compact**, if every net $(x_\alpha)_{\alpha \in I}$ in X has at least one cluster point (recall Pb. 7.2), equivalently, at least one convergent subnet. We say that X is **countably compact**, if every countable open cover of X has a finite subcover.

Sequential compactness and net-compactness clearly share the same intuition. Countable compactness is intuitively similar to compactness. Also, compactness clearly implies countable compactness. But net-compactness does not imply sequential compactness in general: In a net-compact space, every sequence has a convergent subnet, but not necessarily a convergent subsequence.

The relationship between these four versions of compactness is as follows. (We will prove this in the course of proving Thm. 8.7.)

Metric spaces: all the four versions of compactness are equivalent   (8.1a)

Topological spaces: net-compact $\Longleftrightarrow$ compact   (8.1b)

After proving (8.1), we will not use the notions of countable compactness and net-compactness. This is because net-compactness is equivalent to compactness, and countable compactness is more difficult to use than compactness. (Nevertheless, proving/using compactness by proving/using the existence of cluster points is often helpful.)

8.2   Act 1: case studies

8.2.1   Extreme value theorem (EVT)

Lemma 8.9 (**Extreme value theorem**)**.** *Let X be a compact topological space. Let $f : X \to \mathbb{R}$ be a continuous function. Then f attains its maximum and minimum at points*

of X. In particular, $f(X)$ is a bounded subset of $\mathbb{R}$.

We have seen the sequential compactness version of **EVT** (extreme value theorem) in Lem. 3.2. There, we find the point $x \in X$ at which f attains its maximum by first finding a sequence (x_n) such that $f(x_n)$ converges to $\sup f(X)$. Then we choose any convergent subsequence (x_{n_k}), which converges to the desired point x. The same method can be used to prove EVT for net-compact spaces if we replace sequences by nets.

For compact spaces, EVT is proved in a completely different way. In fact, without the tools of sequences and nets, one can not easily find x at which f attains it maximum. The argument is rather indirect:

Proof of Lem. 8.9. It suffices to prove that $f(X)$ is bounded for all continuous maps $f : X \to \mathbb{R}$. Then $a = \sup f(X)$ is in $\mathbb{R}$. If $a \notin f(X)$, we choose a homeomorphism $\varphi : (-\infty, a) \to \mathbb{R}$. So $\varphi \circ f : X \to \mathbb{R}$ is continuous but has no upper bound. This is impossible. $\qquad\square$

Thus, to prove EVT, it remains to prove:

Example 8.10. Assume that X is compact and $f : X \to \mathbb{R}$ is continuous. Then $f(X)$ is a bounded subset of $\mathbb{R}$.

Proof. For each $x \in X$, since f is continuous at x, there exists $U_x \in \mathrm{Nbh}(x)$ such that $|f(p) - f(x)| < 1$ for all $p \in U_x$. In particular, $f(U_x)$ is bounded. Since $X = \bigcup_{x \in X} U_x$ is an open cover of X, by compactness, $X = U_{x_1} \cup \cdots \cup U_{x_n}$ for some $x_1, \ldots, x_n \in X$. Thus $X = f(U_{x_1}) \cup \cdots \cup f(U_{x_n})$ is bounded. $\qquad\square$

The above proof is typical. It suggests that compactness is powerful for proving finiteness properties rather than finding solutions of functions satisfying certain requirements. Thus, if you want to prove a finiteness property using sequential or net-compactness, you have to prove it indirectly. For example, you need to prove by contradiction:

Example 8.11. Assume that X is net-compact or sequentially compact. Assume that $f : X \to \mathbb{R}$ is continuous. Then $f(X)$ is a bounded subset of $\mathbb{R}$.

Proof. Assume that X is net-compact. If $f(X)$ is not bounded above, then there is a sequence (x_α) in X (viewed as a net) such that $\lim_\alpha f(x_\alpha) = +\infty$. By net-compactness, (x_α) has a subnet (x'_β) converging to $x \in X$. So $f(x) = \lim_\beta f(x'_\beta) = +\infty$, impossible. So f is bounded above, and hence bounded below by a similar argument. The case where X is sequentially compact can be proved by a similar method. $\qquad\square$

Let us look at a more complicated example.

8.2.2 Uniform convergence in multivariable functions

Example 8.12. Let X, Y be topological spaces. Assume that Y is compact. Let V be a normed vector space. Choose $f \in C(X \times Y, V)$. For each $x \in X$, let

$$f_x : Y \to V \qquad y \mapsto f(x, y)$$

Equip $C(Y, V)$ with the l^∞-norm. Then the following map is continuous:

$$\Phi(f) : X \to C(Y, V) \qquad x \mapsto f_x \tag{8.2}$$

Remark 8.13. Note that since Y is compact, each $g \in C(Y, V)$ is bounded by EVT (applied to $|g|$). So $C(Y, V) \subset l^\infty(Y, V)$.

The continuity of $\Phi(f)$ means that for each $x \in X$, the following statement holds:

✐ For every $\varepsilon > 0$ there exists $U \in \mathrm{Nbh}_X(x)$ such that for all $p \in U$ and all $y \in Y$ we have $\|f(p, y) - f(x, y)\| < \varepsilon$.

This is clearly a finiteness property. Thus, its sequentially compact version or net-compact version should be proved indirectly. Indeed, when X, Y are metric spaces and Y is sequentially compact, we have proved this in Pb. 3.7 by contradiction. The same method (with sequences replaced by nets) also works for the net-compact case:

Example 8.14. Example 8.12 is true, assuming that Y is net-compact rather than compact.

Proof. Suppose "✐" is not true. Then there is $\varepsilon > 0$ such that for every $U \in \mathrm{Nbh}_X(x)$ there is $x_U \in U$ and $y_U \in Y$ such that $\|f(x_U, y_U) - f(x, y_U)\| \geqslant \varepsilon$. Then $(x_\alpha)_{\alpha \in \mathrm{Nbh}_X(x)}$ is a net converging to x, where $x_\alpha = x_U$ if $\alpha = U$. Since Y is net-compact, (y_α) has a subnet (y_β) converging to some $y \in Y$. Since the subnet (x_β) also converges to x, we have $\lim_\beta f(x_\beta, y_\beta) = f(x, y)$ and $\lim_\beta f(x, y_\beta) = f(x, y)$ by the continuity of f. This contradicts the fact that $\|f(x_\beta, y_\beta) - f(x, y_\beta)\| \geqslant \varepsilon$ for all β. $\qquad \square$

On the other hand, the solution of Exp. 8.12 using open covers is a direct proof:

Proof of Exp. 8.12. "✐" is a finiteness property global over Y. We prove "✐" by first proving it locally, and then passing to the global space Y using the compactness of Y.

Fix $x \in X$. Choose any $\varepsilon > 0$. For each $y \in Y$, since f is continuous at (x, y), there is a neighborhood W of (x, y) such that for every $(p, q) \in W$ we have $\|f(p, q) - f(x, y)\| < \varepsilon/2$. By Exp. 7.72, we can shrink W to a smaller neighborhood of the form $U_y \times V_y$ where $U_y \in \mathrm{Nbh}_X(x)$ and $V_y \in \mathrm{Nbh}_Y(y)$. Then for each $p \in U_y$

and $q \in U_y$ we have $\|f(p,q) - f(x,y)\| < \varepsilon/2$, and hence $\|f(p,q) - f(x,q)\| < \varepsilon$. This proves the special case of "⊘" where Y is replaced by U_y.

Now we pass from local to global in the same way as in the proof of Exp. 8.10. Since $Y = \bigcup_{y \in Y} V_y$ is an open cover of Y, by the compactness of Y, there is a finite subset $F \subset Y$ such that $Y = \bigcup_{y \in F} V_y$. Then "⊘" is true if we let $U = \bigcap_{y \in F} U_y$. □

8.2.3 Conclusions

1. Sequential compactness and net-compactness are useful for finding solutions of a function satisfying some given conditions.

2. Compactness is useful for proving finiteness properties. The proof is usually a local-to-global argument. It is usually a direct argument (rather than proof by contradiction).

3. If one uses sequential/net-compactness to prove a finiteness property, one usually proves it by contradiction: Assume that this finiteness is not true. Find a sequence/net (x_α) that violates this finiteness property, and pass to a convergent subsequence/subnet to find a contradiction.

4. Therefore, for sequential/net-compact spaces, the argument is in the direction of "getting smaller and smaller", opposite to the argument for compact spaces.

Let me emphasize that the proof for sequentially/net-compact spaces is opposite to the one for compact spaces in two aspects: (1) If one argument is direct, the other is a proof by contradiction for the same problem. (2) The former has the intuition of "getting smaller", while the latter local-to-global argument has the intuition of "getting larger".

I have already touched on this phenomenon in Rem. 7.54: *The reason that sequences and nets run in the opposite direction to that of open sets is because closed sets are opposite to open sets*, as proved in Thm. 7.48. Thus, you can expect that the transition between closed and open sets plays a crucial role in the following proof of Thm. 8.7.

8.3 Act 2: "sequentially compact ⇔ countably compact" for metric spaces, just as "net-compact ⇔ compact"

> The road up and the road down is one and the same.
>
> Heraclitus

As mentioned before, our goal of this chapter is to prove "sequentially com-

pact ⇔ compact" for metric spaces. Our strategy is as follows: We reformulate the sequential compactness condition in terms of decreasing chains of closed sets, and reformulate the compactness condition in terms of increasing chains of open sets. Then we relate these two pictures easily using Thm. 7.48.

The following difficulty arises when carrying out this strategy. Sequences are countable by nature, whereas open covers can have arbitrarily large cardinality. Thus, sequences are related to countable decreasing chains, and hence countable open covers. Therefore, the above idea only implies the equivalence

$$\text{sequentially compact} \quad \Longleftrightarrow \quad \text{countably compact} \qquad \text{(for metric spaces)}$$
$$(8.3a)$$

Accordingly, it will only imply

$$\text{net-compact} \quad \Longleftrightarrow \quad \text{compact} \qquad\qquad (8.3b)$$

since there are no constraints on the cardinalities of indexed sets of nets. We will prove (8.3) in this section, and leave the proof of "countably compact ⇔ compact" for metric spaces to Sec. 8.5.

Since the proofs of (8.3a) and (8.3b) are similar, we first discuss (8.3b).

Proposition 8.15. *Let X be a topological space. Then the following are equivalent.*

(1) X is compact.

*(2) (**Increasing chain property**) If $(U_\mu)_{\mu \in I}$ is an increasing net of open subsets of X satisfying $\bigcup_{\mu \in I} U_\mu = X$, then $U_\mu = X$ for some μ.*

*(3) (**Decreasing chain property**) If $(E_\mu)_{\mu \in I}$ is a decreasing net of nonempty closed subsets of X, then $\bigcap_{\mu \in I} E_\mu \neq \varnothing$.*

Here, "increasing net" means $U_\mu \subset U_\nu$ if $\mu \leqslant \nu$, and "decreasing net" means the opposite.

Proof. (1)⇒(2): Assume (1). Then $X = \bigcup_\mu U_\mu$ is an open cover of X. So, by the compactness of X, we have $X = U_{\mu_1} \cup \cdots \cup U_{\mu_n}$ for some $\mu_1, \ldots, \mu_n \in I$. Choose $\mu \in I$ which is $\geqslant \mu_1, \ldots, \mu_n$. Then $X = U_\mu$.

(2)⇒(1): Assume (2). Let $X = \bigcup_{\alpha \in \mathscr{A}} W_\alpha$ be an open cover of X. Let $I = \mathrm{fin}(2^{\mathscr{A}})$. For each $\mu = \{\alpha_1, \ldots, \alpha_n\} \in I$, let $U_\mu = W_{\alpha_1} \cup \cdots \cup W_{\alpha_n}$. Then $(U_\mu)_{\mu \in I}$ is an increasing net of open sets covering X. Thus, by (2), we have $U_\mu = X$ for some μ. This proves (1).

(2)⇔(3): If we let $E_\mu = X \backslash U_\mu$, then (2) says that if (E_μ) is a decreasing net of closed sets whose intersection is $\varnothing$, then $E_\mu = \varnothing$ for some μ. This is the contraposition of (3). □

We now relate decreasing chain property and cluster points of nets using (8.4).

Theorem 8.16. *Let X be a topological space. Then X is net-compact iff X is compact.*

Proof. Assume that X is net-compact. By Prop. 8.15, it suffices to prove that X satisfies the decreasing chain property. Let $(E_\mu)_{\mu \in I}$ be a decreasing net of nonempty closed subsets of X. For each μ we choose $x_\mu \in E_\mu$, which gives a net $(x_\mu)_{\mu \in I}$ in X. The fact that (E_μ) is decreasing implies that $F_\mu \subset E_\mu$ if we set

$$F_\mu = \{x_\nu : \nu \in I, \nu \geqslant \mu\} \tag{8.4a}$$

Thus, the closure $\overline{F}_\mu$ is a subset of E_μ since E_μ is closed. It suffices to prove that $\bigcap_{\mu \in I} \overline{F}_\mu \neq \varnothing$. By Pb. 7.2,

$$\bigcap_\mu \overline{F}_\mu = \{\text{the cluster points of the net } (x_\mu)_{\mu \in I}\} \tag{8.4b}$$

which is nonempty because X is net-compact. This finishes the proof that X is compact.

Now we assume that X is compact. Let $(x_\mu)_{\mu \in I}$ be a net in X. Define F_μ by (8.4a). Then $(\overline{F}_\mu)_{\mu \in I}$ is a decreasing net of nonempty closed subsets. So $\bigcap_\mu \overline{F}_\mu$, the set of cluster points of $(x_\mu)_{\mu \in I}$, is nonempty by the decreasing chain property (cf. Prop. 8.15). So X is net-compact. $\qquad\square$

Proposition 8.17. *Let X be a topological space. Then the following are equivalent.*

(1) X is countably compact.

*(2) (**Increasing chain property**) If $(U_n)_{n \in \mathbb{Z}_+}$ is an increasing sequence of open subsets of X satisfying $\bigcup_{n \in \mathbb{Z}_+} U_n = X$, then $U_n = X$ for some n.*

*(3) (**Decreasing chain property**) If $(E_n)_{n \in \mathbb{Z}_+}$ is a decreasing sequence of nonempty closed subsets of X, then $\bigcap_{n \in \mathbb{Z}_+} E_n \neq \varnothing$.*

Proof. Similar to the proof of Prop. 8.15. $\qquad\square$

Exercise 8.18. Fill in the details of the proof of Prop. 8.17.

Lemma 8.19. *Let X be a metric space. Then X is sequentially compact iff X is countably compact.*

Proof. This lemma can be proved in a similar way as Thm. 8.17. The only difference is that one should use the sequential version of Pb. 7.2, namely, Prop. 7.113. Note that the "(1)⇔(2)" of Prop. 7.113 does not hold for general topological spaces. Thus, sequential compactness is not equivalent to countable compactness for general topological spaces. $\qquad\square$

⋆ **Remark 8.20.** As pointed out in Rem. 7.114, Prop. 7.113 holds more generally for first countable topological spaces (Def. 7.115). Therefore, Lem. 8.19 also holds for such spaces:

$$\text{sequentially compact} \quad \Longleftrightarrow \quad \text{countably compact} \atop \text{(for first countable spaces)} \tag{8.5}$$

8.4 Intermezzo: elementary properties about compactness

Proposition 8.21. *Suppose that $X \subset Y$ where Y is a topological space. The following are true.*

1. *Assume that Y is compact and X is closed in Y. Then X is compact.*

2. *Assume that Y is Hausdorff and X is compact, then X is a closed subset of Y.*

Proof. Part 1. Let (x_α) be a net in X. Since Y is compact, (x_α) has a subnet (x'_β) converging to some $p \in Y$. Since (x'_β) is in X, we have $p \in \overline{X} = X$. So X is compact.

Part 2. To prove $\overline{X} = X$, we choose any net (x_α) in X converging to $p \in Y$ and show that $p \in X$. Indeed, since X is compact, (x_α) has a subnet (x'_β) converging to some $x \in X$. Since (x'_β) also converges to p, we have $p = x$ because Y is Hausdorff. So $p \in X$. $\qquad\square$

I have mentioned that non-Hausdorff spaces are not often used in analysis. Thus, we mainly use the following special case of Prop. 8.21:

Corollary 8.22. *Let Y be a Hausdorff space and $X \subset Y$. If X is compact, then X is closed in Y. If X is closed in Y and if Y is compact, then X is compact.*

Recall that a similar property holds for complete metric spaces, cf. Prop. 3.27.

Theorem 8.23. *Suppose that $f : X \to Y$ is a continuous map of topological spaces where X is compact. Then $f(X)$ is compact. Moreover, if f is injective and X, Y are Hausdorff, then f restricts to a homeomorphism $f : X \to f(X)$.*

Proof. Choose any net $(f(x_\alpha))$ in $f(X)$ where $x_\alpha \in X$. Since X is compact, (x_α) has a subnet (x'_β) converging to some $x \in X$. Then $f(x'_\beta)$ converges to $f(x)$. So $f(X)$ is compact.

Now assume that f is injective and Y is Hausdorff. Then the subspace $f(X)$ is also Hausdorff. By replacing Y by $f(X)$, we assume that f is bijective. To show that f^{-1} is continuous, by Prop. 7.60, it suffices to prove that f is a closed map, i.e., f sends every closed $E \subset X$ to a closed subset $f(E)$. Indeed, since X is compact, E is also compact by Cor. 8.22. So $f(E)$ is compact by the first paragraph. So $f(E)$ is closed in X by Cor. 8.22. $\qquad\square$

The second part of Thm. 8.23 can also be proved in a similar way as Pb. 3.3 by replacing sequences with nets. But that argument relies on the fact that every net in a compact Hausdorff space with only one cluster point is convergent. We leave the proof of this fact to the readers (cf. Pb. 8.1).

Exercise 8.24. The first part of Thm. 8.23 can be viewed as a generalization of extreme value theorem. Why?

Proposition 8.25. *Suppose that X, Y are compact topological spaces. Then $X \times Y$ is compact.*

Proof. Take a net (x_α, y_α) in $X \times Y$. Since X is compact, (x_α) has a convergent subnet (x_{α_β}). Since Y is compact, (y_{α_β}) has a convergent subnet $(y_{\alpha_{\beta_\gamma}})$. So $(x_{\alpha_{\beta_\gamma}}, y_{\alpha_{\beta_\gamma}})$ is a convergent subnet of (x_α, y_α). $\square$

Remark 8.26. Note that if $X \times Y$ is compact, then X, as the image of $X \times Y$ under the projection map, is compact by Thm. 8.23. Therefore, we conclude that $X \times Y$ is compact iff X and Y are compact.

8.5 Act 3: "countably compact $\Leftrightarrow$ compact" for Lindelöf spaces

Assume in this section that X is a topological space. Recall that X is Lindelöf iff every open cover has a countable subcover. Thus, it is obvious that

$$\text{countably compact} \quad \Longleftrightarrow \quad \text{compact} \quad \text{(for Lindelöf spaces)} \tag{8.6}$$

Thus, by Lem. 8.19, to prove that sequentially/countably compact metric spaces are compact, it suffices to prove that they are Lindelöf.

We introduce two related concepts that are more useful than Lindelöf spaces:

Definition 8.27. We say that a topological space X is **separable** if X has a countable dense subset. We say that X is **second countable** if the topology of X has a countable basis (i.e., a basis $\mathcal{B}$ with countably many elements).

Example 8.28. $\mathbb{R}^N$ is separable, since $\mathbb{Q}^N$ is a dense subset.

As we shall immediately see, these two notions are equivalent for metric spaces. It is often easier to visualize and prove separability for concrete examples (such as Exp. 8.28). Indeed, Exp. 8.28 is the typical example that helps us imagine more general separable spaces. However, for general topological spaces, second countability behaves better than separability. The following property gives one reason.

Proposition 8.29. *If Y is a subset of a second countable space X, then Y is second countable.*

Proof. Let $\mathcal{B}$ be a countable basis for the topology of X. Then $\{Y \cap U : U \in \mathcal{B}\}$ is a countable basis for the topology of Y. $\square$

Another reason that second countability is better is because it implies Lindelöf property. This is mainly due to the following fact:

Proposition 8.30. *Let $\mathcal{B}$ be a basis for the topology of X. Then the following are equivalent.*

(1) X *is Lindelöf (resp. compact).*

(2) *If X has open cover $\mathfrak{U}$ where each member of $\mathfrak{U}$ is an element of $\mathcal{B}$, then $\mathfrak{U}$ has a countable (resp. finite) subcover.*

Proof. "(1)$\Rightarrow$(2)" is obvious. Assume (2). Let $\mathfrak{W}$ be an open cover of X. Let

$$\mathfrak{U} = \{U \in \mathcal{B} : U \subset W \text{ for some } W \in \mathfrak{W}\}$$

For each $x \in X$, since there is $W \in \mathfrak{W}$ containing x, and since $\mathcal{B}$ is a basis, there is $U \in \mathcal{B}$ such that $x \in U \subset W$. This proves that $\mathfrak{U}$ is an open cover of X. So $\mathfrak{U}$ has a countable (resp. finite) subcover $\mathfrak{U}_0$. For each $U \in \mathfrak{U}_0$, choose $W_U \in \mathfrak{W}$ containing U. Then $\{W_U : U \in \mathfrak{U}_0\}$ is a countable (resp. finite) subcover of $\mathfrak{W}$. $\qquad\square$

Corollary 8.31. *Every second countable topological space X is Lindelöf.*

Proof. Let $\mathcal{B}$ be a countable basis for the topology of X. Let $\mathfrak{U}$ be an open cover of X such that each member of $\mathfrak{U}$ is in $\mathcal{B}$. Then by discarding duplicated terms, $\mathfrak{U}$ becomes countable. This verifies (2) of Prop. 8.30. $\qquad\square$

Theorem 8.32. *Consider the following statements:*

(1) X *is second countable.*

(2) X *is separable.*

Then (1)$\Rightarrow$(2). If X is metrizable, then (1)$\Leftrightarrow$(2).

Proof. Assume that X is second countable with countable basis $\mathcal{B}$. For each $U \in \mathcal{B}$, choose $x_U \in U$. Then one checks easily that $\{x_U : U \in \mathcal{B}\}$ is dense by checking that it intersects every nonempty open subset of X. This proves (1)$\Rightarrow$(2).

Assume that X is a metric space with countable dense subset E. Let us prove that the countable set

$$\mathcal{B} = \{B_X(e, 1/n) : e \in E, n \in \mathbb{Z}_+\}$$

is a basis for the topology of X. Choose any open $W \subset X$ with $x \in W$. We want to show that some member of $\mathcal{B}$ contains x and is in W. By shrinking W, we may assume that $W = B_X(x, 1/n)$ for some $n \in \mathbb{Z}_+$. Since E is dense, $B(x, 1/2n)$ contains some $e \in E$. So $d(x, e) < 1/2n$. Therefore, $B(e, 1/2n)$ contains x and is inside W by triangle inequality. $\qquad\square$

It can be proved that Lindelöf metric spaces are separable. (Therefore, the three notions agree for metric spaces.) We will not use this fact. So we leave its proof to the readers as an exercise (cf. Pb. 8.12). The following chart is a summary of the relationships between the various topological properties about countability.

$$\text{Topological spaces:} \quad \text{second countable} \implies \left\{ \begin{array}{l} \text{subset is second countable} \\ \text{separable} \\ \text{Lindelöf} \end{array} \right.$$

$$\text{Metric spaces:} \qquad \text{second countable} \iff \text{separable} \iff \text{Lindelöf}$$

<div align="right">(8.7)</div>

Example 8.33. Since $\mathbb{R}^N$ is separable and hence second countable, every subset of $\mathbb{R}^N$ is second countable, and hence is separable.

Theorem 8.34. *Let X be a sequentially compact metric space. Then X is separable, and hence second countable and Lindelöf.*

Proof. We claim that for every $\delta > 0$, there is a finite set $E \subset X$ such that for every $x \in E$, the distance $d(x, E) < \delta$. Suppose that there is no such a finite set for some given number $\delta > 0$. Pick any $x_1 \in X$. Suppose $x_1, \ldots, x_k \in X$ have been constructed. Then there is a point $x_{k+1} \in X$ whose distance to $\{x_1, \ldots, x_k\}$ is $\geqslant \delta$. This defines inductively a sequence $(x_k)_{k \in \mathbb{Z}_+}$ in X such that any two elements have distance $\geqslant \delta$. So (x_k) has no convergent subsequence, contradicting the sequential compactness of X.

Thus, for each $n \in \mathbb{Z}_+$, we can choose a finite $E_n \subset X$ satisfying $d(x, E_n) < 1/n$ for all $x \in X$. Let $E = \bigcup_{n \in \mathbb{Z}_+} E_n$, which is countable. Then for each $x \in X$, $d(x, E) \leqslant d(x, E_n) < 1/n$ for every n, which implies $d(x, E) = 0$ and hence $x \in \overline{E}$ by Pb. 7.3. So E is dense in X. $\qquad \square$

Remark 8.35. The above proof is indirect because it proves the existence of E_n by contradiction but not by explicit construction. However, if X is a bounded closed subset of $\mathbb{R}^N$, one can find an explicit countable basis for the topology of X:

$$\mathcal{B} = \{X \cap B(x, 1/n) : n \in \mathbb{Z}_+, x \in \mathbb{Q}^N\}$$

and hence has an explicit countable dense subset $\{x_U : U \in \mathcal{B}, U \neq \varnothing\}$ where for each U we choose some $x_U \in U$. More generally, if X is a closed subset of the sequentially compact space $[0, 1]^{\mathbb{Z}_+}$, one can find a countable basis for the topology of X and hence a countable dense subset of X in a similar way. (You will be asked to construct them in Pb. 8.13.)

We shall see in Thm. 8.45 that every sequentially compact metric space is homeomorphic to a closed subset of $[0, 1]^{\mathbb{Z}_+}$. Therefore, for any sequentially compact metric space X you will see in the real (mathematical) life, you don't need the indirect construction in the proof of Thm. 8.34 to prove the separability of X. So what is the point of giving an indirect proof of Thm. 8.34? Well, you need Thm. 8.34 to prove Thm. 8.45. $\qquad \square$

Proof of Thm. 8.7. Let X be a metric space. Assume that X is compact. Then X is clearly countably compact, and hence sequentially compact by Lem. 8.19. Conversely, assume that X is sequentially compact. Then by Lem. 8.19 and Thm. 8.34, X is countably compact and Lindelöf, and hence compact. □

⋆ **Remark 8.36.** Since second countable spaces are first countable, by Rem. 8.20 and Cor. 8.31, we have

$$\text{sequentially compact} \iff \text{countably compact} \iff \text{compact}$$
$$\text{(for second countable topological spaces)} \tag{8.8}$$

Relation (8.8) not only generalizes Thm. 8.7, but also tells us what are the crucial properties that ensure the equivalence of compactness and sequential compactness for metric spaces.

8.6 Problems and supplementary material

Let X, Y be topological spaces.

8.6.1 Compactness

Problem 8.1. Let $(x_\alpha)_{\alpha \in I}$ be a net in a compact Hausdorff space X. Prove that (x_α) is convergent iff (x_α) has exactly one cluster point.

Problem 8.2. Let $(x_\alpha)_{\alpha \in I}$ be a net in the compact space $\overline{\overline{\mathbb{R}}}$. Let S be the (automatically nonempty) set of cluster points of (x_α) in $\overline{\overline{\mathbb{R}}}$. Recall that S is a closed subset by Rem. 7.112. For each $\alpha \in I$, define

$$A_\alpha = \inf\{x_\beta : \beta \geqslant \alpha\} \qquad B_\alpha = \sup\{x_\beta : \beta \geqslant \alpha\} \tag{8.9}$$

Then (A_α) is increasing and (B_α) is decreasing. So they converge in $\overline{\overline{\mathbb{R}}}$. Define

$$\liminf_{\alpha \in I} x_\alpha = \sup\{A_\alpha : \alpha \in I\} = \lim_{\alpha \in I} A_\alpha \tag{8.10a}$$
$$\limsup_{\alpha \in I} x_\alpha = \inf\{B_\alpha : \alpha \in I\} = \lim_{\alpha \in I} B_\alpha \tag{8.10b}$$

Prove that

$$\liminf_{\alpha \in I} x_\alpha = \inf S \qquad \limsup_{\alpha \in I} x_\alpha = \sup S \tag{8.11}$$

Note. You will get a quick proof by choosing the right one of the three equivalent definitions of cluster points in Pb. 7.2. A wrong choice will take you much more effort. □

Corollary 8.37. *Let* (x_α) *be a net in* $\overline{\mathbb{R}}$. *Then* (x_α) *converges in* $\overline{\mathbb{R}}$ *iff* $\limsup_\alpha x_\alpha = \liminf_\alpha x_\alpha$.

Proof. $\overline{\mathbb{R}}$ is compact. Therefore, $\limsup_\alpha x_\alpha = \liminf_\alpha x_\alpha$ iff (x_α) has only one cluster point (by Pb. 8.2), iff (x_α) converges (by Pb. 8.1). $\qquad\square$

Problem 8.3. Given a general topological space X, which one implies the other between the conditions of "sequential compactness" and "countable compactness"? Prove your conclusion with details.

Hint. Check the proof of Thm. 8.16, and think about the question: If (x_n) is a sequence in X, what is the inclusion relation between $\bigcap_n \overline{\{x_k : k \geqslant n\}}$ and the set of limits of the convergent subsequences (rather than subnets) of (x_n)? $\qquad\square$

Problem 8.4. Prove Prop. 8.21 using the original definition of compact spaces (i.e. every open cover has a finite subcover) instead of using nets.

⋆ **Problem 8.5.** Prove Prop. 8.25 using the original definition of compact spaces instead of using nets.

Problem 8.6. Assume that Y is compact. Let $(x_\alpha, y_\alpha)_{\alpha \in I}$ be a net in $X \times Y$. Assume that $x \in X$ is a cluster point of (x_α). Prove that there exists $y \in Y$ such that (x, y) is a cluster point of (x_α, y_α).

Problem 8.7. (Tychonoff theorem, countable version) Let $(X_n)_{n \in \mathbb{Z}_+}$ be a sequence of compact topological spaces. Prove that the product space $S = \prod_{n \in \mathbb{Z}_+} X_n$ (equipped with the product topology) is compact using the following hint.

Hint. Let $(f_\alpha)_{\alpha \in I}$ be a net in S where $f_\alpha = (f_\alpha(1), f_\alpha(2), \dots)$. Use Pb. 8.6 to construct inductively an element $x = (x(1), x(2), \dots) \in S$ such that for every $n \in \mathbb{Z}_+$, the element $(x(1), \dots, x(n))$ is a cluster point of $(f_\alpha(1), \dots, f_\alpha(n))_{\alpha \in I}$ in $X_1 \times \cdots \times X_n$. Prove that x is a cluster point of $(f_\alpha)_{\alpha \in I}$ in S. $\qquad\square$

Remark 8.38. The same idea as above can be used to prove the general Tychonoff theorem (the version where the index set $\mathbb{Z}_+$ in Pb. 8.7 is replaced by an arbitrary set) by replacing mathematical induction by Zorn's lemma.

⋆ **Problem 8.8.** Let X be the set of two elements: $X = \{0, 1\}$, viewed as a metric subspace of $\mathbb{R}$. Let $S = X^{[0,1]}$, the product space of uncountably many X, where the index set is the interval $[0, 1]$. S is equipped with the product topology. According to Tychonoff theorem (to be proved in the future), S is compact. Prove that S is not sequentially compact.

Hint. Use binary representations in $[0, 1]$. $\qquad\square$

8.6.2 LCH spaces

Definition 8.39. A subset A of a Hausdorff space X is called **precompact** if its closure $\overline{A}$ is compact. This is equivalent to saying that A is contained in a compact subset of X.

Proof of equivalence. If $\overline{A}$ is compact, then $\overline{A}$ is a compact subset of X (cf. Cor. 8.22) containing A. Conversely, if $A \subset K$ where K is a compact subset of X, then K is closed by Cor. 8.22. Since $\overline{A}$ is the smallest closed subset of X containing A, we have $\overline{A} \subset K$. By Exe. 7.52, $\overline{A}$ is closed in K. By Cor. 8.22, $\overline{A}$ is compact. $\qquad\square$

It is clear that a subset of a precompact set is precompact.

Problem 8.9. Suppose that X is metric space. Let $A \subset X$. Prove that the following are equivalent.

(1) A is precompact, i.e., $\overline{A}$ is compact.

(2) Every sequence in A has a subsequence converging to some point of X.

Definition 8.40. A Hausdorff space X is called a **locally compact Hausdorff (LCH) space** if every point has a precompact neighborhood.

Proposition 8.41. *Let X be an LCH space. Then the closed subsets and the open subsets of X are LCH.*

Proof. Let $E \subset X$ be closed. Let $x \in E$. Then there is $U \in \mathrm{Nbh}_X(x)$ with compact closure $\mathrm{Cl}_X(U)$. So $\mathrm{Cl}_X(U) \cap E$ is compact by Cor. 8.22. Note that $U \cap E \in \mathrm{Nbh}_E(x)$, and $U \cap E$ is a subset of $\mathrm{Cl}_X(U) \cap E$. So $U \cap E$ is precompact in E. This proves that E is LCH.

Next, we let W be an open subset of X. Let $x \in W$. We want to show that there exists $U \in \mathrm{Nbh}_X(x)$ such that $\overline{U} = \mathrm{Cl}_X(U)$ is compact and $\overline{U} \subset W$. Then U is clearly precompact in W.

Since X is LCH, there is $\Omega \in \mathrm{Nbh}_X(x)$ with compact closure $\overline{\Omega} = \mathrm{Cl}_X(\Omega)$. Then every open subset U of Ω containing x has compact closure in X. Thus, it suffices to prove that there exists $U \in \mathcal{I} = \mathrm{Nbh}_\Omega(x)$ such that $\overline{U} \subset W$. Suppose that this is not true. Then for each $U \in \mathcal{I}$ there is $x_U \in \overline{U} \backslash W$. So $(x_U)_{U \in \mathcal{I}}$ is a net in $\overline{\Omega} \backslash W$, where $\overline{\Omega} \backslash W$ is compact by Cor. 8.22. So $(x_U)_{U \in \mathcal{I}}$ has a cluster point $y \in \overline{\Omega} \backslash W$. In particular, $y \neq x$.

Since X is Hausdorff, by Cor. 7.50, there is $V_0 \in \mathrm{Nbh}_X(x)$ such that $y \notin \overline{V}_0$. Let $V = V_0 \cap \Omega$. Then $V \in \mathcal{I}$ and $y \notin \overline{V}$. For every $U \in \mathcal{I}$ satisfying $U \subset V$ we have $x_U \in \overline{V}$. So $(x_U)_{U \in \mathcal{I}}$ is eventually not in the neighborhood $X \backslash \overline{V}$ of y. This is impossible. $\qquad\square$

Definition 8.42. Let X be LCH. Let Y be a metric space. Let $(f_\alpha)_{\alpha \in I}$ be a net in Y^X. Let $f \in Y^X$. We say that (f_α) **converges locally uniformly** to f if the following equivalent conditions are satisfied:

(1) For each $x \in X$, there exists $U \in \mathrm{Nbh}_X(x)$ such that $(f_\alpha|_U)$ converges uniformly to $f|_U$.

(2) For each precompact open subset $W \subset X$, the net $(f_\alpha|_W)$ converges uniformly to $f|_W$.

Problem 8.10. Prove that in Def. 8.42, conditions (1) and (2) are equivalent.

Exercise 8.43. Let X be LCH. Let $\mathcal{V}$ be a normed vector space. Let $(f_\alpha)_{\alpha \in I}$ be a net in $C(X, \mathcal{V})$ converging pointwise to $f : X \to \mathcal{V}$. Assume that the net (f_α) converges locally uniformly on X (clearly to f). Prove that f is continuous.

An example of locally uniform convergence was given in Thm. 4.27.

8.6.3 Countability in topological spaces

Problem 8.11. Let $(X, \mathcal{T})$ be a second countable LCH space. Prove that the topology $\mathcal{T}$ has a countable basis $\mathcal{B}$ whose members are all precompact.

Hint. Use Lindelöf property for open subsets of X. □

★ **Problem 8.12.** Let X be a Lindelöf metric space. Prove that X is separable.

Problem 8.13. Let $(X_n)_{n \in \mathbb{Z}_+}$ be a sequence of topological spaces. Equip $S = \prod_{n \in \mathbb{Z}_+} X_n$ with the product topology.

1. Prove that S is second countable if each X_n is second countable.

2. Prove that S is separable if each X_n is separable

Recall from Pb. 7.13 the basic facts about connected components.

Problem 8.14. Let X be a locally connected topological space. Prove that if X is second countable, then X has countably many connected components. Use this result to show that every open subset of $\mathbb{R}$ is a countable disjoint union of open intervals.

8.6.4 The problem of embedding

Definition 8.44. Let $\mathscr{F}$ be a set of functions $X \to Y$. We say that $\mathscr{F}$ **separates points** of X, if for every distinct $x_1, x_2 \in X$ there exists $f \in \mathscr{F}$ such that $f(x_1) \neq f(x_2)$.

Problem 8.15. Let X be a nonempty compact metric space.

1. Prove that there is a sequence of continuous functions $(f_n)_{n \in \mathbb{Z}_+}$ from X to $[0, 1]$ separating points of X.

2. Prove that

$$\Phi : X \to [0,1]^{\mathbb{Z}_+} \qquad x \mapsto (f_1(x), f_2(x), \dots)$$

gives a homeomorphism $\Phi : X \to \Phi(X)$ where $\Phi(X)$ is a closed subspace of $[0,1]^{\mathbb{Z}_+}$ (equipped with the subspace topology).

The topological space $[0,1]^{\mathbb{Z}_+}$, equipped with the product topology, is called the **Hilbert cube**.

Hint. Part 1: Choose an infinite countable basis $\mathcal{B} = (U_1, U_2, \dots)$ for the topology of X where each U_n is nonempty. (Why can you do so?) Use Urysohn functions (Rem. 7.118) to construct $f_n : X \to [0,1]$ such that $f^{-1}(0) = X \backslash U_n$.

Part 2: Notice Pb. 7.8. □

Theorem 8.45. *Let X be a topological space. The following are equivalent.*

(1) X is a compact metrizable space.

(2) X is homeomorphic to a closed subset of the Hilbert cube $[0,1]^{\mathbb{Z}_+}$.

Proof. (1)⇒(2): By Pb. 8.15. (2)⇒(1): By Cor. 7.76, $[0,1]^{\mathbb{Z}_+}$ is metrizable. By countable Tychonoff theorem (Thm. 3.54 or Pb. 8.7), $[0,1]^{\mathbb{Z}_+}$ is compact. So its closed subsets are compact by Cor. 8.22. □

⋆ **Exercise 8.46.** Let $N \in \mathbb{Z}_+$. Prove that the following are equivalent.

(1) X is a compact Hausdorff space. Moreover, there exist $f_1, \dots, f_N \in C(X, \mathbb{R})$ separating points of X.

(2) X is homeomorphic to a bounded closed subset of $\mathbb{R}^N$.

9 The injection $\Psi : C(X, C(Y, \mathcal{V})) \to C(X \times Y, \mathcal{V})$

In this chapter, unless otherwise stated, X, Y are topological spaces, all normed vectors spaces are over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, and $\mathcal{V}$ is a normed vector space.

Remark 9.1. Let $\Phi : \mathcal{V} \to \mathcal{W}$ be a linear map of normed vector spaces. Recall that Φ is an isometry (in the category of metric spaces) iff $\|\Phi(u) - \Phi(v)\| = \|u - v\|$ for all $u, v \in \mathcal{V}$. By linearity, we have

$$\Phi \text{ is an isometry} \qquad \Longleftrightarrow \qquad \|\Phi(v)\| = \|v\| \text{ for all } v \in \mathcal{V} \tag{9.1}$$

Definition 9.2. A linear map $\Phi : \mathcal{V} \to \mathcal{W}$ of normed vector spaces over $\mathbb{F}$ is a called an **isomorphism of normed vector spaces** if Φ is an isometric isomorphism. If Φ is an isomorphism, and if one of $\mathcal{V}, \mathcal{W}$ is complete, then the other one is also complete. In this case, we call Φ an **isomorphism of Banach spaces**.

Recall that if X is compact, then the norm on $C(X, \mathcal{V})$ is assumed to be the l^∞-norm (cf. Conv. 3.49).

9.1 Ψ is bijective when Y is compact

Theorem 9.3. *Equip $C(Y, \mathcal{V})$ with the uniform convergence topology (cf. Exp. 7.77). Then there is a well-defined injective linear map*

$$\Psi : C(X, C(Y, \mathcal{V})) \to C(X \times Y, \mathcal{V})$$
$$\Psi(F)(x, y) = F(x)(y) \tag{9.2}$$

(for each $F \in C(X, C(Y, \mathcal{V}))$). Moreover, the following are true:

(a) *If Y is compact, then Ψ is a linear isomorphism of vector spaces.*

(b) *If X, Y are compact, then Ψ is an isomorphism of normed vector spaces.*

Proof. Step 1. Write $\mathcal{W} = C(Y, \mathcal{V})$ for simplicity. To prove that (9.2) is a well-defined map, we need to prove that for each $F \in C(X, \mathcal{W})$, the map $f = \Psi(F) : X \times Y \to \mathcal{V}$ sending (x, y) to $f(x, y) = F(x)(y)$ is continuous.

The continuity of $F : X \to \mathcal{W}$ means that if $(x_\alpha)_{\alpha \in I}$ is a net in X converging to x, then $F(x_\alpha)$ converges to $F(x)$, i.e.,

$$\lim_{\alpha \in I} A_\alpha = 0 \tag{9.3a}$$

where each $A_\alpha \in \overline{\mathbb{R}}_{\geq 0}$ is

$$A_\alpha = \sup_{y \in Y} \|f(x_\alpha, y) - f(x, y)\| \tag{9.3b}$$

Now, suppose that $(x_\alpha, y_\alpha)_{\alpha \in I}$ is a net in $X \times Y$ converging to (x, y). Then

$$\|f(x_\alpha, y_\alpha) - f(x, y)\| \leqslant \|f(x_\alpha, y_\alpha) - f(x, y_\alpha)\| + \|f(x, y_\alpha) - f(x, y)\|$$
$$\leqslant A_\alpha + \|f(x, y_\alpha) - f(x, y)\|$$

Since $F(x) : y \in Y \mapsto f(x, y) \in \mathcal{V}$ is continuous, we have $\lim_\alpha \|f(x, y_\alpha) - f(x, y)\| = 0$. Therefore, by Squeeze theorem, we have

$$\lim_{\alpha \in I} \|f(x_\alpha, y_\alpha) - f(x, y)\| = 0 \tag{9.4}$$

Thus, f is continuous at every (x, y).

Step 2. Clearly Ψ is linear and injective. Assume that Y is compact. Then the surjectivity of Ψ follows from Exp. 8.12. This proves (a). Assume that X is also compact. Choose any $F \in C(X, \mathcal{W})$ and write $f = \Psi(F)$. Then

$$\sup_{x \in X} \sup_{y \in Y} \|f(x, y)\| = \sup_{x \in X, y \in Y} \|f(x, y)\|$$

by the easy Lem. 9.4. This proves that Φ is an isometry, and hence proves (b). $\square$

Lemma 9.4. *Let* $g : A \times B \to \overline{\mathbb{R}}$ *be a function where* A, B *are sets. Then*

$$\sup_{a \in A} \sup_{b \in B} g(a, b) = \sup_{(a,b) \in A \times B} g(a, b)$$

Proof. Write $\lambda_a = \sup_{b \in B} g(a, b)$ and $\rho = \sup_{(a,b) \in A \times B} g(a, b)$. Then, clearly $\lambda_a \leqslant \rho$ for each a. So $\sup_a \lambda_a \leqslant \rho$. For each a, b we have $g(a, b) \leqslant \lambda_a$, and hence $g(a, b) \leqslant \sup_a \lambda_a$. Taking sup over a, b yields $\rho \leqslant \sup_a \lambda_a$. $\square$

Remark 9.5. Thanks to Thm. 9.3, we can reduce many problems about multivariable functions to problems about single-variable functions. Here is an example we will study in the future: If $I = [a, b], J = [c, d]$ are compact intervals and $F \in C(I \times J, \mathbb{R})$, then with the help of Thm. 9.3, the Fubini's theorem

$$\int_a^b \int_c^d F(x, y) dx dy = \int_c^d \int_a^b F(x, y) dx dy$$

for Riemann integrals follows directly from the easy general fact

$$\int_a^b \Lambda \circ f(x) dx = \Lambda \left(\int_a^b f(x) dx \right)$$

where $f : [a, b] \to \mathcal{W}$ is a continuous map to a real Banach space $\mathcal{W}$, and $\Lambda : \mathcal{W} \to \mathbb{R}$ is a continuous linear map.

Thm. 9.3 can be used the other way round: We will prove in the future that every Banach space $\mathcal{V}$ is isomorphic to a closed linear subspace of $C(Y, \mathbb{F})$ for some compact Hausdorff space Y. Thus, a problem about continuous maps $X \to \mathcal{V}$ (where $\mathcal{V}$ is an abstract Banach space) can be reduced to a problem about continuous scalar-valued functions $X \times Y \to \mathbb{F}$. $\square$

In the following sections, we give a surprising application of Thm. 9.3: We show that uniform-convergence and equicontinuity, two closely related but different notions, can be understood in the same context.

9.2 Equicontinuity

Let I be a set not necessarily preordered or directed.

Definition 9.6. Assume that Y is a metric space. A family of functions $(f_\alpha)_{\alpha \in I}$ from X to Y is called **equicontinuous at** $x \in X$ if the following equivalent conditions hold:

(1) The function

$$X \to Y^I \qquad x \mapsto (f_\alpha(x))_{\alpha \in I} \tag{9.5}$$

is continuous at x, where Y^I is equipped with the uniform convergence topology (Exp. 7.77).

(2) For every $\varepsilon > 0$, there exists $U \in \mathrm{Nbh}_X(x)$ such that for every $p \in U$ we have

$$\sup_{\alpha \in I} d_Y(f_\alpha(p), f_\alpha(x)) < \varepsilon$$

Clearly, if $(f_\alpha)_{\alpha \in I}$ is equicontinuous at x, then $f_\alpha : X \to Y$ is continuous at x for every $\alpha \in I$. We say that $(f_\alpha)_{\alpha \in I}$ is **(pointwise) equicontinuous** if it is equicontinuous at every point of X.

Proof of equivalence. This is immediate if we choose the uniform convergence metric on Y^I to be

$$d((y_\alpha), (y'_\alpha)) = \min \left\{ \sup_{\alpha \in I} d_Y(y_\alpha, y'_\alpha), 1 \right\}$$

and use (the base version of) Def. 7.56-(2). □

Remark 9.7. Warning: The above definition of equicontinuity is weaker than the one in Rudin's book [Rud-P, Ch. 7], which will be called **uniform equicontinuity** in this course (cf. Def. 10.11).

Example 9.8. Assume that X and Y are metric spaces. Fix $C \geqslant 0$. Then

$$\{f \in Y^X : f \text{ has Lipschitz constant } C\}$$

is an equicontinuous family of functions $X \to Y$.

In the future, we will see that if $f : [a, b] \to \mathbb{R}$ (where $[a, b] \subset \mathbb{R}$) is differentiable and satisfies $|f'(x)| \leqslant C$ for all $x \in [a, b]$, then f has Lipschitz constant C. (For example, if we assume moreover that f' is continuous, then for each $a \leqslant x < y \leqslant b$ we have $|f(y) - f(x)| = |\int_x^y f'| \leqslant \int_x^y |f'| \leqslant C(y - x)$.) Therefore, all such functions form an equicontinuous family of functions. □

Equicontinuity is important for several reasons. First, equicontinuity is closely related to compactness, both under the uniform convergence topology and under the pointwise convergence topology. This is hinted at in Rem. 9.22, and will be explored in more detail in a future chapter. Second, equicontinuity and uniform convergence are symmetric notions:

Remark 9.9. Note that if $x \in \overline{X \backslash x}$, a map $\varphi : X \to Y$ is continuous at x iff $\lim_{p \to x} f(p) = f(x)$, where $\lim_{p \to x} f(p)$ is the limit of a net (cf. Rem. 7.82). Now, assume that Y is a metric space. Then we see that a family $(f_\alpha)_{\alpha \in I}$ in Y^X satisfies that

$$(f_\alpha)_{\alpha \in I} \text{ is equicontinuous at } x \qquad \Longleftrightarrow \qquad \lim_{p \to x} \sup_{\alpha \in I} d(f_\alpha(p), f_\alpha(x)) = 0 \qquad (9.6)$$

If we compare this with

$$f_\alpha \rightrightarrows f \qquad \Longleftrightarrow \qquad \lim_{\alpha \in I} \sup_{x \in X} d(f_\alpha(x), f(x)) = 0 \qquad (9.7)$$

(if I is a directed set and $f \in Y^X$), we see that equicontinuity and uniform convergence are "symmetric about the diagonal line of the Cartesian product $I \times X$": Equicontinuity is a uniform convergence over the index set I, and the uniform convergence $f_\alpha \rightrightarrows f$ is uniform over X.

The symmetry of equicontinuity and uniform convergence will be further studied in Sec. 9.3. (Indeed, we will see that it is better to view "uniform convergence + continuity" and "pointwise convergence + equicontinuity" as symmetric conditions.) As an application, we will see that equicontinuity is equivalent to uniform convergence for any sequence (f_n) of pointwise convergent continuous functions on a compact topological space. (See Cor. 9.26.) Thus, in this case, one can prove the uniform convergence of (f_n) by proving for instance that it converges pointwise and has a uniform Lipschitz constant.

9.3 Uniform convergence and equicontinuity: two faces of Ψ

9.3.1 Main results

In this subsection, we fix a directed set I, and let

$$I^* = I \cup \{\infty_I\}$$

where ∞_I is a new symbol not in I. Write ∞_I as ∞ for simplicity. Equip I^* with the standard topology as in Def. 7.121. Recall that this topology has basis

$$\mathcal{B} = \left\{ \{\alpha\} : \alpha \in I \right\} \cup \left\{ I^*_{\geqslant \alpha} : \alpha \in I \right\}$$

Clearly $I \times X$ is dense in $I^* \times X$. Equip $C(X, \mathcal{V})$ and $C(I^*, \mathcal{V})$ with the uniform convergence topologies.

Throughout this subsection, we fix a net $(f_\alpha)_{\alpha \in I}$ in $\mathcal{V}^X$ and an element $f_\infty \in \mathcal{V}^X$. Define

$$F : I^* \times X \to \mathcal{V} \qquad F(\mu, x) = f_\mu(x) \tag{9.8}$$

The meaning of the title of this section is illustrated by the following theorem.

Theorem 9.10. *We have (a)$\Leftrightarrow$(b) and (1)$\Leftrightarrow$(2) where:*

(a) *$(f_\alpha)_{\alpha \in I}$ converges uniformly to f_∞, and $f_\alpha : X \to \mathcal{V}$ is continuous for each $\alpha \in I$.*

(b) *F gives rise to a continuous map $I^* \to C(X, \mathcal{V})$*

(1) *$(f_\alpha)_{\alpha \in I}$ is equicontinuous and converges pointwise to f_∞.*

(2) *F gives rise to a continuous map $X \to C(I^*, \mathcal{V})$.*

Proof. Assume (a). Then we have $f_\infty \in C(X, \mathcal{V})$ due to Thm. 7.79. So F gives rise to a map $I^* \to C(X, \mathcal{V})$. Since $f_\alpha \rightrightarrows f_\infty$, we see that F is continuous by Pb. 7.6-2. This proves (b).

Assume (b), then the fact that the map $I^* \to C(X, \mathcal{V})$ has range in $C(X, \mathcal{V})$ means precisely that each f_α and f_∞ are continuous. The continuity of the map $I^* \to C(X, \mathcal{V})$ at ∞ means $f_\alpha \rightrightarrows f_\infty$. This proves (a).

(1)$\Rightarrow$(2): Assume (1). The equicontinuity of (f_α) is equivalent to that

$$x \in X \mapsto (f_\alpha(x))_{\alpha \in I} \in \mathcal{V}^I \tag{9.9}$$

is continuous where $\mathcal{V}^I$ is equipped with the uniform convergence topology. Equivalently, for each $x \in X$ and $\varepsilon > 0$ there is $U \in \mathrm{Nbh}_X(x)$ such that for all $p \in U$ and all $\alpha \in I$ we have

$$\|f_\alpha(p) - f_\alpha(x)\| \leqslant \varepsilon \tag{9.10}$$

Since $(f_\alpha)_{\alpha \in I}$ converges pointwise to f_∞, by applying $\lim_{\alpha \in I}$ to (9.10), we see that $\|f_\mu(p) - f_\mu(x)\| \leqslant \varepsilon$ for all $p \in U$ and $\mu \in I^*$. So

$$x \in X \mapsto (f_\mu(x))_{\mu \in I^*} \in \mathcal{V}^{I^*} \tag{9.11}$$

is continuous, where $\mathcal{V}^{I^*}$ is given the uniform convergence metric. By Pb. 7.6-2, the pointwise convergence of $(f_\alpha)_{\alpha \in I}$ to f_∞ is equivalent to the continuity of

$$\mu \in I^* \to f_\mu(x) \in \mathcal{V} \tag{9.12}$$

for each $x \in X$. So the map (9.11) has range inside $C(I^*, \mathcal{V})$.

(2)$\Rightarrow$(1): Assume (2). The continuity of $X \to C(I^*, \mathcal{V})$ implies that of (9.9). So $(f_\alpha)_{\alpha \in I}$ is equicontinuous. Its pointwise convergence to f is due to the continuity of (9.12) for each x, which is clearly true by (2). $\qquad \square$

Remark 9.11. Conditions (a) and (1) in Thm. 9.10 are symmetric. Condition (a) says roughly that F converges uniformly under the limit over I, and converges pointwise under the limit over X. Condition (1) says roughly that F converges uniformly under the limit over X, and pointwise under the limit over I. The next theorem clarifies the relationship between these two symmetric conditions. We will see a similar condition in Thm. 9.28.

Theorem 9.12. *Consider the following statements:*

(1) The function $F : I^* \times X \to \mathcal{V}$ is continuous.

(2) $(f_\alpha)_{\alpha \in I}$ converges uniformly to f_∞, and $f_\alpha : X \to \mathcal{V}$ is continuous for each $\alpha \in I$.

(3) $(f_\alpha)_{\alpha \in I}$ is equicontinuous and converges pointwise to f_∞.

Then we have

$$(2) \Longrightarrow (1) \Longleftarrow (3)$$
$$(2) \Longleftrightarrow (1) \qquad \text{if } X \text{ is compact}$$
$$(1) \Longleftrightarrow (3) \qquad \text{if } I = \mathbb{N}$$

where $\mathbb{N}$ is equipped with the usual order.

Proof. This follows immediately from Thm. 9.10, Thm. 9.3, and the fact that I^* is compact if $I = \mathbb{N}$. $\qquad\square$

Remark 9.13. From the above proof, we see that the equivalence (1)⇔(3) holds in the more general case that I^* is compact. See Pb. 9.1 for an equivalent description of the compactness of I^*.

Example 9.14. Define $f_n : (0,1) \to \mathbb{R}$ by $f_n(x) = x^n$. Then $(f_n)_{n \in \mathbb{Z}_+}$ is equicontinuous (by Exp. 9.8) and pointwise convergent, but not uniformly convergent. Accordingly, $(0,1)$ is not compact.

Example 9.15. Let $I = \mathbb{Z}_+ \times \mathbb{Z}_+$, equipped with the product preorder: $(k_1, n_1) \leqslant (k_2, n_2)$ means $k_1 \leqslant k_2$ and $n_1 \leqslant n_2$. Consider $(f_{k,n})_{(k,n) \in I}$, where $f_{k,n} : [0,1] \to \mathbb{R}$ is defined by $f_{k,n}(x) = \dfrac{x^n}{k}$. Then $\lim_{(k,n) \in I} f_{k,n}$ converges uniformly to 0 by squeeze theorem and $0 \leqslant f_{k,n} \leqslant k^{-1}$. Thus, condition (1) of Thm. 9.12 is satisfied. However, this net of functions is not equicontinuous at $x = 1$. Moreover, for every $(k_0, n_0) \in I$, the net of functions $(f_{k,n})_{(k,n) \in I_{\geqslant (k_0, n_0)}}$ is not equicontinuous at 1.

Proof. Choose any $x \in [0,1)$. Then (recalling Lem. 9.4)

$$\sup_{k \geqslant k_0} \sup_{n \geqslant n_0} |f_{k,n}(1) - f_{k,n}(x)| = \sup_{k \geqslant k_0} \sup_{n \geqslant n_0} \frac{1 - x^n}{k} = \sup_{k \geqslant k_0} \frac{1}{k} = \frac{1}{k_0}$$

$\qquad\square$

In the following, let me give a more explicit description of condition (1) of Thm. 9.12 in terms of (f_α) and f_∞.

Proposition 9.16. *Assume that $f_\alpha : X \to \mathcal{V}$ is continuous for each $\alpha \in I$. Then part (1) of Thm. 9.12 is equivalent to both (1') and (1"), where*

(1') For each $x \in X$, we have

$$\lim_{\substack{(\alpha,p)\in I\times X \\ (\alpha,p)\to(\infty,x)}} f_\alpha(p) = f_\infty(x) \tag{9.13}$$

(1") For each $x \in X$ and $\varepsilon > 0$, there exist $\beta \in I$ and $U \in \mathrm{Nbh}_X(x)$ such that

$$\|f_\alpha(p) - f_\infty(x)\| < \varepsilon \qquad \text{(for all $\alpha \in I_{\geqslant\beta}$ and $p \in U$)} \tag{9.14}$$

Note that one can make sense of the LHS of (9.13) because $I \times X$ is clearly dense in $I^* \times X$.

Proof. The equivalence of (1') and (1") is clear by Def. 7.81. The continuity of f_α for all $\alpha \in I$ means precisely that $F|_{I\times X}$ is continuous. Therefore, (1') is equivalent to part (1) of Thm. 9.12, thanks to the following Thm. 9.17. $\qquad\square$

See Exp. 9.31 for an elementary example satisfying (1') of Prop. 9.16 (or equivalently, (1) of Thm. 9.12), but not satisfying (2) or (3) of Thm. 9.12. (See also Pb. 9.5 and Thm. 9.30 for related facts.)

9.3.2 Proving continuity using limits

Theorem 9.17. *Let $\varphi : X \to Y$ be a map of topological spaces where Y is metrizable. Let A be a dense subset of X. Then the following are equivalent.*

(1) $\varphi : X \to Y$ is continuous.

(2) The restriction $\varphi|_A : A \to Y$ is continuous. Moreover, for each $x \in X \backslash A$, the restriction $\varphi|_{A\cup\{x\}}$ is continuous at x, namely (cf. Def. 7.81) ,

$$\lim_{\substack{p\in A \\ p\to x}} \varphi(p) = \varphi(x) \tag{9.15}$$

Proof. Clearly (2) is equivalent to

$$\varphi|_{A\cup\{x\}} \text{ is continuous at } x \text{ for all } x \in X \tag{9.16}$$

So (1)$\Rightarrow$(2). Assume (2) and that Y is a metric space. Choose any $x \in X$. We want to show that $\varphi : X \to Y$ is continuous at x. Choose any $\varepsilon > 0$. Recall that an open

subset of A is precisely the intersection of A and an open subset of X. Thus, by (9.16), there is $U \in \mathrm{Nbh}_X(x)$ such that for all $p \in A \cap U$ we have

$$d(\varphi(p), \varphi(x)) \leqslant \varepsilon/2 \qquad (9.17)$$

It remains to prove (9.17) for all $p \in U$. Choose any $p \in U$. By Lem. 9.18, $A \cap U$ is dense in U. Thus, there is a net (p_α) in $A \cap U$ converging to p. In particular, for each α we have

$$d(\varphi(p_\alpha), \varphi(x)) \leqslant \varepsilon/2$$

Since $\varphi|_{A \cup \{p\}}$ is continuous (by (9.16)), we have $\lim_\alpha \varphi(p_\alpha) = \varphi(p)$. Thus, applying $\lim_\alpha$ to the above inequality proves (9.17) for $p \in U$. $\qquad \square$

Lemma 9.18. *Suppose that A is a dense subset of X. Let U be an open subset of A. Then $A \cap U$ is dense in U.*

This lemma is clearly false if U is not assumed to be open: simply take $U = X \backslash A$.

First proof. Choose any $x \in U$. We want to find a net in $A \cap U$ converging to x. Since A is dense in X, there is a net $(x_\alpha)_{\alpha \in J}$ in X converging to x. Since U is a neighborhood of x, (x_α) is eventually in U, say $x_\alpha \in U$ whenever $\alpha \geqslant \beta$. Then $(x_\alpha)_{\alpha \in J_{\geqslant \beta}}$ is a net in $A \cap U$ converging to x. $\qquad \square$

Second proof. We want to show that every nonempty open subset of U intersects $A \cap U$. But an open subset of U is precisely an open subset of X contained inside U (Exe. 7.52). So this set (when nonempty) must intersect A because A is dense in X. $\qquad \square$

⋆ **Exercise 9.19.** In Thm. 9.17, weaken the metrizability of Y to the condition that Y is regular. (See the definition below.) Prove the conclusion of Thm. 9.17.

⋆ **Definition 9.20.** A topological space Y is called **regular** if for every $y \in Y$ and every $U \in \mathrm{Nbh}_Y(y)$ there exists $V \in \mathrm{Nbh}_Y(y)$ such that $\mathrm{Cl}_Y(V) \subset U$.

9.3.3 Immediate consequences of Thm. 9.12

The following result is parallel to Thm. 7.79 for uniformly convergent nets of continuous functions.

Corollary 9.21. *Let $(f_\alpha)_{\alpha \in I}$ be an equicontinuous net of functions $X \to \mathcal{V}$. Assume that (f_α) converges pointwise to $f : X \to \mathcal{V}$. Then f is continuous.*

Proof. Write $f_\infty = f$ and define F by (9.8). Then F is continuous by (3)⇒(1) of Thm. 9.12. So $f = f_\infty$ is continuous, since f is the composition of F with the inclusion map $x \in X \mapsto (\infty, x) \in I^* \times X$. $\qquad \square$

Remark 9.22. In the future, we will study the general Tychonoff theorem, which says for example that if $(f_\alpha)_{\alpha \in I}$ is a net of functions $X \to \mathbb{R}^N$ which is pointwise bounded, i.e. $\sup_{\alpha \in I} \|f(x)\| < +\infty$ for all $x \in X$, then (f_α) has a subnet converging poinwise. However, if we assume moreover that each f_α is continuous, we cannot conclude in general that (f_α) has a subnet converging pointwise to a *continuous* function. But we can make such a conclusion when (f_α) is equicontinuous, thanks to Cor. 9.21. Therefore, Cor. 9.21 tells us that equicontinuity is useful for studying the problems of compactness of families of continuous functions (under the pointwise convergence topology). Cf. Thm. 17.7.

In fact, we have a slightly stronger version of Cor. 9.21:

Corollary 9.23. *Let* $(f_\alpha)_{\alpha \in I}$ *be a net of functions* $X \to \mathcal{V}$ *equicontinuous at* x. *Assume that* (f_α) *converges pointwise to* $f : X \to \mathcal{V}$. *Then* f *is continuous at* x.

Proof. Define X_x to be the same as X as a set, but has a different topology: the one generated by the basis

$$\mathcal{B}_x = \mathrm{Nbh}_X(x) \cup \big\{ \{p\} : p \neq x \big\}$$

(cf. Exe. 7.62). Define $g_\alpha : X_x \to \mathcal{V}$ and $g : X_x \to \mathcal{V}$ to be the same as f_α and f. Then (g_α) is a net of equicontinuous functions converging pointwise to g. Therefore, by Cor. 9.21, g is continuous. So f is continuous at x. $\qquad \square$

Remark 9.24. By a similar argument, we can generalize Thm. 7.79 to the following form: Let (f_α) be a net of functions $X \to \mathcal{V}$ converging uniformly to $f : X \to \mathcal{V}$. Suppose that each f_α is continuous at x. Then f is continuous at x.

Example 9.25. In this example, we pretend to know derivatives. Let (f_n) be a sequence of functions $\mathbb{R}_{\geq 0} \to \mathbb{R}$ defined by $f_n(x) = x^{1/n}$. (We understand $0^{\frac{1}{n}} = 0$.) Find all $x \in \mathbb{R}_{\geq 0}$ at which (f_n) is equicontinuous.

Proof. We prove that $\mathbb{R}_{>0}$ is the set of all points at which (f_n) is equicontinuous. First, assume $x > 0$. Choose $0 < a < 1 < b$ such that $a < x < b$. Then, on $[a, b]$, $f'_n(x) = \frac{1}{n} x^{\frac{1}{n}-1}$ is bounded by $C = \max\{a^{-1}, b\}$. So $(f_n|_{[a,b]})$ has Lipschitz constant C by Exp. 9.8. So (f_n) is equicontinuous at x.

Note that (f_n) converges pointwise to f where $f(x) = 1$ if $x > 0$ and $f(0) = 0$. But f is not continuous at 0. So (f_n) is not equicontinuous at 0 due to Cor. 9.23. $\quad \square$

One can also prove that (f_n) is equicontinuous on $(0, 1) \cup (1, +\infty)$ without using derivatives: See Thm. 9.30-1.

Corollary 9.26. *Let* $(f_\alpha)_{\alpha \in I}$ *be a net in* $C(X, \mathcal{V})$ *converging pointwise to* $f \in \mathcal{V}^X$. *Consider the following statements:*

(1) $(f_\alpha)_{\alpha \in I}$ *converges uniformly to* f.

(2) $(f_\alpha)_{\alpha\in I}$ *is equicontinuous.*

Then the following are true.

    1. If $(f_\alpha)_{\alpha\in I}$ *is a sequence* $(f_n)_{n\in\mathbb{Z}_+}$*, then (1)⇒(2).*

    2. If X is compact, then (2)⇒(1).

Proof. Immediate from Thm. 9.12. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

9.4   Uniform convergence and double limits

In this section, we use Thm. 9.12 to study the problem of when the order of double limits can be exchanged. We first give a useful criterion on uniform convergence.

Proposition 9.27. *Assume that* $\mathcal{V}$ *is a Banach space. Let* $(f_\alpha)_{\alpha\in I}$ *be a net in* $C(X,\mathcal{V})$. *Assume that* (f_α) *converges uniformly on a dense subset E of X. Then* (f_α) *converges uniformly on X to some* $f \in C(X,\mathcal{V})$.

The completeness of $\mathcal{V}$ is important here.

Proof. Since E is dense, by (7.9a) (applied to the function $|f_\alpha - f_\beta|$), we have

$$\sup_{x\in X}\|f_\alpha(x) - f_\beta(x)\| = \sup_{x\in E}\|f_\alpha(x) - f_\beta(x)\|$$

where the RHS converges to 0 under $\lim_{\alpha,\beta}$. Thus, (f_α) is a Cauchy net in $\mathcal{V}^X$ where $\mathcal{V}^X$ is equipped with the uniform convergence metric as in Exp. 7.77. So $(f_\alpha)_{\alpha\in I}$ converges uniformly on X to some $f : X \to \mathcal{V}$ by Thm. 7.78. By Thm. 7.79, f is continuous. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

Theorem 9.28 (Moore-Osgood theorem). *Let* $(f_{\alpha,\beta})_{(\alpha,\beta)\in I\times J}$ *be a net in a Banach space* $\mathcal{V}$ *with index set* $I \times J$ *where* I, J *are directed sets. Assume the following conditions:*

(1) For each $\alpha \in I$*, there exists* $f_{\alpha,\infty} \in \mathcal{V}$ *such that*

$$\lim_{\beta\in J} f_{\alpha,\beta} = f_{\alpha,\infty} \tag{9.18}$$

(2) For each $\beta \in J$*, there exists* $f_{\infty,\beta} \in \mathcal{V}$ *such that*

$$\lim_{\alpha\in I}\sup_{\beta\in J}\|f_{\alpha,\beta} - f_{\infty,\beta}\| = 0 \tag{9.19}$$

Then the following are true:

1. *The following limits exist and are equal:*

$$\lim_{(\alpha,\beta)\in I\times J} f_{\alpha,\beta} = \lim_{\alpha\in I} f_{\alpha,\infty} = \lim_{\beta\in J} f_{\infty,\beta} \tag{9.20}$$

2. *If $I = \mathbb{N}$ where $\mathbb{N}$ is equipped with the usual order, then*

$$\limsup_{\beta\in J} \|f_{\alpha,\beta} - f_{\alpha,\infty}\| = 0 \tag{9.21}$$

Conditions (1) and (2) in Thm. 9.28, which say that the limit of $f_{\alpha,\beta}$ is pointwise over one index and uniform in another index, should remind you of conditions (2) and (3) in Thm. 9.12 (cf. Rem. 9.11). In fact, we shall use Thm. 9.12 to understand and prove the Moore-Osgood theorem.

Proof. Part 1: By Thm. 5.29, it suffices to prove that $\lim_{\alpha,\beta} f_{\alpha,\beta}$ converges. Define topological spaces $I^* = I \cup \{\infty_I\}$ and $J^* = J \cup \{\infty_J\}$ as in Subsec. 9.3.1. Define

$$g_\alpha : J^* \to \mathcal{V} \qquad g_\alpha(\nu) = f_{\alpha,\nu}$$

where $g_\alpha(\infty_J) = f_{\alpha,\infty}$. By (1), for each $\alpha \in I$, the function g_α is continuous at ∞_J, and hence $g_\alpha \in C(J^*, \mathcal{V})$. By (2), $(g_\alpha)_{\alpha\in I}$ converges uniformly on J. Since J is a dense subset of J^*, by Prop. 9.27, $(g_\alpha)_{\alpha\in I}$ converges uniformly on J^* to some $g_{\infty_I} : J^* \to \mathcal{V}$. Thus, by Thm. 9.12, the function

$$F : I^* \times J^* \to \mathcal{V} \qquad F(\mu,\nu) = g_\mu(\nu)$$

is continuous. Its continuity at (∞_I, ∞_J) implies that $\lim_{\alpha\in I,\beta\in J} f_{\alpha,\beta} = \lim_{\alpha\in I,\beta\in J} F(\alpha,\beta)$ converges to $F(\infty_I, \infty_J)$.

Part 2: By Thm. 9.12, $(g_\alpha)_{\alpha\in I}$ is an equicontinuous family of functions $J^* \to \mathcal{V}$. Its equicontinuity at ∞_J means precisely (9.21). $\square$

Remark 9.29. Whenever you see a theorem stated in very plain language but proved using a huge machinery, you should always ask yourself if a direct proof is possible. A huge machinery or fancy language is not always necessary for the proof, but often helps to understand the nature of the problem.

Now, since Thm. 9.28 is stated without using the language of topological spaces and continuous maps, it is desirable to have a direct and elementary proof. It could be done by directly translating the above proof (and the proof of the results cited in that proof) into the pure language of nets. However, we prefer to give a simpler proof which is related to, but is not a direct translation of, the above topological proof. $\square$

A direct proof of Thm. 9.28. Part 1: By Thm. 5.29, it suffices to prove that $\lim_{\alpha,\beta} f_{\alpha,\beta}$ converges. Since V is complete, it suffices to prove the Cauchy condition:

(i) For each $\varepsilon > 0$, there exists $(\alpha, \beta) \in I \times J$ such that for all $(\mu, \nu) \in I_{\geqslant \alpha} \times J_{\geqslant \beta}$, we have $\| f_{\alpha,\beta} - f_{\mu,\nu} \| < \varepsilon$.

Choose $\varepsilon > 0$. By condition (2) of Thm. 9.28, $(f_{\alpha,\beta})$, as a net of functions $J \to \mathcal{V}$ with index set I, converges uniformly. Thus, by the Cauchy condition for the uniform convergence metric as in Exp. 7.77 (which is available due to Thm. 7.78), we have:

(ii) There exists $\alpha \in I$ such that for all $\mu \geqslant \alpha$, we have $\sup_{\nu \in J} \| f_{\alpha,\nu} - f_{\mu,\nu} \| < \varepsilon/2$.

Fix α as above. By condition (1), the limit $\lim_{\beta \in J} f_{\alpha,\beta}$ exists. Thus:

(iii) There exists $\beta \in J$ such that for all $\nu \geqslant \beta$ we have $\| f_{\alpha,\beta} - f_{\alpha,\nu} \| < \varepsilon/2$.

Combining (ii) and (iii) and using triangle inequality, we see that for each $\mu \geqslant \alpha$ and $\nu \geqslant \beta$,

$$\| f_{\alpha,\beta} - f_{\mu,\nu} \| \leqslant \| f_{\alpha,\beta} - f_{\alpha,\nu} \| + \| f_{\alpha,\nu} - f_{\mu,\nu} \| < \varepsilon$$

This proves (i).

Part 2: Assume $I = \mathbb{N}$. Since $(f_{\alpha,\beta})$ is a Cauchy net, for each $\varepsilon > 0$ there exist $\alpha_0 \in I, \beta_0 \in J$ such that for all $\alpha \geqslant \alpha_0$ and $\beta, \nu \geqslant \beta_0$ we have $\| f_{\alpha,\beta} - f_{\alpha,\nu} \| < \varepsilon$. Applying $\lim_\nu$ gives

$$\| f_{\alpha,\beta} - f_{\alpha,\infty} \| \leqslant \varepsilon \qquad (\forall \alpha \geqslant \alpha_0, \beta \geqslant \beta_0)$$

Since $I = \mathbb{N}$, there are finitely many α not $\geqslant \alpha_0$. For any such α, by condition (1) of Thm. 9.28, there exists $\beta_\alpha \in J$ such that for all $\beta \geqslant \beta_\alpha$, we have $\| f_{\alpha,\beta} - f_{\alpha,\infty} \| \leqslant \varepsilon$. Choose $\widetilde{\beta}$ greater than or equal to β_0 and all these (finitely many) β_α. Thus, we have $\| f_{\alpha,\beta} - f_{\alpha,\infty} \| \leqslant \varepsilon$ for all $\alpha \in I$ and all $\beta \geqslant \widetilde{\beta}$. This proves (9.21). □

9.5 Problems and supplementary materials

Recall from the beginning of this chapter that X is a topological space and $\mathcal{V}$ is a normed vector space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$.

Problem 9.1. Let I be a directed set. Let $I^* = I \cup \{\infty\}$, equipped with the standard topology as in Subsec. 9.3.1. Prove that the following are equivalent:

(1) I^* is compact.

(2) For each $\alpha \in I$, the complement of $I^*_{\geqslant \alpha} = \{\beta \in I^* : \beta \geqslant \alpha\}$ is a finite set.

Note. Prop. 8.30 can make your proof shorter. □

Problem 9.2. Let (f_n) be a sequence of functions $\mathbb{R}_{\geq 0} \to \mathbb{R}$ defined by $f(x) = x^{\frac{1}{n}}$ (as in Exp. 9.25). Give a direct proof that (f_n) is not equicontinuous at 0 using the definition of equicontinuity. Do not use Cor. 9.23.

Problem 9.3. Give a direct proof of Cor. 9.21 without using Thm. 9.3 (and its consequences) or using Thm. 9.28.

Problem 9.4. Give a direct proof of Cor. 9.26 without using Thm. 9.3 (and its consequences) or using Thm. 9.28.

$\star$ **Problem 9.5.** Let $(f_\alpha)_{\alpha \in I}$ be a net in $C(X, \mathbb{R})$. Assume that $(f_\alpha)_{\alpha \in I}$ is increasing, i.e., $f_\alpha \leq f_\beta$ whenever $\alpha \leq \beta$. Assume that $(f_\alpha)_{\alpha \in I}$ converges pointwise to $f \in C(X, \mathbb{R})$. Prove that for every $x \in X$,

$$\lim_{\substack{\alpha \in I \\ p \to x}} f_\alpha(p) = f(x) \tag{9.22}$$

in the sense of Prop. 9.16. Namely, prove that for every $x \in X$ and $\varepsilon > 0$ there exist $\beta \in I$ and $U \in \mathrm{Nbh}_X(x)$ such that $|f_\alpha(p) - f(x)| < \varepsilon$ for all $\alpha \geq \beta$ and all $p \in U$.

Note. Let $g_\alpha = f - f_\alpha$. Then $(g_\alpha)_{\alpha \in I}$ is a decreasing net of continuous functions converging pointwise to 0. It suffices to prove the easier statement that $\lim_{\alpha \in I, p \to x} g_\alpha(p) = 0$ for every $x \in X$. (Why is this sufficient?) $\qquad\square$

$\star$ **Theorem 9.30.** *Let $(f_\alpha)_{\alpha \in I}$ be a net in $C(X, \mathbb{R})$. Assume that $(f_\alpha)_{\alpha \in I}$ is increasing and converges pointwise to $f \in C(X, \mathbb{R})$. The following statements are true.*

1. *If $(f_\alpha)_{\alpha \in I}$ is a sequence $(f_n)_{n \in \mathbb{Z}_+}$, then $(f_n)_{n \in \mathbb{Z}_+}$ is equicontinuous.*

2. *(**Dini's theorem**) If X is compact, then $(f_\alpha)_{\alpha \in I}$ converges uniformly to f.*

Proof. By Pb. 9.5, $(f_\alpha)_{\alpha \in I}$ and $f_\infty = f$ satisfy (1′) of Prop. 9.16. Therefore, the two statements follow directly from Thm. 9.12. $\qquad\square$

Example 9.31. Let $f_n : (0, 1) \to \mathbb{R}$ be $f_n(x) = x^n$, where $n \in \mathbb{Z}$. Then $(f_n)_{n \in \mathbb{Z}}$ is a decreasing net of continuous functions converging pointwise to 0. Here, $\mathbb{Z}$ is given the usual order "$\leq$". By Pb. 9.5, (f_n) satisfies

$$\lim_{(n,p) \to (+\infty, x)} f_n(p) = f(x)$$

for all $x \in (0, 1)$. (So it satisfies condition (1) of Thm. 9.12.) However, $(f_n)_{n \in \mathbb{Z}}$ is neither equicontinuous (since $\sup_{n \in \mathbb{Z}} |f_n(p) - f_n(x)| = +\infty$ whenever $0 < p < x < 1$) nor converging uniformly to 0 (since $\sup_{x \in (0,1)} |f_n(x)| = 1$ if $n > 0$). Accordingly, $(0, 1)$ is not compact, and $I^* = I \cup \{\infty_I\}$ is not compact if $I = \mathbb{Z}$.

However, if we replace $\mathbb{Z}$ by $\mathbb{Z}_+$, then $(f_n)_{n \in \mathbb{Z}_+}$ is equicontinuous by Exp. 9.8 (applied to any compact subinterval of $(0, 1)$), or by Thm. 9.30. But $(f_n)_{n \in \mathbb{Z}_+}$ is still not uniformly convergent. $\qquad\square$

Problem 9.6. Let $X_1, X_2, \ldots$ be a sequence of nonempty topological spaces. Let $S = \prod_{n \in \mathbb{N}_+} X_n$, equipped with the product topology. Let $f : S \to \mathbb{R}$ be continuous. Fix $(p_n)_{n \in \mathbb{Z}_+} \in S$. For each $n \in \mathbb{Z}_+$, define

$$\varphi_n : S \to S$$
$$(x_1, x_2, \ldots, x_{n-1}, x_n, x_{n+1}, \ldots) \mapsto (x_1, x_2, \ldots, x_{n-1}, p_n, p_{n+1}, \ldots)$$

Prove for every $x_\bullet = (x_n)_{n \in \mathbb{Z}_+}$ that

$$\lim_{\substack{(n, y_\bullet) \in \mathbb{Z}_+ \times S \\ (n, y_\bullet) \to (\infty, x_\bullet)}} f \circ \varphi_n(y_\bullet) = f(x_\bullet)$$

in the sense of Prop. 9.16. Conclude that $(f \circ \varphi_n)_{n \in \mathbb{Z}_+}$ is equicontinuous, and that if each X_n is compact then $(f \circ \varphi_n)$ converges uniformly to f. (Recall that in this case S is compact by the countable Tychonoff theorem, cf. Pb. 8.7.)

10 Extending continuous functions to the closures

10.1 Introduction

Let $\widetilde{X}, Y$ be topological spaces, let $X \subset \widetilde{X}$, and let $f : X \to Y$ be a continuous map. The **extension problem** asks whether f can be extended to a continuous map $\widetilde{f} : \widetilde{X} \to Y$. "Extended" means that $\widetilde{f}|_X = f$. Since we can try to first extend f from X to its closure $\mathrm{Cl}_{\widetilde{X}}(X)$, and then from $\mathrm{Cl}_{\widetilde{X}}(X)$ to $\widetilde{X}$, the extension problem can naturally be divided into two cases: (1) X is dense in $\widetilde{X}$. (2) X is closed in $\widetilde{X}$.

In this chapter, we study the first case. (The second case will be discussed in Sec. 15.4.) Assume that X is dense in $\widetilde{X}$. Then by Prop. 7.63, we know that f can have at most one extension if Y is Hausdorff. So there is essentially no uniqueness issue.

The study of extension problem in this case has a long history. As we have seen in Subsec. 7.6, the limits of functions can be understood in this light: If $x \in \widetilde{X} \backslash X$, then f can be extended to a continuous function on $X \cup \{x\}$ iff $\lim_{p \to x} f(p)$ exists. Of course, this is simply a rephrasing of the definition of $\lim_{p \to x} f(p)$. But the idea of "extending f to $\widetilde{X}$ by first extending f to a slightly larger set with one extra point $\{x\}$" is helpful and can sometimes simplify proofs.

Indeed, recall that in the proof of Prop. 9.16 we used Thm. 9.17, which tells us that if $\lim_{p \to x} f(p)$ converges for all $x \in \widetilde{X} \backslash X$, then f can be extended (necessarily uniquely) to a continuous $\widetilde{f} : \widetilde{X} \to Y$ where Y is assumed metrizable. Thus, the extensibility of f to $\widetilde{X}$ can be checked pointwise. Thm. 9.17 is our first important general result on the extension problem. Let me state Thm. 9.17 in the following equivalent way, which is more convenient for the study of extension problems.

Corollary 10.1. *Let $f : X \to Y$ be a continuous map of topological spaces where Y is metrizable and X is a dense subspace of a topological space $\widetilde{X}$. The following are equivalent:*

(1) There exists a continuous map $\widetilde{f} : \widetilde{X} \to Y$ such that $\widetilde{f}|_X = f$.

(2) For each $x \in \widetilde{X} \backslash X$, the limit $\lim_{p \to x} f(p)$ exists.

Proof. If (1) is true, then $\widetilde{f}|_{X \cup \{x\}}$ is continuous whenever $x \in \widetilde{X} \backslash X$. This proves (2). Conversely, assume (2). Extend f to a map $\widetilde{f} : \widetilde{X} \to Y$ by setting $\widetilde{f}(x) = \lim_{p \to x} f(p)$ if $x \in \widetilde{X} \backslash X$. Then $\widetilde{f}$ is continuous by Thm. 9.17. $\qquad\square$

This chapter will focus on another useful method for extending continuous functions in the setting of metric spaces. A main result (cf. Cor. 10.9) is that if $\widetilde{X}, Y$ are metric spaces and Y is complete, then a sufficient condition for the extensibility of f onto $\widetilde{X}$ is that f is uniformly continuous. Moreover, uniform continuity

is also a necessary condition if $\widetilde{X}$ is compact. (Recall that compact metric spaces are complete, cf. Thm. 3.23.) It is worth mentioning that the extensibility of f is a purely topological question, whereas the uniform continuity of f depends on the metric on $\widetilde{X}$.

10.2 Uniform continuity

Fix metric spaces $\widetilde{X}, Y$ and a dense subset $X \subset \widetilde{X}$.

10.2.1 Basics

We first give some examples of $f \in C(X, Y)$ that cannot be extended to a continuous function on $\widetilde{X}$. Since uniform continuity is a sufficient condition for the extensibility (when Y is complete), these examples are not uniformly continuous. Recall our convention that subsets of $\mathbb{R}$ or $\mathbb{R}^N$ are equipped with the Euclidean metrics.

Example 10.2. Assume $X = (0, +\infty)$, $\widetilde{X} = [0, +\infty)$, $Y = \mathbb{R}$, and $f : X \to Y$ is defined by $f(x) = 1/x$. Then $\lim_{x \to 0} f(x)$ is $+\infty$ in $\overline{\mathbb{R}}$, and hence does not converge in Y. So f cannot be extended onto $\widetilde{X}$.

Example 10.3. $X = (0, 1]$, $\widetilde{X} = [0, 1]$, $Y = \mathbb{R}$, $f : X \to Y$ is defined by $f(x) = \sin(1/x)$. Then $\lim_{n \to \infty} f(x_n)$ equals 0 if the sequence (x_n) in X converging to 0 is defined by $x_n = 1/2n\pi$, and equals 1 if $x_n = 1/(2n + \frac{1}{2})\pi$. Thus, by Rem. 7.85, $\lim_{x \to 0} f(x)$ does not exist in Y. So f cannot be extended onto $\widetilde{X}$.

Definition 10.4. A map $f : X \to Y$ is called **uniformly continuous** if the following equivalent conditions are satisfied:

(1) For every $\varepsilon > 0$ there exists $\delta > 0$ such that for all $x, x' \in X$ we have

$$d(x, x') < \delta \qquad \Longrightarrow \qquad d(f(x), f(x')) < \varepsilon \qquad (10.1a)$$

(2) For every nets $(x_\alpha)_{\alpha \in I}$, $(x'_\alpha)_{\alpha \in I}$ in X (with the same index set I) we have

$$\lim_{\alpha \in I} d(x_\alpha, x'_\alpha) = 0 \qquad \Longrightarrow \qquad \lim_{\alpha \in I} d(f(x_\alpha), f(x'_\alpha)) = 0 \qquad (10.1b)$$

(3) For every sequences $(x_n)_{n \in \mathbb{Z}_+}$, $(x'_n)_{n \in \mathbb{Z}_+}$ in X we have

$$\lim_{n \to \infty} d(x_n, x'_n) = 0 \qquad \Longrightarrow \qquad \lim_{n \to \infty} d(f(x_n), f(x'_n)) = 0 \qquad (10.1c)$$

Uniformly continuous maps are clearly continuous. Def. 10.4-(2) says that uniformly continuous functions are those sending Cauchy-equivalent nets (cf. Def. 5.36) to Cauchy-equivalent nets.

Proof of equivalence. (1)⇒(2): Assume (1). Choose nets $(x_\alpha)_{\alpha \in I}, (x'_\alpha)_{\alpha \in I}$ in X. Choose any $\varepsilon > 0$. Choose $\delta > 0$ such that (10.1a) holds. If $d(x_\alpha, x'_\alpha) \to 0$, then $d(x_\alpha, x'_\alpha) < \delta$ for sufficiently large α. So $d(f(x_\alpha), f(x'_\alpha)) < \varepsilon$ for sufficiently large α. This proves (2).

(2)⇒(3): Obvious.

¬(1) ⇒ ¬(3): Assume that (1) is false. Then there exists $\varepsilon > 0$ such that for all $\delta > 0$ there exist $x, x' \in X$ with $d(x, x') < \delta$ such that $d(f(x), f(x')) \geqslant \varepsilon$. Thus, for each $n \in \mathbb{Z}_+$, there exist $x_n, x'_n \in X$ such that $d(x_n, x'_n) < 1/n$ and $d(f(x_n), f(x'_n)) \geqslant \varepsilon$. The sequences $(x_n), (x'_n)$ imply that (10.1c) is false. □

Corollary 10.5. *Assume that $f : X \to Y$ is uniformly continuous. Let $(x_\alpha)_{\alpha \in I}$ be a Cauchy net in X. Then $(f(x_\alpha))_{\alpha \in I}$ is a Cauchy net in Y.*

Proof. Apply Def. 10.4-(2) to the nets $(x_\alpha)_{(\alpha, \beta) \in I^2}$ and $(x_\beta)_{(\alpha, \beta) \in I^2}$ of X. □

10.2.2 Extensibility of uniformly continuous functions

The following theorem can be viewed as the uniform continuity version of Prop. 9.27. In particular, both results assume the completeness of the codomain.

Theorem 10.6. *Let $f : X \to Y$ be uniformly continuous, and assume that Y is complete. Then there exists a (necessarily unique) uniformly continuous $\widetilde{f} : \widetilde{X} \to Y$ satisfying $\widetilde{f}|_X = f$.*

Proof. Choose any $x \in \widetilde{X}$. Since X is dense in $\widetilde{X}$, we can choose a sequence (x_n) in X converging to x in $\widetilde{X}$. In particular, (x_n) is a Cauchy sequence. Therefore, by Cor. 10.5, $(f(x_n))$ is a Cauchy sequence in Y, which converges to some point $\widetilde{f}(x) \in Y$ by the completeness of Y. If $x \in X$, we assume that (x_n) is the constant sequence x. This shows that $\widetilde{f}(x) = f(x)$ if $x \in X$.

We have constructed a function $\widetilde{f} : \widetilde{X} \to Y$ satisfying $\widetilde{f}|_X = f$. Let us prove that $\widetilde{f}$ is uniformly continuous. Choose any $\varepsilon > 0$. Since f is uniformly continuous, there is $\delta > 0$ such that for all $p, q \in X$ we have

$$d(p, q) < 2\delta \qquad \Longrightarrow \qquad d(f(p), f(q)) < \varepsilon/2$$

Choose any $x, x' \in \widetilde{X}$ satisfying $d(x, x') < \delta$. By our construction of $\widetilde{f}$, there are sequences (x_n) in X converging to x and (x'_n) in X converging to x' such that $f(x_n) \to \widetilde{f}(x)$ and $f(x'_n) \to \widetilde{f}(x')$. Thus, there exist $N \in \mathbb{Z}_+$ such that for all $n \geqslant N$ we have

$$d(x, x_n) < \frac{\delta}{2} \qquad d(x', x'_n) < \frac{\delta}{2} \qquad d(\widetilde{f}(x), f(x_n)) < \frac{\varepsilon}{4} \qquad d(\widetilde{f}(x'), f(x'_n)) < \frac{\varepsilon}{4}$$

Choose $n = N$. Then by triangle inequality, we have $d(x_n, x'_n) < 2\delta$, and hence $d(f(x_n), f(x'_n)) < \varepsilon/2$. So $d(\widetilde{f}(x), \widetilde{f}(x')) < \varepsilon$ by triangle inequality again. □

We now study the other direction. The following theorem implies that if a continuous $f : X \to Y$ can be extended to a continuous $\widetilde{f} : \widetilde{X} \to Y$, and if $\widetilde{X}$ is compact, then f is uniformly continuous.

Theorem 10.7. *Suppose that $f : X \to Y$ is continuous and X is compact. Then f is uniformly continuous.*

In the same spirit as in Sec. 8.2, we give two proofs for this theorem, one using sequences and the other using open covers.

First proof. Assume that f is not uniformly continuous. By Def. 10.4-(3), there exist sequences (x_n) and (x'_n) in X such that $\lim_{n \to \infty} d(x_n, x'_n) = 0$, and that $d(f(x_n), f(x'_n)) \not\to 0$. The latter means that there is $\varepsilon > 0$ such that $d(f(x_n), f(x'_n))$ is frequently $\geqslant \varepsilon$. Thus, by passing to a subsequence, we may assume that $d(f(x_n), f(x'_n)) \geqslant \varepsilon$ for all n. Since $X \times X$ is sequentially compact, by replacing (x_n, x'_n) with a convergent subsequence, we assume that $x_n \to x$ and $x'_n \to x'$ where $x, x' \in X$. Since $d(x_n, x'_n) \to 0$, we must have $x = x'$. By the continuity of f, $f(x_n)$ and $f(x'_n)$ converge to $f(x) = f(x')$, contradicting the fact that $d(f(x_n), f(x'_n)) \geqslant \varepsilon$ for all n. $\qquad\square$

To prove Thm. 10.7 using open covers, we prove a more general result instead. The following theorem is useful for proving properties of the form "there exists $\delta > 0$ such that for all $x, x' \in X$ satisfying $d(x, x') < \delta$, we have ...".

Theorem 10.8 (Lebesgue number lemma). *Assume that the metric space X is compact. Let $\mathfrak{U} \subset 2^X$ be an open cover of X. Then there exists $\delta > 0$ satisfying the following conditions:*

- *For every $x \in X$ there exists $U \in \mathfrak{U}$ such that $B_X(x, \delta) \subset U$.*

The number δ in Thm. 10.8 is called a **Lebesgue number of** $\mathfrak{U}$. In the following proof, we follow the local-to-global strategy as in Sec. 8.2.

Proof. Choose any $p \in X$. Then there is $U \in \mathfrak{U}$ containing p. So there is $\delta_p > 0$ such that $B(p, 2\delta_p) \subset U$. Therefore, there exists $V_p \in \mathrm{Nbh}_X(p)$ such that for every x in V_p we have $B(x, \delta_p) \subset U$. (Simply take $V_p = B(p, \delta_p)$.) This solves the problem locally: for each $x \in V_p$, the ball $B(x, \delta_p)$ is a subset of some member of $\mathfrak{U}$.

Since $X = \bigcup_{p \in X} V_p$ and since X is compact, there is a finite subset $E \subset X$ such that $X = \bigcup_{p \in E} V_p$. Take $\delta = \min\{\delta_p : p \in E\}$. For each $x \in X$, choose $p \in E$ such that $x \in V_p$. Then $B(x, \delta_p)$ is a subset of some member of $\mathfrak{U}$ by the last paragraph. So the same is true for $B(x, \delta)$. $\qquad\square$

Of course, similar to the examples studied in Sec. 8.2, Thm. 10.8 can also be proved by contradiction and by using sequential compactness. See Pb. 10.2.

Second proof of Thm. 10.7. Let us verify Def. 10.4-(1). Choose any $\varepsilon > 0$. For each $x \in X$, the set $U_x = f^{-1}(B_Y(f(x), \varepsilon/2))$ is a neighborhood of x by Prop. 7.60. So $\{U_x : x \in X\}$ is an open cover of X. Let δ be a Lebesgue number of $\mathfrak{U}$. Choose any $x, y \in X$ satisfying $d(x, y) < \delta$. Then $B_X(x, \delta) \subset U_z$ for some $z \in X$. So $x, y \in B_X(x, \delta)$ and hence $x, y \in U_z$. Therefore,

$$d(f(x), f(y)) \leqslant d(f(x), f(z)) + d(f(z), f(y)) < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

$\square$

Corollary 10.9. *Choose $f \in C(X, Y)$. Consider the following statements:*

(1) f is uniformly continuous.

(2) There exists $\widetilde{f} \in C(\widetilde{X}, Y)$ such that $\widetilde{f}|_X = f$.

Then (1)$\Rightarrow$(2) if Y is complete, and (2)$\Rightarrow$(1) if $\widetilde{X}$ is compact.

Proof. Immediate from Thm. 10.6 and Thm. 10.7. $\square$

Example 10.10. Let $D = B_{\mathbb{C}}(0, 1) = \{z \in \mathbb{C} : |z| < 1\}$ and $S^1 = \{z \in \mathbb{C} : |z| = 1\}$. Let $f : D \to Y$ be a continuous function where Y is a complete metric space. Then f is uniformly continuous iff $\lim_{w \to z} f(w)$ exists for every $z \in S^1$.

Proof. By Cor. 10.1, the limit $\lim_{w \to z} f(w)$ exists for every $z \in S^1$ iff f can be extended to a continuous function $\widetilde{f} : \overline{D} = D \cup S^1 \to Y$. By Cor. 10.9, this is equivalent to that f is uniformly continuous (because $\overline{D}$ is compact and Y is compete). $\square$

10.2.3 Uniform equicontinuity

Although the notion of uniform equicontinuity will rarely be used in our notes, it is used in many textbooks. So let me give a brief account of uniform equicontinuity.

Definition 10.11. Let $(f_\alpha)_{\alpha \in I}$ be a family of functions $X \to Y$. (Here, the index set I is not necessarily directed.) Define a metric d on Y^I in a similar way as (7.19), namely, if $\mathbf{y}, \mathbf{y}' \in Y^I$ then

$$d(\mathbf{y}, \mathbf{y}') = \min \left\{ 1, \sup_{\alpha \in I} d_Y(\mathbf{y}(\alpha), \mathbf{y}'(\alpha)) \right\}$$

We say that $(f_\alpha)_{\alpha \in I}$ is **uniformly equicontinuous** if the map

$$X \to Y^I \qquad x \mapsto (f_\alpha(x))_{\alpha \in I} \tag{10.2}$$

is uniformly continuous with respect to the metric d. Clearly, this is equivalent to saying that:

- For every $\varepsilon > 0$ there exists $\delta > 0$ such that for every $x, x' \in X$ satisfying $d_{(}x, x') < \delta$, we have

$$\sup_{\alpha \in I} d_Y(f_\alpha(x), f_\alpha(x')) < \varepsilon$$

Uniformly equicontinuous families of functions are equicontinuous, because uniformly continuous functions are continuous. Conversely, we have:

Proposition 10.12. *Assume that X is compact and $(f_\alpha)_{\alpha \in I}$ is a family of functions $X \to Y$. Then $(f_\alpha)_{\alpha \in I}$ is equicontinuous iff it is uniformly equicontinuous.*

Proof. "$\Leftarrow$" is obvious, as mentioned above. "$\Rightarrow$" follows immediately by applying Thm. 10.7 to the continuous map (10.2). $\square$

10.3 Completion of metric spaces

Fix a metric space X in this section. We are going to apply uniform continuity to the study of completions of metric spaces. Roughly speaking, a completion of X is a complete metric space $\widehat{X}$ containing X as a dense subspace. However, completions are not unique, but are unique "up to equivalence". So we want to show that two completions $\widehat{X}, \widetilde{X}$ of the same metric space X are equivalent. However, it is confusing to view X as a subset of $\widetilde{X}$ and $\widetilde{X}$ simultaneously. A better approach is to consider (automatically injective) isometries $\varphi : X \to \widehat{X}, \psi : X \to \widetilde{X}$, and to show that φ and ψ are equivalent using the language of commutative diagrams (cf. Sec. 1.2).

Definition 10.13. A **completion** of the metric space X is an isometry $\varphi : X \to \widehat{X}$ where $\widehat{X}$ is a complete metric space, and $\varphi(X)$ is dense in $\widehat{X}$. We sometimes just say that $\widehat{X}$ is a completion of X.

Thus, if A is a dense subset of a complete metric space B, then the inclusion map $A \hookrightarrow B$ is a completion. Therefore, $\mathbb{R}$ is a completion of $\mathbb{Q}$, and $[0, 1]$ is a completion of $(0, 1), [0, 1), [0, 1] \cap \mathbb{Q}$.

Example 10.14. Let A be a dense subset of a metric space X. Suppose that $\varphi : X \to \widehat{X}$ is a completion of X. Then $\varphi|_A : A \to \widehat{X}$ is clearly a completion of A.

Example 10.15. Let X be a subset of a complete metric space Y. Then $X \hookrightarrow \mathrm{Cl}_Y(X)$ is a completion of X because every closed subset of Y is complete (cf. Prop. 3.27), and hence $\mathrm{Cl}_Y(X)$ is complete. For example, $\{(x, y) \in \mathbb{R}^2 : x \geqslant 0\}$ is a completion of both $A = \{(x, y) \in \mathbb{R}^2 : x > 0\}$ and $B = A \cap \mathbb{Q}^2$.

We want to prove that every metric space X has a completion $\widehat{X}$. First, we need a lemma, which can be viewed as analogous to Pb. 8.9.

Lemma 10.16. *Suppose that X is a dense subspace of a metric space $\widehat{X}$. Suppose that every Cauchy sequence in X converges to an element of $\widehat{X}$. Then $\widehat{X}$ is complete.*

Proof. Let (x_n) be a Cauchy sequence in $\widehat{X}$. Since X is dense, there exists $x'_n \in X$ such that $d(x_n, x'_n) < 1/n$. So (x'_n) is Cauchy-equivalent to (x_n). Thus (x'_n) is a Cauchy sequence by Exe. 5.37. By assumption, (x'_n) converges to some $x \in X$. So (x'_n) also converges to x by Exe. 5.37. □

Theorem 10.17. *Every metric space X has a completion $\varphi : X \to \widehat{X}$. Moreover, any completion $\psi : X \to \widetilde{X}$ is **equivalent** (also called **isomorphic**) to φ in the sense that there is an isometric isomorphism of metric spaces $\Phi : \widehat{X} \to \widetilde{X}$ such that the following diagram commutes:*

$$
\begin{array}{ccc}
 & X & \\
\varphi \swarrow & & \searrow \psi \\
\widehat{X} & \xrightarrow[\simeq]{\Phi} & \widetilde{X}
\end{array}
\tag{10.3}
$$

Recall that the commutativity of (10.3) means that $\psi = \Phi \circ \varphi$.

Proof of existence. The construction of $\varphi : X \to \widehat{X}$ is similar to the construction of $\mathbb{R}$ from $\mathbb{Q}$ in Ch. 6. Let $\mathscr{C}$ be the set of Cauchy sequences in X. Let $\widehat{X} = \mathscr{C}/\sim$ be the quotient set (cf. Def. 1.18) where $\sim$ is the Cauchy-equivalence relation: $(x_n) \sim (y_n)$ iff $\lim_{n\to\infty} d(x_n, y_n) = 0$. We let $[x_n]_{n\in\mathbb{Z}_+}$ or simply let $[x_n]$ denote the equivalence class of (x_n) in $\widehat{X}$. The map φ is defined by

$$
\varphi : X \to \widehat{X} \qquad x \mapsto [x]_{n\in\mathbb{Z}_+}
$$

where $[x]_{n\in\mathbb{Z}_+}$ is the equivalence class of the constant sequence $(x, x, \dots)$.

Step 1: Let us define a metric $d_{\widehat{X}}$ on $\widehat{X}$. Note that if $(x_n), (y_n) \in \mathscr{C}$, then by triangle inequality,

$$
|d(x_m, y_m) - d(x_n, y_n)| \leqslant d(x_m, x_n) + d(y_m, y_n)
$$

where the RHS converges to 0 as $m, n \to +\infty$. Therefore, the LHS also converges to 0. This shows that $(d(x_n, y_n))_{n\in\mathbb{Z}_+}$ is a Cauchy sequence in $\mathbb{R}_{\geqslant 0}$, and hence converges. Therefore, we define

$$
d_{\widehat{X}} : \widehat{X} \times \widehat{X} \to \mathbb{R}_{\geqslant 0}
$$
$$
d_{\widehat{X}}([x_n], [y_n]) = \lim_{n\to\infty} d(x_n, y_n)
$$

This is well-defined: If $[x_n] = [x'_n]$, and $[y_n] = [y'_n]$, then

$$
|d(x_n, y_n) - d(x'_n, y'_n)| \leqslant d(x_n, x'_n) + d(y_n, y'_n)
$$

which converge to 0 as $n \to \infty$. So $(d(x'_n, y'_n))_{n \in \mathbb{Z}_+}$ and $(d(x_n, y_n))_{n \in \mathbb{Z}_+}$ are Cauchy-equivalent, and hence converge to the same number.

Clearly $d_{\widehat{X}}([x_n], [y_n]) = 0$ iff $(x_n) \sim (y_n)$ iff $[x_n] = [y_n]$. And clearly $d_{\widehat{X}}([x_n], [y_n]) = d_{\widehat{X}}([y_n], [x_n])$. If $[x_n], [y_n], [z_n]$ are in $\mathscr{C}$, applying $\lim_{n \to \infty}$ to

$$d(x_n, z_n) \leqslant d(x_n, y_n) + d(y_n, z_n)$$

yields $d_{\widehat{X}}([x_n], [z_n]) \leqslant d_{\widehat{X}}([x_n], [y_n]) + d_{\widehat{X}}([y_n], [z_n])$. So $d_{\widehat{X}}$ is a metric.

Step 2. The map $\varphi : X \to \widehat{X}$ is clearly an isometry. Let us show that it has dense range. Choose any $[x_n]_{n \in \mathbb{Z}_+} \in \widehat{X}$. We shall show that $\varphi(x_k) = [x_k, x_k, \dots]$ approaches $[x_n]_{n \in \mathbb{Z}_+}$ as $k \to \infty$.

For each k, we have

$$d_{\widehat{X}}(\varphi(x_k), [x_n]_{n \in \mathbb{Z}_+}) = \lim_{n \to \infty} d(x_k, x_n) \tag{10.4}$$

where the RHS converges because $d_{\widehat{X}}$ is defined. Since $(x_n)_{n \in \mathbb{Z}_+}$ is a Cauchy sequence in X, we have

$$\lim_{k, n \to \infty} d(x_k, x_n) = 0 \tag{10.5}$$

Therefore, by (10.5) and the convergence of the RHS of (10.4), we can use Thm. 5.29 to conclude that

$$\lim_{k \to \infty} d_{\widehat{X}}(\varphi(x_k), [x_n]_{n \in \mathbb{Z}_+}) = \lim_{k \to \infty} \lim_{n \to \infty} d(x_k, x_n) = \lim_{k, n \to \infty} d(x_k, x_n) = 0$$

Step 3. It remains to prove that $\widehat{X}$ is complete. By Lem. 10.16 (applied to $\varphi(X) \subset \widehat{X}$) and the fact that φ is an isometry, it suffices to prove that for every Cauchy sequence $(x_k)_{k \in \mathbb{Z}_+}$ in X, the sequence $(\varphi(x_k))_{k \in \mathbb{Z}_+}$ converges in $\widehat{X}$. But this is true: we have shown in Step 2 that $(\varphi(x_k))_{k \in \mathbb{Z}_+}$ converges to $[x_n]_{n \in \mathbb{Z}_+}$. $\qquad\square$

Proof of equivalence. Suppose that $\psi : X \to \widetilde{X}$ is another completion. The map

$$\Phi : \varphi(X) \to \psi(X) \qquad \varphi(x) \mapsto \psi(x)$$

is well-defined since φ is injective. Moreover, Φ is an isometry since φ and ψ are isometries. In particular, Φ is uniformly continuous. Therefore, by Cor. 10.9, Φ can be extended to a uniformly continuous map $\Phi : \widehat{X} \to \widetilde{X}$. Clearly $\psi = \Phi \circ \varphi$. The continuous map

$$\widehat{X} \times \widehat{X} \to \mathbb{R}$$
$$(p, q) \mapsto d_{\widetilde{X}}(\Phi(p), \Phi(q)) - d_{\widehat{X}}(p, q)$$

is zero on the dense subset $\varphi(X) \times \varphi(X)$ of its domain. Therefore it is constantly zero by Prop. 7.63. This proves that Φ is an isometry.

It remains to prove that Φ is surjective. Since $\widehat{X}$ is complete and Φ restricts to an isometric isomorphism of metric spaces $\widehat{X} \to \Phi(\widehat{X})$, $\Phi(\widehat{X})$ is a complete metric subspace of $\widehat{X}$. Thus, by Prop. 3.27, $\Phi(\widehat{X})$ is a closed subset of $\widetilde{X}$. But $\Phi(\widehat{X})$ is dense in $\widetilde{X}$ since it contains $\psi(X)$. Therefore $\Phi(\widehat{X}) = \widetilde{X}$. $\qquad\square$

The proof of Thm. 10.17 is complete.

10.4 Why did Hausdorff believe in completion?

Thm. 10.17, the existence and uniqueness of completion of an arbitrary metric space, was proved by Hausdorff in his 1914 work introducing Hausdorff topological spaces [Hau14, Sec. 8.8, p.315]. The construction of completion using equivalence classes of Cauchy sequences is quite abstract: Although it mimics Cantor's construction of real numbers (cf. Ch. 6), its main application is not in the realm of finite-dimensional geometric objects, but in the world of function spaces. But what is the practical significance of the equivalence classes of Cauchy sequences of functions? In concrete analysis problems about functions, these objects are much more difficult to deal with than functions themselves.

Strangely enough, the only nontrivial examples we have now is $\mathbb{Q} \hookrightarrow \mathbb{R}$. Besides this, we do not yet have any exciting new metric spaces arising from completion. For example: we know that $[0, 1]$ is the completion of $(0, 1)$ under the Euclidean metric, and that $C([0,1], \mathbb{R})$ is the completion of the set of real polynomials $\mathbb{R}[x]$ under the norm $\sup_{x \in [0,1]} |f(x)|$. (The density of the set of polynomials in $C([0,1], \mathbb{R})$ is due to Weierstrass.) But $[0, 1]$ and $C([0,1], \mathbb{R})$ are examples we are already familiar with.

Mathematicians do not generalize just for the sake of generalization. They want to solve problems by generalizing old concepts to a broader context. Moreover, mathematicians do not randomly choose a way of generalization and then build a huge theory. Instead, they develop a theory only in the direction that has already proved useful in solving explicit problems. Thus, Hausdorff proved Thm. 10.17 because he was already convinced of the importance of abstract completion by certain powerful examples. For the moment, we are not ready to study these examples rigorously. (We will do this in the next semester.) But I want to give an informal introduction to one of these examples which historically has paved the way for many important ideas in analysis.

10.4.1 The Fourier series method in integral equations

In the years of 1900-1907, many important progress has been made in integral equations, which originated from the study of Dirichlet problems (finding solutions of harmonic equation $\Delta\varphi(x, y) = 0$ with given boundary condition,

where $\Delta = (\partial_x)^2 + (\partial_y)^2$). For example, one asks if there is $\lambda \in \mathbb{R}$ and a single-variable complex valued function $f : [0, 2\pi] \to \mathbb{C}$ satisfying the eigenvalue problem $Tf = \lambda f$ where

$$(Tf)(x) = \int_0^{2\pi} K(x, y)f(y)dx$$

and $K : [0, 2\pi]^2 \to \mathbb{R}$ is given. (Cf. also Sec. 2.1.)

Hilbert studied this problem using Fourier series. In the modern language, the theory of Fourier series claims that $C([0, 2\pi], \mathbb{C})$, under the L^2-norm

$$\|f\|_{L^2} = \sqrt{\int_0^{2\pi} |f(x)|^2 \cdot \frac{dx}{2\pi}} \tag{10.6}$$

has a completion

$$\boxed{\begin{aligned} \Phi : C([0, 2\pi], \mathbb{C}) &\to l^2(\mathbb{Z}, \mathbb{C}) \\ \Phi(f) &= \widehat{f} \end{aligned}} \tag{10.7}$$

where $l^2(\mathbb{Z}, \mathbb{C})$ is the space of all $g : \mathbb{Z} \to \mathbb{C}$ satisfying $\sqrt{\sum_{n \in \mathbb{Z}} |g(n)|^2} < +\infty$, and

$$\widehat{f} : \mathbb{Z} \to \mathbb{C} \qquad \widehat{f}(n) = \int_0^{2\pi} f(x)e^{-inx} \cdot \frac{dx}{2\pi}$$

gives the **Fourier coefficients** of f.

Hilbert studied the eigenvalue problem $Tf = \lambda f$ by transforming it into a linear algebra problem on $l^2(\mathbb{Z}, \mathbb{C})$ (so that T becomes an $\infty \times \infty$ discrete matrix $\widehat{T}$), finding the possible eigenvectors $\widehat{f}$ and eigenvalues λ satisfying

$$\widehat{T}\widehat{f} = \lambda \widehat{f}$$

and returning to the original problems by finding the function f whose Fourier coeffients are $\widehat{f}$. It is easy to return: if $f \in C([0, 2\pi], \mathbb{C})$, then one gets f from $\widehat{f}$ by the formula

$$f(x) = \sum_{n \in \mathbb{Z}} \widehat{f}(n)e^{inx} \tag{10.8}$$

where the RHS is called the **Fourier series** of f.

Here comes the crucial point: the range of Φ, namely $\{\widehat{f} : f \in C([0, 1], \mathbb{C})\}$, is not the whole space $l^2(\mathbb{Z}, \mathbb{C})$ but only its dense subspace. This remains true if we enlarge $C([0, 1], \mathbb{C})$ to the space of Riemann integrable functions $\mathscr{R}([0, 1], \mathbb{C})$. However, the eigenvectors of $\widehat{T}$ found by Hilbert were only known to be elements

of $l^2(\mathbb{Z}, \mathbb{C})$. If we use (10.8) to find the eigenvectors of the original T, namely, if for an arbitrary $g \in l^2(\mathbb{Z}, \mathbb{C})$ we write

$$\Psi(g) = \sum_{n \in \mathbb{Z}} g(n)e^{\mathbf{i}nx} \tag{10.9}$$

then $\Psi(g)$ is not necessarily continuous or even Riemann integrable. It seems that, in the light of the proof of Thm. 10.17, the only way to make sense of (10.9) is as follows:

- One views $\Psi(g)$ as the Cauchy-equivalence class of the Cauchy sequence $(s_n)_{n \in \mathbb{Z}_+}$ in $C([0, 2\pi], \mathbb{C})$ under the L^2-norm, where $s_n(x) = \sum_{k=-n}^{n} g(k)e^{\mathbf{i}kx}$.

But how can we understand $\Psi(g)$ as an actual function on $[0, 2\pi]$ solving the eigenvalue problem $T\Psi(g) = \lambda\Psi(g)$?

Therefore, (10.7) and (10.9) are the very first example of abstract completion, and also one of the most important examples motivating Hausdorff's study of completion in general.

10.4.2   Riesz-Fischer theorem

Hilbert established these results by 1906, the same year Fréchet defined metric spaces. The story was finished by Riesz and Fischer, who proved in 1907 that $\Psi(g)$ can actually be represented by a **Lebesgue measurable** $f : [0, 2\pi] \to \mathbb{C}$ satisfying $\|f\|_{L^2} < +\infty$, where the norm $\|f\|_{L^2}$ is defined by (10.6) using Lebesgue integral instead of Riemann integral. Let $L^2([0, 2\pi], \mathbb{C})$ denote the space of all such functions, and view $C([0, 2\pi], \mathbb{C})$ as its subspace. Then using the language of Thm. 10.17, we have a commutative diagram

$$\begin{array}{ccc}
& C([0, 2\pi], \mathbb{C}) & \\
{\scriptstyle \Phi}\swarrow & & \searrow{\scriptstyle \iota} \\
l^2(\mathbb{Z}, \mathbb{C}) & \xrightarrow[\simeq]{\Psi} & L^2([0, 2\pi], \mathbb{C})
\end{array} \tag{10.10}$$

where ι is the inclusion map. Thus:

- The abstract completion $\Phi : C([0, 2\pi], \mathbb{C}) \to l^2(\mathbb{Z}, \mathbb{C})$ is equivalent to the concrete completion $C([0, 2\pi], \mathbb{C}) \subset L^2([0, 2\pi], \mathbb{C})$, where "concrete" means that it is function-theoretic.

This equivalence of abstract and concrete completions, connecting Hilbert's algebraic approach to integral equations and Lebesgue's function-theoretic approach

to measure theory, is the original form of **Riesz-Fischer theorem** (proved in 1907),[1] one of the most significant theorems in early 20th century.

Thanks to Riesz-Fischer theorem, people were convinced that abstract completions of function spaces could deepen one's understanding of analysis in much the same way that Lebesgue's measure theory broadened one's understanding of functions by harmonizing with Hilbert's l^2 theory. So what is the hesitation in trying to prove an abstract theorem like Thm. 10.17?

Hausdorff proved the existence and uniqueness of completion in general, but it was Riesz-Fischer theorem that proved the value of abstract completion.

10.5 Completion of normed vector spaces

Fix a normed vector space V over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$.

Definition 10.18. A **completion of the normed vector space** V is a *linear* isometry $\varphi : V \to \widehat{V}$ such that $\widehat{V}$ is a Banach space, and $\varphi(V)$ is dense in $\widehat{V}$.

Thus, if $\varphi : V \to \widehat{V}$ is a completion, then V is isomorphic to $\varphi(V)$ as normed vector spaces, and hence can be viewed as equivalently a dense normed subspace of $\widehat{V}$.

Theorem 10.19. *Every normed vector space V has a completion $\varphi : V \to \widehat{V}$. Morevoer, every completion $\psi : V \to \widetilde{V}$ is **isomorphic** to φ in the sense that there is an isomorphism of Banach spaces (cf. Def. 9.2) $\Phi : \widehat{V} \to \widetilde{V}$ such that the following diagram commutes:*

$$\begin{array}{ccc} & V & \\ {\scriptstyle \varphi} \swarrow & & \searrow {\scriptstyle \psi} \\ \widehat{V} & \xrightarrow[\simeq]{\ \Phi\ } & \widetilde{V} \end{array} \qquad (10.11)$$

Proof of existence. By Thm. 10.17, we have a completion $\varphi : V \to \widehat{V}$ in the context of metric spaces. So $\widehat{V}$ is a completion of metric space with metric $d_{\widehat{V}}$, and φ is an isometry of metric spaces with dense range. We need to show that $\widehat{V}$ is a complete normed vector space, and that φ is linear with dense range. Indeed, we shall identify V with $\varphi(\widehat{V})$ via φ so that V is the metric subspace of $\widehat{V}$. We shall extend the structure of normed vector space from V to $\widehat{V}$. In the following, the convergence in $\widehat{V}$ and the continuity of the maps about $\widehat{V}$ are understood using the metric $d_{\widehat{V}}$.

---

[1]More precisely, the original form of Riesz-Fischer theorem says that the Ψ in (10.10) (whose explicit description is as in (10.9)) is an isometric isomorphism. As a consequence, the space L^2 is complete, because it is fairly easy to show that l^2 is complete. However, most modern analysis textbooks present Riesz-Fischer theorem in the following simple form: "L^2 is a complete metric space".

The map

$$+ : V \times V \to V \qquad (u, v) \mapsto u + v$$

is Lipschitz continuous, and hence uniformly continuous. Therefore, by Cor. 10.9, it can be extended (uniquely) to a continuous map $+ : \widehat{V} \times \widehat{V} \to \widehat{V}$. Similarly, if $\lambda \in \mathbb{F}$, the Lipschitz continuous map

$$V \to V \qquad v \mapsto \lambda \cdot v$$

can be extended to a continuous map $\widehat{V} \to \widehat{V}$. Thus, we have defined the addition $+$ and the scalar multiplication $\cdot$ for $\widehat{V}$. Similarly, the map

$$\| \cdot \| : V \to \mathbb{R}_{\geqslant 0} \qquad v \mapsto \|v\|$$

is uniformly continuous and hence can be extended to a map $\widehat{V} \to \mathbb{R}_{\geqslant 0}$.

We want to show that the above addition, scalar multiplication, and norm function make $\widehat{V}$ a normed vector space. For example, suppose that $\lambda \in \mathbb{F}$. We want to prove that $\lambda(u + v) = \lambda u + \lambda v$ for all $u, v \in \widehat{V}$, and we know that this is true when $u, v \in V$. Indeed, since $V \times V$ is dense in $\widehat{V} \times \widehat{V}$, and since the following two continuous maps

$$(u, v) \in \widehat{V} \times \widehat{V} \mapsto \lambda(u + v) \in \widehat{V}$$
$$(u, v) \in \widehat{V} \times \widehat{V} \mapsto \lambda u + \lambda v \in \widehat{V}$$

are equal on $V \times V$, these two maps are the same by Prop. 7.63. The same argument proves that $\widehat{V}$ is a vector space.

Since the following continuous map

$$\varphi : (u, v) \in \widehat{V} \times \widehat{V} \mapsto \|u\| + \|v\| - \|u + v\| \in \mathbb{R}$$

satisfies $\varphi(V \times V) \subset \mathbb{R}_{\geqslant 0}$, by $\varphi(\widehat{V} \times \widehat{V}) \subset \overline{\varphi(V \times V)}$ (due to Prop. 7.63 again), we conclude that $\varphi(\widehat{V} \times \widehat{V}) \subset \mathbb{R}_{\geqslant 0}$. So $\|u + v\| \leqslant \|u\| + \|v\|$ for all $u, v \in \widehat{V}$. A similar argument shows $\|\lambda v\| = |\lambda| \cdot \|v\|$. Finally, suppose that $v \in \widehat{V}$ satisfies $\|v\| = 0$. Pick a sequence (u_n) in V satisfying $d_{\widehat{V}}(u_n, v) \to 0$. Since $\| \cdot \| : \widehat{V} \to \mathbb{R}$ is continuous,

$$\lim_{n \to \infty} d_V(u_n, 0) = \lim_{n \to \infty} \|u_n\| = \big\| \lim_{n \to \infty} u_n \big\| = \|v\| = 0$$

Therefore (u_n) converges to 0 in V. This proves $v = 0$. So $\widehat{V}$ is normed.

Since the metric on V is induced by the norm of V, the map

$$\widehat{V} \times \widehat{V} \to \mathbb{R} \qquad (u, v) \mapsto d_{\widehat{V}}(u, v) - \|u - v\|$$

is zero on the dense subset $V \times V$ of $\widehat{V} \times \widehat{V}$. Since this map is continuous, it is constantly zero. So the complete metric $d_{\widehat{V}}$ on $\widehat{V}$ (arising from the metric-space-completion of V) equals the metric induced by the norm of $\widehat{V}$. So latter is complete. Therefore, $\widehat{V}$ is a Banach space. Since V is dense in $\widehat{V}$ under $d_{\widehat{V}}$, V is dense in $\widehat{V}$ under the norm of $\widehat{V}$. $\qquad\square$

Proof of uniqueness. By Thm. 10.17, there is a unique isometric isomorphism of metric spaces $\Phi : \widehat{V} \to \widetilde{V}$ such that $\psi = \Phi \circ \varphi$. Since φ and ψ are linear injections, Φ is a linear isomorphism when restricted to $\varphi(V) \to \psi(V)$. Since Φ is continuous and $\varphi(V)$ is dense in $\widehat{V}$, we conclude that Φ is linear thanks to the following property. $\qquad\square$

Proposition 10.20. *Let $T : V \to W$ be a continuous map of normed vector spaces. Assume that V_0 is a dense linear subspace of V. Assume that $T|_{V_0} : V_0 \to W$ is linear. Then T is linear.*

Proof. This is same as the proof of the existence part of Thm. 10.19. Choose any $\alpha, \beta \in \mathbb{F}$. Then the following continuous map

$$(u, v) \in V \times V \mapsto T(\alpha u + \beta v) - (\alpha T(u) + \beta T(v))$$

is zero on the dense subset $V_0 \times V_0$. So it is zero on $V \times V$. $\qquad\square$

Exercise 10.21. Let V and W be normed vector spaces over $\mathbb{F}$, where $\mathbb{F} = \mathbb{R}$ (resp. $\mathbb{F} = \mathbb{C}$). Let $T : V \to W$ be a continuous map. Assume that V_0 is a dense $\mathbb{K}$-linear subspace of V, where $\mathbb{K} = \mathbb{Q}$ (resp. $\mathbb{K} = \mathbb{Q} + i\mathbb{Q}$). Assume that the restriction $T|_{V_0} : V_0 \to W$ is $\mathbb{K}$-linear. Prove that $T : V \to W$ is $\mathbb{F}$-linear.

Remark 10.22. So far in this course, we have proved a lot of results about functions whose codomains are normed vector spaces. Some results assume that these spaces are Banach (i.e. complete), some do not. Thanks to Thm. 10.19, we can assume that all these results are stated only for Banach spaces, and then check whether they also hold for normed vector spaces in general (which is not difficult). This will make us easier to remember theorems.

For example, suppose that we know that Thm. 7.79 holds only for Banach spaces: Namely, suppose we know that for any Banach space V and topological space X, if (f_α) is a net in $C(X, V)$ converging uniformly to $f : X \to V$, then f is continuous. Then we know that this result also holds when V is a normed vector space. To see this, consider the completion $V \subset \widehat{V}$. Then (f_α) is a net of continuous functions $X \to \widehat{V}$ converging uniformly to some $f : X \to \widehat{V}$ (satisfying $f(X) \subset V$). Then $f : X \to \widehat{V}$ is continuous. Hence $f : X \to V$ is continuous.

Consider Prop. 9.27 as another example. It tells us that if V is a Banach space and (f_α) is a net in $C(X, V)$ converging uniformly on a dense subset $E \subset X$, then (f_α) converges uniformly on X. Now assume that V is only a normed vector

space, and take completion $V \subset \widehat{V}$. Then (f_α) is a net in $C(X, \widehat{V})$ converges uniformly on E to a function $E \to V$. Thus, by Prop. 9.27, it converges uniformly to a function $f : X \to \widehat{V}$. Although we know $f(E) \subset V$ by assumption, we do not know whether $f(X) \subset V$. So we cannot prove the normed vector space version of Prop. 9.27. $\qquad \square$

10.6 Bounded linear maps

In this section, we consider normed vector spaces over a given field $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$.

We have seen that uniform continuity is crucial to the construction of addition and scalar multiplication in Thm. 10.19. Indeed, uniform continuity is ubiquitous in the world of normed vector spaces: We shall see that every continuous linear map of normed vector spaces is uniformly continuous.

Definition 10.23. Let $T : V \to W$ be a linear map of normed vector spaces. The **operator norm** $\|T\|$ is defined to be

$$\|T\| \overset{\text{def}}{=\!=\!=} \sup_{v \in \overline{B}_V(0,1)} \|Tv\| \tag{10.12}$$

Remark 10.24. $\|T\|$ is the smallest number in $\overline{\mathbb{R}}_{\geqslant 0}$ satisfying

$$\|Tv\| \leqslant \|T\| \cdot \|v\| \qquad (\forall v \in V) \tag{10.13}$$

Proof. (10.13) is clearly true when $v = 0$. Assume $v \neq 0$. Since $v/\|v\| \in \overline{B}_V(0, 1)$, we have $\|T(v/\|v\|)\| \leqslant \|T\|$. This proves (10.13).

Now suppose that $C \in \overline{\mathbb{R}}_{\geqslant 0}$ satisfies that $\|Tv\| \leqslant C\|v\|$ for all v. Then for each $v \in \overline{B}_V(0, 1)$ we have $\|Tv\| \leqslant C$. So $\|T\| \leqslant C$. $\qquad \square$

Proposition 10.25. *Let $T : V \to W$ be a linear map of normed vector spaces. Then the following are equivalent:*

(1) T is Lipschitz continuous.

(2) T is continuous.

(3) T is continuous at 0.

(4) $\|T\| < +\infty$.

Moreover, if one of these conditions holds, then T has Lipschitz constant $\|T\|$.

Proof. Clearly (1)$\Rightarrow$(2) and (2)$\Rightarrow$(3). Suppose (3) is true. Note that $T0 = 0$. So there is $\delta > 0$ such that $Tv \in \overline{B}_W(0, 1)$ for all $v \in \overline{B}_V(0, \delta)$. Namely, for all $v \in V$

satisfying $\|v\| \leqslant \delta$ we have $\|Tv\| = \|Tv - T0\| \leqslant 1$. Thus, if $\|v\| \leqslant 1$, then $\|\delta v\| \leqslant \delta$. So

$$\|Tv\| = \delta^{-1}\|T(\delta v)\| \leqslant \delta^{-1}$$

This proves $\|T\| \leqslant \delta^{-1}$. So (4) is proved.

Assume (4). By Rem. 10.24, for each $u, v \in V$ we have

$$\|Tu - Tv\| = \|T(u - v)\| \leqslant \|T\| \cdot \|u - v\|$$

This proves that T has Lipschitz constant $\|T\|$. $\qquad \square$

Due to Prop. 10.25-(4), we make the following definition:

Definition 10.26. Let V, W be normed vector spaces over $\mathbb{F}$. We call $T : V \to W$ to be a **bounded linear map** if T is a continuous linear map. We write

$$\mathfrak{L}(V, W) = \{\text{bounded linear maps } V \to W\} \qquad \mathfrak{L}(V) = \mathfrak{L}(V, V) \qquad (10.14)$$

Thus, the word "bounded" means that the linear map T is bounded on $\overline{B}_V(0, 1)$, but not that T is bounded on V.

Proposition 10.27. $\mathfrak{L}(V, W)$ *is a linear subspace of W^V, and the operator norm $\| \cdot \|$ is a norm on $\mathfrak{L}(V, W)$.*

Proof. By Rem. 10.24, for each linear $S, T : V \to W$ and $\lambda \in \mathbb{F}$, and for each $v \in V$ we have

$$\|(S + T)v\| \leqslant \|Sv\| + \|Tv\| \leqslant (\|S\| + \|T\|)\|v\|$$
$$\|\lambda Tv\| = |\lambda| \cdot \|Tv\| \leqslant |\lambda| \cdot \|T\| \cdot \|v\|$$

Thus, by Rem. 10.24 again, we have

$$\|S + T\| \leqslant \|S\| + \|T\| \qquad \|\lambda T\| \leqslant |\lambda| \cdot \|T\| \qquad (10.15)$$

The proposition now follows easily from the above inequalities. (Notice Rem. 3.34) $\qquad \square$

Since Lipschitz continuous functions are uniformly continuous, we have:

Proposition 10.28. *Let V_0 be a dense linear subspace of a normed vector space V. Let W be a Banach space. Let $T_0 : V_0 \to W$ be a bounded linear map. Then there is a unique bounded linear map $T : V \to W$ such that $T|_{V_0} = T_0$.*

Proof. Uniqueness is clear from the density of V_0. By Prop. 10.25, T_0 is uniformly continuous. Therefore, by Cor. 10.9, T_0 can be extended to a continuous map $T : V \to W$, which is linear by Prop. 10.20. $\qquad \square$

10.7 Problems and supplementary material

Problem 10.1. Give a direct proof of Thm. 10.7 using open covers instead of using subsequences. Do not use Lebesgue numbers.

The following Pb. 10.2 gives another proof that sequentially compact metric spaces are compact. Therefore, do not use this fact in your solution of Pb. 10.2.

$\star$ **Problem 10.2.** Let X be a sequentially compact metric space. Let $\mathfrak{U} \subset 2^X$ be an open cover of X.

1. Prove that $\mathfrak{U}$ has a Lebesgue number. Namely, prove that there exists $\delta > 0$ such that for every $x \in X$ there is $U \in \mathfrak{U}$ satisfying $B_X(x, \delta) \subset U$.

2. In our proof that X is separable (cf. Thm. 8.34), we showed that for every $\delta > 0$ there exists a finite set $E \subset X$ such that $d(x, E) < \delta$ for all $x \in X$. Use this fact and Part 1 to prove that $\mathfrak{U}$ has a finite subcover.

Definition 10.29. Two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on a vector space $\mathbb{V}$ over $\mathbb{R}$ or $\mathbb{C}$ are called **equivalent** if there exist $\alpha, \beta > 0$ such that for all $v \in V$ we have

$$\|v\|_1 \leqslant \alpha \|v\|_2 \qquad \|v\|_2 \leqslant \beta \|v\|_1$$

Clearly, two equivalent norms induce equivalent metrics, and hence induce the same topology.

Problem 10.3. Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. Let $\|\cdot\|$ be the Euclidean norm on $\mathbb{F}^n$. Let $\nu : \mathbb{F}^n \to \mathbb{R}_{\geqslant 0}$ be a norm on $\mathbb{F}^n$.

1. Prove that there exists $\alpha > 0$ such that $\nu(\mathbf{x}) \leqslant \alpha \|\mathbf{x}\|$ for all $\mathbf{x} \in \mathbb{F}^n$. In particular, show that ν is continuous (under the Euclidean topology).

2. Let $\beta = \inf\{\nu(\mathbf{x}) : \mathbf{x} \in \mathbb{F}^n, \|\mathbf{x}\| \leqslant 1\}$. Prove that $\beta > 0$. Prove that $\|\mathbf{x}\| \leqslant \beta^{-1} \cdot \nu(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{F}^n$.

The above problem proves

Theorem 10.30. *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. Then any norm on $\mathbb{F}^n$ is equivalent to the Euclidean norm. In particular, the operator norm on $\mathbb{F}^{m \times n}$ (if we view an $m \times n$ matrix as an element of $\mathfrak{L}(\mathbb{F}^n, \mathbb{F}^m)$) is equivalent to the Euclidean norm.*

11 Derivatives

11.1 Basic properties of derivatives

Fix a Banach space V over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. Let Ω be a nonempty open subset of $\mathbb{C}$. The following assumption will be often considered:

Definition 11.1. Let $f : [a, b] \to V$ (where $-\infty < a < b < +\infty$) be a map with variable t, and let $x \in [a, b]$. The **derivative** of f at x is

$$f'(x) \equiv \frac{df}{dt}(x) \stackrel{\text{def}}{=\!=} \lim_{\substack{t \in [a,b]\setminus\{x\} \\ t \to x}} \frac{f(t) - f(x)}{t - x} = \lim_{\substack{h \in [a-x, b-x]\setminus\{0\} \\ h \to 0}} \frac{f(x + h) - f(x)}{h}$$

provided that the limits converge. In other words (cf. Def. 7.81),

- $f'(x)$ converges to $v \in V$ iff for every $\varepsilon > 0$ there exists $\delta > 0$ such that for every $t \in [a, b]$ satisfying $0 < |t - x| < \delta$ we have

$$\left\| \frac{f(t) - f(x)}{t - x} - v \right\| < \varepsilon$$

If $f'(x)$ exists for some x, we say that f is **differentiable** at x. If $f'(x)$ exists for every $x \in [a, b]$, we say that f is a **differentiable function** and view f' as a function $[a, b] \to V$.

Derivatives on intervals $[a, b), (a, b], (a, b)$ are understood in a similar way. $\square$

Definition 11.2. Let E be a subset of $\mathbb{R}^n$. Let $f : E \to V$ be a function with variables $t_1, \ldots, t_n$, and $\mathbf{x} = (x_1, \ldots, x_n) \in E$. Let $1 \leqslant i \leqslant n$. Suppose that there are a, b satisfying $-\infty < a < x_i < b < +\infty$ such that $(x_1, \ldots, x_{i-1}, t, x_{i+1}, \ldots, x_n)$ belongs to E for all $t \in [a, b]$. The derivative of the function $t_i \mapsto f((x_1, \ldots, x_{i-1}, t_i, x_{i+1}, \ldots, x_n)$ at $t = x_i$ is denoted by

$$\frac{\partial f}{\partial t_i}(\mathbf{x}) \equiv \partial_i f(\mathbf{x})$$

and called the **partial derivative** of f at $\mathbf{x}$ with respect to the variable t_i.

Definition 11.3. If V is over $\mathbb{C}$, and if $f : \Omega \to V$ and $z \in \Omega$, we define the **derivative** of f at z to be

$$f'(z) = \lim_{\substack{w \in \Omega\setminus\{z\} \\ w \to z}} \frac{f(w) - f(z)}{w - z}$$

provided that the RHS exists, and simply write it as

$$\lim_{w \to z} \frac{f(w) - f(z)}{w - z} = \lim_{\zeta \to 0} \frac{f(z + \zeta) - f(z)}{\zeta}$$

Convention 11.4. Unless otherwise stated, when talking about derivatives of a function defined on an interval I, we always assume that I is inside $\mathbb{R}$ and has at least two points. When talking about derivatives of a function $f : \Omega \to V$, we always assume that V is over $\mathbb{C}$.

Definition 11.5. Given a function $f : E \to V$ where E is an interval in $\mathbb{R}$ with at least two points or $E = \Omega$, if $n \in \mathbb{N}$ and $x \in E$, we define the **n-th derivative** $f^{(n)}(x)$ inductively by $f^{(0)} = f$ and $f^{(n)}(x) = (f^{(n-1)})'(x)$ if $f^{(n-1)}$ exists on some neighborhood of x with respect to E. f'', f''', f'''', ... mean $f^{(2)}$, $f^{(3)}$, $f^{(4)}$, ...

The n-th partial derivative on the i-th variable is written as $\partial_i^n f$. $\qquad\square$

It is desirable to use sequences or nets to study derivatives. For that purpose, the following lemma is useful:

Lemma 11.6. *Let E be an interval in $\mathbb{R}$ with at least two elements, or let $E = \Omega$. Let $z \in E$. Let $f : E \to V$. Let $v \in V$. The following are equivalent.*

(1) We have $f'(z) = v$.

(2) For any sequence $(z_n)_{n \in \mathbb{Z}_+}$ in $E \backslash \{z\}$ converging to z, we have $\displaystyle\lim_{n \to \infty} \frac{f(z_n) - f(z)}{z_n - z} = v$.

(3) For any net $(z_\alpha)_{\alpha \in \mathcal{I}}$ in E converging to z, we have $\displaystyle\lim_{\alpha \in \mathcal{I}} \varphi(z_\alpha) = v$, where $\varphi : E \to V$ is defined by

$$\varphi(w) = \begin{cases} \dfrac{f(w) - f(z)}{w - z} & \text{if } w \neq z \\ v & \text{if } w = z \end{cases} \tag{11.1}$$

Proof. The equivalence of (1)$\Leftrightarrow$(2) is due to Def. 7.81-(3m). Also, by Def. 7.81-(1), that $f'(z) = v$ is equivalent to that the function φ in (11.1) is continuous at z. So it is equivalent to (3) by Def. 7.56-(1). $\qquad\square$

Proposition 11.7. *Let E be an interval in $\mathbb{R}$ or $E = \Omega$. If $f : E \to V$ is differentiable at $z \in E$, then f is continuous at z.*

Proof. We consider the case $f : \Omega \to V$; the other case is similar. Choose any sequence (z_n) in $\Omega \backslash \{z\}$ converging to z. By Lem. 11.6, we have $\lim_n \frac{f(z_n) - f(z)}{z_n - z} = v$. Since $\lim_n (z_n - z) = 0$, by the continuity of scalar multiplication (Prop. 3.38), we have

$$\lim_{n \to \infty} f(z_n) - f(z) = 0 \cdot v = 0$$

Thus, by Def. 7.81-(3m), we obtain $\lim_{w \neq z, w \to z} f(w) = f(z)$, which means by Def. 7.81-(1) that f is continuous at z. $\qquad\square$

Proposition 11.8. *Let E be an interval in $\mathbb{R}$ or $E = \Omega$, and $x \in E$. Suppose that f, g are functions $E \to V$ such that $f'(x)$ and $g'(x)$ exist. Then $(f + g)'(x)$ exists, and*

$$(f + g)'(x) = f'(x) + g'(x) \tag{11.2a}$$

If λ is a function $E \to \mathbb{F}$ such that $\lambda'(x)$ exists, then $(\lambda f)'(x)$ exists and satisfies the **Leibniz rule**

$$(\lambda f)'(x) = \lambda'(x) f(x) + \lambda(x) f'(x) \tag{11.2b}$$

Assume moreover that λ does not have value 0. Then

$$\left(\frac{1}{\lambda} \cdot f \right)'(x) = \frac{-\lambda'(x) f(x) + \lambda(x) f'(x)}{\lambda(x)^2} \tag{11.2c}$$

Proof. The first formula is easy. To compute the second one, we choose any sequence (x_n) in $[a, b] \backslash \{x\}$ or $\Omega \backslash \{x\}$ converging to x. Then

$$\frac{\lambda(x_n) f(x_n) - \lambda(x) f(x)}{x_n - x} = \frac{(\lambda(x_n) - \lambda(x))}{x_n - x} \cdot f(x_n) + \lambda(x) \frac{(f(x_n) - f(x))}{x_n - x}$$

which, by Lem. 11.6 and Prop. 3.38 and the continuity of f at x (Prop. 11.7), converges to the RHS of (11.2b). This proves (11.2b), thanks to Lem. 11.6.

The third formula will follow from the second one if we can prove that $1/\lambda$ has derivative $-\frac{\lambda'(x)}{\lambda(x)^2}$ at x. This is not hard: Choose any sequence $x_n \to x$ but $x_n \neq x$. Then

$$\left(\frac{1}{\lambda(x_n)} - \frac{1}{\lambda(x)} \right) \Big/ (x_n - x) = -\frac{\lambda(x_n) - \lambda(x)}{x_n - x} \cdot \frac{1}{\lambda(x_n) \lambda(x)}$$

converges to $-\lambda'(x) \cdot \frac{1}{\lambda(x)^2}$ as $n \to \infty$. Here, we have used the continuity of λ at x and Prop. 3.38 again. □

Example 11.9. The derivative of a constant function is 0. Thus, by Leibniz rule, if λ is a scalar, and if $f'(z)$ exists, then $(\lambda f)'(z) = \lambda \cdot f'(z)$.

Example 11.10. The identity map $f : z \in \mathbb{C} \mapsto z \in \mathbb{C}$ has derivative $\lim_{w \to z} \frac{w-z}{w-z} = 1$. Thus, by induction and Prop. 11.8, we have $(z^n)' = nz^{n-1}$ if $n \in \mathbb{Z}_+$. If $-n \in \mathbb{Z}_+$, then when $z \neq 0$ we have

$$(z^n)' = (1/z^{-n})' = -\frac{(z^{-n})'}{z^{-2n}} = -\frac{-nz^{-n-1}}{z^{-2n}} = nz^{n-1}$$

We conclude that $(z^n)' = nz^{n-1}$ whenever $n \in \mathbb{N}$, or whenever $n = -1, -2, \ldots$ and $z \neq 0$. The same conclusion holds for the real variable function x^n.

Example 11.11. Let $f : z \in \mathbb{C} \mapsto \overline{z} \in \mathbb{C}$. We claim that for every $z \in \mathbb{C}$, the limit

$$f'(z) = \lim_{w \to z} \frac{\overline{w} - \overline{z}}{w - z} = \lim_{h \to 0} \overline{h}/h \tag{11.3}$$

does not exist with the help of Rem. 7.85: Take $h_n = 1/n$. Then $h_n \to 0$ and $\overline{h_n}/h_n = 1 \to 1$ as $n \to \infty$. Take $h_n = \mathbf{i}/n$. Then $h_n \to 0$ and $\overline{h_n}/h_n = -\mathbf{i}/\mathbf{i} = -1 \to -1$ as $n \to \infty$. So $f'(z)$ does not exist.

Theorem 11.12 (Chain rule). *Let Ω, Γ be nonempty open subsets of $\mathbb{C}$. Assume that $f : \Omega \to \Gamma$ is differentiable at $z \in \Omega$, and that $g : \Gamma \to V$ is differentiable at $f(z)$. Then $g \circ f$ is differentiable at z, and*

$$(g \circ f)'(z) = g'(f(z)) \cdot f'(z) \tag{11.4}$$

The same conclusion holds if Ω is replaced by an interval in $\mathbb{R}$, or if both Ω and Γ are replaced by intervals in $\mathbb{R}$.

Recall Conv. 11.4 for the assumption on the field $\mathbb{F}$.

Proof. Define a function $A : \Gamma \to \mathbb{C}$ by

$$A(\zeta) = \begin{cases} \dfrac{g(\zeta) - g \circ f(z)}{\zeta - f(z)} & \text{if } \zeta \neq f(z) \\ g'(f(z)) & \text{if } \zeta = f(z) \end{cases}$$

Choose any sequence (z_n) in $\Omega \backslash \{z\}$ converging to z. Then

$$\frac{g \circ f(z_n) - g \circ f(z)}{z_n - z} = A(f(z_n)) \cdot \frac{f(z_n) - f(z)}{z_n - z}$$

By Prop. 11.7, f is continuous at z. So $\lim_n f(z_n) = f(z)$. Thus, by Lem. 11.6-(3), the above expression converges to $g'(f(z))f'(z)$ as $n \to \infty$. $\square$

Proposition 11.13. *Let Ω, Γ be nonempty open subsets of $\mathbb{C}$. Let $f : \Omega \to \Gamma$ be a bijection. Let $z \in \Omega$. Suppose that $f'(z)$ exists and $f'(z) \neq 0$. Suppose also that $f^{-1} : \Gamma \to \Omega$ is continuous at $f(z)$. Then f^{-1} is differentiable at $f(z)$, and*

$$(f^{-1})'(f(z)) = \frac{1}{f'(z)} \tag{11.5}$$

The same conclusion holds if Ω and Γ are replaced by intervals of $\mathbb{R}$.

Proof. Choose any sequence (w_n) in $\Gamma \backslash \{f(z)\}$ converging to $f(z)$. Then, as $n \to \infty$, we have $f^{-1}(w_n) \to f^{-1}(f(z)) = z$ since f^{-1} is continuous at $f(z)$, and hence

$$\frac{f^{-1}(w_n) - f^{-1}(f(z))}{w_n - f(z)} = \frac{f^{-1}(w_n) - z}{f(f^{-1}(w_n)) - f(z)} \tag{11.6}$$

converges to $1/f'(z)$ by Lem. 11.6 and the continuity of the map $\zeta \in \mathbb{C}^\times \mapsto 1/\zeta \in \mathbb{C}$. This proves $(f^{-1})'(f(z)) = 1/f'(z)$, thanks to Lem. 11.6. $\square$

11.2 Rolle's and Lagrange's mean value theorems (MVT)

Fix a Banach space V over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. Assume $-\infty < a < b < +\infty$.

11.2.1 MVTs

Definition 11.14. Let $f : X \to \mathbb{R}$ where X is a topological space. We say that f has a **local maximum** (resp. **local minimum**) at $x \in X$, if there exists $U \in \mathrm{Nbh}_X(x)$ such that $f|_U$ attains its maximum (resp. minimum) at x. The word "**local extremum**" refers to either local maximum or local minimum.

You must know that derivatives can be used to find the monotonicity of real-valued real variable functions. Here is the precise statement:

Proposition 11.15. *Assume that $f : [a, b] \to \mathbb{R}$ is differentiable at $x \in [a, b]$, and $f'(x) > 0$. Then there exists $\delta > 0$ such that for any $y \in (x - \delta, x + \delta) \cap [a, b]$, we have*

$$y > x \Rightarrow f(y) > f(x) \qquad y < x \Rightarrow f(y) < f(x) \tag{11.7}$$

We leave it to the readers the find the analogous statement for the case $f'(x) < 0$.

Proof. Let $A = f'(x) > 0$. Then there exists $\delta > 0$ such that for all $y \in (x - \delta, x + \delta) \cap [a, b]$ not equal to x, we have $\left| \dfrac{f(y) - f(x)}{y - x} - A \right| < \dfrac{A}{2}$, and hence $\dfrac{f(y) - f(x)}{y - x} > \dfrac{A}{2}$. This proves (11.7). $\square$

Corollary 11.16. *Assume that $f : [a, b] \to \mathbb{R}$ has a local extremum at $x \in (a, b)$. Assume that $f'(x)$ exists. Then $f'(x) = 0$.*

Proof. If $f'(x)$ is a non-zero number, then either $f'(x) > 0$ or $(-f)'(x) > 0$. In either case, Prop. 11.15 indicates that f cannot have a local extremum at x. $\square$

From Prop. 11.15, it is clear that if $f' > 0$ on (a, b), then f is strictly increasing. However, to prove that if $f' \geqslant 0$ then f is increasing, we need more preparation.

Lemma 11.17 (Rolle's MVT). *Suppose that $f \in C([a, b], \mathbb{R})$ is differentiable on (a, b). Suppose moreover that $f(a) = f(b)$. Then there exists $x \in (a, b)$ such that $f'(x) = 0$.*

Proof. If f is constant than $f' = 0$. Suppose that f is not constant. Then there is $x \in (a, b)$ at which f attains its maximum (if $f(t) > f(a)$ for some $t \in (a, b)$) or minimum (if $f(t) < f(b)$ for some $t \in (a, b)$). So $f'(x) = 0$ by Cor. 11.16. $\square$

Example 11.18. Rolle's MVT does not hold for vector-valued functions. Especially, it does hot hold for functions to $\mathbb{C} \simeq \mathbb{R}^2$. Consider $f : [0, 2\pi] \to \mathbb{C}$ defined by $f(t) = e^{\mathbf{i}t} = \cos t + \mathbf{i}\sin t$. We assume that $\sin' = \cos$ and $\cos' = -\sin$ are proved. Then $f(0) = f(2\pi) = 1$, whereas $f'(t) = -\sin t + \mathbf{i}\cos t$ is never zero.

Theorem 11.19 (Lagrange's MVT). *Suppose that $f \in C([a,b], \mathbb{R})$ is differentiable on (a,b). Then there is $x \in (a,b)$ such that*

$$f'(x) = \frac{f(b) - f(a)}{b - a} \tag{11.8}$$

Proof. The case $f(a) = f(b)$ is just Rolle's MVT. When $f(a) \neq f(b)$, we can "shift the function f vertically" so that its two end points have the same height. Technically, we consider $g(x) = f(x) - kx$ where $k = \frac{f(b) - f(a)}{b - a}$ is the slope of the interval from $(a, f(a))$ to $(b, f(b))$. Then $g(a) = g(b)$. By Rolle's MVT, there is $x \in (a,b)$ such that $g'(x) = 0$, i.e. $f'(x) - k = 0$. □

11.2.2 Applications of MVTs

Corollary 11.20. *Assume that $f \in C([a,b], \mathbb{R})$ is differentiable on (a,b). Then*

$$f' \geqslant 0 \text{ on } (a,b) \qquad \Longleftrightarrow \qquad f \text{ is increasing on } [a,b] \tag{11.9}$$

Moreover, if $f' > 0$ on (a,b), then f is strictly increasing on $[a,b]$.

Proof. If $f'(x) < 0$ for some $x \in (a,b)$, then Prop. 11.15 implies that x has a neighborhood on which f is not increasing. Conversely, suppose that $f'(x) \geqslant 0$ for all $x \in (a,b)$. Choose $x, y \in [a,b]$ satisfying $a \leqslant x < y \leqslant b$. Then by Lagrange's MVT, there is $z \in (x,y)$ such that $f(y) - f(x) = f'(z)(y - x) \geqslant 0$. So f is increasing on $[a,b]$. We have finished proving (11.9).

Suppose that $f' > 0$ on (a,b). Then by Prop. 11.15, f is strictly increasing on (a,b). If $a < x < b$, then there is $\varepsilon > 0$ such that $f(a + 1/n) \leqslant f(x) - \varepsilon$ for sufficiently large n. Since f is continuous at a, by taking $\lim_{n \to \infty}$ we get $f(a) \leqslant f(x) - \varepsilon$ and hence $f(a) < f(x)$. A similar argument proves $f(b) > f(x)$. So f is strictly increasing on $[a,b]$. □

Corollary 11.21. *Suppose that $f \in C([a,b], \mathbb{R})$ is differentiable on (a,b), and $f'(x) \neq 0$ for all $x \in (a,b)$. Then f is strictly monotonic (i.e. strictly increasing or strictly decreasing). In particular, by Cor. 11.20, either $f' \geqslant 0$ on (a,b) or $f' \leqslant 0$ on (a,b).*

Proof. Rolle's MVT implies $f(b) - f(a) \neq 0$, and similarly, $f(y) - f(x) \neq 0$ whenever $a \leqslant x < y \leqslant b$. Therefore f is injective. The strict monotonicity of f follows from the following general fact: □

Proposition 11.22. *Let $-\infty \leqslant a < b \leqslant +\infty$, and let $f : [a,b] \to \overline{\mathbb{R}}$ be a continuous injective function. Then f is strictly monotonic.*

Proof. Choose a strictly increasing homeophism $\varphi : [0,1] \to [a,b]$. It suffices to prove that $g = f \circ \varphi : [0,1] \to \overline{\mathbb{R}}$ is strictly monotonic. Assume for simplicity that $g(0) \leqslant g(1)$. Let us show that g is increasing. Then the injectivity implies that g is strictly increasing.

We claim that for every $x \in (0, 1)$ we have $g(0) \leqslant g(x) \leqslant g(1)$. Suppose the claim is true. Then for every $0 < x < y < 1$ we have $g(0) \leqslant g(y) \leqslant g(1)$. Applying the claim to the interval $[0, y]$ shows $g(0) \leqslant g(x) \leqslant g(y)$. So f is increasing.

Let us prove the claim. Suppose the claim is not true. Then there is $x \in (0, 1)$ such that either $g(0) \leqslant g(1) < g(x)$ or $g(x) < g(0) \leqslant g(1)$. In the first case, by intermediate value theorem (applied to $f|_{[0,x]}$), there is $p \in [0, x]$ such that $g(p) = g(1)$. So g is not injective since $p < 1$. This is impossible. Similarly, the second case is also impossible. $\qquad\square$

Besides monotonicity, the uniqueness of antiderivatives is another classical application of MVT.

Definition 11.23. An **antiderivative** of a function $f : E \to V$ is a differentiable function $g : E \to V$ satisfying $g' = f$ on E.

The renowned fact that any two antiderivatives g_1, g_2 of $f : [a, b] \to V$ differ by a constant is immediate from the following fact (applied to $g_1 - g_2$):

Corollary 11.24. *Suppose that $f \in C([a, b], V)$ is differentiable on (a, b) and satisfies $f' = 0$ on (a, b). Then f is a constant function.*

Proof for $V = \mathbb{F}^N$. Since $\mathbb{C}^N \simeq \mathbb{R}^{2N}$, it suffices to prove the case $V = \mathbb{R}^N$. Since a sequence in $\mathbb{R}^N$ converges to a vector iff each component of the sequence converges to the corresponding component of the vector, it suffices to prove the case $f : [a, b] \to \mathbb{R}$.

Choose any $x \in (a, b]$. By Lagrange's MVT, there exists $y \in (a, x)$ such that $0 = f'(y) = \dfrac{f(x) - f(a)}{x - a}$. This proves $f(x) = f(a)$ for all $x \in (a, b]$. $\qquad\square$

Now we discuss the general case.

Remark 11.25. When V is not necessarily finite-dimensional, the method of reducing Cor. 11.24 to the case of scalar-valued functions is quite subtle: How should we understand the "components" of an element v in the Banach space V? In fact, in the general case, we should view

$$\langle \varphi, v \rangle \equiv \langle v, \varphi \rangle := \varphi(v)$$

as a component of v, where φ is an element inside the **dual (Banach) space** of V, defined by

$$V^* = \mathfrak{L}(V, \mathbb{F}) \tag{11.10}$$

An element φ in V^* is called a **bounded linear functional** on V. (In general, a **linear functional** on a vectors space W over a field $\mathbb{K}$ is simply a linear map $W \to \mathbb{K}$.)

The vector space V^*, equipped with the operator norm, is indeed a Banach space. (This is why we call V^* the dual *Banach* space.) For the moment we do not need this fact. And we will discuss this topic in a later chapter.

Remark 11.26. In the future, we will prove that V^* separates points of V. (Recall Def. 8.44.) By linearity, this is equivalent to saying that for any $v \in V$ we have

$$v = 0 \qquad \Longleftrightarrow \qquad \langle \varphi, v \rangle = 0 \text{ for all } \varphi \in V^* \tag{11.11}$$

In fact, in the future we will prove the famous **Hahn-Banach extension theorem**, which implies (cf. Cor. 16.6) that if $v \in V$ then

$$\exists \varphi \in V^* \backslash \{0\} \qquad \text{such that} \qquad \langle \varphi, v \rangle = \|\varphi\| \cdot \|v\| \tag{11.12}$$

where $\|\varphi\|$ is the operator norm of φ. (Note that, by Rem. 10.24, we have $|\langle \varphi, v \rangle| \leqslant \|\varphi\| \cdot \|v\|$ for all $v \in V, \varphi \in V^*$. It is nontrivial that "$\leqslant$" can be "$=$" for some $\varphi \neq 0$.)

Proof of Cor. 11.24 assuming Hahn-Banach. Let us prove that $f(x) = f(a)$ for any $x \in [a, b]$. Since V^* separates points of V, it suffices to choose an arbitrary $\varphi \in V^*$ and prove that $\varphi \circ f(x) = \varphi \circ f(a)$. Indeed, since φ is continuous, for each sequence (x_n) in $[a, b] \backslash \{x\}$ converging to x, in view of Lem. 11.6 we have

$$\lim_{n \to \infty} \frac{\varphi \circ f(x_n) - \varphi \circ f(x)}{x_n - x} = \varphi \left(\lim_{n \to \infty} \frac{f(x_n) - f(x)}{x_n - x} \right) = \varphi(f'(x)) = \varphi(0) = 0$$

if $a < x < b$. Thus $(\varphi \circ f)' = 0$ on (a, b). Therefore, by the finite-dimensional version of Cor. 11.24, we have that $\varphi \circ f$ is constant on $[a, b]$. □

Hahn-Banach theorem is extremely useful for reducing a problem about vector-valued functions to one about scalar-valued functions. In this course, we will also use Hahn-Banach to prove another fun fact: every Banach space over $\mathbb{F}$ is isormorphic to a closed linear subspace of $C(X, \mathbb{F})$ where X is a compact Hausdorff space. However, in the following section we would like to give an elementary proof of Cor. 11.24 without using Hahn-Banach.

11.3   Finite-increment theorem

Fix a Banach space V over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. Assume $-\infty < a < b < +\infty$.
Cor. 11.24 follows immediately from the following theorem by taking $g = 0$.

Theorem 11.27. *Suppose that $f \in C([a, b], V)$ and $g \in C([a, b], \mathbb{R})$ are differentiable on (a, b). Assume that for every $x \in (a, b)$ we have $\|f'(x)\| \leqslant g'(x)$. Then*

$$\|f(b) - f(a)\| \leqslant g(b) - g(a) \tag{11.13}$$

Proof. By continuity, it suffices to prove $\|f(\beta) - f(\alpha)\| \leqslant g(\beta) - g(\alpha)$ for all α, β satisfying $a < \alpha < \beta < b$. Therefore, by replacing $[a, b]$ by $[\alpha, \beta]$, it suffices to assume that f, g are differentiable on $[a, b]$.

Step 1. For each $\varepsilon > 0$, consider the condition on $x \in [a, b]$:

$$\|f(x) - f(a)\| \leqslant g(x) - g(a) + \varepsilon(x - a) \tag{11.14}$$

The set

$$E_\varepsilon = \big\{x \in [a, b] : x \text{ satisfies (11.14)}\big\}$$

is nonempty because $a \in E_\varepsilon$. One checks easily that E_ε is a closed subset of $[a, b]$: This is because, for example, E_ε is the inverse image of the closed subset $(-\infty, 0]$ of $\mathbb{R}$ under the map

$$[a, b] \to \mathbb{R} \qquad x \mapsto \|f(x) - f(a)\| - g(x) + g(a) - \varepsilon(x - a) \tag{11.15}$$

We now fix $x = \sup E_\varepsilon$. Then $x \in E_\varepsilon$ because E_ε is closed. We shall prove that $x = b$. Then the fact that $b \in E_\varepsilon$ for all $\varepsilon > 0$ proves (11.13).

Step 2. Suppose that $x \neq b$. Then $a \leqslant x < b$. We shall prove that there exists $y \in (x, b)$ such that

$$\|f(y) - f(x)\| \leqslant g(y) - g(x) + \varepsilon(y - x) \tag{11.16}$$

Add this inequality to (11.14) (which holds because $x \in E_\varepsilon$). Then by triangle inequality, we obtain $y \in E_\varepsilon$, contradicting $x = \sup E_\varepsilon$.

Let us prove the existence of such y. For each $y \in (x, b)$, define $v(y) \in V, \lambda(y) \in \mathbb{R}$ such that

$$f(y) - f(x) = f'(x)(y - x) + v(y)(y - x)$$
$$g(y) - g(x) = g'(x)(y - x) + \lambda(y)(y - x)$$

The definition of $f'(x)$ and $g'(x)$ implies that $v(y) \to 0$ and $\lambda(y) \to 0$ as $y \to x$. Therefore, there exists $y \in (x, b)$ such that $\|v(u)\| < \varepsilon/2$ and $|\lambda(y)| < \varepsilon/2$. Thus

$$\|f(y) - f(x)\| - (g(y) - g(x))$$
$$\leqslant (\|f'(x)\| - g'(x))(y - x) + (\|v(y)\| - \lambda(y)) \cdot (y - x)$$
$$\leqslant 0 \cdot (y - x) + \Big(\frac{\varepsilon}{2} + \frac{\varepsilon}{2}\Big) \cdot (y - x) = \varepsilon(y - x)$$

This proves (11.16). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Remark 11.28. If we apply Thm. 11.27 to the special case that $f = 0$, we see that if $g \in C([a, b], \mathbb{R})$ satisfies $g' \geqslant 0$ on (a, b), then g is increasing. This gives another proof of (11.9) besides the one via Lagrange's MVT.

An important special case of Thm. 11.27 is:

Corollary 11.29 (**Finite-increment theorem**). *Suppose that $f \in C([a, b], V)$ is differentiable on (a, b), and that there exists $M \in \mathbb{R}_{\geqslant 0}$ such that $\|f'(x)\| \leqslant M$ for all $x \in (a, b)$. Then*

$$\|f(b) - f(a)\| \leqslant M|b - a| \tag{11.17}$$

Proof. Choose $g(x) = Mx$ in Thm. 11.27. □

It is fairly easy to prove finite-increment theorem for complex-variable functions.

Definition 11.30. A subset E of a real vector space is called **convex**, if for every $x, y \in E$, the interval

$$[x, y] = \{tx + (1 - t)y : t \in [0, 1]\} \tag{11.18}$$

is a subset of E.

When talking about convex subsets of $\mathbb{C}$, we view $\mathbb{C}$ as $\mathbb{R}^2$. Then, for example, all open disks in $\mathbb{C}$ are convex.

Corollary 11.31 (**Finite-increment theorem**). *Assume $\mathbb{F} = \mathbb{C}$. Let Ω be a nonempty open convex subset of $\mathbb{C}$. Let $f : \Omega \to V$ be differentiable on Ω. Choose any $z_1, z_2 \in \Omega$. Suppose that there exists $M \in \mathbb{R}_{\geqslant 0}$ such that $\|f'(z)\| \leqslant M$ for all z in the interval $[z_1, z_2]$. Then*

$$\|f(z_2) - f(z_1)\| \leqslant M|z_2 - z_1| \tag{11.19}$$

Proof. Define $\gamma : [0, 1] \to \mathbb{C}$ by $\gamma(t) = (1 - t)z_1 + tz_2$. Then $\gamma([0, 1]) \subset \Omega$ because Ω is convex. By chain rule (Thm. 11.12), we have

$$(f \circ \gamma)'(t) = f'(\gamma(t)) \cdot \gamma'(t) = (z_2 - z_1)f'(\gamma(t))$$

whose norm is bounded by $M|z_2 - z_1|$. Applying Cor. 11.29 to $f \circ \gamma$ finishes the proof. □

Example 11.32. Let X be an interval in $\mathbb{R}$, or let $X = \Omega$ where Ω is convex. Let $\mathscr{A}$ be the set of differentiable functions $f : \Omega \to V$ satisfying $\|f'\|_{l^\infty} \leqslant M$. Then elements of $\mathscr{A}$ have Lipschitz constant M by Finite-increment theorems. So $\mathscr{A}$ is equicontinuous.

11.4 Commutativity of derivatives and limits

Fix a Banach space V over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. Let Ω be a nonempty open subset of $\mathbb{C}$. Recall Conv. 11.4. Recall Def. 8.42 for the meaning of locally uniform convergence.

11.4.1 Main theorem

Theorem 11.33. *Let X be Ω or an interval in $\mathbb{R}$. Let $(f_\alpha)_{\alpha \in I}$ be a net of differentiable functions $X \to V$. Suppose that the following are true:*

(a) *The net $(f_\alpha)_{\alpha \in I}$ converges pointwise to some $f : X \to V$.*

(b) *The net $(f'_\alpha)_{\alpha \in I}$ converges uniformly to some $g : X \to V$.*

Then f is differentiable on X, and $f' = g$.

Proof. We prove the case $X = \Omega$. The other case is similar. Choose any $z \in \Omega$. By shrinking Ω, we assume that Ω is an open disk centered at z. We know $\lim_{w \to z} \frac{f_\alpha(w) - f_\alpha(z)}{w - z}$ converges to $f'_\alpha(z)$ for each α. Therefore, if we can show that $\frac{f_\alpha(w) - f_\alpha(z)}{w - z}$ converges uniformly (over all $w \in \Omega \backslash \{z\}$) under $\lim_\alpha$, then it must converge uniformly to $\frac{f(w) - f(z)}{w - z}$ since it converges pointwise to $\frac{f(w) - f(z)}{w - z}$ by (a). Then, by Moore-Osgood Thm. 9.28, we have

$$\lim_{w \to z} \lim_\alpha \frac{f_\alpha(w) - f_\alpha(z)}{w - z} = \lim_\alpha \lim_{w \to z} \frac{f_\alpha(w) - f_\alpha(z)}{w - z} = \lim_\alpha f'_\alpha(z) = g(z)$$

finishing the proof.

To prove the uniform convergence, by the Cauchy condition on $V^{\Omega \backslash \{z\}}$ (equipped with a complete uniform convergence metric as in Exp. 7.77), it suffices to prove that

$$\sup_{w \in \Omega \backslash \{z\}} \left\| \frac{f_\alpha(w) - f_\alpha(z)}{w - z} - \frac{f_\beta(w) - f_\beta(z)}{w - z} \right\| \tag{11.20}$$

converges to 0 under $\lim_{\alpha, \beta \in I}$. Applying Cor. 11.31 to the function $f_\alpha - f_\beta$, we see

$$\| f_\alpha(w) - f_\beta(w) - f_\alpha(z) + f_\beta(z) \| \leqslant |w - z| \cdot \sup_{\zeta \in [z, w]} \| f'_\alpha(\zeta) - f'_\beta(\zeta) \|$$

$$\leqslant |w - z| \cdot \| f'_\alpha - f'_\beta \|_{l^\infty(\Omega, V)} \tag{11.21}$$

Thus, we have

$$(11.20) \leqslant \| f'_\alpha - f'_\beta \|_{l^\infty(\Omega, V)}$$

where the RHS converges to 0 under $\lim_{\alpha, \beta}$ due to (b). This finishes the proof. $\square$

We didn't assume the uniform convergence of (f_α) in Thm. 11.33 because it is often redundant:

Lemma 11.34. *Assume that either X is a bounded interval in $\mathbb{R}$, or $X = \Omega$ where Ω is assumed to be bounded and convex. Then under the assumptions in Thm. 11.33, the net $(f_\alpha)_{\alpha \in I}$ converges uniformly to f.*

Proof. We already know that (f_α) converges pointwise to f. In fact, we shall only use the fact that $\lim_\alpha f_\alpha(z) = f(z)$ for some $z \in X$. Let z be such a point. Motivated by the proof of Thm. 11.33, let us prove the Cauchy condition that

$$\lim_{\alpha,\beta\in I} \sup_{w\in\Omega} \|f_\alpha(w) - f_\beta(w) - f_\alpha(z) + f_\beta(z)\| = 0$$

Then $(f_\alpha - f_\alpha(z))$ converges uniformly, and hence (f_α) converges uniformly. Indeed, the Cauchy condition follows easily from (11.21), in which $\sup_{w\in\Omega} |w - z|$ is a finite number because X is bounded. □

Thus, whether or not Ω satisfies boundedness and convexity, the net (f_α) must converge locally uniformly to f. Knowing the locally uniform convergence is often enough for applications. And here is another proof of this fact:

Another proof *that (f_α) converges locally uniformly to f.* We consider the case $X = \Omega$; the other case is similar. For each $z \in X$, choose a convex precompact $U \in \mathrm{Nbh}_X(x)$. (Namely, $\mathrm{Cl}_X(U)$ is compact.) By (b) of Thm. 11.33, there is $\mu \in I$ such that $\sup_{x\in\overline{U}} \|f'_\alpha(x) - f'_\mu(x)\| \leqslant 1$ for all $\alpha \geqslant \mu$. Let $h_\alpha = f_\alpha - f_\mu$, and replace I by $I_{\geqslant\mu}$. By finite-increment Thm. 11.31, for each $x, y \in U$ we have

$$\|h_\alpha(x) - h_\alpha(y)\| \leqslant \|x - y\|$$

i.e., $(h_\alpha|_U)_{\alpha\in I}$ has uniform Lipschitz constant 1, and hence is an equicontinuous set of functions. Choose a closed ball B centered at z such that $B \subset U$. Then, since $(h_\alpha|_B)_{\alpha\in I}$ is equicontinuous and converges pointwise to $f - f_\mu$ (on B), by Cor. 9.26, $(f_\alpha - f_\mu)_{\alpha\in I}$ converges uniformly on B to $f - f_\mu$. Thus f_α converges uniformly on B to f. □

11.4.2   An interpretation of Thm. 11.33 in terms of Banach spaces

We give a more concise way of understanding the two conditions in Thm. 11.33.

Corollary 11.35. *Let X be Ω or an interval in $\mathbb{R}$. Define*

$$l^{1,\infty}(X, V) = \{f \in V^X : f \text{ is differentiable on } X \text{ and } \|f\|_{l^{1,\infty}} < +\infty\}$$

where $\|\cdot\|_{l^{1,\infty}} = \|\cdot\|_{1,\infty}$ is defined by

$$\|f\|_{1,\infty} = \|f\|_{l^\infty} + \|f'\|_{l^\infty} = \sup_{x\in X} \|f(x)\| + \sup_{x\in X} \|f'(x)\|$$

Then $l^{1,\infty}(X, V)$ is a Banach space with norm $\|\cdot\|_{1,\infty}$.

Proof. It is a routine check that $\|f\|_{l^{1,\infty}}$ defines a norm on $l^{1,\infty}(X,V)$. We now prove that $l^{1,\infty}(X,V)$ is complete. Let (f_n) be a Cauchy sequence in $l^{1,\infty}(X,V)$. So (f_n) and (f_n') are Cauchy sequences in l^∞, converging uniformly to $f, g \in V^X$ respectively. By Lem. 11.33, f is differentiable, and $f' = g$. (In particular, $\|f'\|_\infty < +\infty$.) So $f_n' \rightrightarrows f'$. Thus

$$\|f_n - f\|_{1,\infty} = \|f_n - f\|_\infty + \|f_n' - f'\|_\infty \to 0$$

$\square$

Remark 11.36. Thm. 11.33 and Cor. 11.35 are almost equivalent. We have proved Cor. 11.35 using Thm. 11.33. But we can also prove a slightly weaker version of Thm. 11.33 using Cor. 11.35 as follows: In the setting of Thm. 11.33, assume that each $f_\alpha : X \to V$ is differentiable, and that

$$f_\alpha \rightrightarrows f \qquad f_\alpha' \rightrightarrows g \tag{11.22}$$

where $f, g \in l^\infty(X,V)$. Then Cor. 11.35 implies that f is differentiable and $f' = g$.

$\star$ *Proof.* By (11.22), there is $\beta \in I$ such that $\|f_\alpha - f\|_\infty < +\infty$ and $\|f_\alpha' - g\| < +\infty$ for all $\alpha \geqslant \beta$. Thus, by replacing I with I_β, we assume that f_α and f_α' are bounded on X. So $f_\alpha \in l^{1,\infty}(X,V)$.

By (11.22), both (f_α) and (f_α') converge in $l^\infty(X,V)$. So they are Cauchy nets under the l^∞-norm. So (f_α) is a Cauchy net in $l^{1,\infty}(X,V)$. By Cor. 11.35, (f_α) converges to $\widetilde{f} \in l^{1,\infty}(X,V)$ under the $l^{1,\infty}$ norm. In particular, $f_\alpha \rightrightarrows \widetilde{f}$ and $f_\alpha' \rightrightarrows \widetilde{f}'$. Since, by assumption, we have $f_\alpha \rightrightarrows f$ and $f_\alpha' \rightrightarrows g$, we therefore get $f = \widetilde{f}$ and $g = \widetilde{f}'$. $\square$

It is not surprising that Thm. 11.33 can be rephrased in terms of the completeness of a normed vector space. After all, our proof of Thm. 11.33 uses Cauchy nets in an essential way. In the future, we will use the completeness of $l^{1,\infty}$ to understand the commutativity of derivatives and other limit processes.

11.5 Derivatives of power series

This section is a continuation of the previous one. We shall use Thm. 11.33 to compute derivatives of power series.

11.5.1 General result

Corollary 11.37. *Assume that V is over $\mathbb{C}$. Let $f(z) = \sum_{n=0}^\infty v_n z^n$ be a power series in V. Assume that its radius of convergence is $R > 0$. Then f is differentiable on $B_{\mathbb{C}}(0,R)$ and*

$$f'(z) = \sum_{n=0}^\infty n v_n z^{n-1}$$

Note that since $\lim_n \sqrt[n]{n} = 1$, $(\sqrt[n]{\|v_n\|})_{n \in \mathbb{Z}_+}$ and $(\sqrt[n]{n\|v_n\|})_{n \in \mathbb{Z}_+}$ have the same cluster points. Therefore

$$\limsup_{n \to \infty} \sqrt[n]{\|v_n\|} = \limsup_{n \to \infty} \sqrt[n]{n\|v_n\|} \tag{11.23}$$

So $\sum v_n z^n$ and $\sum_{n=0}^{\infty} n v_n z^{n-1}$ have the same radius of convergence.

First proof. For each $n \in \mathbb{N}_+$, let $g_n(z) = v_0 + v_1 z + \cdots + v_n z^n$. Then by Thm. 4.27, for each $0 < r < R$, the sequences (g_n) (resp. (g_n')) converges uniformly to f (resp. converges uniformly to $h(z) = \sum_0^{\infty} n v_n z^{n-1}$) on $B_{\mathbb{C}}(0, r)$. Therefore, by Thm. 11.33, we have $f'(z) = h(z)$ for all $z \in B_{\mathbb{C}}(0, r)$, and hence all $z \in B_{\mathbb{C}}(0, R)$ since r is arbitrary. $\qquad\square$

Second proof. Choose any $0 < r < R$, and let $X_r = B_{\mathbb{C}}(0, r)$. Consider $\sum v_n z^n$ as a series in the Banach space $l^{1,\infty}(X_r, \mathbb{C})$ (cf. Cor. 11.35). The norm of each term is $\|v_n\| + (n+1)\|v_{n+1}\|$. So the radius of convergence is R. Thus $\sum v_n z^n$ converges in $l^{1,\infty}(X_r, \mathbb{C})$ to some $g \in l^{1,\infty}(X_r, \mathbb{C})$. This means that, on X_r, $\sum v_n z^n$ converges uniformly to g (and hence $f = g$ on X_r) and $\sum n v_n z^{n-1}$ converges uniformly to g'. So $f'(z) = g'(z) = \sum n v_n z^{n-1}$ for all $z \in X_r$. $\qquad\square$

11.5.2 Examples

Example 11.38. By Cor. 11.37, the function $\exp : \mathbb{C} \to \mathbb{C}$ is differentiable, and

$$\frac{d}{dz} e^z = e^z$$

Thus, if $\gamma : [a, b] \to \mathbb{C}$ is differentiable, then by chain rule, $\exp \circ \gamma$ is differentiable on $[a, b]$, and

$$\frac{d}{dt} e^{\gamma(t)} = e^{\gamma(t)} \cdot \frac{d}{dt} \gamma(t)$$

For example:

$$(e^{\alpha t})' = \alpha e^{\alpha t} \qquad (e^{t^2 + \mathbf{i} t})' = (2t + \mathbf{i}) e^{t^2 + \mathbf{i} t}$$

Example 11.39. Define functions $\sin : \mathbb{C} \to \mathbb{C}$ and $\cos : \mathbb{C} \to \mathbb{C}$ by

$$\cos z = \frac{e^{\mathbf{i} z} + e^{-\mathbf{i} z}}{2} \qquad \sin z = \frac{e^{\mathbf{i} z} - e^{-\mathbf{i} z}}{2\mathbf{i}}$$

It follows from $e^z e^w = e^{z+w}$ that

$$\cos(z + w) = \cos z \cos w - \sin z \sin w \qquad \sin(z + w) = \sin z \cos w + \cos z \sin w \tag{11.24}$$

By chain rule, we have $(e^{\alpha z})' = \alpha e^{\alpha z}$. So

$$\sin' z = \cos z \qquad \cos' z = -\sin z$$

Remark 11.40. From $e^z = \sum_{n \in \mathbb{N}} z^n / n!$, it is clear that the complex conjugate of e^z is

$$\overline{e^z} = \overline{\sum_{n=0}^{\infty} \frac{z^n}{n!}} = \sum_{n=0}^{\infty} \frac{\overline{z}^n}{n!} = e^{\overline{z}}$$

Here, we have exchanged the order of conjugate and infinite sum, because the function $z \in \mathbb{C} \mapsto \overline{z}$ is continuous. Thus, if $t \in \mathbb{R}$, then

$$\overline{e^{\mathbf{i}t}} = e^{-\mathbf{i}t} \qquad |e^{\mathbf{i}t}|^2 = \overline{e^{\mathbf{i}t}} e^{\mathbf{i}t} = e^{-\mathbf{i}t} e^{\mathbf{i}t} = 1$$

It follows that

$$\cos t = \mathrm{Re}(e^{\mathbf{i}t}) \qquad \sin t = \mathrm{Im}(e^{\mathbf{i}t})$$
$$(\cos t)^2 + (\sin t)^2 = (\mathrm{Re}(e^{\mathbf{i}t}))^2 + (\mathrm{Im}(e^{\mathbf{i}t}))^2 = |e^{\mathbf{i}t}|^2 = 1$$

It also follows from $|e^{\mathbf{i}t}| = 1$ that if $a, b \in \mathbb{R}$ then, by $e^{a+b\mathbf{i}} = e^a e^{b\mathbf{i}}$, we have

$$e^{a+b\mathbf{i}} = e^a \qquad \in \mathbb{R}_{>0}$$

Example 11.41. By Prop. 11.13, the function $\log : \mathbb{R}_{>0} \to \mathbb{R}$ is differentiable, and

$$(\log x)' = \frac{1}{x}$$

Example 11.42. We have

$$\lim_{t \to 0} \frac{\log(1 + t)}{t} = (\log x)'|_{x=1} = 1$$

and hence $\lim_{x \to +\infty} x \log(1 + 1/x) = 1$ by Def. 7.81-(3) (since $\lim_{x \to +\infty} 1/x = 0$ as a net limit). Taking exponential, and using the continuity of $\exp$ at 1, we get

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

Example 11.43. If $a > 0$ and $z \in \mathbb{C}$, recalling $a^z = e^{z \log a}$, we use chain rule to find

$$\frac{d}{dz} a^z = a^z \log a$$

Similarly, if $\alpha \in \mathbb{C}$ and $x > 0$, then the chain rule gives the derivative of x^α:

$$\frac{d}{dx} x^\alpha = \alpha \cdot x^{\alpha - 1}$$

Example 11.44. Compute $\displaystyle\sum_{n=0}^{\infty} \frac{n^2}{3^n}$

Proof. The series $f(z) = \sum_{n=0}^{\infty} \dfrac{z^n}{3^n}$ has radius of convergence 3. When $|z| < 3$, it converges absolutely to $(1 - \frac{z}{3})^{-1} = 3(3 - z)^{-1}$. So, by Cor. 11.37, when $|z| < 3$ we have

$$\sum_{n=0}^{\infty} \frac{nz^{n-1}}{3^n} = f'(z) = 3(3 - z)^{-2}$$

$$\sum_{n=0}^{\infty} \frac{n(n - 1)z^{n-2}}{3^n} = f''(z) = 6(3 - z)^{-3}$$

So the value of the original series equals

$$f''(1) + f'(1) = \frac{3}{2}$$

$\square$

11.5.3 A proof of (generalized) Leibniz rule

We end this section by giving a fun proof of Leibniz rule for higher derivatives. For simplicity, we consider only scalar-valued functions.

Proposition 11.45 (Leibniz rule). *Let X be either a nonempty interval in $\mathbb{R}$ (resp. a nonempty open subset of $\mathbb{C}$). Let f, g be functions from X to $\mathbb{R}$ (resp. to $\mathbb{C}$). Let $z \in X$. Suppose that $f^{(n)}(z)$ and $g^{(n)}(z)$ exist. Then*

$$(fg)^{(n)}(z) = \sum_{j=0}^{n} \binom{n}{j} f^{(n-j)}(z) g^{(j)}(z) \tag{11.25}$$

The above Leibniz rule is usually proved using the formula

$$\binom{n+1}{j} = \binom{n}{j-1} + \binom{n}{j} \tag{11.26}$$

where $n \in \mathbb{N}$ and $1 \leqslant j \leqslant n$. In the following, we give a proof without using this formula. We need the fact that the function

$$\mathbb{C}^n \to C(\mathbb{R}, \mathbb{C}) \qquad (a_0, \dots, a_{n-1}) \mapsto p(x) = \sum_{j=0}^{n-1} a_j x^j \tag{11.27}$$

is injective: this is because $j! \cdot a_j = f^{(j)}(0)$ by Exp. 11.10.

Proof of Prop. 11.45. By induction on n and by the classical Leibniz rule (Prop. 11.8), we have

$$(fg)^{(n)}(z) = \sum_{j=0}^{n} C_{n,j} \cdot f^{(n-j)}(z)g^{(j)}(z) \tag{11.28}$$

where each $C_{n,j}$ is an integer independent of f and g. (In particular, $C_{n,j}$ is independent of whether the variables are real of complex.) Thus, to determine the value of $C_{n,j}$, we can use some special functions.

We consider $f(x) = e^{sx}$ and $g(x) = e^{tx}$ where $s, t \in \mathbb{R}$. Recall that we have proved that $(e^{\alpha t})' = \alpha e^{\alpha t}$ using chain rule and the derivative formula for exponentials. So (11.28) reads

$$(s+t)^n \cdot e^{(s+t)x} = \sum_{j=0}^{n} C_{n,j} \cdot s^{n-j} t^j \cdot e^{(s+t)x}$$

Taking $x = 0$ gives

$$(s+t)^n = \sum_{j=0}^{n} C_{n,j} \cdot s^{n-j} t^j \tag{11.29}$$

Comparing this with the binomial formula (4.5), and noticing the injectivity of (11.27) (first applied to (11.29) for each fixed t, where (11.29) is viewed as a polynomial of s; then applied to each coefficient before s^{n-j}, which is a polynomial of t), we immediately get $C_{n,j} = \binom{n}{j}$. $\qquad\square$

11.6   Problems and supplementary material

Let Ω be nonempty open subset of $\mathbb{C}$. Let V be a Banach space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. Assume that $\mathbb{F} = \mathbb{C}$ if we take derivatives with respect to complex variables.

Definition 11.46. Let X be either Ω or an interval in $\mathbb{R}$. Define

$$C^n(X, V) = \{f \in V^X : f, f', \dots, f^{(n)} \text{ exist and are continuous}\} \qquad (\text{if } n \in \mathbb{N})$$
$$C^\infty(X, V) = \bigcap_{n \in \mathbb{N}} C^n(X, V)$$

If $X \subset \mathbb{R}$, elements in $C^\infty(X, V)$ are called **smooth functions**.

Problem 11.1. Let X be either Ω or an interval in $\mathbb{R}$. For each $n \in \mathbb{N}$, define

$$l^{n,\infty}(X, V) = \{f \in V^X : f, f', \dots, f^{(n)} \text{ exist, and } \|f\|_{n,\infty} < +\infty\}$$

where $\|f\|_{n,\infty} = \|f\|_{l^{n,\infty}}$ is defined by

$$\|f\|_{n,\infty} = \|f\|_\infty + \|f'\|_\infty + \cdots + \|f^{(n)}\|_\infty$$

(In particular, we understand $l^{0,\infty}$ as l^∞.) Clearly $\|\cdot\|_{n,\infty}$ is a norm. We have proved that $l^{n,\infty}(X, V)$ is complete when $n = 0, 1$.

1. Prove by induction on n that $l^{n,\infty}(X,V)$ is complete for every n.

2. Prove that for each $n \in \mathbb{N}$, $C^n(X,V) \cap l^{n,\infty}(X,V)$ is a closed subset of $l^{n,\infty}(X,V)$ (and hence, is a Banach space by Prop. 3.27). Prove that if X is compact then $C^n(X,V) \subset l^{n,\infty}(X,V)$.

Convention 11.47. Unless otherwise stated, when X is compact, we always choose $l^{n,\infty}$ to be the norm on $C^n(X,V)$.

12 More on derivatives

12.1 Cauchy's MVT and L'Hôpital's rule

The goal of this section is to prove L'Hôpital's rule. For that purpose, we first need to prove Cauchy's MVT.

12.1.1 Main theorems

Theorem 12.1 (Cauchy's MVT). *Let $-\infty < a < b < +\infty$. Let $f, g \in C([a, b], \mathbb{R})$ be differentiable on (a, b). Then there exists $x \in (a, b)$ such that*

$$f'(x)(g(b) - g(a)) = g'(x)(f(b) - f(a))$$

In particular, if $g' \neq 0$ on (a, b), then g is injective (Cor. 11.21), and we can write the above formula as

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

Proof. Cauchy's MVT specializes to Lagrange's MVT if we set $g(x) = x$. Moreover, in the proof of Lagrange's MVT (Thm. 11.19), we applied Rolle's MVT to the function $f(x) - kx$ where k is the slope of a line segment. Motivated by this observation, we consider the function $\psi(x) = f(x) - kg(x)$. If one wants $\psi(a) = \psi(b)$, one then solves that $k = \frac{f(b)-f(a)}{g(b)-g(a)}$. But we would rather consider $(g(b)-g(a))\psi$ in order to avoid the issue that the denominator is possibly zero. So we set

$$h(x) = (g(b) - g(a))f(x) - (f(b) - f(a))g(x)$$

Clearly $h(a) = h(b)$. By Rolle's MVT, there exists $x \in (a, b)$ such that

$$0 = h'(x) = (g(b) - g(a))f'(x) - (f(b) - f(a))g'(x)$$

$\square$

Theorem 12.2 (L'Hôpital's rule). *Let $-\infty \leqslant a < b \leqslant +\infty$. Let $f, g \in C((a, b), \mathbb{R})$ be differentiable on (a, b). Assume that g' is nowhere zero on (a, b). (So g is strictly monotonic, cf. Prop. 11.22.) Assume*

$$\lim_{x \to a} \frac{f'(x)}{g'(x)} = A \qquad \textit{exists in } \overline{\mathbb{R}} \tag{12.1}$$

Assume that one of the following two cases are satisfied:

$$\lim_{x \to a} f(x) = \lim_{x \to a} g(x) = 0 \qquad \text{(Case } \tfrac{0}{0})$$

$$\lim_{x \to a} |g(x)| = +\infty \qquad\qquad \text{(Case } \tfrac{*}{\infty}\text{)}$$

Then we have $\lim_{x \to a} \dfrac{f(x)}{g(x)} = A$. *The same conclusion holds if "$x \to a$" is replaced by* "$x \to b$".

Remark 12.3. Since g is strictly monotonic, there is at most one $x \in (a, b)$ such that $g(x) = 0$. So $\lim_{x \to a} f(x)/g(x)$ means the limit over $x \in (a, b) \backslash g^{-1}(0)$. Alternatively, one can assign an arbitrary value to $f(x)/g(x)$ when $g(x) = 0$, and understand $\lim_{x \to a}$ as a limit over $x \in (a, b)$.

Example 12.4. Compute $\lim_{x \to +\infty} \dfrac{x^n}{e^x}$

Proof. By L'Hôpital's rule in the case $\tfrac{*}{\infty}$, we have

$$\lim_{x \to +\infty} \frac{x^n}{e^x} = \lim_{x \to +\infty} \frac{nx^{n-1}}{e^x} = \lim_{x \to +\infty} \frac{n(n-1)x^{n-2}}{e^x} = \cdots = \lim_{x \to +\infty} \frac{n!}{e^x} = 0$$

where the convergence of the limit is derived from right to left. $\qquad\square$

12.1.2 Proof of L'Hôpital's rule

We divide the proof of L'Hôpital's rule into several steps. Also, we only treat the case $x \to a$, since the other case is similar. In the following, (a, b) means an interval, but not an element in the Cartesian product (which will be written as $a \times b$).

Step 1. We let $\dfrac{f(x) - f(y)}{g(x) - g(y)}$ take value $\dfrac{f'(x)}{g'(x)}$ if $x = y$. In this step, we prove

$$\lim_{x \times y \to a \times a} \frac{f(x) - f(y)}{g(x) - g(y)} = A \tag{12.2}$$

where $x \times y$ is defined on $(a, b)^2 = (a, b) \times (a, b)$. In view of Def. 7.81-(3m), we pick any sequence $x_n \times y_n$ in $(a, b)^2$. By Cauchy's MVT, there is $\xi_n \in [x_n, y_n]$ (if $x_n \leqslant y_n$) or $\xi_n \in [y_n, x_n]$ (if $x_n \geqslant y_n$) such that

$$f'(\xi_n)(g(x_n) - g(y_n)) = g'(\xi_n)(f(x_n) - f(y_n)).$$

So we have

$$\frac{f'(\xi_n)}{g'(\xi_n)} = \frac{f(x_n) - f(y_n)}{g(x_n) - g(y_n)}$$

Since $\lim_n x_n = \lim_n y_n = a$, we clearly have $\lim_n \xi_n = a$. Therefore, by (12.1) (this is the only place where we use (12.1)), we have

$$\lim_{n \to \infty} \frac{f(x_n) - f(y_n)}{g(x_n) - g(y_n)} = \lim_{n \to \infty} \frac{f'(\xi_n)}{g'(\xi_n)} = A$$

This proves (12.2). $\qquad\square$

Step 2. It follows from Def. 7.81-(3) that if (x_n) and (y_k) are sequences in (a, b) converging to a, then

$$\lim_{n,k\to\infty} \frac{f(x_n) - f(y_k)}{g(x_n) - g(y_k)} = A \tag{12.3}$$

In other words, we apply Def. 7.81-(3) to the net $(x_n \times y_k)_{n\times k\in\mathbb{Z}_+^2}$ which replaces the sequence $(x_n \times y_n)_{n\in\mathbb{Z}_+}$ in Step 1. We shall use (12.3) to prove the two cases of L'Hôpital's rule.

Let us prove the case $\frac{0}{0}$. This is the easier case. Choose any sequence (x_n) in (a, b) converging to a. We want to prove that $\lim_n f(x_n)/g(x_n) = A$. So we choose any sequence (y_k) in (a, b) converging to a, and we know that the double limit (12.3) exists. Moreover, if n is fixed, then $\lim_k f(y_k) = \lim_k g(y_k) = 0$ by assumption. Thus

$$\lim_{k\to\infty} \frac{f(x_n) - f(y_k)}{g(x_n) - g(y_k)} = \frac{f(x_n)}{g(x_n)} \tag{12.4}$$

Thus, by Thm. 5.29, when $n \to \infty$, the RHS of the above equation converges to (12.3). This finishes the proof for the case $\frac{0}{0}$. $\qquad\square$

Step 3. Finally, we address the (more difficult) case $\frac{*}{\infty}$. Assume $\lim_{x\to a} |g(x)| = +\infty$. Again, we choose a sequence (x_n) in (a, b) converging to a. To prove $f(x_n)/g(x_n) \to A$, one may want to pick any sequence (y_k) in (a, b), and compute the limit on the LHS of (12.4). Unfortunately, in this case, we do not know whether this limit converges or not: As one can compute, it is equal to $\lim_{k\to\infty} f(y_k)/g(y_k)$, whose convergence is part of the result we need to prove!

It is not hard to address this issue: Since $\overline{\mathbb{R}}$ is (sequentially) compact, by Thm. 3.15, it suffices to prove that any cluster point $B \in \overline{\mathbb{R}}$ of $(f(x_n)/g(x_n))_{n\in\mathbb{Z}_+}$ is equal to A. Thus, we let (y_k) be any subsequence of (x_n) such that $(f(y_k)/g(y_k))_{k\in\mathbb{Z}_+}$ converges to B. Let us prove $A = B$ using the same method as in Step 2. We compute that

$$\frac{f(x_n) - f(y_k)}{g(x_n) - g(y_k)} = \frac{g(y_k)}{g(x_n) - g(y_k)} \cdot \frac{f(x_n) - f(y_k)}{g(y_k)} \tag{12.5}$$

Since $\lim_k |g(y_k)| = +\infty$, we have $\lim_{k\to\infty} C/g(y_k) \to 0$ for any $C \in \mathbb{R}$ independent of k. Therefore, as $k \to \infty$, the first factor on the RHS of (12.5) converges to -1, and the second factor converges to $-B$. It follows that

$$\lim_{n\to\infty} \lim_{k\to\infty} \frac{f(x_n) - f(y_k)}{g(x_n) - g(y_k)} = \lim_{n\to\infty} B = B$$

Therefore $A = B$ by Thm. 5.29. $\qquad\square$

The proof of L'Hôpital's rule is now complete.

Remark 12.5. The above proof can be easily translated into a language without double limits. We consider the case of $\frac{*}{\infty}$ and assume for example that $-\infty < A < +\infty$ and $a \in \mathbb{R}$, and sketch the proof as follows.

By (12.1) and Cauchy' MVT, for every $\varepsilon > 0$ there is $\delta > 0$ such that for all $a < x, y < a + \delta$ (where $x \neq y$) we have

$$A - \varepsilon \leqslant \frac{f(x) - f(y)}{g(x) - g(y)} \leqslant A + \varepsilon$$

Choose any sequence (x_n) in (a, b) converging to a. Let B be any cluster point of $(f(x_n)/g(x_n))_{n \in \mathbb{Z}_+}$ in $\overline{\mathbb{R}}$. We need to prove that $A = B$. Let $(y_k)_{k \in \mathbb{Z}_+}$ be any subsequence of (x_n) such that $\lim_k f(y_k)/g(y_k) = B$. Then there is $N \in \mathbb{N}$ such that for all $n \geqslant N, k \geqslant N$ satisfying $x_n \neq y_k$, we have

$$A - \varepsilon \leqslant \frac{f(x_n) - f(y_k)}{g(x_n) - g(y_k)} \leqslant A + \varepsilon$$

For each $n \geqslant N$, apply $\lim_{k \to \infty}$ to the above inequality, and notice $\lim_k |g(y_k)| = +\infty$. Then we get $A - \varepsilon \leqslant B \leqslant A + \varepsilon$, finishing the proof. $\qquad \square$

12.2 Trigonometric functions and π

In this section, we prove that $\sin$, $\cos$, and π satisfy the properties we learned in high schools. Some of them have already been proved in Subsec. 11.5.2. We leave the proof of the basic properties of the other trigonometric functions to the reader.

Let x be a real variable. Recall that $\sin, \cos : \mathbb{R} \to \mathbb{R}$ are determined by the fact that $e^{ix} = \cos x + i \sin x$. In particular, that $|e^{ix}| = 1$ implies $(\cos x)^2 + (\sin x)^2 = 1$. We have proved that

$$\sin' = \cos \qquad \cos' = -\sin$$

Since $e^{i \cdot 0} = 1$, we have

$$(\sin x)'|_{x=0} = \cos 0 = 1 \qquad (\cos x)'|_{x=0} = -\sin 0 = 0$$

In particular, since $\sin' = \cos$ is strictly positive on a neighborhood of 0, by Cor. 11.20, $\sin$ is strictly increasing on that neighborhood.

We shall define $\frac{\pi}{2}$ to be the smallest positive zero $\cos$. However, we must first prove the existence of this number:

Lemma 12.6. *There exists $x \geqslant 0$ such that $\cos x = 0$.*

Proof. Suppose this is not true. Then by $\cos 0 = 1$ and intermediate value theorem, we have $\cos x > 0$ for all $x \geqslant 0$. In other words, $\sin' > 0$ on $\mathbb{R}_{\geqslant 0}$. Thus, by Cor.

[11.20](#), $\sin$ is strictly increasing on $\mathbb{R}_{\geqslant 0}$. Therefore $A = \lim_{x \to +\infty} \sin x$ exists in $\overline{\mathbb{R}}$. Since $\sin 0 = 0$, we must have $A \in \overline{\mathbb{R}}_{>0}$. By L'Hôpital's rule in case $\frac{*}{\infty}$, we have

$$\lim_{x \to +\infty} \frac{\cos x}{x} = \lim_{x \to +\infty} -\sin x = -A < 0$$

contradicting the fact that $\cos x > 0$ if $x \geqslant 0$. $\qquad\square$

Definition 12.7. We define the number π to be

$$\pi = 2 \cdot \inf \left(\mathbb{R}_{\geqslant 0} \cap \cos^{-1}(0) \right) = 2 \cdot \inf\{x \in \mathbb{R}_{\geqslant 0} : \cos x = 0\}$$

Note that $\left(\mathbb{R}_{\geqslant 0} \cap \cos^{-1}(0) \right)$ is a closed subset of $\mathbb{R}$. So its infimum belongs to this set. Therefore, $\frac{\pi}{2}$ is the smallest $x \geqslant 0$ satisfying $\cos(x/2) = 0$.

Proposition 12.8. *We have*

$$\sin 0 = 0 \qquad \sin \frac{\pi}{2} = 1 \qquad \cos 0 = 1 \qquad \cos \frac{\pi}{2} = 0 \tag{12.6}$$

On $(0, \pi/2)$, $\sin$ and $\cos$ are strictly positive. On $[0, \pi/2]$, $\sin$ is strictly increasing, and $\cos$ is strictly decreasing.

Proof. All the formulas in [(12.6)](#), except $\sin(\pi/2) = 1$, has been proved. Since $(\sin \frac{\pi}{2})^2 = 1 - (\cos \frac{\pi}{2})^2$, we have $\sin \frac{\pi}{2} = \pm 1$.

By the definition of π, we know that $\sin' = \cos$ is > 0 on $[0, \frac{\pi}{2})$. So $\sin$ is strictly increasing on $[0, \frac{\pi}{2}]$. Thus, $\sin 0 = 0$ implies that $\sin > 0$ on $(0, \frac{\pi}{2}]$. In particular, $\sin \frac{\pi}{2} = 1$. Since $\cos' = -\sin$ is < 0 on $(0, \frac{\pi}{2})$, $\cos$ is strictly decreasing on $[0, \frac{\pi}{2}]$. $\qquad\square$

Proposition 12.9. *We have $\sin(-x) = -\sin x$ and $\cos(-x) = \cos x$.*

Proof. Immediate from $e^{-\mathrm{i}x} = 1/e^{\mathrm{i}x}$ and the definitions of $\sin$ and $\cos$. $\qquad\square$

From what we have proved, we know that the graph of $\sin$ and $\cos$ on $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ looks as follows.
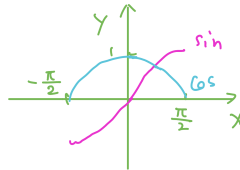
Proposition 12.10. *We have*

$$\sin x = \cos\left(x - \frac{\pi}{2}\right) = \cos\left(\frac{\pi}{2} - x\right) \qquad \cos(x) = \sin\left(\frac{\pi}{2} - x\right)$$

Proof. Immediate from [(11.24)](#) and [(12.6)](#). $\qquad\square$

Thus, the graph of $\sin$ is the rightward translation of that of $\cos$ by $\frac{\pi}{2}$. Therefore, the graph of $\sin$ on $[-\frac{\pi}{2}, 0]$ is translated to the graph of $\cos$ on $[-\pi, -\frac{\pi}{2}]$. That $\cos(x) = \cos(-x)$ gives us the graph of $\cos$ on $[\frac{\pi}{2}, \pi]$. Thus, we know the graph of $\cos$ on $[-\pi, \pi]$.

Theorem 12.11 (Euler's formula). *We have $e^{i\pi} = -1$, and hence $e^{2i\pi} = e^{i\pi}e^{i\pi} = 1$.*

Proof. We have $e^{i\pi/2} = \cos(\frac{\pi}{2}) + i\sin(\frac{\pi}{2}) = i$. Hence $e^{i\pi} = i^2 = -1$. □

Proposition 12.12. *We have $\sin x = \sin(x + 2\pi)$ and $\cos x = \cos(x + 2\pi)$*

Proof. Immediate from (11.24) and that $1 = e^{2i\pi} = \cos(2\pi) + i\sin(2\pi)$. □

Thus, $\cos$ and $\sin$ have period 2π. This completes the graphs of $\cos x$ and $\sin x = \cos(x - \frac{\pi}{2})$ on $\mathbb{R}$.

The fact that 2π is the length of the unit circle will be proved in Exp. 13.39.

12.3 Taylor's theorems

Assume throughout this section that $-\infty < a < b < +\infty$ and V is a Banach space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$.

In this section, we generalize MVTs and finite-increment theorem to higher derivatives. These generalizations are all under the name "Taylor theorem". Recall Def. 11.46 and Pb. 11.1 for the meaning of $l^{n,\infty}, C^n, C^\infty$. We first discuss the generalization of finite-increment theorem, which can be applied to vector-valued functions.

12.3.1 Taylor's theorems for vector-valued functions

Definition 12.13. Let X be a normed vector space. Let $A \subset X$. Let $a \in \mathrm{Cl}_X(A) \backslash A$ (or more generally, assume $a \in \mathrm{Cl}_X(A \backslash \{a\})$). Let $f : A \to V$. Let $r \in \mathbb{R}_{\geqslant 0}$.

- We write $f(x) = o(\|x - a\|^r)$ if $\lim\limits_{x \to a} \dfrac{f(x)}{\|x - a\|^r} = 0$.

- We write $f(x) = O(\|x - a\|^r)$ if $\limsup\limits_{x \to a} \dfrac{\|f(x)\|}{\|x - a\|^r} < +\infty$ where $\limsup$ is the limit superior of a net (cf. Rem. 7.83 and Pb. 8.2). In other words, there exists $U \in \mathrm{Nbh}_X(a)$ such that $\sup_{x \in A \cap U} \|f(x)\| / \|x - a\|^r < +\infty$.

When $r = 0$, we simply write $o(1)$ and $O(1)$. The two symbols o, O are called **Landau symbols**.

In this section, we will frequently use the formula

$$S_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x - a)^k \tag{12.7}$$

whenever the RHS makes sense.

Theorem 12.14 (Taylor's theorem, Peano form). *Let $f : [a, b] \to V$ and $n \in \mathbb{Z}_+$. Assume that $f', f'', \ldots, f^{(n)}$ exist at a (resp. at b). Then for each $x \in (a, b)$ we have*

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x - a)^k + o((x - a)^n) \tag{12.8a}$$

resp.

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(b)}{k!}(x - b)^k + o((x - b)^n) \tag{12.8b}$$

Expressions of the form (12.8a) are called **Taylor expansions** of f with **center** a.

Remark 12.15. In order for $f^{(n)}(a)$ to exist, it is assumed that $f^{(n-1)}$ exists on a neighborhood of a.

Proof of Peano-form. Since the two cases are similar, we only treat the case at a. Define **remainder** $g(x) = f(x) - S_n(x)$. Then

$$g(a) = g'(a) = \cdots = g^{(n)}(a) = 0 \tag{12.9}$$

It suffices to prove Taylor's theorem for g, i.e.

$$g(x) = o((x - a)^n) \tag{12.10}$$

We prove this by induction. The case $n = 1$ is obvious from the definition of derivatives, which says

$$\frac{g(x) - g(a)}{x - a} - g'(a) = o(1) \tag{12.11}$$

and hence $g(x) = (x - a)o(1) = o(x - a)$.

Now assume $n \geqslant 2$. Assume that Peano form has been proved for case $n - 1$. Applying this result to g'. We then get $g'(x) = o((x - a)^{n-1})$. Since $g''(a)$ exists, g' exists on $[a, c]$ where $a < c < b$. In particular, g is continuous on $[a, c]$. Therefore, by finite-increment Thm. 11.29, if $x \in (a, c)$, then

$$\|g(x)\| \leqslant (x - a) \cdot \sup_{a < t < x} \|g'(t)\|$$

If we can prove $\sup_{a \leqslant t \leqslant x} \|g'(t)\| = o((x - a)^{n-1})$, then we immediately have $g(x) = o((x - a)^n)$. Thus, the proof of Peano form is finished by the next lemma. $\qquad\square$

Lemma 12.16. *Assume that $f : [a, b] \to V$ satisfies $f(x) = o((x - a)^r)$. Define*

$$\widetilde{f}(x) = \sup_{a < t < x} \|f(t)\|$$

Then $\widetilde{f}(x) = o((x - a)^r)$.

Proof. Choose any $\varepsilon > 0$. Since $f(x) = o((x-a)^r)$, we know that there is $c \in (a, b)$ such that for all $a < x < c$ we have $\|f(x)\| \leqslant \varepsilon |x - a|^r$. Thus

$$|\widetilde{f}(x)| \leqslant \sup_{a < t < x} \varepsilon |t - a|^r = \varepsilon |x - a|^r$$

$\square$

Among all the versions of Taylor's theorem discussed in this section, the following one is most useful. (Another useful version, the integral form of Taylor's theorem, will be proved in Thm. 13.32.) Note that if we assume $f \in C^{(n+1)}([a, b], V)$ in Thm. 12.14, then Thm. 12.17 immediately implies Thm. 12.14.

Theorem 12.17 (Higher order finite-increment theorem). *Let $n \in \mathbb{N}$ and $f \in C^n([a, b], V)$. Assume that $f^{(n+1)}$ exists everywhere on (a, b). Then, for every $x \in (a, b]$ resp. $x \in [a, b)$, we have respectively*

$$\left\| f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k \right\| \leqslant \frac{(x-a)^{n+1}}{(n+1)!} \cdot \sup_{a<t<x} \|f^{(n+1)}(t)\| \tag{12.12a}$$

$$\left\| f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(b)}{k!}(x-b)^k \right\| \leqslant \frac{(b-x)^{n+1}}{(n+1)!} \cdot \sup_{x<t<b} \|f^{(n+1)}(t)\| \tag{12.12b}$$

Proof. We only prove the first formula: applying the first formula to $\widetilde{f}(x) = f(-x)$ implies the second one. We prove (12.12a) by induction on n. Moreover, we shall prove (12.12a) for the case $x = b$. The general case follows by restricting f to $[a, x]$. When $n = 0$, (12.12a) is the content of (classical) finite-increment Thm. 11.29. Assume case $n - 1$ has been proved ($n \in \mathbb{Z}_+$). In case n, take $g(x) = f(x) - S_n(x)$. Then (12.9) is true. Let

$$M = \sup_{a<t<b} \|f^{(n+1)}(t)\| = \sup_{a<t<b} \|g^{(n+1)}(t)\|$$

By case $n - 1$, for each $a \leqslant x \leqslant b$ we have

$$\|g'(x)\| \leqslant \frac{(x-a)^n}{n!} \cdot M = h'(x) \qquad \text{where} \ \ h(x) = \frac{M(x-a)^{n+1}}{(n+1)!}$$

Thus, by Thm. 11.27, we have

$$\|g(b) - g(a)\| \leqslant h(b) - h(a) = \frac{M(b-a)^{n+1}}{(n+1)!}$$

This proves (12.12a) for g, and hence for f. $\square$

12.3.2 Taylor's theorem for real-valued functions

Theorem 12.18 (Taylor's theorem, Lagrange form). *Let $n \in \mathbb{N}$ and $f \in C^n([a,b], \mathbb{R})$. Assume that $f^{(n+1)}$ exists everywhere on (a,b). Then for every $x \in (a,b]$ resp. $x \in [a,b)$, there exists $\xi \in (a,x)$ resp. $\eta \in (x,b)$ such that, respectively,*

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k + \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1} \tag{12.13a}$$

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(b)}{k!}(x-b)^k + \frac{f^{(n+1)}(\eta)}{(n+1)!}(x-b)^{n+1} \tag{12.13b}$$

Proof. We only prove (12.13a). Applying (12.13a) to $\widetilde{f}(x) = f(-x)$ (defined on $[-b,-a]$) implies the second formula. Again, it suffices to prove (12.13a) for $g(x) = f(x) - S_n(x)$, which satisfies (12.9). Thus, we want to find $\xi \in (a,x)$ satisfying

$$g(x) = \frac{g^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}$$

This can be proved by applying Cauchy's MVT repeatedly:

$$\frac{g(x)}{(x-a)^{n+1}} \stackrel{\exists x_1 \in (a,x)}{=\!=\!=\!=} \frac{g'(x_1)}{(n+1)(x_1-a)^n} \stackrel{\exists x_2 \in (a,x_1)}{=\!=\!=\!=} \frac{g''(x_2)}{(n+1)n(x_2-a)^{n-1}}$$

$$= \cdots \stackrel{\exists x_n \in (a,x_{n-1})}{=\!=\!=\!=} \frac{g^{(n)}(x_n)}{(n+1)!(x_n-a)} \stackrel{\exists \xi \in (a,x_n)}{=\!=\!=\!=} \frac{g^{(n+1)}(\xi)}{(n+1)!}$$

$$\square$$

We will mainly use Thm. 12.17 instead of the Lagrange form, since the latter does not apply directly to vector valued functions. However, Thm. 12.17 can be derived from the Lagrange form and Hahn-Banach theorem. We have used the fact that V^* separates points of V to prove that a function $[a,b] \to V$ is constant if its derivative exists and is constantly zero. (See Cor. 11.24.) However, to prove Thm. 12.17, we need the stronger fact that for every $v \in V$ there exists a nonzero $\varphi \in V^*$ such that $\langle \varphi, v \rangle = \|\varphi\| \cdot \|v\|$. (See Rem. 11.26.) By scaling φ, we can assume that $\|\varphi\| = 1$ and $\langle \varphi, v \rangle = \|v\|$.

We now give the second proof of Thm. 12.17:

Proof of Thm. 12.17 using Lagrange form and Hahn-Banach. The Lagrange form clearly implies Thm. 12.17 in the special case that $V = \mathbb{R}$. Now consider the general case, and view V as a real Banach space if it was originally over $\mathbb{C}$. Take $g(x) = f(x) - S_n(x)$. We need to prove $\|g(x)\| \leqslant \frac{(x-a)^{n+1}}{(n+1)!} M$ where $M = \sup_{a < t < x} \|g^{(n+1)}(t)\|$.

By Hahn-Banach (Rem. 11.26), there exists $\varphi \in V^*$ with $\|\varphi\| = 1$ such that $\varphi \circ g(x) = \|g(x)\|$. Applying the one dimensional special case to $\varphi \circ g$, we have

$$\|g(x)\| = \varphi \circ g(x) \leqslant \frac{(x-a)^{n+1}}{(n+1)!} \cdot \sup_{a<t<x} \left|(\varphi \circ g)^{(n+1)}(t)\right| \qquad (12.14)$$

provided that $\varphi \circ g \in C^n([a,b], V)$ and that $(\varphi \circ g)^{(n+1)}$ exists on (a,b).

By the continuity of φ, we have

$$(\varphi \circ g)'(t) = \lim_{s \to t} \frac{\varphi \circ g(s) - \varphi \circ g(t)}{s-t} = \varphi\left(\lim_{s \to t} \frac{g(s) - g(t)}{s-t}\right) = \varphi(g'(t)) \qquad (12.15)$$

Applying this formula repeatedly, we see that $\varphi \circ g \in C^n([a,b], V)$, that $(\varphi \circ g)^{(n+1)}$ exists on (a,b), and that $(\varphi \circ g)^{(n+1)} = \varphi \circ g^{(n+1)}$. By Rem. 10.24, we have

$$\sup_{a<t<x} \left|(\varphi \circ g)^{(n+1)}(t)\right| = \sup_{a<t<x} \left|\langle \varphi, g^{(n+1)}(t)\rangle\right| \leqslant \sup_{a<t<x} \|g^{(n+1)}(t)\| = M$$

Combining this result with (12.14) finishes the proof. $\qquad\square$

When V is $\mathbb{R}^N$ equipped with the Euclidean norm, the Hahn-Banach theorem (in the form of Rem. 11.26) is very easy: Define the linear map $\varphi : \mathbb{R}^N \to \mathbb{R}$ sending each $v \in \mathbb{R}^N$ to its dot product with $g(x)/\|g(x)\|$. Then φ satisfies $\|\varphi\| = 1$ and $\varphi \circ g(x) = \|g(x)\|$. So there is nothing mysterious in the above proof. (You are encouraged to compare this proof with the one of [Rud-P, Thm. 5.19].)

12.4 Functions approximated by their Taylor series

Taylor's theorems do not imply that a smooth function f on an interval of $\mathbb{R}$ can be approximated by its **Taylor series** (with **center** a):

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k \qquad (12.16)$$

The following is a typical example:

Example 12.19. Define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} \exp\left(-\dfrac{1}{x^2}\right) & (\text{if } x > 0) \\ 0 & (\text{if } x \leqslant 0) \end{cases}$$

Then $f^{(n)}(0) = 0$ for all $n \in \mathbb{N}$. So the Taylor series of f at 0 is 0, which cannot approximate $f(x)$ when $x > 0$.

On the contrary, if f is defined on an open subset of $\mathbb{C}$, and if f' exists everywhere on its domain, then $f^{(n)}$ exists for all n, and f can be approximated locally uniformly by its Taylor series. This is a deep result in complex analysis. In fact, a thorough understanding of power series is impossible without the help of complex analysis.

In the following, we show that some important real variable functions can be approximated by their Taylor series. Actually, the proof can be simplified by complex analysis, since these examples are the restriction of some differentiable complex variable functions (aka **holomorphic functions**) to the real line.

Example 12.20. Consider the Taylor expansion of $\log : \mathbb{R}_{>0} \to \mathbb{R}$ at $x = 1$. By induction on $n \in \mathbb{Z}_+$, one computes that

$$\log^{(n)}(x) = (-1)^{n-1}\frac{(n-1)!}{x^n}$$

Therefore, its Taylor expansion in order n is

$$\sum_{k=1}^{n} \frac{(-1)^{k-1}}{k}(x-1)^k + R_{n+1}(x)$$

where $R_{n+1}(x)$ is the remainder. To show that $\log x$ is approximated uniformly (on certain domain) by its Taylor series, one need to show that R_{n+1} converges uniformly to 0 on that domain.

We would like to prove that for every $0 < r < 1$ we have $R_{n+1} \rightrightarrows 0$ on $[1 - r, 1+r]$. This would imply that series on the RHS of the following formula (whose radius of convergence is 1) converges uniformly to f on $[1 - r, 1 + r]$:

$$\log(x) = \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k}(x-1)^k \tag{12.17}$$

However, using the Taylor's theorems proved in the previous section, one can prove the uniform convergence only when $0 < r \leqslant 1/2$. The general case of $0 < r < 1$ should be proved by another method. $\qquad\square$

Proof for the case $0 < r \leqslant \frac{1}{2}$. In fact, we shall prove the uniform convergence on $[1 - r, 2]$. By Thm. 12.17, for all $x \in [1 - r, 1 + r]$ we have

$$|R_{n+1}(x)| \leqslant \frac{|x-1|^{n+1}}{(n+1)!} \cdot \sup_{\substack{1 \leqslant t \leqslant x \\ \text{or } x \leqslant t \leqslant 1}} \left|\log^{(n+1)}(t)\right| = \frac{|x-1|^{n+1}}{n+1} \cdot \sup_{\substack{1 \leqslant t \leqslant x \\ \text{or } x \leqslant t \leqslant 1}} \frac{1}{t^{n+1}}$$

where the sup is over $1 \leqslant t \leqslant x$ or $x \leqslant t \leqslant 1$, depending on whether $1 \leqslant x \leqslant 1 + r$ or $1 - r \leqslant x \leqslant 1$.

If $1 \leqslant x \leqslant 1 + r$, then $1 \leqslant t \leqslant x$ implies $1/t^{n+1} \leqslant 1$. So

$$|R_{n+1}(x)| \leqslant \frac{(x-1)^{n+1}}{n+1} \leqslant \frac{r^{n+1}}{n+1}$$

where the RHS converges to 0 as $n \to \infty$ whenever $0 < r \leqslant 1$. In particular, we get the renowned formula

$$\log 2 = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \tag{12.18}$$

If $1 - r \leqslant x \leqslant 1$, then $x \leqslant t \leqslant 1$ implies $1/t^{n+1} \leqslant 1/x^{n+1}$. Thus

$$|R_{n+1}(x)| \leqslant \frac{1}{n+1}\left(\frac{1}{x} - 1\right)^{n+1} \leqslant \frac{1}{n+1}\left(\frac{1}{1-r} - 1\right)^{n+1} = \frac{1}{n+1}\left(\frac{r}{1-r}\right)^{n+1}$$

If $0 < r \leqslant 1/2$, then $0 < r/(1-r) \leqslant 1$, and hence the RHS above converges to 0 as $n \to \infty$. This proves that $R_{n+1} \rightrightarrows 0$ on $[1 - r, 1 + r]$ when $0 < r \leqslant 1/2$. $\qquad \square$

Proof for the case $0 < r < 1$. Assume $0 < r < 1$. For the convenience of discussion, we prove instead that

$$\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} x^k \tag{12.19}$$

the Taylor series of $\log(1 + x)$ at 0, converges uniformly to $\log(1 + x)$ on $[-r, r]$

The radius of convergence of the series (12.19) is 1. Thus, by Thm. 4.27, (12.19) converges locally uniformly on $(-1, 1)$ to some function $f : (-1, 1) \to \mathbb{R}$. By Cor. 11.37, f is differentiable on $(-1, 1)$, and the series

$$\sum_{k=0}^{\infty} (-1)^k x^k \tag{12.20}$$

converges locally uniformly on $(-1, 1)$ to f'. But we clearly have

$$(12.20) = \frac{1}{1 - (-x)} = \frac{1}{1 + x} \tag{12.21}$$

Thus $f'(x) = \frac{d}{dz}(\log(1+x))$. Now, (12.19) clearly implies that $f(0) = 0 = \log(1+0)$. Therefore, by Cor. 11.24, we obtain $f(x) = \log(1 + x)$ on $(-1, 1)$. In other words, (12.19) converges locally uniformly on $(-1, 1)$ (and hence uniformly on $[-r, r]$) to $\log(1 + x)$. $\qquad \square$

Example 12.21. Let $\alpha \in \mathbb{C}$. Then

$$\sum_{k=0}^{\infty} \binom{\alpha}{k} x^k \tag{12.22}$$

the Taylor series of $(1 + x)^\alpha = \exp(\alpha \log(1 + x))$ at $x = 0$, converges locally uniformly on $(-1, 1)$ to $(1 + x)^\alpha$.

Proof. Using Prop. 4.23, one easily checks that the radius of convergence of (12.22) is 1. So Thm. 4.27 implies that (12.22) converges locally uniformly on $(-1, 1)$ to some $f : (-1, 1) \to \mathbb{C}$. Let

$$g(x) = (1 + x)^\alpha$$

Our goal is to prove $f = g$.

It is clear that $(1 + x)g'(x) = \alpha g(x)$. By Cor. 11.37, the series

$$\sum_{k=0}^{\infty} k \binom{\alpha}{k} x^{k-1} = \sum_{k=0}^{\infty} (k + 1) \binom{\alpha}{k+1} x^k \tag{12.23}$$

converges locally uniformly on $(-1, 1)$ to f'. Thus, by Cor. 5.60, we have

$$(1 + x)f'(x) = \sum_{k=0}^{\infty} \left(k \binom{\alpha}{k} + (k + 1) \binom{\alpha}{k+1} \right) x^k$$

$$= \sum_{k=0}^{\infty} \alpha \binom{\alpha}{k} x^k = \alpha f(x)$$

Therefore, f satisfies the same differential equation as g, namely $(1 + x)f' = \alpha f$. Moreover, we clearly have $f(0) = 1 = g(0)$. So we have $f = g$ on $(-1, 1)$ if we apply the next lemma to $f - g$. $\qquad \square$

Lemma 12.22. *Let I be an interval of $\mathbb{R}$ (with at least two points) such that $0 \in I$. Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. Let $\varphi \in C(I, \mathbb{F}^{n \times n})$. Assume that $f \in C^1(I, \mathbb{F}^n)$ satisfies*

$$f'(x) = \varphi(x) \cdot f(x) \qquad f(0) = 0 \tag{12.24}$$

Then for all $x \in I$ we have $f(x) = 0$.

Proof. We prove this lemma for the case that I is an open interval. The proof for the other types of intervals is similar. Also, we equip $\mathbb{F}^{n \times n}$ with the operator norm by viewing $n \times n$ matrices as elements of $\mathfrak{L}(\mathbb{F}^n)$. (Recall that the operator norm is equivalent to the Eucliden norm, see Thm. 10.30.) Let

$$\Omega = \{x \in I : f(x) = 0\} = f^{-1}(0)$$

which is a closed subset of I (since the inverse image under a continuous map of a closed set is closed). Clearly Ω is nonempty since $0 \in \Omega$. If we can prove that Ω is open, then we have $\Omega = I$ because I is connected. This will finish the proof.

So let us prove that any $p \in \Omega$ is an interior point of Ω with respect to I. Choose $r > 0$ such that $[p - r, p + r] \subset I$. Let

$$C = \sup_{-r \leqslant x \leqslant r} \|\varphi(x)\|$$

which is a finite number by extreme value theorem. Let $\delta = \min\{\frac{1}{2C}, r\}$ which is > 0. It suffices to prove that

$$M = \sup_{x \in [p-\delta, p+\delta]} \|f(x)\|$$

is zero. (Note that $M < +\infty$ again by EVT.) Then we will have $[p - \delta, p + \delta] \subset \Omega$, finishing the proof.

For each $x \in [p - \delta, p + \delta]$, by Rem. 10.24 for operator norms, we have

$$\|f'(x)\| = \|\varphi(x)f(x)\| \leqslant \|\varphi(x)\| \cdot \|f(x)\| \leqslant CM$$

Thus, by finite-increment theorem, we have for all $x \in [p - \delta, p + \delta]$ that

$$\|f(x)\| = \|f(x) - f(p)\| \leqslant CM|x - p| \leqslant CM\delta \leqslant \frac{M}{2}$$

Applying $\sup_{x \in [p-\delta, p+\delta]}$ to $\|f(x)\|$, we get $M \leqslant M/2$. So $M = 0$. $\qquad \square$

Remark 12.23. The above Lemma can also be proved in a similar way as Thm. 11.27: Let $x \in I$. Assume WLOG that $x > 0$. Let $E = f^{-1}(0) \cap [0, x]$, which is a closed subset of $[0, x]$. So $t = \sup E$ is in $[0, x]$. To show $f(x) = 0$, one just need to prove $t = x$. If not, then $t < x$. Then, as in the above proof, one can find $\delta > 0$ such that $\varphi(s) = 0$ whenever $s \in [t, t + \delta]$, impossible.

12.5   Convex functions

In this section, we fix an interval $I \subset \mathbb{R}$ containing at least two points.

Definition 12.24. Let $\mathbf{x} = (x_1, x_2)$, $\mathbf{y} = (y_1, y_2)$, $\mathbf{z} = (z_1, z_2)$ be three points of $\mathbb{R}^2$ satisfying $x_1 < y_1 < z_1$. We say that the ordered triple $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ is a **convex triple** if the following equivalent conditions are satisfied:

(1) $\mathbf{y}$ is below or on the interval $[\mathbf{x}, \mathbf{z}]$. In other words, we have $y_2 \leqslant tx_2 + (1-t)z_2$ if $t \in [0, 1]$ is such that $y_1 = tx_1 + (1 - t)z_1$.

(2) We have $\dfrac{y_2 - x_2}{y_1 - x_1} \leqslant \dfrac{z_2 - y_2}{z_1 - y_1}$.

(3) We have $\dfrac{y_2 - x_2}{y_1 - x_1} \leqslant \dfrac{z_2 - x_2}{z_1 - x_1}$.

(4) We have $\dfrac{z_2 - x_2}{z_1 - x_1} \leqslant \dfrac{z_2 - y_2}{z_1 - y_1}$.

Proof of equivalence. The equivalence of these statements is clear from the picture, and is not hard to check rigorously using inequalities. We leave the details to the reader. $\qquad \square$

Recall the definition of convex subsets in real vector spaces (Def. 11.30).

Definition 12.25. A function $f : I \to \mathbb{R}$ is called **convex** if the following equivalent conditions are true:

(1) The set

$$D_f = \{(x, y) \in \mathbb{R}^2 : x \in I, y \geqslant f(x)\} \tag{12.25}$$

is a convex subset of $\mathbb{R}^2$.

(1) For any three points $x < y < z$ of I, the points $(x, f(x)), (y, f(y)), (z, f(z))$ form a convex triple.

Proof of equivalence. Again, the equivalence is clear from the picture. The details are left to the readers. $\qquad\square$

The convexity of differentiable functions is easy to characterize:

Theorem 12.26. *Let $f : I \to \mathbb{R}$ be differentiable. Then f is convex iff f' is increasing. In particular, if f'' exists on I, then f is convex iff $f'' \geqslant 0$ on I.*

Proof. Assume that f' is increasing. Choose $x < y < z$ in I. By Lagrange's MVT, there exist $\xi \in (x, y)$ and $\eta \in (y, z)$ such that the slope of the interval $[(x, f(x)), (y, f(y))]$ equals $f'(\xi)$, and that the slope of $[(y, f(y)), (z, f(z))]$ equals $f'(\eta)$. Since $f'(\xi) \leqslant f'(\eta)$, the points $(x, f(x)), (y, f(y)), (z, f(z))$ form a convex triple. This proves that f is convex.

Now assume that f is convex. Choose any $x < y$ in I. We need to prove that $f'(x) \leqslant f'(y)$. Choose any $\varepsilon > 0$. Then there exist x_1 and y_1 such that $x < x_1 < y_1 < y$, that $f'(x) - \varepsilon$ is $\leqslant$ the slope k_1 of $[(x, f(x)), (x_1, f(x_1))]$, and that $f'(y) + \varepsilon$ is $\geqslant$ the slope k_2 of $[(y_1, f(y_1)), (y, f(y))]$. Since f is convex, k_1 is $\leqslant$ the slope k' of $[(x_1, f(x_1)), (y_1, f(y_1))]$, and $k' \leqslant k_2$. Therefore $k_1 \leqslant k_2$. Thus $f'(x) - \varepsilon \leqslant f'(y) + \varepsilon$. Since $\varepsilon > 0$ is arbitrary, we get $f'(x) \leqslant f'(y)$.

When f'' exists, the equivalence between $f'' \geqslant 0$ and the increasing monotonicity of f' is due to Cor. 11.20. $\qquad\square$

The most important general property about convex functions is:

Theorem 12.27 (Jensen's inequality). *Let $f : I \to \mathbb{R}$ be a convex function. Let $n \in \mathbb{Z}_+$ and $x_1, \ldots, x_n \in I$. Choose $t_1, \ldots, t_n \in [0, 1]$ such that $t_1 + \cdots + t_n = 1$. Then*

$$f(t_1 x_1 + \cdots + t_n x_n) \leqslant t_1 f(x_1) + \cdots + t_n f(x_n) \tag{12.26}$$

Proof. The points $(x_1, f(x_1)), \ldots, (x_n, f(x_n))$ are in $D_f = (12.25)$. Therefore, by Lem. 12.28, the point

$$t_1(x_1, f(x_1)) + \cdots + t_n(x_n, f(x_n)) = (t_1 x_1 + \cdots + t_n x_n, t_1 f(x_1) + \cdots + t_n f(x_n))$$

is in the convex set D_f. $\qquad\square$

Lemma 12.28. *Let V be a real vector space. Let E be a convex subset of V. Let $n \in \mathbb{Z}_+$ and $p_1, \ldots, p_n \in E$. Any **convex combination** of $p_1, \ldots, p_n$ (i.e., any point of the form $t_1 p_1 + \cdots + t_n p_n$ where $t_1, \ldots, t_n \in [0, 1]$ and $t_1 + \cdots + t_n = 1$) is inside E.*

The geometric meaning of this lemma is clear: If $V = \mathbb{R}^2$ (which is the main case we are interested in), then $t_1 p_1 + \cdots + t_n p_n$ stands for an arbitrary point inside the polygon with vertices $p_1, \ldots, p_n$. So this point is clearly inside E.

Proof. We prove by induction on n. The case $n = 1$ is obvious. Suppose case $n - 1$ has been proved $n \geqslant 2$. Consider case n. It is trivial when $t_1 + \cdots + t_{n-1} = 0$. So let us assume $t_1 + \cdots + t_{n-1} > 0$. By case $n - 1$, the point $q = \lambda_1 p_1 + \cdots + \lambda_{n-1} p_{n-1}$ is in E where

$$\lambda_i = \frac{t_i}{t_1 + \cdots + t_{n-1}}$$

Since E is convex, the point $(t_1 + \cdots + t_{n-1}) q + t_n p_n$ (which equals $t_1 p_1 + \cdots + t_n p_n$) is in E. $\qquad\square$

Example 12.29. Since $(-\log x)'' = x^{-2}$ is positive on $\mathbb{R}_{>0}$, by Thm. 12.26, $-\log$ is a convex function on $\mathbb{R}_{>0}$. Therefore, if $0 < \lambda_1, \ldots, \lambda_n \leqslant 1$ and $\lambda_1 + \cdots + \lambda_n = 1$, then by Jensen's inequality, for each $x_1, \ldots, x_n > 0$ we have $-\log(\sum_i \lambda_i x_i) \leqslant -\lambda_i \log x_i$. Taking exponentials, we get

$$x_1^{\lambda_1} \cdots x_n^{\lambda_n} \leqslant \lambda_1 x_1 + \cdots + \lambda_n x_n$$

for all $x_1, \ldots, x_n > 0$, and hence for all $x_1, \ldots, x_n \geqslant 0$. In particular, we get the **inequality of arithmetic and geometric means**

$$(x_1 \cdots x_n)^{\frac{1}{n}} \leqslant \frac{x_1 + \cdots + x_n}{n} \tag{12.27}$$

and the **Young's inequality**

$$x^{\frac{1}{p}} y^{\frac{1}{q}} \leqslant \frac{x}{p} + \frac{y}{q} \tag{12.28}$$

if $p, q > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$.

Definition 12.30. Let $p \in [1, +\infty]$. We say that $q \in [1, +\infty]$ is the **Hölder conjugate** of p if $\frac{1}{p} + \frac{1}{q} = 1$.

12.6 Hölder's and Minkowski's inequalities; l^p spaces

In this section, we use Young's inequality to prove two inequalities that are of vital importance to the development of modern analysis:

Theorem 12.31. *Let* $x_1, \ldots, x_n, y_1, \ldots, y_n \in \mathbb{R}_{\geqslant 0}$. *Let* $p, q \in (1, +\infty)$ *satisfy* $\frac{1}{p} + \frac{1}{q} = 1$. *Then the following inequalities (called respectively **Hölder's inequality** and **Minkowski's inequality**) are true:*

$$x_1 y_1 + \cdots + x_n y_n \leqslant \sqrt[p]{x_1^p + \cdots + x_n^p} \cdot \sqrt[q]{y_1^q + \cdots + y_n^q} \tag{12.29a}$$

$$\sqrt[p]{(x_1 + y_1)^p + \cdots + (x_n + y_n)^p} \leqslant \sqrt[p]{x_1^p + \cdots + x_n^p} + \sqrt[p]{x_1^p + \cdots + x_n^p} \tag{12.29b}$$

In the special case that $p = q = 2$, Hölder's inequality is called **Cauchy-Schwarz inequality**

$\star\star$ *Proof.* Assume WLOG that $x_i > 0$ and $y_j > 0$ for some i, j. Young's inequality implies that for each i,

$$\frac{x_i}{\sqrt[p]{\sum_k x_k^p}} \cdot \frac{y_i}{\sqrt[q]{\sum_k y_k^q}} \leqslant \frac{x_i^p}{p \sum_k x_k^p} + \frac{y_i^q}{q \sum_k y_k^q}$$

Taking sum over i gives

$$\sum_i \frac{x_i}{\sqrt[p]{\sum_k x_k^p}} \cdot \frac{y_i}{\sqrt[q]{\sum_k y_k^q}} \leqslant \frac{1}{p} + \frac{1}{q} = 1$$

This proves Hölder's inequality.

Notice that $pq = p + q$. Let $z_i = x_i + y_i$. By Hölder's inequality, we have

$$x_1 z_1^{p-1} + \cdots + x_n z_n^{p-1} \leqslant \left(\sum_k x_k^p \right)^{\frac{1}{p}} \cdot \left(\sum_k z_k^{(p-1)q} \right)^{\frac{1}{q}} = \left(\sum_k x_k^p \right)^{\frac{1}{p}} \cdot \left(\sum_k z_k^p \right)^{\frac{1}{q}}$$

and similarly

$$y_1 z_1^{p-1} + \cdots + y_n z_n^{p-1} \leqslant \left(\sum_k y_k^p \right)^{\frac{1}{p}} \cdot \left(\sum_k z_k^p \right)^{\frac{1}{q}}$$

Adding up these two inequalities, we get

$$\sum_k z_k^p \leqslant \left(\left(\sum_k x_k^p \right)^{\frac{1}{p}} + \left(\sum_k y_k^p \right)^{\frac{1}{p}} \right) \cdot \left(\sum_k z_k^p \right)^{\frac{1}{q}}$$

Dividing both sides by $\left(\sum_k z_k^p \right)^{\frac{1}{q}}$ proves Minkowski's inequality. $\qquad\square$

Minkowski's inequality is often used in the following way:

Theorem 12.32. *Let* X *be a set, and let* V *be a normed vector space over* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. *Let* $1 \leqslant p < +\infty$. *For each* $f \in V^X$, *define the* l^p**-norm**

$$\|f\|_p \equiv \|f\|_{l^p} = \left(\sum_{x \in X} \|f(x)\|^p \right)^{\frac{1}{p}} \tag{12.30}$$

Then for each $f, g \in V^X$ and $\lambda \in \mathbb{F}$ we have

$$\|f + g\|_p \leqslant \|f\|_p + \|g\|_p \qquad \|\lambda f\|_p = |\lambda| \cdot \|f\|_p \qquad (12.31)$$

*In particular, $\| \cdot \|_p$ is a norm on the **l^p-space***

$$l^p(X, V) = \{f \in V^X : \|f\|_{l^p} < +\infty\} \qquad (12.32)$$

If V is Banach space, then so is $l^p(X, V)$.

Recall $0 \cdot (\pm \infty) = 0$.

Proof. Choose any $A \in \mathrm{fin}(2^X)$. By Minkowski's and triangle inequality, we have

$$\sum_A |f + g|^p \leqslant \sum_A (|f| + |g|)^p \leqslant \left(\left(\sum_A |f|^p \right)^{\frac{1}{p}} + \left(\sum_B |f|^p \right)^{\frac{1}{p}} \right)^p$$

$$\leqslant \left(\left(\sum_X |f|^p \right)^{\frac{1}{p}} + \left(\sum_X |f|^p \right)^{\frac{1}{p}} \right)^p = (\|f\|_p + \|g\|_p)^p$$

Taking $\lim_{A \in \mathrm{fin}(2^X)}$ gives $\|f + g\|_p^p \leqslant (\|f\|_p + \|g\|_p)^p$, proving the first inequality of (12.31). The second of (12.31) clearly holds by taking $\lim_{A \in \mathrm{fin}(2^X)}$ of

$$\sum_A |\lambda f|^p = |\lambda| \sum_A |f|^p$$

and noting that the multiplication map $t \in \overline{\mathbb{R}}_{\geqslant 0} \mapsto |\lambda| t \in \overline{\mathbb{R}}_{\geqslant 0}$ is continuous.

Suppose that V is complete. Let (f_n) be a Cauchy sequence in $l^p(X, V)$. Then for each $x \in X$, $(f_n(x))$ is a Cauchy sequence in V, converging to some element $f(x) \in V$. By Cauchyness, for each $\varepsilon > 0$ there is $N \in \mathbb{Z}_+$ such that for all $n, k \geqslant N$ we have $\|f_n - f_k\|_p \leqslant \varepsilon$, and hence

$$\sum_{x \in A} \|f_n(x) - f_k(x)\|^p \leqslant \varepsilon^p$$

for all $A \in \mathrm{fin}(2^X)$. Taking $\lim_k$ gives $\sum_A |f_n - f|^p \leqslant \varepsilon^p$. Taking $\lim_{A \in \mathrm{fin}(2^X)}$ gives $\|f_n - f\|_p \leqslant \varepsilon$ for all $n \in \geqslant N$. (In particular, $\|f\|_p \leqslant \|f_N\| + \varepsilon < +\infty$, and hence $f \in l^p(X, V)$.) This proves that (f_n) converges to f under the l^p-norm. So $l^p(X, V)$ is complete. $\square$

Theorem 12.33. *Let X be a set. Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. Let $p, q \in [1, +\infty]$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Then there is a linear isometry*

$$\Psi : l^p(X, \mathbb{F}) \to l^q(X, \mathbb{F})^* \qquad f \mapsto \Psi(f) \qquad (12.33a)$$

where for each $g \in l^q(X, \mathbb{F})$, the value of $\Psi(f)$ at g is

$$\langle \Psi(f), g \rangle = \sum_{x \in X} f(x) g(x) \qquad (12.33b)$$

where the RHS converges.

In fact, we will see in Thm. 17.28 that if $q < +\infty$ (and hence $p > 1$), Ψ is surjective, and hence is an isomorphism of normed vector spaces. (This is not a difficult fact. So you can prove it yourself.)

Proof. We treat the case $1 < p, q < +\infty$, and leave the case that $p \in \{1, +\infty\}$ to the readers. Assume $f \in l^p$ and $g \in l^q$, then for each $A \in \mathrm{fin}(2^X)$ we have by Hölder's inequality that

$$\left\| \sum_A fg \right\| \leqslant \sum_A |fg| \leqslant \left(\sum_A |f|^p \right)^{\frac{1}{p}} \left(\sum_A |g|^q \right)^{\frac{1}{q}} \leqslant \|f\|_p \cdot \|g\|_q$$

Applying $\lim_{A \in \mathrm{fin}(2^X)}$ to the second term above implies that $\sum_X |fg| < +\infty$. So $\sum_X fg$ converges absolutely, and hence converges. So (12.33b) makes sense, and hence Ψ is a well-defined linear map. Applying $\lim_{A \in \mathrm{fin}(2^X)}$ to the first term above implies that

$$|\langle \Psi(f), g \rangle| \leqslant \|f\|_p \cdot \|g\|_q$$

Thus, by Rem. 10.24, we obtain $\|\Psi(f)\| \leqslant \|f\|_p$.

To show $\|\Psi(f)\| = \|f\|_p$, assume WLOG that $f \neq 0$, and define $g : X \to \mathbb{F}$ to be $g = \frac{\overline{f}}{|f|} \cdot |f|^{p-1}$, where $g(x)$ is understood as 0 if $f(x) = 0$. Then

$$\langle \Psi(f), g \rangle = \sum_X |f|^p = \|f\|_p^p$$

and $\|g\|_q^q = \sum_X |f|^{pq-q} = \sum_X |f|^p = \|f\|_p^p$. So

$$\|f\|_p \cdot \|g\|_q = \|f\|_p^{1+p/q} = \|f\|_p^p = \langle \Psi(f), g \rangle$$

This proves $\|\Psi(f)\| = \|f\|_p$ by Rem. 10.24. So Ψ is a linear isometry. $\qquad \square$

I do not want to deviate too much from the main purpose of this section, which is to show some important applications of convex functions. I will therefore stop my discussion about l^p spaces, and continue this topic in the future. The crucial role of l^p spaces and their continuous versions (namely L^p spaces) in modern analysis is a long story. The study of these objects constitutes a major part of the second half of this course. Let me mention just one point: the compactness of $\overline{B}_{l^2(\mathbb{Z}, \mathbb{C})}(0, 1)$ under the pointwise convergence topology (by viewing it as a subspace of $\mathbb{C}^{\mathbb{Z}}$) was the most important reason for Hilbert and Schmidt to study the Hilbert space $l^2(\mathbb{Z}, \mathbb{C})$, and it was the crucial property that allowed them to fully solve the eigenvalue problem in integral equations (cf. Subsec. 10.4.1).

12.7   Problems and supplementary material

Let $-\infty < a < b < +\infty$. Let V be a Banach space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$.

Problem 12.1. Let f be the function in Exp. 12.19. Prove that $f^{(n)}(0) = 0$ for all $n \in \mathbb{N}$.

Problem 12.2. Let $n \in \mathbb{N}$. Let $f : [a, b] \to V$ such that $f, f', \ldots, f^{(n)}$ exist everywhere on $[a, b]$. Use the higher order finite-increment theorem to prove that

$$\left\| f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!} (x - a)^k \right\| \leqslant \frac{(x - a)^n}{n!} \cdot \sup_{a < t < x} \| f^{(n)}(t) - f^{(n)}(a) \| \tag{12.34}$$

Remark 12.34. Setting $n = 1$ and dividing both sides by $x - a$, we obtain an especially useful formula for any differentiable $f : [a, b] \to V$ as follows.

$$\left\| \frac{f(x) - f(a)}{x - a} - f'(a) \right\| \leqslant \sup_{a < t < x} \| f'(t) - f'(a) \| \tag{12.35}$$

Problem 12.3. Use Thm. 11.33 to prove the following theorem. (Formula (12.35) might be helpful.)

Theorem 12.35. *Let $I = [a, b]$ and $J = [c, d]$ be intervals in $\mathbb{R}$ with at least two points. Let $f : I \times J \to V$ be a function such that $\partial_1 f, \partial_2 f, \partial_2 \partial_1 f$ exist on $I \times J$, and that $\partial_2 \partial_1 f$ is continuous. Then $\partial_1 \partial_2 f$ exists on $I \times J$ and equals $\partial_2 \partial_1 f$.*

⋆ **Problem 12.4.** Prove that, for $x \in (-1, 1]$,

$$\arctan x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{2k + 1} = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + - \cdots, \tag{12.36}$$

and hence (**Leibniz formula**)

$$\frac{\pi}{4} = \sum_{k=0}^{\infty} \frac{(-1)^k}{2k + 1} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + - \cdots$$

Hint. Use the method in the second proof of Exp. 12.20 to prove (12.36) for $x \in (-1, 1)$. To prove (12.36) for $x = 1$, show that the series on the RHS of (12.36) converges uniformly on $[0, 1]$ to a continuous function using Dirichlet's test for uniform convergence (Thm. 4.29). ☐

⋆ **Problem 12.5.** Let $n \in \mathbb{N}$. Find the radius of convergence of

$$f(z) = \sum_{k=n}^{\infty} \binom{k}{n} z^k \tag{12.37}$$

Find the explicit formula of $f(z)$.

Problem 12.6. Prove the following higher order finite-increment theorem for complex variables. (For simplicity, it suffices prove the case that $z_0 = 0$.)

Theorem 12.36 (Higher order finite-increment theorem). *Let V be a Banach space over $\mathbb{C}$. Let $R > 0$ and $z_0 \in \mathbb{C}$. Let $f : \Omega \to V$ where*

$$\Omega = B_{\mathbb{C}}(z_0, R) = \{z \in \mathbb{C} : |z - z_0| < R\}$$

Assume that $f', f'', \ldots, f^{(n+1)}$ exist everywhere on Ω. Then for every $z \in \Omega$ we have

$$\left\| f(z) - \sum_{k=0}^{n} \frac{f^{(k)}(z_0)}{k!}(z - z_0)^k \right\| \leqslant \frac{|z - z_0|^{n+1}}{(n+1)!} \cdot \sup_{\xi \in [z_0, z]} \| f^{(n+1)}(\xi) \| \qquad (12.38)$$

where $[z_0, z] = \{tz + (1 - t)z_0 : 0 \leqslant t \leqslant 1\}$.

$\star$ **Problem 12.7.** Let V be a Banach space over $\mathbb{C}$. Assume that the power series $f(z) = \sum_{k=0}^{\infty} v_k z^k$ (where $v_k \in V$) has radius of convergence $R > 0$. Choose any $z_0 \in B_{\mathbb{C}}(0, R)$. Prove that there exists a neighborhood $\Omega \in \mathrm{Nbh}_{\mathbb{C}}(z_0)$ contained inside $B_{\mathbb{C}}(0, R)$, such that the Taylor series of f at z_0 converges uniformly on Ω to f. Namely, prove that the series $g(z)$ converges locally uniformly to f on Ω where

$$g(z) = \sum_{k=0}^{\infty} \frac{f^{(k)}(z_0)}{k!}(z - z_0)^k \qquad (12.39)$$

Hint. Use Thm. 12.36. Pb. 12.5 might also be helpful. $\qquad \square$

$\star$ **Problem 12.8.** Let W be a real vector space. Let A be a nonempty subset of W. Define the **convex hull** of A to be the set of all convex combinations (recall Lem. 12.28) of elements of A, i.e.

$$\mathrm{Cvh}(A) = \big\{\text{convex combinations of } v_1, \ldots, v_n : n \in \mathbb{Z}_+ \text{ and } v_1, \ldots, v_n \in A\big\} \qquad (12.40)$$

Prove that $\mathrm{Cvh}(A)$ is the smallest convex set containing A. In other words, prove that $\mathrm{Cvh}(A)$ is a convex set containing A, and that $\mathrm{Cvh}(A) \subset B$ if B is a convex subset of W containing A.

Problem 12.9. Prove Thm. 12.33 for the case $p = 1$ and $p = +\infty$.

13 Riemann integrals

13.1 Introduction: the origin of integral theory in Fourier series

13.1.1 Antiderivatives VS. approximation by areas of rectangles

Modern people can easily appreciate the importance of giving a rigorous definition to the integral $\int_a^b f(x)dx$ where f is a general (say, continuous) function. And you may already know that this integral is defined by taking the limit of Riemann sums, which are "areas of some rectangles". This idea sounds so natural to you, because you know that by the time Calculus was invented, people knew very well that $\int_a^b f$ means the area of the region between the graph of f and the x-axis, and that this area can be approximated by the areas of some rectangles or triangles. So why did Riemann integral not appear until 19th century, more than a hundred years after Newton's and Leibniz's invention of calculus? And why was the inadequacy of defining integrals by means of antiderivatives not recognized until 19th century?

Beyond the superficial reason that 19th century is the century in which people began to pay attention to the foundations of calculus, there is a deeper reason: the study of **partial differential equations (PDE)** and the introduction of **Fourier series** caused a drastic change in the general view of what functions are. This change of view motivated people to search for a general definition of integrals, in particular one that does not use antiderivatives.

Before the systematic study of PDEs (i.e., before 19th century), functions only mean analytic functions, which means that they can be approximated by their Taylor series. But if f is such a function, then $f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$. So the values of f are all determined by $f(a), f'(a), f''(a), \ldots$, and hence are determined by $f|_I$ where I is an arbitrary nonempty open interval. (Technically, such functions are called **(real) analytic functions**.) Therefore, periodic functions (except trigonometric functions) were not accepted, because they are not determined by their values on a small interval. (A function with expression $f(x) = x$ on $[0,1]$ should also have the same expression on $\mathbb{R}$. So it must not be the periodic function $f(x) = x - n$ where $x \in (n, n+1]$.) The function Exp. 12.19 was not accepted, because it was not determined by $f(0), f'(0), f''(0), \ldots$.

Therefore, in these old days, people did not worry about the rigorous definition of integrals, because they can simply define integrals using antiderivatives instead of using the areas of rectangles to approximate the integrals, which is practically less convenient than finding the antiderivatives. In particular, one understood

$$\int_a^b \sum a_n(x-c)^n dx \overset{\text{def}}{=\!=} F(b) - F(a) \qquad \text{where} \qquad F(x) = \sum \frac{a_n}{n+1}(x-c)^{n+1}$$

13.1.2 PDEs and periodic functions

The wonderful dream was shattered when people started working on PDEs. The simplest such example is **wave equation**

$$\partial_x^2 f(x,t) - \partial_t^2 f(x,t) = 0 \tag{13.1}$$

This equation describes the vibration of a string: at time t, the shape of the string is the graph of the function $x \in [a,b] \mapsto f(x,t) \in \mathbb{R}$. (Here, we assume that the two end points of the string are $(a,0)$ and $(b,0)$.)

Using the trick

$$\partial_x^2 - \partial_t^2 = 4\partial_u\partial_v \qquad \text{where} \qquad u = x+t, v = x-t$$

it is not hard to find the general solution of the wave equation:

$$f(x,t) = \frac{1}{2}\big(g(x+t) + g(x-t)\big) + \frac{1}{2}\int_{x-t}^{x+t} h(s)ds \tag{13.2}$$

where $g(x) = f(x,0)$ and $h(x) = \partial_t f(x,0)$. In particular, if we assume that $h = 0$ (i.e., at time $t = 0$ the string is held in place and does not vibrate), then the solution is

$$f(x,t) = \frac{1}{2}\big(g(x+t) + g(x-t)\big) \tag{13.3}$$

Here comes the trouble. If we assume that the two end points of the string are always pinned at $(0,0)$ and $(1,0)$ respectively, then we have $f(0,t) = f(1,t) = 0$ for all $t \in \mathbb{R}$. Translating this condition to g, we get $g(t) = -g(-t) = -g(2-t)$ and hence g is a function with period 2, totally unacceptable! Worse still, the derivatives of periodic functions might have points of discontinuities.

13.1.3 Fourier series

The next important progress was made by Fourier in the study of **heat equation**

$$\partial_t f(x,t) - \partial_x^2 f(x,t) = 0 \tag{13.4}$$

where x is defined on a closed interval (say $[-\pi, \pi]$) representing a thin rod, $f(x,t)$ is the temperature of the point x of this rod at time t. Fourier solved the problem by **separation of variables**: He first assumed $f(x,t) = u(x)v(t)$. Then the heat equation implies $u''(x)v(t) = u(x)v'(t)$, and hence

$$\frac{u''(x)}{u(x)} = \frac{v'(t)}{v(t)} \tag{13.5}$$

The LHS is independent of t, and the RHS is independ of x. So (13.5) should be a constant $-\lambda$. The solution is then $u(x) = e^{i\sqrt{\lambda}x}$ and $v(t) = e^{-\lambda t}$. If one assumes that the temperature at the two end points $-\pi, \pi$ are equal, then $\sqrt{\lambda}$ must be an integer $n \in \mathbb{Z}$. So $f(x,t) = e^{inx-n^2 t}$. Taking infinite linear combinations, Fourier found the general solution

$$f(x,t) = \sum_{n=-\infty}^{+\infty} a_n e^{inx-n^2 t} \tag{13.6}$$

The initial temperature $g(x) = f(x,0)$ is a **Fourier series**

$$g(x) = \sum_{n=-\infty}^{+\infty} a_n e^{inx} \tag{13.7}$$

and, in particular, a function with period 2π (since each e^{inx} is so).

In order to use (13.6) to determine the solution $f(x,t)$ when the initial condition $f(x,t) = g(x)$ is given, one must first find the values of these **Fourier coefficients** a_n in terms of g. In fact, since

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ikx} \cdot e^{-inx} dx = \delta_{k,n} \tag{13.8}$$

(recall (1.2)), it is not hard to guess the formula

$$a_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} g(x) e^{-inx} dx \tag{13.9}$$

since this formula is true when $g(x) = e^{ikx}$.

13.1.4   From Riemann integrals to Lebesgue integrals

Here comes the question that is closely related to integral theory: What is the meaning of the integral (13.9) if g is no longer a real analytic function?

To be more precise, in 19th century there was a long debate about what are (good) functions, and what functions have **Fourier expansions**, i.e., expansions of the form (13.7). Fourier himself believed that all "functions" can be approximated by trigonometric functions (equivalently, functions of the form e^{inx}). But Lagrange did not, partly due to his (and many people's) insistence that functions must be real analytic (cf. [Kli, Sec. 28.2]). The doubt on whether many functions have Fourier expansions is quite understandable. From the modern perspective, we know that many Lebesgue measurable functions are not approximated by their Fourier series pointwise or uniformly, but are approximated under some

other norms (e.g. the L^2-norm $\|f\|_{L^2} = \sqrt{\int_{-\pi}^{\pi} |f|^2}$). Indeed, there are many continuous functions $[-\pi, \pi] \to \mathbb{R}$ whose Fourier series diverge on a dense subset of $[-\pi, \pi]$ (cf. [Rud-R, Thm. 5.12]). Therefore, Fourier's pioneering view that all "reasonable" functions can be approximated by Fourier series is incorrect unless we define what "approximation" means in a new and appropriate way.

In order to understand which functions have Fourier series expansions (in whatever sense), the first step is to understand for which function g the integral (13.9) makes sense. Therefore, the history of extending the class of integrable functions from continuous functions to Riemann integrable functions and finally to Lebesgue integrable functions is also part of the history of understanding which functions can have Fourier expansions and which functions are "reasonable".

The goal of this chapter is to learn Riemann integrals. The construction of Riemann integrals is much easier than Lebesgue theory. But compared to the latter, Riemann integrals have some serious drawbacks.

For example, suppose that a sequence of functions (f_n) on $[-\pi, \pi]$ converges to f in some sense. It is natural to ask whether the Fourier coefficients of (f_n) converge to those of f. (A typical example is $f_n(x) = \sum_{k=-n}^{n} a_k e^{ikx}$. Then this question asks whether the n-th Fourier coefficient of the series $\sum_{k=-\infty}^{+\infty} a_k e^{ikx}$ is a_n.) In view of (13.9), this problem is reduced to the problem of showing $\int f = \lim_n \int f_n$. If (f_n) is a sequence of Riemann integrable functions converging uniformly to f, then f is Riemann integrable, and $\int f = \lim \int f_n$. (See Cor. 13.21.) However, if (f_n) only converges pointwise to f, then f is not necessarily Riemann integrable. Even if f is Riemann integrable, one does not have a useful criterion for $\int f = \lim \int f_n$ in the framework of Riemann integrals.

But uniform convergence often does not hold in application, and especially in Fourier theory and PDEs. For example, let f be the function on $\mathbb{R}$ with period 2π defined by $f(x) = x$ if $-\pi < x < \pi$ and $f(\pi) = 0$. Then the Fourier series

$$\sum_{n=1}^{\infty} (-1)^{n-1} \frac{2}{n} \sin(nx) \qquad (13.10)$$

converges pointwise to f. (See also Pb. 4.2.) It does not converge uniformly to f, because the uniform limit of a sequence of continuous functions is continuous, but f is not continuous.

Lebesgue's integral theory will provide a more satisfying answer to the above problem. We will learn it in the second semester.

13.2   Riemann integrability and oscillation

In this section, we fix a Banach space V over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, and let $I = [a, b]$ where $-\infty < a < b < +\infty$. We understand $|I|$ as $b - a$.

13.2.1   Riemann integrals

Definition 13.1. A **partition** of the interval $I = [a, b]$ is defined to be an element of the form

$$\sigma = \{a_0, a_1, \ldots, a_n \in I : a_0 = a < a_1 < a_2 < \cdots < a_n = b, n \in \mathbb{Z}_+\} \qquad (13.11)$$

Equivalently, this partition can be written as

$$I = I_1 \cup I_2 \cup \cdots \cup I_n \qquad I_j = [a_{j-1}, a_j] \qquad (13.12)$$

If $\sigma, \sigma' \in \text{fin}(2^I)$ are partitions of I, we say that σ' is a **refinement** of σ (or that σ' is **finer than** σ), if $\sigma \subset \sigma'$. In this case, we also write

$$\sigma \prec \sigma'$$

We define $\mathcal{P}(I)$ or simply $\mathcal{P}$ to be

$$\mathcal{P}(I) = \{\text{partitions of } I\}$$

Remark 13.2. If $\sigma, \sigma' \in \mathcal{P}(I)$, then clearly $\sigma \cup \sigma' \in \mathcal{P}(I)$ and $\sigma, \sigma' \prec \sigma \cup \sigma$. Therefore, $\prec$ is a partial order on $\mathcal{P}(I)$. We call $\sigma \cup \sigma'$ the **common refinement** of σ and σ'.

Definition 13.3. A **tagged partition** of I is an ordered pair

$$(\sigma, \xi_\bullet) = \left(\{a_0 = a < a_1 < \cdots < a_n = b\}, (\xi_1, \ldots, \xi_n)\right)$$

where $\sigma \in \mathcal{P}(I)$ and $\xi_j \in I_j = [a_{j-1}, a_j]$ for all $1 \leqslant j \leqslant n$. The set

$$\mathcal{Q}(I) = \{\text{tagged partitions of } I\}$$

equipped with the preorder $\prec$ defined by

$$(\sigma, \xi_\bullet) \prec (\sigma', \xi'_\bullet) \qquad \Longleftrightarrow \qquad \sigma \subset \sigma' \qquad (13.13)$$

is a directed set.

Definition 13.4. Let $f : I \to V$. For each $(\sigma, \xi_\bullet) \in \mathcal{Q}(I)$, define the **Riemann sum**

$$S(f, \sigma, \xi_\bullet) = \sum_{j \geqslant 1} f(\xi_j)(a_j - a_{j-1})$$

The **Riemann integral** is defined to be the limit of the net $(S(f, \sigma, \xi_\bullet))_{(\sigma, \xi_\bullet) \in \mathcal{Q}(I)}$:

$$\int_a^b f \equiv \int_a^b f(x)dx = \lim_{(\sigma, \xi_\bullet) \in \mathcal{Q}(I)} S(f, \sigma, \xi_\bullet)$$

When the RHS exists, we way that f is **Riemann integrable** on I. We let

$$\mathscr{R}(I, V) = \{\text{Riemann integrable } f \in V^I\}$$

Convention 13.5. There are several equivalent ways to write the integrals:

$$\int_a^b f = -\int_b^a f = \int_{[a,b]} f$$

if $a < b$. Also, if $a = b$, the above terms are understood to be 0, and all functions on $[a, b]$ are considered Riemann integrable.

Remark 13.6. It is clear that $\int_a^b f = v \in V$ iff for every $\varepsilon > 0$ there exists $\sigma_0 \in \mathcal{P}(I)$ such that for any partition $\sigma = \{a_0 < \cdots < a_n\}$ finer than σ_0, and for any $\xi_j \in [a_{j-1}, a_j]$ (for all $1 \leqslant j \leqslant n$) we have

$$\left\| v - \sum_{j \geqslant 1} f(\xi_j)(a_j - a_{j-1}) \right\| < \varepsilon$$

There is not need to tag σ_0.

13.2.2 Riemann integrability and strong Riemann integrability

In this subsection, we give a useful criterion for Riemann integrability.

Definition 13.7. Let A be a nonempty subset of a metric space Y. The **diameter** of A is defined to be

$$\mathrm{diam}(A) = \sup_{x,y \in A} d(x, y)$$

If $f : X \to Y$ is a map where X is a set, and if $E \subset X$, the **oscillation** of f on E is defined to be $\mathrm{diam} f(E)$.

The following lemma allows us to control the difference of Riemann sums by means of the oscillation.

Lemma 13.8. *Let* $f : I = [a, b] \to V$ *and* $M = \mathrm{diam}(f(I))$. *Then for each* $(\sigma, \xi_\bullet), (\sigma', \xi'_\bullet) \in \mathcal{Q}(I)$ *we have*

$$\| S(f, \sigma, \xi_\bullet) - S(f, \sigma', \xi'_\bullet) \| \leqslant M(b - a) \tag{13.14}$$

Proof. Write $\sigma \cup \sigma' = \{a_0 = a < a_1 < \cdots < a_n = b\}$. Then there exist $\gamma_1, \ldots, \gamma_n \in \{\xi_1, \xi_2, \ldots\}$ such that

$$S(f, \sigma, \xi_\bullet) = \sum_{i=1}^n f(\gamma_i)(a_i - a_{i-1})$$

(Here is how to choose γ_i: Let k be the unique number such that the k-th subinterval of σ contains $[a_{i-1}, a_i]$. Then let $\gamma_i = \xi_k$.) Similarly, there exist $\gamma_1', \ldots, \gamma_n' \in \{\xi_1', \xi_2', \ldots\}$ such that

$$S(f, \sigma', \xi_\bullet') = \sum_{i=1}^{n} f(\gamma_i')(a_i - a_{i-1})$$

Then

$$\|S(f, \sigma, \xi_\bullet) - S(f, \sigma', \xi_\bullet')\| \leqslant \sum_{i=1}^{n} \|f(\gamma_i) - f(\gamma_i')\| \cdot (a_i - a_{i-1})$$

$$\leqslant \sum_{i=1}^{n} M(a_i - a_{i-1}) \leqslant M(b - a)$$

$\square$

Definition 13.9. Let $f : I \to V$. For each $\sigma = \{a_0 < \cdots < a_n\} \in \mathcal{P}(I)$, write $I_i = [a_{i-1}, a_i]$ and define the **oscillation of f on σ** to be

$$\omega(f, \sigma) = \sum_{j=1}^{n} \mathrm{diam}(f(I_j)) \cdot (a_j - a_{j-1}) \tag{13.15}$$

Exercise 13.10. Show that if $\sigma' \supset \sigma$, then $\omega(f, \sigma') \leqslant \omega(f, \sigma)$. Therefore, $(\omega(f, \sigma))_{\sigma \in \mathcal{P}(I)}$ is a decreasing net in $\overline{\mathbb{R}}_{\geqslant 0}$.

Now, Lem. 13.8 can be upgraded to the following version.

Lemma 13.11. Let $f : I \to V$. Let $\sigma \in \mathcal{P}(I)$. Choose $(\sigma', \xi_\bullet'), (\sigma'', \xi_\bullet'') \in \mathcal{Q}(I)$ such that $\sigma \subset \sigma'$ and $\sigma \subset \sigma''$. Then

$$\|S(f, \sigma', \xi_\bullet') - S(f, \sigma'', \xi_\bullet'')\| \leqslant \omega(f, \sigma) \tag{13.16}$$

Proof. Write $\sigma = \{a_0 < \cdots < a_n\}$ and $I_j = [a_{j-1}, a_j]$. Let $S(f, \sigma', \xi_\bullet')_{I_j}$ be the **restriction of** $S(f, \sigma', \xi_\bullet')$ **to** I_j. Namely,

$$S(f, \sigma', \xi_\bullet')_{I_j} = \sum_{\substack{\text{all } k \text{ such that} \\ [a_{k-1}, a_k] \subset I_j}} f(\xi_k)(a_k - a_{k-1}) \tag{13.17}$$

Define $S(f, \sigma'', \xi_\bullet'')_{I_j}$ in a similar way. Then, by Lem. 13.8 we have

$$\|S(f, \sigma', \xi_\bullet')_{I_j} - S(f, \sigma'', \xi_\bullet'')_{I_j}\| \leqslant \mathrm{diam}(f(I_j))|I_j|$$

Thus, (13.16) follows immediately from triangle inequality and

$$\sum_{j=1}^{n} S(f, \sigma', \xi_\bullet')_{I_j} = S(f, \sigma', \xi_\bullet') \qquad \sum_{j=1}^{n} S(f, \sigma'', \xi_\bullet'')_{I_j} = S(f, \sigma'', \xi_\bullet'')$$

$\square$

Definition 13.12. We say that $f : I \to V$ is **strongly Riemann integrable** if

$$\inf_{\sigma \in \mathcal{P}(I)} \omega(f, \sigma) = 0$$

Theorem 13.13. *Let $f : I \to V$. Consider the following statements:*

(1) $f \in \mathscr{R}(I, V)$.

(2) f is strongly Riemann integrable.

Then (2)$\Rightarrow$(1). If V is $\mathbb{R}^N$ or $\mathbb{C}^N$, then (1)$\Leftrightarrow$(2).

When V is infinite-dimensional, Riemann integrable functions are not necessarily strongly Riemann integrable. See Pb. 13.3.

Proof. Assume (2). Choose any $\varepsilon > 0$. Then there exists $\sigma \in \mathcal{P}(I)$ such that $\omega(f, \sigma) < \varepsilon$. By Lem. 13.11, for every $(\sigma', \xi'_\bullet), (\sigma'', \xi''_\bullet) \in \mathcal{Q}(I)$ satisfying $\sigma \subset \sigma'$ and $\sigma \subset \sigma''$, we have $\|S(f, \sigma', \xi'_\bullet) - S(f, \sigma'', \xi''_\bullet)\| < \varepsilon$. This proves that $(S(f, \sigma, \xi_\bullet))_{(\sigma, \xi_\bullet) \in \mathcal{Q}(I)}$ is a Cauchy net in V, and hence $f \in \mathscr{R}(I, V)$.

Now assume that V is $\mathbb{R}^N$ or $\mathbb{C}^N$. Since $\mathbb{C}^N \simeq \mathbb{R}^{2N}$, it suffices to consider the case $V = \mathbb{R}^N$. Since a net in $\mathbb{R}^N$ converges iff each component of this net converge, and since the strong Riemann integrability can be checked componentwisely, it suffices to prove (1)$\Rightarrow$(2) for the case $V = \mathbb{R}$.

So let us assume $f \in \mathscr{R}(I, \mathbb{R})$. Then $(S(f, \sigma, \xi_\bullet))_{(\sigma, \xi_\bullet) \in \mathcal{Q}(I)}$ is a Cauchy net in $\mathbb{R}$. Thus, there exists $\sigma = \{a_0 < \cdots < a_n\} \in \mathcal{P}(I)$ such that for all $(\sigma', \xi'_\bullet), (\sigma'', \xi''_\bullet) \in \mathcal{Q}(I)$ satisfying $\sigma \subset \sigma'$ and $\sigma \subset \sigma''$, we have $\|S(f, \sigma', \xi'_\bullet) - S(f, \sigma'', \xi''_\bullet)\| < \varepsilon$. To prove (2), we only need to take $\sigma' = \sigma'' = \sigma$. Thus, for any tags $\xi'_\bullet, \xi''_\bullet$ of σ, we have

$$\|S(f, \sigma, \xi''_\bullet) - S(f, \sigma, \xi'_\bullet)\| < \varepsilon$$

Write $I_j = [a_{j-1}, a_j]$. Then

$$\mathrm{diam}(f(I_j)) = \sup f(I_j) - \inf f(I_j)$$

So there exist $\xi'_j, \xi''_j \in I_j$ such that

$$f(\xi'_j) \leqslant \inf f(I_j) + \varepsilon \qquad f(\xi''_j) \geqslant \sup f(I_j) - \varepsilon$$

It follows that $\mathrm{diam}(f(I_j)) - 2\varepsilon \leqslant f(\xi''_j) - f(\xi'_j)$. Thus

$$\omega(f, \sigma) = \sum_j \mathrm{diam}(f(I_j)) \cdot |I_j| \leqslant \sum_j (f(\xi''_j) - f(\xi'_j)) \cdot |I_j| + \sum_j 2\varepsilon \cdot |I_j|$$

$$= S(f, \sigma, \xi''_\bullet) - S(f, \sigma, \xi'_\bullet) + 2(b - a)\varepsilon < \varepsilon + 2(b - a)\varepsilon$$

Since $\varepsilon > 0$ is arbitrary, we have $\inf_{\sigma \in \mathcal{P}(I)} \omega(f, \sigma) = 0$. $\qquad\square$

Example 13.14. Every continuous function $f : I \to V$ is strongly Riemann integrable, and hence Riemann integrable.

Proof. Since $I = [a,b]$ is compact, by Thm. 10.7, $f \in C(I,V)$ is uniformly continuous. Therefore, for every $\varepsilon > 0$, there exists $n \in \mathbb{Z}_+$ such that for all $x, y \in I$ satisfying $|x - y| \leqslant 1/n$, we have $\|f(x) - f(y)\| < \varepsilon$. Let $\sigma = \{a_0 < a_1 < \cdots < a_n\}$ be the partition of I such that $|I_j| = a_j - a_{j-1} = 1/n$. Then $\operatorname{diam}(f(I_j)) \leqslant \varepsilon$, and hence

$$\omega(f, \sigma) \leqslant \sum_j \varepsilon \cdot |I_j| = \varepsilon(b - a)$$

Since ε is arbitrary, we get $\inf_{\sigma \in \mathcal{P}(I)} \omega(f, \sigma) = 0$. $\square$

Example 13.15. The **Dirichlet function** $\chi_{\mathbb{Q}}$ is not (strongly) Riemann integrable on $I = [a,b]$, since for every $\sigma \in \mathcal{P}(I)$ we have $\omega(\chi_{\mathbb{Q}}, \sigma) = b - a$.

13.3 Basic properties of Riemann integrals

Let $I = [a,b]$ be a compact interval in $\mathbb{R}$. Let V be a Banach space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. We begin this section with the following fundamental fact, which will be used to prove Fubini's theorem, the second fundamental theorem calculus, and much more.

Theorem 13.16. *Let W be also a Banach space over $\mathbb{F}$. Let $T \in \mathfrak{L}(V, W)$. Then for every $f \in \mathscr{R}(I, V)$, we have $T \circ f \in \mathscr{R}(I, W)$ and*

$$T\left(\int_a^b f \right) = \int_a^b T \circ f \tag{13.18}$$

In other words, we have a commutative diagram

$$\begin{array}{ccc} \mathscr{R}(I,V) & \xrightarrow{\;T\circ\;} & \mathscr{R}(I,W) \\ {\scriptstyle \int_I} \downarrow & & \downarrow {\scriptstyle \int_I} \\ V & \xrightarrow{\quad T \quad} & W \end{array} \tag{13.19}$$

where the top arrow denotes the composition map $f \mapsto T \circ f$.

Proof. By linearity, we have

$$S(T \circ f, \sigma, \xi_\bullet) = T\big(S(f, \sigma, \xi_\bullet)\big) \tag{13.20}$$

Since T is continuous, the limit over $(\sigma, \xi_\bullet) \in \mathcal{Q}(I)$ of (13.20) is

$$T\Big(\lim_{(\sigma,\xi_\bullet) \in \mathcal{Q}(I)} S(f, \sigma, \xi_\bullet) \Big) = T\left(\int_a^b f \right)$$

This proves that $\lim S(T \circ f, \sigma, \xi_\bullet)$ converges to $T(\int_a^b f)$, i.e., $\int_a^b T \circ f$ exists and equals $T(\int_a^b f)$. $\square$

Remark 13.17. Let $f : I \to V$ and $\varepsilon > 0$. To simplify the following discussion, we say that f is ε-**dominated** by $\sigma \in \mathcal{P}(I)$, if for all $(\sigma', \xi'_\bullet), (\sigma'', \xi''_\bullet) \in \mathcal{Q}([a, b])$ satisfying $\sigma \subset \sigma', \sigma''$, we have $\|S(f, \sigma', \xi'_\bullet) - S(f, \sigma'', \xi''_\bullet)\| < \varepsilon$.

Thus, by Cauchy condition, if f is ε-dominated by some partition for every $\varepsilon > 0$, then $f \in \mathscr{R}(I, V)$. Moreover,

$$\left\| \int_a^b f - S(f, \sigma, \xi_\bullet) \right\| \leqslant \varepsilon \tag{13.21}$$

for every $(\sigma, \xi_\bullet) \in \mathcal{Q}(I)$ such that f is ε-dominated by σ. $\qquad\square$

13.3.1 The bounded linear map of integral operator

Proposition 13.18. *Let $f \in \mathscr{R}(I, V)$ and $g \in \mathscr{R}(I, \mathbb{R})$. Assume that $|f| \leqslant g$, i.e., $\|f(x)\| \leqslant g(x)$ for all $x \in I$. Then $\left\| \int_a^b f \right\| \leqslant \int_a^b g$.*

Proof. Apply $\lim_{(\sigma, \xi_\bullet) \in \mathcal{Q}(I)}$ to the obvious inequality $\|S(f, \sigma, \xi_\bullet)\| \leqslant S(g, \sigma, \xi_\bullet)$. $\qquad\square$

Corollary 13.19. *Assume that $f : I \to V$ is strongly Riemann integrable. Then $|f| : I \to \mathbb{R}$ is (strongly) Riemann integrable, and $\left\| \int_a^b f \right\| \leqslant \int_a^b |f|$*

Proof. By triangle inequality, for every $\sigma \in \mathcal{P}(I)$ we have $\omega(|f|, \sigma) \leqslant \omega(f, \sigma)$. So $|f|$ is strongly integrable. The rest of the corollary follows from Prop. 13.18. $\qquad\square$

Theorem 13.20. *$\mathscr{R}(I, V)$ is a closed linear subspace of $l^\infty(I, V)$. So $\mathscr{R}(I, V)$ is a Banach space under the l^∞-norm. Moreover, the map*

$$\int : \mathscr{R}(I, V) \to V \qquad f \mapsto \int_a^b f \tag{13.22}$$

is a bounded linear map with operator norm $b - a$ if we equip $\mathscr{R}(I, V)$ with the l^∞-norm.

Proof. By the basic properties of limits of nets, we know that if $f, g \in \mathscr{R}(I, V)$ and $\alpha, \beta \in \mathbb{F}$, then $\alpha f + \beta g \in \mathscr{R}(I, V)$, and

$$\int_a^b (\alpha f + \beta g) = \alpha \int_a^b f + \beta \int_a^b g \tag{13.23}$$

This proves that $\mathscr{R}(I, V)$ is a linear subspace of V^I, and that (13.22) is linear.

Let us prove $\mathscr{R}(I, V) \subset l^\infty(I, V)$. Choose $f \in \mathscr{R}(I, V)$. Then f is 1-dominated by some $\sigma = \{a_0 < \cdots < a_n\} \in \mathcal{P}(I)$. Fix any tag $\xi_\bullet$ on σ. Choose any $x \in X$. Let $[a_{i-1}, a_i]$ be the subinterval containing x. Let $\eta_\bullet = (\xi_1, \ldots, \xi_{i-1}, x, \xi_{i+1}, \ldots, \xi_n)$. Then $\|S(f, \sigma, \eta_\bullet) - S(f, \sigma, \xi_\bullet)\| < 1$ implies that $\|f(x) - f(\xi_i)\| \leqslant 1/(a_i - a_{i-1})$. So

$$\|f\|_{l^\infty} \leqslant \max \left\{ \|f(\xi_i)\| + (a_i - a_{i-1})^{-1} : 1 \leqslant i \leqslant n \right\} < +\infty$$

Let (f_n) be a sequence in $\mathscr{R}(I, V)$ converging to $f \in l^\infty(I, V)$. Choose any $\varepsilon > 0$. Then there is n such that $\|f - f_n\|_\infty < \varepsilon$. Since f_n is Riemann integrable, f_n is ε-dominated by some $\sigma \in \mathcal{P}(I)$. By triangle inequality, f is $(\varepsilon + (b-a)\varepsilon)$-dominated by σ. Since ε is arbitrary, we conclude that $f \in \mathscr{R}(I, V)$. This proves that $\mathscr{R}(I, V)$ is closed, and hence is Banach by Prop. 3.27.

Let us prove the claim about the operator norm. Choose any $f \in \mathscr{R}(I, V)$. Let $M = \|f\|_\infty < +\infty$. Then $|f| \leqslant M$. It is easy to see that $\int_a^b M = M(b-a)$. Thus, by Prop. 13.18, we have $\|\int_a^b f\| \leqslant M(b-a) = \|f\|_\infty \cdot (b-a)$, where "$\leqslant$" becomes "$=$" if we let f be a nonzero constant function. This proves that (13.22) has operator norm $b - a$ thanks to Rem. 10.24. $\qquad \square$

Corollary 13.21. *Let $(f_\alpha)_{\alpha \in \mathscr{I}}$ be a net in $\mathscr{R}(I, V)$ converging uniformly to $f \in V^I$. Then $f \in \mathscr{R}(I, V)$ and $\lim_{\alpha \in \mathscr{I}} \int_a^b f_\alpha = \int_a^b f$.*

Proof. This is immediate from Thm. 13.20, which implies that $\mathscr{R}(I, V)$ is closed in $l^\infty(I, V)$ (and hence closed in V^I since $l^\infty(I, V)$ is closed in V^I), and that the map (13.22) is continuous. $\qquad \square$

13.3.2 Some criteria for Riemann integrability

Proposition 13.22. *Let $f, g : I \to V$. Suppose that $\{x \in I : f(x) \neq g(x)\}$ is a finite set. Suppose that $f \in \mathscr{R}(I, V)$. Then $g \in \mathscr{R}(I, V)$, and $\int_a^b f = \int_a^b g$.*

Proof. By Thm. 13.20, it suffices to prove that $g - f$ is Riemann integrable and $\int_a^b (g - f) = 0$. This is easy to show, since $g - f$ is zero outside finitely many points. $\qquad \square$

Proposition 13.23. *Let $f : [a, b] \to V$. Let $c \in [a, b]$. Then $f \in \mathscr{R}([a, b], V)$ iff $f|_{[a,c]} \in \mathscr{R}([a, c], V)$ and $f|_{[c,b]} \in \mathscr{R}([c, b], V)$. Moreover, if $f \in \mathscr{R}([a, b], V)$, then*

$$\int_a^b f = \int_a^c f + \int_c^b f \tag{13.24}$$

Proof. This is obvious when $c = a$ or $c = b$ (recall Conv. 13.5). So we assume $a < c < b$.

First, we assume $f \in \mathscr{R}([a, b], V)$. By Cauchy condition, for each $\varepsilon > 0$, f is ε-dominated by some $\sigma \in \mathcal{P}(I)$. By enlarging σ, we assume that $c \in \sigma$. Then it is easy to see that $f|_{[a,c]}$ is ε-dominated by $\sigma \cap [a, c]$, and $f|_{[c,b]}$ is ε-dominated by $\sigma \cap [c, b]$. So $f|_{[a,c]}$ and $f|_{[c,b]}$ are Riemann integrable.

Now assume that $f|_{[a,c]}$ and $f|_{[c,b]}$ are Riemann integrable. Choose any $\varepsilon > 0$. Then $f|_{[a,c]}$ is ε-dominated by some $\tau \in \mathcal{P}([a, c])$, and $f|_{[c,b]}$ is ε-dominated by some $\varrho \in \mathcal{P}([c, b])$. Then f is 2ε-dominated by $\sigma = \tau \cup \varrho$. This proves that $f \in$

$\mathscr{R}(I, V)$. Let $\alpha_\bullet$ be a tag on $[a, c]$, and let $\beta_\bullet$ be a tag on $[c, b]$. Then $\xi_\bullet = (\alpha_\bullet, \beta_\bullet) = (\alpha_1, \alpha_2, \ldots, \beta_1, \beta_2, \ldots)$ is a tag on $[a, b]$, and

$$S(f, \sigma, \xi_\bullet) = S(f|_{[a,c]}, \tau, \alpha_\bullet) + S(f|_{[c,b]}, \varrho, \beta_\bullet)$$

By Rem. 13.17, we have

$$\left\| \int_a^b f - S(f, \sigma, \xi_\bullet) \right\| \leqslant 2\varepsilon$$

$$\left\| \int_a^c f - S(f|_{[a,c]}, \tau, \alpha_\bullet) \right\| \leqslant \varepsilon$$

$$\left\| \int_c^b f - S(f|_{[c,b]}, \varrho, \beta_\bullet) \right\| \leqslant \varepsilon$$

Therefore, the difference of the LHS and the RHS of (13.24) has norm $\leqslant 4\varepsilon$. This proves (13.24) since ε can be arbitrary. $\square$

Example 13.24. Let $f \in l^\infty(I, V)$. Suppose that there exist $\sigma = \{a_0 < a_1 < \cdots < a_n\} \in \mathcal{P}(I)$ such that $f|_{(a_{j-1}, a_j)} : (a_{j-1}, a_j) \to V$ is continuous for each $1 \leqslant j \leqslant n$. Then $f \in \mathscr{R}(I, V)$.

Proof. By Prop. 13.23, it suffices to prove that each $f|_{[a_{j-1}, a_j]}$ is Riemann integrable. Thus, we assume WLOG that $M := \|f\|_\infty < +\infty$, and that f is continuous when restricted to (a, b). Choose any $\varepsilon > 0$. Choose $\delta > 0$ such that $M\delta < \varepsilon$ and $a + \delta < b - \delta$. Let $J = [a + \delta, b - \delta]$. Then $f|_J$ is continuous, and hence Riemann integrable by Exp. 13.14. So $f|_J$ is ε-dominated by some $\varrho \in \mathcal{P}(J)$. Since $\operatorname{diam}(f(I)) \leqslant 2M$, by Lem. 13.8, $f|_{[a, a+\delta]}$ is 2ε-dominated by $\{a, a + \delta\}$, and $f|_{[b-\delta, b]}$ is 2ε-dominated by $\{b - \delta, b\}$. So f is 5ε-dominated by $\sigma = \varrho \cup \{a, b\}$. Since $\varepsilon > 0$ is arbitrary, we conclude $f \in \mathscr{R}(I, V)$ by Cauchy condition. $\square$

Example 13.25. The function $f : [0, 1] \to \mathbb{R}$ defined by $f(x) = \sin(1/x)$ if $0 < x \leqslant 1$ and $f(0) = 0$ is Riemann integrable, although f is not uniformly continuous on $(0, 1]$.

Example 13.26. Let $I_1, \ldots, I_n$ be intervals inside I. Choose $v_1, \ldots, v_n \in V$ and set $f = \sum_{j=1}^n v_j \chi_{I_j}$. Then $f \in \mathscr{R}(I, V)$.

Proof. The case of arbitrary n follows from the case $n = 1$ by linearity and Thm. 13.20. Assume $n = 1$, and let $c = \inf I_1$ and $d = \sup I_1$. Then the restriction of f to $[a, c]$ (resp. $[c, d]$ and $[d, b]$) equals a constant function except possibly at the end points of the interval. So $f|_{[a,c]}, f|_{[c,d]}, f|_{[d,b]}$ are Riemann integrable by Prop. 13.22. So f is Riemann integrable by Prop. 13.23. $\square$

Example 13.27. Let $f \in C(I, V)$. For each $n \in \mathbb{Z}_+$, choose a tag $\xi_{\bullet, n}$ for the partition $\sigma = \{a, a + |I|/n, a + 2|I|/n, \ldots, a + (n-1)|I|/n, b\}$ of I. Then

$$\lim_{n \to \infty} \frac{b - a}{n} \sum_{i=1}^n f(\xi_{i,n}) = \int_a^b f \qquad (13.25)$$

Proof. Let $f_n = \sum_{i=1}^{n} f(\xi_{i,n}) \cdot \chi_{J_{i,n}}$ where $J_{i,n} = \left(a + \frac{i-1}{n}|I|, a + \frac{i}{n}|I|\right]$ if $i > 1$, and $J_{1,n} = [a, a + |I|/n]$. Then $f_n \in \mathscr{R}(I, V)$ by Exp. 13.26. Moreover, by Prop. 13.23, the LHS of (13.25) equals $\lim_n \int_a^b f_n$.

By Thm. 10.7, f is uniformly continuous. So for every $\varepsilon > 0$ there exists $N \in \mathbb{Z}_+$ such that for all $n \geqslant N$ and all $x, y \in \mathrm{Cl}_{\mathbb{R}}(J_{i,n})$ we have $\|f(x) - f(y)\| < \varepsilon$. Thus, for all $n \geqslant N$ we have $\|f - f_n\|_{l^\infty} < \varepsilon$. Therefore $f_n \rightrightarrows f$. Thus, by Cor. 13.21 we have $\int_a^b f_n \to \int_a^b f$. This proves (13.25). $\qquad\square$

Example 13.28. By Thm. 13.30, we have $\int_1^2 x^{-1}dx = \log 2$. Thus, by Exp. 13.27, we have $\lim_{n\to\infty} \sum_{i=1}^{n} \frac{2-1}{n} \cdot (1 + i/n)^{-1} = \log 2$, namely

$$\lim_{n\to\infty} \left((n+1)^{-1} + (n+2)^{-1} + \cdots + (2n)^{-1}\right) = \log 2$$

13.4   Integrals and derivatives

Let $I = [a, b]$ be an interval in $\mathbb{R}$. Let V be a Banach space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$.

13.4.1   Fundamental theorems of calculus (FTC)

There are two versions of FTC. Roughly speaking, the first FTC says that integrals give antiderivatives. The second FTC says that antiderivatives give integrals. These two FTC are equivalent when the function f to be integrated is continuous. Otherwise, there is a subtle difference (which I can never remember) between these two theorems.

Theorem 13.29 (First FTC). *Let $f \in \mathscr{R}(I, V)$. Define*

$$F : I \to V \qquad F(x) = \int_a^x f \tag{13.26}$$

Then $F \in C(I, V)$. If f is continuous at x, then $F'(x) = f(x)$.

In particular, if $f \in C(I, V)$, then $F' = f$. Thus, by Cor. 11.24, the antiderivatives of f are precisely of the form $F(x) + v_0$ where $v_0 \in V$ is viewed as a constant function.

Recall Conv. 13.5.

Proof. Recall by Thm. 13.20 that $\|f\|_\infty < +\infty$. For each $x, y \in [a, b]$ we have

$$\|F(y) - F(x)\| = \left\| \int_a^y f - \int_a^x f \right\| = \left\| \int_x^y f \right\| \leqslant \|f\|_\infty \cdot |y - x|$$

So F has Lipschitz constant $\|f\|_\infty$. Now suppose that f is continuous at x. Then for every $\varepsilon > 0$, there exists $U \in \mathrm{Nbh}_I(x)$ such that $\|f(x) - f(y)\| < \varepsilon$ for every $y \in U$. Thus, for each $y \in U \backslash \{x\}$, since $\int_x^y f(x)dt = f(x)(y - x)$, we have

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

$$\leqslant |y - x|^{-1} \int_{[x,y]} \|f(t) - f(x)\| dt \leqslant |y - x|^{-1} \int_{[x,y]} \varepsilon \, dt = \varepsilon$$

This proves $F'(x) = f(x)$. $\qquad\square$

Theorem 13.30 (Second FTC)**.** *Let $f \in \mathcal{R}(I, V)$. Assume that $F : I \to V$ is differentiable and $F' = f$. Then*

$$\int_a^b f = F|_a^b \overset{\mathrm{def}}{=\!=\!=} F(b) - F(a) \tag{13.27}$$

This theorem is easy when $f \in C(I, V)$: In this case, by Thm. 13.29, we have $F(x) = v_0 + \int_a^x f$ for some $v_0 \in V$. Then (13.27) follows immediately from Prop. 13.23. The proof for the general case is more difficult:

Proof assuming Hahn-Banach. Since $\mathbb{R}$ is a subfield of $\mathbb{C}$, we may view V as a real Banach space. We first consider the special case that $V = \mathbb{R}$. Let $A = \int_a^b f$. Choose any $\varepsilon > 0$. Since $f \in \mathcal{R}(I, \mathbb{R})$, there exists $\sigma = \{a_0 < \cdots < a_n\} \in \mathcal{P}(I)$ such that for every tag $\xi_\bullet$ on σ, we have $|A - S(f, \sigma, \xi_\bullet)| < \varepsilon$. By Lagrange's MVT (Thm. 11.19), there exists $\xi_i \in (a_{i-1}, a_i)$ such that

$$F(a_i) - F(a_{i-1}) = f(\xi_i)(a_i - a_{i-1})$$

Thus, we have a tag $\xi_\bullet$ such that $S(f, \sigma, \xi_\bullet) = F(b) - F(a)$. Hence $|A - F(b) + F(a)| < \varepsilon$. Since ε is arbitrary, we get $A = F(b) - F(a)$.

The case $V = \mathbb{R}^N$ can be reduced to the above special case easily. We now consider the general case that V is a Banach space over $\mathbb{R}$. Similar to (12.15), for every $\varphi \in V^* = \mathfrak{L}(V, \mathbb{R})$, we have that $\varphi \circ F$ is differentiable, and that $(\varphi \circ F)' = \varphi \circ f$. Note that $\varphi \circ f \in \mathcal{R}(I, \mathbb{R})$ by Thm. 13.16. Apply the one-dimensional special case to $\varphi \circ f$. Then by Thm. 13.16, we have

$$\varphi\left(\int_a^b f\right) = \int_a^b \varphi \circ f = \varphi \circ F|_a^b = \varphi(F(b) - F(a))$$

By Hahn-Banach theorem, V^* separates points of V. (See Rem. 11.26.) Therefore $\int_a^b f = F(b) - F(a)$. $\qquad\square$

13.4.2 Applications of FTC: integration by parts

Proposition 13.31 (Integration by parts)**.** *Let $f \in C^1(I, V)$ and $g \in C^1(I, \mathbb{F})$. Then*

$$\int_a^b f'g = (fg)|_a^b - \int_a^b fg' \tag{13.28}$$

Proof. $(fg)|_a^b = \int_a^b (fg)' = \int_a^b f'g + \int_a^b fg'$. $\qquad\square$

Theorem 13.32 (Taylor's theorem, integral form). *Let* $n \in \mathbb{N}$ *and* $f \in C^{n+1}([a, b], V)$. *Then for every* $x \in [a, b]$ *we have*

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x - a)^k + \int_a^x \frac{f^{(n+1)}(t)}{n!}(x - t)^n dt \qquad (13.29)$$

Proof. When $n = 0$, (13.29) is just FTC. We now prove (13.29) by induction. Suppose that case $n - 1$ has been proved, then by integration by parts,

$$f(x) - \sum_{k=0}^{n-1} \frac{f^{(k)}(a)}{k!}(x - a)^k = \int_a^x \frac{f^{(n)}(t)}{(n-1)!}(x - t)^{n-1} dt$$

$$= \int_a^x -\frac{f^{(n)}(t)}{n!}\partial_t(x - t)^n dt$$

$$= -\frac{f^{(n)}(t)}{n!}(x - t)^n \Big|_{t=a}^x + \int_a^x \frac{f^{(n+1)}(t)}{n!}(x - t)^n dt$$

$$= \frac{f^{(n)}(a)}{n!}(x - a)^n + \int_a^x \frac{f^{(n+1)}(t)}{n!}(x - t)^n dt$$

$\square$

Exercise 13.33. Use the integral form of Taylor's theorem to give a quick proof higher order finite-increment Thm. 12.17 under the assumption that $f \in C^{n+1}([a, b], V)$. (This assumption is slightly stronger than that in Thm. 12.17, but is enough for applications.)

In the case that $V = \mathbb{R}$, the integral form of Taylor's theorem actually implies a slightly weaker (but useful enough) version of Lagrange form. This relies on the following easy fact:

Proposition 13.34 (Mean value theorem). *Let* $f, g \in C([a, b], \mathbb{R})$ *such that* $g(x) \geqslant 0$ *for all* $x \in (a, b)$. *Then there exists* $\xi \in [a, b]$ *such that*

$$\int_a^b fg = f(\xi) \int_a^b g \qquad (13.30)$$

Moreover, ξ *can be chosen to be in* (a, b) *if* $g(x) > 0$ *for all* $x \in (a, b)$.

Proof. Let $m = \inf f(I)$ and $M = \sup f(I)$. Then $mg \leqslant fg \leqslant Mg$, and hence $m \int_I g \leqslant \int_I fg \leqslant \int_I Mg$. So there is $y \in [m, M]$ such that $\int_I fg = y \int_I g$. By extreme value theorem, we have $m, M \in f(I)$. Thus, by intermediate value property, we have $y \in f(I)$. So $y = f(\xi)$ for some $\xi \in I$.

Now assume $g(x) > 0$ for all $a < x < b$. Let $\varphi(x) = \int_a^x fg$ and $\psi(x) = \int_a^x g$. By Cauchy's MVT, there exists $\xi \in (a, b)$ such that $g(\xi)(\varphi(b) - \varphi(a)) = f(\xi)g(\xi)(\psi(b) - \psi(a))$. This finishes the proof. $\square$

Remark 13.35. Let $f \in C^{n+1}([a,b], \mathbb{R})$. By the above mean value theorem, for each $x \in (a,b]$ there exists $\xi \in (a,x)$ such that

$$\int_a^x \frac{f^{(n+1)}(t)}{n!}(x-t)^n dt = f^{(n+1)}(\xi) \int_a^x \frac{(x-t)^n}{n!} dt = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}$$

Thus, the integral form of Taylor's theorem implies Lagrange's form (Thm. 12.18) in the case that $f \in C^{n+1}(I, \mathbb{R})$.

13.4.3 Application of FTC: change of variables

Proposition 13.36 (Change of variables). *Let $\Phi \in C^1(I, \mathbb{R})$ such that $\Phi(I) \subset J = [c,d]$. Let $f \in C(J, V)$. Then*

$$\int_{\Phi(a)}^{\Phi(b)} f = \int_a^b (f \circ \Phi) \cdot \Phi' \tag{13.31}$$

Proof. Define $F : J \to V$ and $G : I \to V$ by $F(y) = \int_{\Phi(a)}^y f$ and $G(x) = F \circ \Phi(x) = \int_{\Phi(a)}^{\Phi(x)} f$. Then, by chain rule and FTC, $G' = (F' \circ \Phi) \cdot \Phi' = (f \circ \Phi) \cdot \Phi'$. By FTC, we have $G' = H'$ where $H : I \to V$ is defined by

$$H(x) = \int_a^x (f \circ \Phi) \cdot \Phi'$$

Thus, since $G(a) = F(a) = 0$, by Cor. 11.24 we have $G = H$. Then $G(b) = H(b)$ finishes the proof. $\qquad \square$

The change of variable formula allows us to define the length of a curve:

Definition 13.37. Let γ be a C^1-**curve** in V, i.e. $\gamma \in C^1([a,b], V)$. Its **length** is defined to be

$$l(\gamma) = \int_a^b \|\gamma'(t)\| dt$$

Proposition 13.38. *The length of γ is invariant under a reparametrization. Namely, if $f : [c,d] \to [a,b]$ is a bijection and is in $C^1([c,d], \mathbb{R})$, then $l(\gamma) = l(\gamma \circ f)$.*

Proof. By Prop. 11.22, f is either increasing or decreasing. We prove the case that f is decreasing; the other case is similar. Then $f(c) = b$ and $f(d) = a$. By chain rule, we have

$$l(\gamma \circ f) = \int_c^d |(\gamma \circ f)'| = \int_c^d |(\gamma' \circ f) \cdot f'| = \int_c^d (|\gamma'| \circ f) \cdot |f'|$$

which equals $-\int_c^d (|\gamma'| \circ f) \cdot f' = \int_d^c (|\gamma'| \circ f) \cdot f'$ because $f' \leqslant 0$ by Cor. 11.20. This expression equals $\int_a^b |\gamma'| = l(\gamma)$ by Prop. 13.36. $\qquad \square$

Example 13.39. The upper half circle $S_+^1 = \{z \in \mathbb{C} : |z| = 1, \mathrm{Im}\, z \geqslant 0\}$ has a bijective C^∞-parametrization $\gamma : [0, \pi] \to \mathbb{C}, \gamma(t) = e^{it}$. Its length is π.

Proof. We have $l(\gamma) = \int_0^\pi |\gamma'| = \int_0^\pi 1 = \pi$. So it remains to prove that γ restricts to a bijection $[0, \pi] \to S_+^1$. Clearly $\gamma([0, \pi]) \subset S^1 = \{z \in \mathbb{C} : |z| = 1\}$. Note that $\gamma(t) = \cos(t) + \mathbf{i}\sin(t)$. In Sec. 12.2, we have proved that $\sin(x) \geqslant 0$ when $x \in [0, \pi]$. So $\gamma([0, \pi]) \subset S_+^1$.

We have also proved in Sec. 12.2 that $\cos : [0, \pi/2] \to \mathbb{R}$ is a decreasing (continuous) function such that $\cos(0) = 1$ and $\cos(\pi/2) = 0$, and that $\cos(x) = \cos(\pi - x)$. The last relation shows that $\cos : [0, \pi] \to \mathbb{R}$ is decreasing, and $\cos(0) = 1, \cos(\pi) = -1$. Therefore, by the intermediate value theorem, we see that $\cos$ sends $[0, \pi]$ bijectively to $[-1, 1]$. Since the projection map onto the x-axis sends S_+^1 bijectively to $[-1, 1]$, we conclude that γ sends $[0, \pi]$ bijectively to S_+^1. $\qquad\square$

Remark 13.40. You may wonder if the above argument really proves that π is the length of S_+^1: Suppose that $\lambda : [a, b] \to \mathbb{C}$ is another C^∞ map restricting to a bijection $[a, b] \to S_+^1$, how can we show that $\int_0^\pi |\gamma'| = \int_a^b |\lambda'|$? Clearly, there is a bijection $f : [a, b] \to [0, \pi]$ such that $\lambda = \gamma \circ f$. Thus, by Prop. 13.38, the two integrals are equal if $f \in C^\infty$, or at least if $f \in C^1$. However, it seems that there is no general argument ensuring that $f \in C^1$.

In the next semester, we will learn that $f \in C^\infty$ if the two C^∞-parametrizations λ, γ satisfy that λ' and γ' are nowhere zero. (Clearly γ' is nowhere zero if $\gamma(t) = e^{it}$.) Such parametrizations are called **(smooth) immersions**. $\qquad\square$

13.5 Problems and supplementary material

Let $I = [a, b]$ where $-\infty < a < b < +\infty$.

Problem 13.1. Let $f \in l^\infty(I, \mathbb{R})$. Recall from Subsec. 13.2.1 that the refinements of partitions define preoders on $\mathcal{P}(I)$ and $\mathcal{Q}(I)$ so that they are directed sets. For each $\sigma = \{a_0 < a_1 < \cdots < a_n\} \in \mathcal{P}(I)$, define the **upper Darboux sum** and the **lower Darboux sum**

$$\overline{S}(f, \sigma) = \sum_{i=1}^n M_i(a_i - a_{i-1}) \qquad \underline{S}(f, \sigma) = \sum_{i=1}^n m_i(a_i - a_{i-1})$$

$$\text{where} \qquad M_i = \sup_{\xi \in [a_{i-1}, a_i]} f(\xi) \qquad m_i = \inf_{\xi \in [a_{i-1}, a_i]} f(\xi)$$

Prove that

$$-(b - a)\|f\|_\infty \leqslant \underline{S}(f, \sigma) < \overline{S}(f, \sigma) \leqslant (b - a)\|f\|_\infty$$

Choose any tag $\xi_\bullet$ on σ. Prove that

$$\overline{S}(f, \sigma) = \sup_{(\sigma', \xi_\bullet') > (\sigma, \xi_\bullet)} S(f, \sigma', \xi_\bullet') \qquad \underline{S}(f, \sigma) = \inf_{(\sigma', \xi_\bullet') > (\sigma, \xi_\bullet)} S(f, \sigma', \xi_\bullet') \qquad (13.32)$$

Pb. 13.1 immediately implies:

Theorem 13.41. *Let $f \in l^\infty(I, \mathbb{R})$. Define the **upper Darboux integral** and the **lower Darboux integral** to be*

$$\overline{\int_a^b} f = \inf_{\sigma \in \mathcal{P}(I)} \overline{S}(f, \sigma) \qquad \underline{\int_a^b} f = \sup_{\sigma \in \mathcal{P}(I)} \underline{S}(f, \sigma)$$

which are elements of $[-(b-a)\|f\|_\infty, (b-a)\|f\|_\infty]$. Then we have (recalling Pb. 8.2)

$$\overline{\int_a^b} f = \limsup_{(\sigma, \xi_\bullet) \in \mathcal{Q}(I)} S(f, \sigma, \xi_\bullet)$$

$$\underline{\int_a^b} f = \liminf_{(\sigma, \xi_\bullet) \in \mathcal{Q}(I)} S(f, \sigma, \xi_\bullet)$$

Therefore, by Cor. 8.37, we have

$$f \in \mathcal{R}(I, \mathbb{R}) \qquad \Longleftrightarrow \qquad \overline{\int_a^b} f = \underline{\int_a^b} f \tag{13.33}$$

Moreover, if $f \in \mathcal{R}(I, \mathbb{R})$, then $\int_a^b f = \overline{\int_a^b} f = \underline{\int_a^b} f$.

Problem 13.2. Let $\mathcal{V}$ be a vector space over $\mathbb{C}$. Since $\mathbb{R}$ is a subfield of $\mathbb{C}$, $\mathcal{V}$ can be viewed as a real normed vector space. Let $\Lambda : \mathcal{V} \to \mathbb{R}$ be a $\mathbb{R}$-linear map. Recall $\mathbf{i} = \sqrt{-1}$. Define a map

$$\Lambda_{\mathbb{C}} : \mathcal{V} \to \mathbb{C} \qquad \Lambda_{\mathbb{C}}(v) = \Lambda(v) - \mathbf{i}\Lambda(\mathbf{i}v) \tag{13.34}$$

1. Prove that $\Lambda_{\mathbb{C}}$ is $\mathbb{C}$-linear.

2. Given a $\mathbb{C}$-linear $\Phi : \mathcal{V} \to \mathbb{C}$, we define its **real part**

$$\mathrm{Re}\Phi : \mathcal{V} \to \mathbb{R} \qquad v \mapsto \mathrm{Re}\left(\Phi(v)\right) \tag{13.35}$$

   Then clearly $\mathrm{Re}(\Phi)$ is $\mathbb{R}$-linear. Prove that $\Phi \mapsto \mathrm{Re}\Phi$ is a bijection from the set of $\mathbb{C}$-linear maps $\mathcal{V} \to \mathbb{C}$ to the set of $\mathbb{R}$-linear maps $\mathcal{V} \to \mathbb{R}$, and that its inverse is the map $\Lambda \mapsto \Lambda_{\mathbb{C}}$ defined by (13.34).

3. Assume that $\mathcal{V}$ is a (non-necessarily complete) normed $\mathbb{C}$-vector space. For each $\mathbb{C}$-linear $\Phi : \mathcal{V} \to \mathbb{C}$, prove the following equation about operator norms:

$$\|\Phi\| = \|\mathrm{Re}\Phi\| \tag{13.36}$$

   (Hint: One of "$\leqslant$" and "$\geqslant$" is obvious. To prove the other one, for each $v \in \mathcal{V}$, find some $\theta \in \mathbb{R}$ such that $e^{\mathbf{i}\theta}\Phi(v) \in \mathbb{R}$.)

Remark 13.42. The above problem shows how to extend a real-valued integral to a complex-valued one. For example, suppose that we define real-valued Riemann integrals using Darboux integrals. Suppose we have proved that the map $f \in \mathscr{R}(I, \mathbb{R}) \mapsto \int_a^b f$ is $\mathbb{R}$-linear with operator norm $(b - a)$. Then, applying Pb. 13.2 to the $\mathbb{R}$-linear map

$$\Lambda : \mathscr{R}(I, \mathbb{C}) \to \mathbb{R} \qquad f \mapsto \int_a^b \operatorname{Re} f(t) dt$$

gives a $\mathbb{C}$-linear map

$$\int_a^b : \mathscr{R}(I, \mathbb{C}) \to \mathbb{C}$$

$$\int_a^b f = \int_a^b \operatorname{Re} f + \mathbf{i} \int_a^b \operatorname{Im} f$$

since $\operatorname{Re}(\mathbf{i} f) = -\operatorname{Im} f$. Moreover, this linear map has operator norm $(b - a)$. This defines the complex integral operator by means of real Darboux integrals. In the next semester, we will use the same method to extend real-valued Lebesgue integrals to complex-valued ones. Pb. 13.2 will also be used to prove the Hahn-Banach extension theorem.

Exercise 13.43. Let $u, v \in C(I, \mathbb{R})$. Find the $\mathbb{C}$-linear map $C(I, \mathbb{C}) \to \mathbb{C}$ whose real part is $f \in C(I, \mathbb{C}) \mapsto \int_I (u \operatorname{Re} f + v \operatorname{Im} f) \in \mathbb{R}$.

⋆ **Problem 13.3.** Let $V = l^\infty([0, 1], \mathbb{R})$, equipped with the l^∞-norm. Define

$$f : [0, 1] \to V \qquad f(x) = \chi_{[x,1]}$$

Then, for every $x \neq y$ in $[0, 1]$ we have $\|f(x) - f(y)\|_{l^\infty([0,1], \mathbb{R})} = 1$. This implies $\omega(f, \sigma) = 1$ for every $\sigma \in \mathcal{P}(I)$. So f is not strongly Riemann integrable on any closed subintegral of $[a, b]$.

Define $F : [0, 1] \to V$ such that for each $x \in [0, 1]$,

$$F(x) : [0, 1] \to \mathbb{R} \qquad t \mapsto \min\{t, x\}$$

Choose any $x \in [0, 1]$. Prove that $f \in \mathscr{R}([0, x], V)$ and $\int_0^x f = F(x)$. (In particular, $\int_0^1 f = \mathrm{id}_{[0,1]}$.) Prove that $F'(x)$ does not exist. □

Problem 13.4. Let V be a Banach space. Use the fundamental theorem of calculus to give another proof that $C^1(I, V)$ is complete under the $l^{1,\infty}$-norm. (Do not use Thm. 11.33.)

⋆ **Theorem 13.44 (Gronwall's inequality).** *Let $f \in C([a, b], \mathbb{R}_{\geqslant 0})$. Let $\alpha \in \mathbb{R}_{\geqslant 0}$, and $\beta \in C([a, b], \mathbb{R}_{\geqslant 0})$. Assume that for each $x \in [a, b]$ we have*

$$f(x) \leqslant \alpha + \int_a^x \beta(t) f(t) dt \tag{13.37}$$

Then for each $x \in [a, b]$ we have

$$f(x) \leqslant \alpha \cdot \exp\left(\int_a^x \beta(t)dt\right) \tag{13.38}$$

In particular, if β is a constant, then Gronwall's inequality reads

$$f(x) \leqslant \alpha \cdot e^{\beta(t-a)}$$

$\star$ **Problem 13.5.** Prove Gronwall's inequality. Hint: Let $g(x)$ be the RHS of (13.37). Show that $\exp(-\int_a^x \beta) \cdot g(x)$ is a decreasing function.

$\star$ **Remark 13.45.** Gronwall's inequality is often used in the following way. Let V be a Banach space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. Let $f \in C(I, V)$, $v \in V$, and $\beta \in C(I, \mathbb{R}_{\geqslant 0})$. Suppose that for all $x \in [a, b]$ we have

$$\|f(x) - v\| \leqslant \int_a^x \beta(t)\|f(t)\|dt \tag{13.39}$$

Applying Gronwall's inequality to $|f|$ and $\alpha = \|v\|$, we see that for every $x \in [a, b]$,

$$\|f(x)\| \leqslant \|v\| \cdot \exp\left(\int_a^x \beta(t)dt\right) \tag{13.40}$$

$\star$ **Problem 13.6.** Let V be a Banach space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. Assume that $\varphi \in C(I \times V, V)$ has Lipschitz constant $L \in \mathbb{R}_{\geqslant 0}$ over the second variable, i.e., for each $t \in [a, b]$ and $u, v \in V$ we have

$$\|\varphi(t, u) - \varphi(t, v)\| \leqslant L\|u - v\| \tag{13.41}$$

Use Gronwall's inequality to solve the following problems.

1. Let $f_1, f_2 : I \to V$ be differentiable and satisfying the differential equation

    $$f_i'(t) = \varphi(t, f_i(t)) \qquad (\forall t \in I)$$

    with the same initial condition $f_1(a) = f_2(a)$. Prove that $f_1 = f_2$ on I.

2. Let X be a topological space. Let $f : I \times X \to V$ be a function such that $\partial_1 f$ exists everywhere, and that

    $$\partial_1 f(t, x) = \varphi(t, f(t, x)) \qquad (\forall t \in I, x \in X)$$

    Assume that the function $f(a, \cdot) : X \to V$ (sending x to $f(a, x)$) is continuous. Prove that $f \in C(I \times X, V)$.

233

Remark 13.46. In practice, V is often $\mathbb{R}^N$, and φ is a "smooth function", i.e. a function whose (mixed) partial derivatives of all orders exist and are continuous. It is also common that φ is independent of $t \in I$. (Indeed, a t-dependent differential equation $f' = \varphi(t, f)$ can be transformed into a t-independent one $(t, f)' = (1, \varphi(t, f))$.) However, sometimes f is not defined on $\mathbb{R}^N$, but on a closed subset of $\mathbb{R}^N$, for example, on a closed ball. In this case, the uniqueness and the continuity of the solutions of differential equations can be treated by extending φ to a smooth function on $\mathbb{R}^N$ that is zero outside a compact set. (Then the Lipschitz continuity of this extended function will follow automatically.)

In fact, by the **smooth Tietze extension theorem** (cf. [Lee, Lem. 2.26]), every smooth function $A \to \mathbb{R}^k$ (where A is a compact subset of a smooth real manifold M (such as a Euclidean space, an n-dimensional sphere, etc.)) can be extended to a smooth function $M \to \mathbb{R}^k$ vanishing outside a compact set. We will study this theorem in the second semester. (The case $A \subset \mathbb{R}$ will be proved in Exp. 15.30.) $\square$

14 More on Riemann integrals

14.1 Commutativity of integrals and other limit processes

Let $I = [a, b]$ and $J = [c, d]$ where $-\infty < a < b < +\infty$ and $-\infty < c < d < +\infty$. Let V be a Banach space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$.

14.1.1 Fubini's theorem

Theorem 14.1 (Fubini's theorem for Riemann integrals). *Let $f \in C(I \times J, V)$. Then $\int_I \int_J f = \int_J \int_I f$. More precisely,*

$$\int_I \left(\int_J f(x, y) dy \right) dx = \int_J \left(\int_I f(x, y) dx \right) dy \tag{14.1}$$

Our strategy is to view $\int_I f(x, y) dx$ as the integral of the function $I \to C(J, V)$ sending x to $f(x, \cdot)$. Then Fubini's theorem follows from the commutativity of integrals and the bounded linear map $\int_J : C(J, V) \to V$ (cf. Thm. 13.16). We first make some general discussion before giving the rigorous proof.

Let Y be a compact topological space. By Thm. 9.3, we have a canonical equivalence $C(I \times Y, V) \simeq C(I, C(Y, V))$ by viewing $f \in C(I \times Y, V)$ as a map

$$\Phi(f) : I \to C(Y, V) \tag{14.2a}$$

where

$$\Phi(f)(x) = f(x, \cdot) : Y \to V \qquad y \mapsto f(x, y) \tag{14.2b}$$

Recall that the space of continuous functions on a compact space is equipped with the l^∞-norm, and $C(Y, V)$ is complete since V is complete (Cor. 3.50).

Lemma 14.2. *The integral $\int_I \Phi(f)$, which is an element of $C(Y, V)$, is the function*

$$\int_I f(x, \cdot) dx : Y \to V \qquad y \mapsto \int_I f(x, y) dx \tag{14.3}$$

In other words, for every $y \in Y$ we have

$$\left(\int_I \Phi(f) \right)(y) = \int_I f(x, y) dx \tag{14.4}$$

Consequently, the function $\int_I f(x, \cdot) dx$ is continuous.

It is easy to show that $\int_I f(x, \cdot) dx$ is continuous without assuming that Y is compact. See Exe. 14.8.

Proof. For each $y \in Y$, define linear map

$$\varphi_y : C(Y, V) \to V \qquad g \mapsto g(y)$$

Then this linear map is clear bounded (with operator norm 1). Thus, by Thm. 13.16, we have

$$\left(\int_I \Phi(f)(x)dx \right)(y) = \varphi_y \left(\int_I \Phi(f)(x)dx \right) = \int_I \varphi_y \big(\Phi(f)(x) \big) dx = \int_I f(x, y)dx$$

$\square$

Proof of Thm. 14.1. By Thm. 13.20, the integral operator $\int_J : C(J, V) \to V$ is a bounded linear map. Therefore, by Thm. 13.16 (and in particular (13.19)), we have a commutative diagram

$$
\begin{array}{ccc}
C(I, C(J, V)) & \xrightarrow{\ \int_J \circ\ } & C(I, V) \\
{\scriptstyle \int_I} \downarrow & & \downarrow {\scriptstyle \int_I} \\
C(J, V) & \xrightarrow{\ \int_J\ } & V
\end{array}
$$

where the top arrow is the map sending $\Phi(f)$ to $\int_J \circ \Phi(f)$, i.e., sending $x \mapsto f(x, \cdot)$ to $x \mapsto \int_J f(x, y)dy$. Thus, the direction $\xrightarrow{}\downarrow$ sends $\Phi(f)$ to $\int_I \int_J f(x, y)dydx$. By Lem. 14.2, the left downward arrow sends $\Phi(f)$ to the function $y \in J \mapsto \int_I f(x, y)dx$. So $\downarrow\xrightarrow{}$ sends $\Phi(f)$ to $\int_J \int_I f(x, y)dxdy$. Thus, the commutativity of the above diagram proves Fubini's theorem. $\square$

Definition 14.3. Let $I_1, \ldots, I_N$ be compact intervals in $\mathbb{R}$. Let $B = I_1 \times \cdots \times I_N$ and $f \in C(B, V)$. We define the **Riemann integral** of f to be

$$\int_B f = \int_{I_1} \cdots \int_{I_N} f(x_1, \ldots, x_N)dx_N \cdots dx_1$$

Then, by Fubini's theorem, for any bijection $\sigma : \{1, \ldots, N\} \to \{1, \ldots, N\}$ we have

$$\int_B f = \int_{I_{\sigma(1)}} \cdots \int_{I_{\sigma(N)}} f(x_1, \ldots, x_N)dx_{\sigma(N)} \cdots dx_{\sigma(1)}$$

14.1.2 Commutativity of integrals and derivatives

Recall from Cor. 11.35 that $l^{1,\infty}(J, V)$ is a Banach space. So its closed linear subspace $C^1(J, V)$ (cf. Pb. 11.1) is a Banach space under the $l^{1,\infty}$-norm $\|g\|_{1,\infty} = \|g\|_\infty + \|g'\|_\infty$.

Let $f : I \times J \to V$. Consider f as a map $\Psi(f) : I \to V^J$ sending x to

$$\Psi(f)(x) = f(x, \cdot) : J \to V \qquad y \mapsto f(x, y) \tag{14.5}$$

Let ∂_J be the (partial) derivative with respect the variable $y \in J$. Clearly

$$\partial_J\big(\Psi(f)(x)\big) = \partial_J f(x, \cdot) \tag{14.6}$$

The linear map of derivative

$$\partial_J : C^1(J, V) \to C(J, V) \qquad g \mapsto g' = \partial_J g \tag{14.7}$$

is clearly bounded (with operator norm $\leqslant 1$).

Proposition 14.4. *Let X be a topological space. Let $f : X \times J \to V$, and define $\Psi(f) : X \to V^J$ by (14.5). Then the following are equivalent.*

  (1) $\Psi(f)$ is an element of $C(X, C^1(J, V))$. In other words:

    (1a) For each $x \in X$ we have $\Psi(f)(x) \in C^1(J, V)$.

    (1b) The function $x \in X \mapsto \Psi(f)(x) \in C^1(J, V)$ is continuous.

  (2) $\partial_J f$ exists everywhere on $X \times J$. Moreover, we have $f, \partial_J f \in C(X \times J, V)$.

Proof. (1a) means that the functions $y \mapsto f(x, y)$ and $y \mapsto \partial_J f(x, y)$ exist and are continuous. (1b) is equivalent to that the maps

$$x \in X \mapsto \Psi(f)(x) = f(x, \cdot) \in C(J, V)$$
$$x \in X \mapsto \partial_J \Psi(f)(x) = \partial_J f(x, \cdot) \in C(J, V)$$

are continuous (under the l^∞-norm). By Thm. 9.3, this is equivalent to that f and $\partial_J f$ are continuous maps $X \times J \to V$. So (1)⇔(2). □

    We return to the setting of $f : I \times J \to V$.

Lemma 14.5. *The integral $\int_I \Psi(f)$, which is an element of $C^1(J, V)$, is the function*

$$\int_I f(x, \cdot) dx : J \to V \qquad y \mapsto \int_I f(x, y) dx \tag{14.8}$$

Consequently, the function $\int_I f(x, \cdot) dx$ is in $C^1(J, V)$.

Proof. This lemma can be proved in the same way as Lem. 14.2, using the fact that for every $y \in J$, the linear map $g \in C^1(J, V) \to g(y)$ is bounded. □

Theorem 14.6. *Let $f : I \times J \to V$. Assume that $\partial_J f$ exists everywhere on $I \times J$. Assume moreover that $f, \partial_J f \in C(I \times J, V)$. Then for each $y \in J$, the LHS of (14.9) exists and equals the RHS of (14.9), where*

$$\partial_J \int_I f(x, y) dx = \int_I \partial_J f(x, y) dx \tag{14.9}$$

First proof. Again, by Thm. 13.16, we have a commutative diagram

$$
\begin{array}{ccc}
C(I, C^1(J, V)) & \xrightarrow{\ \partial_J\circ\ } & C(I, C(J, V)) \\
{\scriptstyle\int_I}\big\downarrow & & \big\downarrow{\scriptstyle\int_I} \\
C^1(J, V) & \xrightarrow{\ \ \partial_J\ \ } & C(J, V)
\end{array}
$$

By Prop. 14.4, $\Psi(f)$ is an element of $C(I, C^1(J, V))$. The map $\Psi(f) : I \to C^1(J, V)$, composed with ∂_J, gives $x \in I \mapsto \partial_J f(x, \cdot)$. By Lem. 14.2, the direction $\to\downarrow$ sends $\Psi(f)$ to $\int_I \partial_J f(x, \cdot) dx$. By Lem. 14.5, $\int_I \Psi(f)$ equals $\int_I f(x, \cdot)$. In particular, $\int_I f(x, \cdot)$ is a C^1-function. So $\downarrow\to$ sends $\Psi(f)$ to $\partial_J \int_I f(x, \cdot) dx$. This finishes the proof. $\qquad\square$

Second proof. Fix any $y \in J$. In view of Cor. 13.21, it suffices to prove that the limit of the net of functions $(\varphi_p)_{p\in J\setminus\{y\}}$ from I to V converges uniformly to $\partial_J f(\cdot, y)$ under $\lim_{p\to y}$, where

$$
\varphi_p(x) = \frac{f(x, p) - f(x, y)}{p - y}
$$

By Rem. 12.34, we have

$$
\|\varphi_p(x) - \partial_J f(x, y)\| \leqslant A(x, p) := \sup_{q\in[p,y]\cup[y,p]} \|\partial_J f(x, q) - \partial_J f(x, y)\|
$$

Since $\partial_J f$ is continuous, it can be viewed as a continuous map $J \to C(I, V)$ by Thm. 9.3. Thus, for every $\varepsilon > 0$ there exists $\delta > 0$ such that for every $p \in J$ satisfying $|p - y| \leqslant \delta$, we have $\sup_{x\in I} \|\partial_J f(x, q) - \partial_J f(x, y)\| \leqslant \varepsilon$ for all $q \in [p, y] \cup [y, p]$, and hence $\sup_{x\in I} A(x, p) \leqslant \varepsilon$. This proves that $A(\cdot, p)$ converges uniformly to 0 (as a net of functions $I \to \mathbb{R}$) as $p \to y$, finishing the proof. $\qquad\square$

14.1.3 Commutativity of partial derivatives

We write $\partial_I f(x, y)$ as $\partial_1 f(x, y)$ and $\partial_J f(x, y)$ as $\partial_2 f(x, y)$. In the following, we use Thm. 14.6 to give a new proof of Thm. 12.35 on the commutativity of ∂_1 and ∂_2. The idea is as follows. Suppose we know that A, B are linear operators on a vector space such that A is invertible and $A^{-1}B = BA^{-1}$. Then one deduces $A^{-1}BA = BA^{-1}A = B$ and hence $BA = AA^{-1}BA = AB$. Now, Thm. 14.6 says that ∂_J commutes with $\int_I$, the inverse of ∂_I (in a vague sense). So one can use a similar algebraic argument to prove that ∂_J commutes with ∂_I.

For the reader's convenience, we reproduce the content of Thm. 12.35 as follows, along with some additional conclusions (i.e. the continuity of $\partial_1 f$ and $\partial_2 f$).

Theorem 14.7. *Let $f : I \times J \to V$ be a function such that $\partial_1 f, \partial_2 f, \partial_2\partial_1 f$ exist on $I \times J$, and that $\partial_2\partial_1 f$ is continuous. Then $\partial_1 f$ and $\partial_2 f$ are continuous on $I \times J$. Moreover, $\partial_1\partial_2 f$ exists on $I \times J$ and equals $\partial_2\partial_1 f$. (So $\partial_1\partial_2 f$ is also continuous.)*

In concrete examples, if one can show the continuity of $\partial_2\partial_1 f$, one can also prove easily that $\partial_1 f, \partial_2 f$ are continuous. Thus, it does not hurt much to assume in this theorem that $\partial_1 f, \partial_2 f$ are also continuous.

Proof. Since $\partial_2\partial_1 f$ exists, the function $y \in J \mapsto \partial_1 f(a, y)$ is differentiable and in particular continuous. Thus, the continuity of $\partial_1 f$ follows easily from

$$\partial_1 f(x, y) = \partial_1 f(a, y) + \int_a^x \partial_2\partial_1 f(u, y)du$$

and the continuity of $\partial_2\partial_1 f$. (See Exe. 14.8.) Similarly, one can prove that $\partial_2 f$ is continuous if one can prove $\partial_1\partial_2 f = \partial_2\partial_1 f$.

Let $F(x, y) = \int_a^x \partial_2\partial_1 f(u, y)du$. By FTC, we have $\partial_1 F = \partial_2\partial_1 f$. On the other hand, by Thm. 14.6 and the continuity of $\partial_1 f, \partial_2\partial_1 f$, we have

$$F(x, y) = \partial_2 \int_a^x \partial_1 f(u, y)du = \partial_2(f(x, y) - f(a, y))$$

Therefore $\partial_1\partial_2 f$ exists and equals $\partial_1 F = \partial_2\partial_1 f$. $\qquad\square$

Exercise 14.8. Let Y be a topological space. Let $f \in C(I \times Y, V)$. Prove that the following function is continuous:

$$I \times Y \to V \qquad (x, y) \mapsto \int_a^x f(u, y)du \tag{14.10}$$

Exercise 14.9. Use Fubini's Thm 14.1 to prove Thm. 14.6.

14.2   Lebesgue's criterion for Riemann integrability

Fix $I = [a, b]$ where $-\infty < a < b < +\infty$. Let V be a Banach space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. The goal of this section is to prove:

Theorem 14.10 (Lebesgue's criterion). *Let $f : I \to V$. Then the following are equivalent.*

(1) f is strongly Riemann integrable.

(2) f is bounded (i.e. $\|f\|_{l^\infty} < +\infty$). Moreover, the set of discontinuities

$$\{x \in I : f \text{ is not continuous at } x\} \tag{14.11}$$

is a null set.

Definition 14.11. A subset E of $\mathbb{R}$ is called a **(Lebesgue) null set** (or a set of **(Lebesgue) measure zero**) if for every $\varepsilon > 0$ there exist countably many closed intervals $I_1, I_2, \ldots$ such that $E \subset \bigcup_i I_i$, and that $\sum_i |I_i| < \varepsilon$. Here, $|I_i|$ is the length of I_i.

The word "countably many" can be omitted, because the sum of uncountably many strictly positive reals numbers must be $+\infty$ due to Pb. 5.3.

Remark 14.12. Covering E by open intervals instead of closed ones does not change the definition of null sets. This is because any open/closed interval can be stretched by a factor of $1 + \delta$ to a larger closed/open interval, where δ is any given positive number.

Proposition 14.13. *A countable union of null subsets of $\mathbb{R}$ is a null set.*

Proof. If $E = E_1 \cup E_2 \cup \cdots$ where each E_i is a null set, then for each ε, E_i can be covered by countably many closed intervals of total length $< 2^{-i}\varepsilon$. So E can be covered by by countably many closed intervals of total length $< \varepsilon$. $\qquad\square$

Example 14.14. Every interval with at least two points is not a null set.

Proof. Since every such interval contains a closed invertal with at least two points, it suffices to prove that the latter is not null. Thus, let us prove that I is not null. Suppose I is covered by some open intervals. Then, since I is compact, I is covered by finitely many of these open intervals, say $U_1, \ldots, U_n$. It is easy to see that $\sum_{i=1}^{n} |U_i| \geqslant |I| = b - a$. (For example, let $f = \sum_i \chi_{U_i}$. Then $f \geqslant \chi_I$. So $\sum_i |U_i| = \sum_i \int \chi_{U_i} = \int f \geqslant \int_a^b 1 = b - a$.) $\qquad\square$

Before proving Lebesgue's criterion, let us first see some useful applications. Recall that if $V = \mathbb{R}^N$ then Riemann integrability is equivalent to strong Riemann integrability (Thm. 13.13).

Corollary 14.15. *Let $f \in \mathscr{R}(I, \mathbb{R}^n)$. Let Ω be a subset of $\mathbb{R}^n$ containing $f(I)$. Let $g \in C(\Omega, \mathbb{R}^m)$ such that $\|g\|_\infty < +\infty$. Then $g \circ f \in \mathscr{R}(I, \mathbb{R}^m)$*

Note that if Ω is compact, we automatically have $\|g\|_\infty < +\infty$.

Proof. Clearly $g \circ f$ is bounded. The set of discontinuities of $g \circ f$ is null since it is a subset of the set of discontinuities of f, where the latter is a null set. $\qquad\square$

Corollary 14.16. *Let $f \in \mathscr{R}(I, \mathbb{F}^{m \times n})$ and $g \in \mathscr{R}(I, \mathbb{F}^{n \times k})$. Then $fg \in \mathscr{R}(I, \mathbb{F}^{m \times k})$.*

Proof. This is immediate from Lebesgue's criterion. $\qquad\square$

14.2.1 Proof of Lebesgue's criterion

The first step of proving Lebesgue's criterion is to express the set of discontinuities by the oscillation.

Definition 14.17. Let X be a topological space. Let Y be a metric space. The **oscillation** of a function $f : X \to Y$ at $x \in X$ is defined to be

$$\omega(f, x) = \inf_{U \in \mathrm{Nbh}_X(x)} \mathrm{diam}(f(U))$$

Proposition 14.18. *Let X be a topological space, let Y be a metric space, and let $f : X \to Y$. Then f is continuous at $x \in X$ iff $\omega(f, x) = 0$.*

Proof. Assume that f is continuous at x. Then for every $\varepsilon > 0$, there exists $U \in \mathrm{Nbh}(x)$ such that $d(f(p), f(x)) < \varepsilon/2$ for every $p \in X$. Then clearly $\mathrm{diam}(f(U)) \leqslant \varepsilon$. So $\omega(f, x) \leqslant \varepsilon$. Since $\varepsilon > 0$ is arbitrary, we get $\omega(f, x) = 0$.

Conversely, suppose $\omega(f, x) = 0$. Then for every $\varepsilon > 0$ there exists $U \in \mathrm{Nbh}(x)$ such that $\mathrm{diam}(f(U)) < \varepsilon$. So for every $p \in U$ we have $d(f(p), f(x)) < \varepsilon$. $\qquad\square$

Proof of Thm. 14.10, part 1. Let us prove (1)$\Rightarrow$(2). Assume that f is strongly Riemann integrable. Choose any $\varepsilon > 0$. Let us prove that

$$\Omega_\varepsilon(f) = \{x \in I : \omega(f, x) \geqslant \varepsilon\} \tag{14.12}$$

is a null set. Then the set of discontinuities, which is $\bigcup_{n \in \mathbb{Z}_+} \Omega_{1/n}(f)$, is a null set by Prop. 14.13.

Since $\inf_{\sigma \in \mathcal{P}(I)} \omega(f, \sigma) = 0$, for every $\delta > 0$, there exists $\sigma = \{a_0 < \cdots < a_n\} \in \mathcal{P}(I)$ such that, with $I_i = [a_{i-1}, a_i]$, we have

$$\sum_{i=1}^{n} \mathrm{diam}(f(I_i)) \cdot |I_i| < \delta\varepsilon \tag{14.13}$$

Note that if $x \in I \backslash \sigma$ is in some I_i where $\mathrm{diam}(f(I_i)) < \varepsilon$, then clearly $\omega(f, x) < \varepsilon$. Thus, if $\omega(f, x) \geqslant \varepsilon$ (i.e., if $x \in \Omega_\varepsilon(f)$), then either $x \in \sigma$, or $x \in I_i$ for some I_i such that $\mathrm{diam}(f(I_i)) \geqslant \varepsilon$. We conclude

$$\Omega_\varepsilon(f) \subset \{a_0, \ldots, a_n\} \cup \left(\bigcup_{k \in K} I_k \right)$$

where $K = \{1 \leqslant k \leqslant n : \mathrm{diam}(f(I_k)) \geqslant \varepsilon\}$. Clearly $\{a_0, \ldots, a_n\}$ can be covered by some intervals with total length $< \delta$. But (14.13) implies that $\sum_{k \in K} \varepsilon |I_k| < \delta\varepsilon$ and hence $\sum_{k \in K} |I_k| < \delta$. So $\Omega_\varepsilon(f)$ can be covered by intervals with total length $< 2\delta$ for every $\delta > 0$. Thus $\Omega_\varepsilon(f)$ is a null set. $\qquad\square$

To prove the other direction, we need some preparation.

Lemma 14.19. *Let $f : X \to Y$ where X is a topological space and Y is a metric space. Then for every $\varepsilon > 0$, $\Omega_\varepsilon(f) = \{x \in X : \omega(f, x) \geqslant \varepsilon\}$ is a closed subset of X.*

Proof. Let us prove that each $x \in X \backslash \Omega_\varepsilon(f)$ is an interior point. (Recall Prop. 77.43.) Indeed, since $\inf_{U \in \mathrm{Nbh}(x)} \mathrm{diam}(f(U)) < \varepsilon$, there exists $U \in \mathrm{Nbh}(x)$ such that $\mathrm{diam}(f(U)) < \varepsilon$. Then for each $p \in U$ we have $\inf_{V \in \mathrm{Nbh}(p)} \mathrm{diam}(f(V)) < \varepsilon$ since $U \in \mathrm{Nbh}(p)$. So $U \subset X \backslash \Omega_\varepsilon(f)$. $\qquad\square$

Lemma 14.20. *Let $\varepsilon > 0$. Suppose that for every $x \in I$ we have $\omega(f, x) < \varepsilon$. (Namely, suppose $\Omega_\varepsilon(f) = \varnothing$.) Then there exists a partition $I = I_1 \cup \cdots \cup I_n$ satisfying $\mathrm{diam}(f(I_i)) < \varepsilon$ for all i.*

Proof. For every $x \in I$, there exists $U_x \in \mathrm{Nbh}_I(x)$ such that $\mathrm{diam}(f(U_x)) < \varepsilon$. Choose $n \in \mathbb{Z}_+$ such that $1/n$ is a Lebesgue number of the open cover $\mathcal{U} = \{U_x : x \in I\}$ of I. (Recall Thm. 10.8.) Dividing I into $2n$ subintervals with the same length gives the desired partition. $\qquad\square$

Proof of Thm. 14.10, part 2. Let us prove (2)$\Rightarrow$(1). Suppose that $M = \|f\|_\infty$ is $< +\infty$, and that the set of discontinuities is a null set. Then for each $\varepsilon > 0$, $\Omega_\varepsilon(f)$ is a null set. By Lem. 14.19 and Heine-Borel, $\Omega_\varepsilon(f)$ is compact. Thus, for every $\delta > 0$, $\Omega_\varepsilon(f)$ can be covered by finitely many closed intervals with total length $< \delta$. Let Δ be the union of these closed intervals. Then $\Omega_\varepsilon(f) \subset \Delta$. Clearly, both Δ and $J = I \backslash \mathrm{Int}(\Delta)$ can be written as disjoint unions of finitely many closed intervals, which are their connected components. And Δ has length $|\Delta| < \delta$.

Since $J \cap \Omega_\varepsilon(f) = \varnothing$, applying Lem. 14.20 to each of the components $J_1, J_2, \ldots$ of J, we see that J_i has a partition ϱ_i such that the oscillation of f (recall Def. 13.7) on each subinterval cut out by ϱ_i is $< \varepsilon$. Let $\sigma = \varrho_1 \cup \varrho_2 \cup \cdots \cup \{a, b\}$, which is a partition of I. Then σ divides I into subintervals $I_1, I_2, \ldots$ such that either $I_j \subset J$ or $I_j \subset \Delta$.

- If $I_j \subset J$, then I_j belongs to a subinterval cut out by one of $\varrho_1, \varrho_2, \ldots$ in $J_1, J_2, \ldots$. This implies $\mathrm{diam}(f(I_j)) < \varepsilon$ by the construction of $\varrho_1, \varrho_2, \ldots$.

- If $I_j \subset \Delta$, then I_j is a component of Δ. We have $\mathrm{diam}(f(I_j)) \leqslant 2M$. Moreover, the total length of such I_j is equal to $|\Delta|$, which is $< \delta$.

From the discussion of these two cases, we see that

$$\omega(f, \sigma) = \sum_j \mathrm{diam}(f(I_j)) \cdot |I_j| \leqslant \varepsilon \cdot (b - a) + 2M \cdot \delta$$

Since $b-a$ and M are fixed numbers and ε, δ are arbitrary, we see that $\inf_\sigma \omega(f, \sigma) = 0$. So f is strongly Riemann integrable. $\qquad\square$

14.3   Improper integrals

In this section, all intervals in $\mathbb{R}$ are assumed to contain at least two points. Let I be an interval in $\mathbb{R}$ with $a = \inf I$ and $b = \sup I$. So $-\infty \leqslant a < b \leqslant +\infty$. Let V be a Banach space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$.

Definition 14.21. We define

$$\mathscr{R}(I, V) = \{f \in V^I : f|_J \in \mathscr{R}(J, V) \text{ for every compact interval } J \subset I\}$$

Define the **improper integral**

$$\int_I f \equiv \int_a^b f \xlongequal{\mathrm{def}} \lim_{\substack{u \to a \\ v \to b}} \int_u^v f = \lim_J \int_J f \tag{14.14}$$

Here, the last limit is over the directed set {compact intervals in I} with preorder "$\subset$". If the above limit exists, we say that $\int_I f$ **exists** or **converges**.

When $f \in \mathscr{R}(I, \mathbb{R})$ and takes values in $\mathbb{R}_{\geqslant 0}$, we write $f \in \mathscr{R}(I, \mathbb{R}_{\geqslant 0})$. Then $\int_I f$ clearly exists in $\overline{\mathbb{R}}_{\geqslant 0}$. We write $\int_I f < +\infty$ if $\int_I f$ converges in $\mathbb{R}$. $\qquad\square$

In the case that a or b is in I, when taking the limit over $u \to a$ and $v \to b$ in (14.14), it is immaterial whether u, v can take values a, b or not, due to the following easy observation:

Lemma 14.22. *Let $f \in \mathscr{R}(I, V)$. If $I = [a, b]$, then the meanings of $\mathscr{R}(I, V)$ and $\int_I f$ are the same as before. If $I = [a, b)$ resp. $I = (a, b]$, then*

$$\int_I f = \lim_{v \to b} \int_a^v f \qquad resp. \qquad \int_I f = \lim_{u \to a} \int_u^b f$$

where the convergence of the LHS is equivalent to that of the RHS.

Proof. Assume $I = [a, b]$. Then I is a compact interval. The new and old meanings of $\mathscr{R}(I, V)$ are the same by Prop. 13.23. Let $f \in \mathscr{R}(I, V)$. Then by Thm. 13.20, we have $\|f\|_\infty < +\infty$. If $a \leqslant u < v \leqslant b$, then Thm. 13.20 and Prop. 13.23 show that

$$\left\| \int_a^b f - \int_u^v f \right\| = \left\| \int_a^u f + \int_v^b f \right\| \leqslant \|f\|_\infty \cdot ((u - a) + (b - v))$$

which converges to 0 as $u \to a$ and $v \to b$, whether u, v take values a, b or not.

Similarly, if $I = [a, b)$, then since f is Riemann integrabe on the compact subinterval $J = [a, (a + b)/2]$, we have $M := \|f|_J\|_\infty < +\infty$. This implies that for all u, v such that $a \leqslant u < v < b$ and $u \leqslant (a + b)/2$,

$$\left\| \int_a^v f - \int_u^v f \right\| = \left\| \int_a^u f \right\| \leqslant M(u - a)$$

which converges to 0 as $u \to a$ and $v \to b$. So $(\int_a^v f)_{u,v}$ and $(\int_u^v f)_{u,v}$ are Cauchy-equivalent nets. So their convergences and values are equivalent. The case $I = (a, b]$ is similar. $\qquad\square$

Remark 14.23. Let $f \in \mathscr{R}(I, V)$. The Cauchy condition for the convergence of $\int_a^b f$ is easy to describe. In view of $\int_u^v - \int_{u'}^{v'} = \int_u^{u'} + \int_v^{v'}$ (Prop. 13.23), we have:

- For every $\varepsilon > 0$, there exist $u_0 < v_0$ in I such that for all $u, u', v, v' \in I$ satisfying $a \leqslant u, u' < u_0$ and $v_0 < v, v' \leqslant b$ we have

$$\left\| \int_u^{u'} f \right\| < \varepsilon \qquad \left\| \int_v^{v'} f \right\| < \varepsilon$$

Definition 14.24. Let $f \in \mathscr{R}(I, V)$. We say that $\int_I f$ **converges absolutely** (or that f is **absolutely integrable on** I) if there exists $g \in \mathscr{R}(I, \mathbb{R})$ such that $|f| \leqslant g$ (i.e., $\|f(x)\| \leqslant g(x)$ for all $x \in I$, in particular $g(x) \geqslant 0$) and that $\int_I g < +\infty$. We let

$$\mathscr{R}^1(I, V) = \{ f \in \mathscr{R}(I, V) : \textstyle\int_I f \text{ converges absolutely} \} \tag{14.15}$$

which is clearly a linear subspace of V^I. By the Cauchy condition in Rem. 14.23, it is clear that $\int_I f$ converges if $\int_I f$ converges absolutely. Though $\mathbb{R}_{\geqslant 0}$ is not a Banach space, we still write

$$\mathscr{R}^1(I, \mathbb{R}_{\geqslant 0}) = \{ f \in \mathscr{R}(I, \mathbb{R}_{\geqslant 0}) : \textstyle\int_I f \text{ converges absolutely} \}$$

which is clearly the set of all $f \in \mathscr{R}(I, \mathbb{R}_{\geqslant 0})$ satisfying $\int_I f < +\infty$.

The superscript 1 of $\mathscr{R}^1$ has a similar meaning as that of l^1, but is different from that of C^1.

Remark 14.25. Assume that $f : I \to V$ is strongly integrable on each compact subinterval of I. (This is the case, for example, when $f \in C(I, V)$, or when $f \in \mathscr{R}(I, V)$ and $V = \mathbb{F}^N$.) Then $|f| : x \in I \mapsto \|f(x)\| \in \mathbb{R}_{\geqslant 0}$ is an element of $\mathscr{R}(I, \mathbb{R}_{\geqslant 0})$ by Cor. 13.19. Thus, in this case,

$$f \in \mathscr{R}^1(I, V) \qquad \Longleftrightarrow \qquad \int_I |f| < +\infty$$

Example 14.26. We have

$$\int_1^{+\infty} x^{-2} dx = \lim_{v \to +\infty} \int_1^v x^{-2} dx = \lim_{v \to +\infty} (-x^{-1})|_1^v = 1 < +\infty$$

Therefore $\displaystyle\int_1^{+\infty} \frac{e^{ix}}{x^2} dx$ converges absolutely, and hence converges.

The following proposition generalizes Prop. 13.18.

Proposition 14.27. Let $f \in \mathscr{R}^1(I, V)$ and $g \in \mathscr{R}^1(I, \mathbb{R}_{\geqslant 0})$ such that $|f| \leqslant g$. Then $\left\| \int_I f \right\| \leqslant \int_I g$.

Proof. Apply the limit over $u \to a, v \to b$ to $\| \int_u^v f \| \leqslant \int_u^v g$. $\qquad\square$

The next proposition shows that improper integrals are helpful for studying series:

Proposition 14.28. Let $f \in \mathscr{R}([1, +\infty), \mathbb{R}_{\geqslant 0})$ be decreasing. Then we have

$$\int_1^{+\infty} f < +\infty \qquad \Longleftrightarrow \qquad \sum_{n=1}^{+\infty} f(n) < +\infty \tag{14.16}$$

Proof. Since f is decreasing, we clearly have $g \leqslant f \leqslant h$ where g, h are defined by the series in $l^\infty([0, +\infty), \mathbb{R})$:

$$g = \sum_{n=1}^{+\infty} f(n+1) \cdot \chi_{[n,n+1)} \qquad h = \sum_{n=1}^{+\infty} f(n) \cdot \chi_{[n,n+1)}$$

One computes easily that $\int_1^{+\infty} g = \sum_{n=2}^{+\infty} f(n)$ and $\int_1^{+\infty} h = \sum_{n=1}^{+\infty} f(n)$ in $\overline{\mathbb{R}}_{\geqslant 0}$. Thus

$$\sum_{n=2}^{+\infty} f(n) \leqslant \int_1^{+\infty} f \leqslant \sum_{n=1}^{+\infty} f(n)$$

The proposition follows easily. $\qquad\square$

Example 14.29. Let $s > 0$. The function $f : [2, +\infty) \to +\infty$ defined by $f(x) = \dfrac{1}{x(\log x)^s}$ is decreasing, and

$$\int_2^x f = \begin{cases} (1-x)^{-1}\big((\log x)^{1-s} - (\log 2)^{1-s}\big) & \text{if } s \neq 1 \\ \log(\log x) - \log(\log 2) & \text{if } s = 1 \end{cases}$$

So $\sum_{n=2}^\infty f(n) < +\infty$ iff $\int_2^{+\infty} f < +\infty$ iff $s > 1$.

14.4 Commutativity of improper integrals and other limit processes

In this section, all intervals in $\mathbb{R}$ are assumed to contain at least two points. Let I be an interval in $\mathbb{R}$ with $a = \inf I$ and $b = \sup I$. Let $J = [c, d]$ be a compact interval. Let V be a Banach space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$.

The goal of this section is to generalize the main results of Sec. 14.1 to improper integrals. Namely, we shall prove the commutativity of $\int_I$ with $\int_J$ and with ∂_J under reasonable assumptions. There are three ways to achieve this goal:

- We know that $\int_I$ is the limit of integrals over compact intervals. Therefore, the problem is reduced to that of proving that $\lim_{u \to a, v \to b}$ can be moved inside in $\lim \int_J \int_u^v$ and in $\lim \partial_J \int_u^v$.

- We generalize Thm. 13.16 to improper integrals, and use this generalized version to prove the commutativity in a similar way as in Sec. 14.1.

- The commutativity of $\int_I$ with ∂_J can also be studied by generalizing Cor. 13.21 to improper integrals, similar to the second proof of Thm. 14.6.

We will use the second approach because its proof is more conceptual and involves fewer technical calculations, thus making it easier for us to remember the conditions of the theorems to be proved. Nevertheless, we will also give the appropriate generalization of Cor. 13.21, which is helpful for future application. Recall Def. 14.24 for the meaning of $\mathscr{R}^1$.

14.4.1 Commutativity of improper integrals and bounded linear maps

Theorem 14.30. *Let W be also a Banach space over $\mathbb{F}$. Let $T \in \mathcal{L}(V, W)$. Then for every $f \in \mathcal{R}^1(I, V)$ we have $T \circ f \in \mathcal{R}^1(I, W)$ and*

$$T\left(\int_a^b f \right) = \int_a^b T \circ f \tag{14.17}$$

In other words, we have a commutative diagram

$$
\begin{array}{ccc}
\mathcal{R}^1(I,V) & \xrightarrow{\ T\circ\ } & \mathcal{R}^1(I,W) \\
{\scriptstyle \int_I}\Big\downarrow & & \Big\downarrow{\scriptstyle \int_I} \\
V & \xrightarrow{\quad T \quad} & W
\end{array}
\tag{14.18}
$$

Proof. By Thm. 13.16, $T \circ f$ is Riemann integrable on compact subintervals of I. Since $f \in \mathcal{R}^1$, there exists $g \in \mathcal{R}(I, \mathbb{R}_{\geqslant 0})$ such that $|f| \leqslant g$ and $\int_I g < +\infty$. Let $M = \|T\|$ be the operator norm, which is a finite number. By Rem. 10.24, for each $x \in I$ we have $\|T \circ f(x)\| \leqslant M \|f(x)\| \leqslant M g(x)$, and hence $|T \circ f| \leqslant Mg$. This proves $T \circ f \in \mathcal{R}^1(I, W)$. In particular, the integrals of $f, T \circ f$ over I converge. Thus, by the continuity of T and the commutativity of T with $\int_u^v$ (when $a < u < v < b$) due to Thm. 13.16, we have

$$T\left(\int_I f \right) = T\left(\lim_{\substack{u\to a\\v\to b}} \int_u^v f \right) = \lim_{\substack{u\to a\\v\to b}} T\left(\int_u^v f \right) = \lim_{\substack{u\to a\\v\to b}} \int_u^v T \circ f = \int_I T \circ f$$

$\square$

14.4.2 Fubini's theorem

Recall that $J = [c, d]$ is a compact interval but I is not necessarily compact.

Lemma 14.31. *Lem. 14.2 holds verbatim to the current case that I is not necessarily compact.*

Proof. We can prove this general case in the same way as Lem. 14.2, using the commutativity of $\int_I$ and the bounded map $g \in C(Y, V) \mapsto g(y) \in V$, which is available thanks to Thm. 14.30. $\square$

Theorem 14.32 (Fubini's theorem for improper integrals). *Let $f \in C(I \times J, V)$. Assume that there exists $h \in \mathcal{R}^1(I, \mathbb{R}_{\geqslant 0})$ satisfying*

$$\|f(x, y)\| \leqslant h(x) \qquad (\forall x \in I, y \in J) \tag{14.19}$$

Then the functions $\int_J f(\cdot, y)dy : I \to V$ and $\int_I f(x, \cdot)dx : J \to V$ are continuous, and the equation

$$\int_I \left(\int_J f(x, y)dy \right) dx = \int_J \left(\int_I f(x, y)dx \right) dy \tag{14.20}$$

holds where the integral over I on the LHS converges absolutely.

Proof. Let $T : C(J,V) \to V$ be the bounded linear map $\int_J$. By Thm. 14.30, we have a commutative diagram

$$
\begin{array}{ccc}
\mathscr{R}^1(I, C(J,V)) & \xrightarrow{\;T\circ\;} & \mathscr{R}^1(I,V) \\
{\scriptstyle\int_I}\downarrow & & \downarrow{\scriptstyle\int_I} \\
C(J,V) & \xrightarrow{\quad T \quad} & V
\end{array}
\tag{14.21}
$$

Since J is compact, by Thm. 9.3, we can view f as a continuous function $\Phi(f) :$ $I \to C(J,V)$ sending x to $f(x,\cdot)$. Then (14.19) says that $\Phi(f) \leqslant g$. Since $g \in \mathscr{R}^1$, we conclude $\Phi(f) \in \mathscr{R}^1(I, C(J,V))$.

By Lem. 14.31, $\int_I \Phi(f)$ equals the function $\int_I f(x,\cdot)dx$. In particular, $\int_I f(x,\cdot)dx$ is continuous since $\int_I \Phi(f) \in C(J,V)$ by (14.21). The continuity of $\int_J f(\cdot, y)dy$ is similar, or is even easier because J is compact.

Clearly $T(\int_I \Phi(f))$ is $\int_J \int_I f$. By the top arrow of (14.21), $T \circ \Phi(f)$ belongs to $\mathscr{R}^1(I,V)$. Since $T \circ \Phi(f)$ is the function sending x to $\int_J f(x,\cdot)dy$, the integral of this function over I is absolutely convergent, and $\int_I T \circ \Phi(f) = \int_I \int_J f$. Thus, the commutativity of (14.21) proves (14.20). □

The continuity of $\int_I f(x,\cdot)dx$ can be generalized: see Cor. 14.36.

14.4.3 Commutativity of improper integrals and partial derivatives

Lemma 14.33. *Lemma 14.5 holds verbatim to the current case that I is not necessarily compact.*

Proof. Again, this is proved in the same way as Lem. 14.2 by applying Thm. 14.30 to the bounded linear functional $g \in C^1(J,V) \to g(y)$ (where $y \in J$). □

Theorem 14.34. *Let $f : I \times J \to V$. Assume that $\partial_J f$ exists everywhere on $I \times J$. Assume that $f, \partial_J f \in C(I \times J, V)$. Assume moreover that there exists $h \in \mathscr{R}^1(I, \mathbb{R}_{\geqslant 0})$ satisfying*

$$
\|f(x,y)\| \leqslant h(x) \quad and \quad \|\partial_J f(x,y)\| \leqslant h(x) \qquad (\forall x \in I, y \in J)
\tag{14.22}
$$

Then for each $y \in J$, the LHS of (14.23) exists and equals the RHS of (14.23), where

$$
\partial_J \int_I f(x,y)dx = \int_I \partial_J f(x,y)dx
\tag{14.23}
$$

Proof. Let $T : C^1(J,V) \to C(J,V)$ be the bounded linear map ∂_J. By Thm. 14.30, we have a commutative diagram

$$
\begin{array}{ccc}
\mathscr{R}^1(I, C^1(J,V)) & \xrightarrow{\;T\circ\;} & \mathscr{R}^1(I, C(J,V)) \\
{\scriptstyle\int_I}\downarrow & & \downarrow{\scriptstyle\int_I} \\
C^1(J,V) & \xrightarrow{\quad T \quad} & C(J,V)
\end{array}
\tag{14.24}
$$

Define $\Psi(f) : X \rightarrow V^J$ sending x to $f(x, \cdot)$. By Prop. 14.4, $\Psi(f)$ belongs to $C(I, C^1(J, V))$. By (14.22), $\Psi(f)$ belongs to $\mathscr{R}^1(I, C^1(J, V))$. As in the first proof of Thm. 14.6, one can use Lem. 14.31 and 14.33 (the improper version of Lem. 14.2 and 14.5) to show that $\int_I T \circ \Psi(f) = T \int_I \Psi(f)$ (which is a consequence of the commutativity of (14.24)) implies (14.23). $\qquad\square$

14.4.4 Commutativity of improper integrals and net limits

Theorem 14.35. *Let $(f_\alpha)_{\alpha \in \mathscr{I}}$ be a net in $\mathscr{R}^1(I, V)$. Let $f \in V^I$. Assume that the following conditions are true:*

(1) On every compact subinterval of I, the net (f_α) converges uniformly to f.

(2) There exists $g \in \mathscr{R}^1(I, \mathbb{R}_{\geq 0})$ such that $|f_\alpha| \leq g$ for all $\alpha \in \mathscr{I}$.

Then $f \in \mathscr{R}^1(I, V)$, and $\displaystyle\int_I f = \lim_{\alpha \in \mathscr{I}} \int_I f_\alpha$.

Proof. By Cor. 13.21, $f \in \mathscr{R}(I, V)$. Since (f_α) converges pointwise to f, we clearly have $|f| \leq g$. So $f \in \mathscr{R}^1(I, V)$. Choose any $\varepsilon > 0$. Since $\lim_{u \to a, v \to b} \int_I g$ converges, there exists u, v such that $a < u < v < b$ and

$$\int_a^u g + \int_v^b g = \int_I g - \int_u^v g < \varepsilon$$

Thus, by Prop. 14.27, for each α we have $\| \int_a^u f_\alpha \| + \| \int_v^b f_\alpha \| < \varepsilon$ and $\| \int_a^u f \| + \| \int_v^b f \| < \varepsilon$. Since f_α converges uniformly to f on $[u, v]$, by Cor. 13.21, we get $\lim_\alpha \| \int_u^v f_\alpha - \int_u^v f \| = 0$. Thus

$$\left\| \int_I f - \int_I f_\alpha \right\| = \left\| \int_a^u f + \int_v^b f - \int_a^u f_\alpha - \int_v^b f_\alpha + \int_u^v (f - f_\alpha) \right\|$$
$$\leq 2\varepsilon + \left\| \int_u^v (f - f_\alpha) \right\|$$

where the $\limsup_\alpha$ of the RHS is 2ε. Thus $\| \int_I f - \int_I f_\alpha \|$ converges to 0 under $\limsup_\alpha$, and hence under $\lim_\alpha$. $\qquad\square$

It is a good practice to prove Thm. 14.34 using Thm. 14.35. (See Pb. 14.2.)

Corollary 14.36. *Let Y be a topological space. Let $f \in C(I \times Y, V)$. Assume that there exists $h \in \mathscr{R}^1(I, \mathbb{R}_{\geq 0})$ such that*

$$\|f(x, y)\| \leq h(x) \qquad (\forall x \in I, y \in Y) \tag{14.25}$$

Then the map $\int_I f(x, \cdot) dx$ (sending $y \in Y$ to $\int_I f(x, y) dx$) is continuous.

Proof. By Def. 7.56-(1), we need to prove that for every net (y_α) in Y converging to y, we have $\int_I f(x, y)dx = \lim_\alpha \int_I f(x, y_\alpha)dx$. By Thm. 14.35, it suffices to prove that $\lim_\alpha f(\cdot, y_\alpha)$ converges uniformly on any compact subinterval $[u, v] \subset I$ to $f(\cdot, y)$. This is true because, by Thm. 9.3, the restriction of f to $[u, v] \times Y$ can be viewed as an element of $C(Y, C([u, v], V))$. $\qquad\square$

Example 14.37. Compute $F(t) = \displaystyle\int_0^{+\infty} e^{-tx} \cdot \frac{\sin x}{x} dx$ for all $t > 0$.

Solution. We first compute $F'(t)$. Choose any $\delta > 0$. Then for any $t \geqslant \delta$, the norms of $e^{-tx}\frac{\sin x}{x}$ and $\partial_t(e^{-tx}\frac{\sin x}{x}) = -e^{-tx}\sin x$ are both bounded by $e^{-\delta x}$, where the latter is absolutely integrable since $\int_0^{+\infty} e^{-\delta x}dx = \delta^{-1} < +\infty$. Thus, by Thm. 14.34, we have

$$F'(t) = \int_0^{+\infty} -e^{-tx}\sin x\, dx = \int_0^{+\infty} \frac{e^{(-t+\mathbf{i})x} - e^{(-t-\mathbf{i})x}}{2\mathbf{i}}dx$$
$$= \frac{t\sin x + \cos x}{1 + t^2}e^{-tx}\Big|_{x=0}^{+\infty} = -\frac{1}{1 + t^2}$$

for all $t \geqslant \delta$ and $\delta > 0$, and hence for all $t > 0$. Therefore $F(t) = C - \arctan t$ for some $C \in \mathbb{R}$. Let us determine C using $C = \lim_{t\to+\infty} F(t) + \frac{\pi}{2}$.

To compute $\lim_{t\to+\infty} F(t)$, it suffices to assume $t \geqslant 1$. Then $|e^{-tx}\frac{\sin x}{x}| \leqslant e^{-x}$ where $\int_0^{+\infty} e^{-x} < +\infty$. Moreover, on every compact interval $[a, b]$ in $(0, +\infty)$ (where $0 < a < b < +\infty$), $\lim_{t\to+\infty} e^{-tx}\frac{\sin x}{x}$ converges uniformly to 0 since $|e^{-tx}\frac{\sin x}{x}| \leqslant e^{-ta}$. Thus, the assumptions in Thm. 14.35 are satisfied, and hence

$$\lim_{t\to+\infty} F(t) = \int_0^{+\infty} \lim_{t\to+\infty} e^{-tx} \cdot \frac{\sin x}{x} dx = 0$$

This proves $C = \frac{\pi}{2}$, and hence $F(t) = \frac{\pi}{2} - \arctan t$. $\qquad\square$

14.5   Convolutions and smooth/polynomial approximation

Fix a Banach space V over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$.

Definition 14.38. Let X be a topological space. The **support** of a function $f \in V^X$ is defined to be the closure

$$\mathrm{Supp}(f) = \overline{\{x \in X : f(x) \neq 0\}}$$

Define $C_c(X, V)$ to be the set of continuous functions with **compact support**

$$C_c(X, V) = \{f \in C(X, V) : \mathrm{Supp}(f) \text{ is compact}\}$$

Unless otherwise stated, $C_c(X, V)$ is equipped with the l^∞-norm.

14.5.1 Convolutions and approximation of identity

A goal of this section is to show that the elements of $C_c(\mathbb{R}, V)$ can be approximated by smooth functions with compact supports, i.e., elements in

$$C_c^\infty(\mathbb{R}, V) = C^\infty(\mathbb{R}, V) \cap C_c(\mathbb{R}, V)$$

Indeed, we will do more. We shall prove the celebrated **Weierstrass approximation theorem**, which implies that elements in $C_c(\mathbb{R}, V)$ can be approximated uniformly by polynomials on compact intervals. The proof we will give is also due to Weierstrass, which relies on an important construction called convolution: If $f : \mathbb{R} \to V$ and $g : \mathbb{R} \to \mathbb{F}$, then their **convolution** is a functions $f * g : \mathbb{R} \to V$ defined by

$$(f * g)(x) \equiv (g * f)(x) = \int_{\mathbb{R}} f(x - y)g(y)dy \tag{14.26a}$$

whenever the above integral converges for every $x \in \mathbb{R}$. Taking $y = x - t$, we get $\int_u^v f(x - y)g(y)dy = -\int_{x-u}^{x-v} f(t)g(x - t)dt = \int_{x-v}^{x-u} f(t)g(x - t)dt$. Letting $u \to -\infty, v \to +\infty$, we get

$$(f * g)(x) = \int_{\mathbb{R}} f(y)g(x - y)dy \tag{14.26b}$$

The main idea of doing approximation via convolutions is as follows. Suppose that g is smooth, then by (14.26b), one should expect that $(f * g)^{(n)}(x) = \int_{\mathbb{R}} f(y)(\partial_x)^n g(x - y)dy = \int_{\mathbb{R}} f(y)g^{(n)}(x - y)dy$ to be true. Thus $f * g$ is expected to be smooth. Moreover, if g can be approximated by polynomials, for example, if $g(x) = \sum a_n x^n$ on $\mathbb{R}$, then it is expected that $f * g$ can be approximated by $f * g_n$ where $g_n(x) = a_n x^n$, and it is easy to see that $f * g_n = a_n \int_{\mathbb{R}} f(y)(x - y)^n dy$ is a polynomial of x.

To summarize, one advantage of convolution is that whenever g has a good property, the same is in general true for $f * g$. Another key property of convolution is that f can be approximated by $f * g$ in some sense. The meaning of "approximation" will depend on the analytic property of f, e.g. whether f is continuous, continuous with compact support, or only integrable. In this section, we will only be interested in the case that $f \in C_c(\mathbb{R}, V)$. In this case, the integrals in (14.26) are actually over compact intervals, which makes our lives easier.

We shall show that $f \in C_c(\mathbb{R}, V)$ can be approximated uniformly by $f * g$. More precisely, we shall show that for every absolutely convergent $g \in C(\mathbb{R}, \mathbb{R}_{\geqslant 0})$ satisfying $\int_{\mathbb{R}} g = 1$, if we define $g_\varepsilon \in C(\mathbb{R}, \mathbb{R})$ (where $\varepsilon \in \mathbb{R}_{>0}$) by

$$g_\varepsilon(x) = \frac{1}{\varepsilon} g\left(\frac{x}{\varepsilon}\right) \tag{14.27}$$

then $(f * g_\varepsilon)$ converges uniformly to f as $\varepsilon \to 0$. This method is extremely useful and is used widely in analysis. (The assumption that $g \geqslant 0$ is not necessary. We assume $g \geqslant 0$ only for simplifying discussions.)

Lemma 14.39. *Choose $g \in C(\mathbb{R}, \mathbb{R}_{\geqslant 0})$ satisfying $\int_{\mathbb{R}} g = 1$, and define g_ε by (14.27). Then $\int_{\mathbb{R}} g_\varepsilon = 1$ for each $\varepsilon > 0$. Moreover, for every $\delta > 0$ we have*

$$\lim_{\varepsilon \to 0} \int_\delta^{+\infty} g_\varepsilon = \lim_{\varepsilon \to 0} \int_{-\infty}^{-\delta} g_\varepsilon = 0 \tag{14.28}$$

Proof. By the change of variable formula we have $\int_u^v \varepsilon^{-1} g(\varepsilon^{-1} x) dx = \int_{u/\varepsilon}^{v/\varepsilon} g(y) dy$. This gives $\int_{\mathbb{R}} g_\varepsilon = 1$. Another change of variable shows $\int_\delta^{+\infty} g_\varepsilon = \int_{\delta/\varepsilon}^{+\infty} g = \int_0^{+\infty} g - \int_0^{\delta/\varepsilon} g$, which clearly converges to 0 as $\varepsilon \to 0$. This proves the first half of (14.28). The second half is similar. $\qquad \square$

Proposition 14.40. *Let $f \in C_c(\mathbb{R}, V)$. Choose $g \in C(\mathbb{R}, \mathbb{R}_{\geqslant 0})$ satisfying $\int_{\mathbb{R}} g = 1$, and define g_ε by (14.27) for each $\varepsilon > 0$. Then $(f * g_\varepsilon)$ converges uniformly on $\mathbb{R}$ to f as $\varepsilon \to 0$.*

In other words, the convolution operator $f \mapsto f * g_\varepsilon$ converges pointwise to the identity map when $\varepsilon \to 0$.

Proof. Let $M = \|f\|_\infty$, which is $< +\infty$. Since $\int_{\mathbb{R}} g_\varepsilon = 1$, for each $x \in \mathbb{R}$ we have $\int_{\mathbb{R}} f(x) g_\varepsilon(y) dy = f(x)$. Thus

$$\|(f * g_\varepsilon)(x) - f(x)\| = \left\| \int_{\mathbb{R}} (f(x-y) - f(x)) g_\varepsilon(y) dy \right\|$$

$$\leqslant \int_{\mathbb{R}} \|f(x-y) - f(x)\| \cdot g_\varepsilon(y) dy \tag{14.29}$$

By Thm. 10.7 and the compactness of $\mathrm{Supp}(f)$, f is uniformly continuous on $\mathbb{R}$. Therefore, for every $e > 0$, there exists $0 < \delta < 1$ such that $\|f(x-y) - f(x)\| \leqslant e$ for all $x \in \mathbb{R}$ and $y \in (-\delta, \delta)$. Thus, if we let $J_\delta = \mathbb{R} \backslash (-\delta, \delta)$, then

$$(14.29) \leqslant \int_{J_\delta} \|f(x-y) - f(x)\| \cdot g_\varepsilon(y) dy + e \cdot \int_{-\delta}^\delta g_\varepsilon(y) dy$$

$$\leqslant 2M \cdot \int_{J_\delta} g_\varepsilon(y) dy + e$$

which converges to e under $\limsup_{\varepsilon \to 0}$ by Lem. 14.39. Since e is arbitrary, we conclude that $\lim_{\varepsilon \to 0} (14.29) = 0$. $\qquad \square$

Remark 14.41. The uniform continuity of a function $f : \mathbb{R} \to V$ is equivalent to the continuity of $F : \mathbb{R} \to C(\mathbb{R}, V)$ defined by $F(t)(x) = f(x-t)$. Note that the continuity of F follows from Thm. 9.3. This means that Prop. 14.40 can also be proved by Thm. 9.3.

Remark 14.42. In analysis, there are two especially important classes of $g \in C(\mathbb{R}, \mathbb{R}_{\geqslant 0})$ satisfying $\int_{\mathbb{R}} g = 1$. The first class consists of $g \in C_c(\mathbb{R}, \mathbb{R}_{\geqslant 0})$ satisfying $\int_{\mathbb{R}} g = 1$. Although functions with compact supports are often easy to use, they cannot be approximated by their Taylor series on $\mathbb{R}$. (See Exp. 12.19 for a related example.) Instead, we should consider another type of function:

14.5.2 Polynomial approximation

Example 14.43. Define the **Gauss function** $g(x) = \frac{1}{\sqrt{\pi}}e^{-x^2}$. Then $g \in C(\mathbb{R}, \mathbb{R}_{\geqslant 0})$ and $\int_{\mathbb{R}} g = 1$. Thus, g satisfies the assumptions in Prop. 14.40.

Although g does not have compact support, it is a real analytic function. So we can use $f * g_\varepsilon$ to prove Weierstrass approximation theorem.

Proof. It is easy to see that $\int_1^{+\infty} 1/x^2 dx < +\infty$. Thus, since $e^{x^2} \geqslant x^2$, we get $\int_{\mathbb{R}} e^{-x^2} dx < +\infty$.

It is not so easy to show that

$$\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi} \tag{14.30}$$

This integral is called **Gauss integral**. In the next semester, we will use the change of variable formula for double integrals to prove (14.30). A more elementary (but also more complicated) proof is given in Pb. 14.3. For the purpose of this section, knowing $\int_{\mathbb{R}} e^{-x^2} dx < +\infty$ is enough, since we can define g to be e^{-x^2} divided by its integral on $\mathbb{R}$. $\square$

We now show that if g is the Gauss function, then $f * g_\varepsilon$ can be approximated uniformly by polynomials on every compact interval.

Lemma 14.44. *Let $h(x) = \sum_{n=0}^{\infty} a_n x^n$ have radius of convergence $+\infty$, where $a_n \in \mathbb{F}$ for each a_n. Let $h_n(x) = \sum_{j=0}^{n} a_j x^j$ be the partial sum. Let $f \in C_c(\mathbb{R}, V)$. Then $\lim_{n \to \infty} f * h_n$ converges uniformly on compact intervals to $f * h$, and each $f * h_n$ is in*

$$V[x] = \{v_0 + v_1 x + \cdots + v_k x^k : k \in \mathbb{N}, v_0, v_1, \ldots, v_k \in V\} \tag{14.31}$$

Note that we do not need the assumption $\int_{\mathbb{R}} |h| < +\infty$ in this Lemma. So letting $h(x) = e^x$ is also OK. Clearly, if g is the Gauss functions, then g_ε satisfies the assumptions on h. This implies that if $f \in C_c(\mathbb{R}, V)$, then $f * g_\varepsilon$ can be approximated uniformly on compact intervals by polynomials with coefficients in V. Combining this fact with Prop. 14.40, we see that f can be approximated uniformly by polynomials on compact intervals.

Proof. Note that $(f * h)(x) = \int_{\mathbb{R}} h(x - y)f(y)dy$ is well-defined since it is equals $\int_{-a}^{a} h(x - y)f(y)dy$ if $a > 0$ and if $[-a, a]$ contains $\text{Supp}(f)$. (So this is a usual Riemann integral.) Similarly, $(f * h_n)(x) = \int_{-a}^{a} h_n(x - y)f(y)dy$. Since h_n is a polynomial, it is obvious that $f * h_n$ is also a polynomial.

We now show that $f * h_n$ converges uniformly to $f * h$ on $[-b, b]$ for any $b > 0$. Let $c = a + b$. Then

$$\sup_{x \in [-b,b]} \|(f * h)(x) - (f * h_n)(x)\| \leqslant \sup_{x \in [-b,b]} \int_{-a}^{a} |h(x - y) - h_n(x - y)| \cdot \|f(y)\| dy$$

$$\leqslant A_n \int_{-a}^{a} \|f(y)\| dy$$

where $A_n = \sup_{t \in [-c,c]} \|h(t) - h_n(t)\|$. By Thm. 4.27, h_n converges uniformly on $[-c, c]$ to h. So $\lim_{n \to \infty} A_n = 0$. $\qquad\square$

Theorem 14.45 (Weierstrass approximation theorem). *Let $I = [a, b]$ where $-\infty < a < b < +\infty$. Then $V[x]$ is dense in $C(I, V)$ under the l^∞-norm.*

Proof. Choose any $f \in C(I, V)$. Then f can be extended to a continuous function $\mathbb{R} \to V$ with compact support: For example, we let $f(x) = (x - a + 1)f(a)$ if $x \in [a - 1, a]$, let $f(x) = (b + 1 - x)f(b)$ if $b \in [b, b + 1]$, and let $f(x) = 0$ if $x < a - 1$ or $x > b + 1$. Let g be the Gauss function. For every $e > 0$, by Prop. 14.40, there exists $\varepsilon > 0$ such that $\|f - f * g_\varepsilon\|_{l^\infty(\mathbb{R},V)} < e/2$. By Lem. 14.44, there exists a polynomial $p \in V[x]$ such that $\|f * g_\varepsilon - p\|_{l^\infty(I,V)} < e/2$. So $\|f - p\|_{l^\infty(I,V)} < e$. $\qquad\square$

Corollary 14.46. *Let $I = [a, b]$ where $-\infty < a < b < +\infty$. Then $C([0, 1], \mathbb{R})$ is separable.*

Proof. $\mathbb{Q}[x]$, the set of polynomials with coefficients in $\mathbb{Q}$, is countable. By Thm. 14.45, $\mathbb{Q}[x]$ is dense in $C([0, 1], \mathbb{R})$. $\qquad\square$

The Weierstrass approximation theorem will be generalized to **Stone-Weierstrass theorem** (cf. Thm. 15.9). Accordingly, Cor. 14.46 will be substantially generalized as an application of (the proof of) Stone-Weierstrass theorem: we will show that $C(X, \mathbb{R})$ is separable if X is a compact metric space. (See Thm. 15.37)

14.5.3 Smooth approximation

By the Weierstrass approximation theorem, we know that any $f \in C_c(\mathbb{R}, V)$ can be approximated uniformly by polynomials on compact intervals. However, unless $f = 0$, f cannot be approximated by polynomials uniformly on $\mathbb{R}$. (If $p, q \in V[x]$ are different, then $\|p - q\|_{l^\infty(X,V)} = +\infty$. So any Cauchy sequence in $V[x]$ under the $l^\infty(X, V)$-norm is eventually constant. So its limit is a polynomial, which does not have compact support unless when it is zero.) Nevertheless, we shall show that f can be approximated by smooth compactly supported functions uniformly on $\mathbb{R}$. This task is not difficult: one simply pick a nonzero $g \in C_c^\infty(\mathbb{R}, \mathbb{R})$ satisfying $g \geqslant 0$. Dividing g by $g/\int_{\mathbb{R}} g$, we may assume that $\int_{\mathbb{R}} g = 1$. Then it can be shown that $f * g_\varepsilon \in C_c^\infty(\mathbb{R}, V)$. By Prop. 14.40, $\lim_{\varepsilon \to 0} f * g_\varepsilon$ converges uniformly on $\mathbb{R}$ to f. This finishes the proof. However, we must first prove the existence of such g:

Proposition 14.47. *Let $0 < a < b$. Then there exists $g \in C_c^\infty(\mathbb{R}, \mathbb{R})$ such that $g(\mathbb{R}) = [0, 1]$, that $g^{-1}(1) = [-a, a]$, and that $g^{-1}(0) = (-\infty, -b] \cup [b, +\infty)$.*

Proof. Let $\alpha : \mathbb{R} \to \mathbb{R}_{\geqslant 0}$ be the smooth function in Exp. 12.19. Then α is increasing and $\alpha^{-1}(0) = (-\infty, 0]$. Define $\beta : \mathbb{R} \to \mathbb{R}_{\geqslant 0}$ by $\beta(x) = \alpha(x + b)\alpha(-x + b)$. Then $\beta^{-1}(0) = (-\infty, -b] \cup [b, +\infty)$. In particular, $\beta \in C_c^\infty(\mathbb{R}, \mathbb{R})$. Define $\gamma : \mathbb{R} \to \mathbb{R}_{\geqslant 0}$ by $\gamma(x) = \alpha(x - a) + \alpha(-x - a)$. Then $\gamma^{-1}(0) = [-a, a]$. Note that $\beta(x) + \gamma(x) > 0$ for all $x \in \mathbb{R}$, since $\beta^{-1}(0) \cap \gamma^{-1}(0) = \varnothing$.

Then $g = \beta/(\beta + \gamma)$ is a desired function. $\qquad\square$

Corollary 14.48. *Let I be an interval in $\mathbb{R}$ with $a = \inf I$ and $b = \sup I$ satisfying $-\infty \leqslant a < b \leqslant +\infty$. Then $C_c^\infty(I, V)$ is dense in $C_c(I, V)$ under the l^∞-norm.*

In the following, we only prove the case that $I = (a, b)$. The other cases can be reduced to this case: For example, the case $[a, b)$ is implied by the case $(a - 1, b)$.

First proof. Let $f \in C_c(\mathbb{R}, V)$ supported in $[u, v]$ where $a < u < v < b$. By Prop. 14.47, there exists a nonzero smooth $g \in C_c(\mathbb{R}, \mathbb{R}_{\geqslant 0})$ supported in $[-1, 1]$ such that $\int_\mathbb{R} g = 1$. Since g_ε is supported in $[-\varepsilon, \varepsilon]$, it is easy to see that $f * g_\varepsilon$ is supported in $[u - \varepsilon, v + \varepsilon]$. By Lem. 14.2, $f * g_\varepsilon$ is continuous . Moreover, one can check that $f * g_\varepsilon$ is smooth (cf. Pb. 14.4). So $f * g_\varepsilon \in C_c^\infty(\mathbb{R}, V)$, and $f * g$ is supported in I when $\varepsilon < \min\{u - a, b - v\}$. By Prop. 14.40, $\lim_{\varepsilon \to 0} f * g_\varepsilon$ converges uniformly on $\mathbb{R}$ to f. $\qquad\square$

Second proof. Let $f \in C_c(\mathbb{R}, V)$ supported in $[u, v]$ where $a < u < v < b$. Choose any $e > 0$. Choose any positive $\varepsilon < \min\{u - a, b - v\}$. By Weierstrass approximation Thm. 14.45, there exists $p \in V[x]$ such that $\|f(x) - p(x)\| \leqslant e$ for all $x \in [u - \varepsilon, v + \varepsilon]$. In particular, since $f = 0$ on $J = [u - \varepsilon, u] \cup [v, v + \varepsilon]$, we have $\|p(x)\| \leqslant e$ for all $x \in J$.

By Prop. 14.47, there exists $h \in C_c^\infty(\mathbb{R}, \mathbb{R})$ such that $h(\mathbb{R}) = [0, 1]$, that $h^{-1}(1) = [u, v]$, and that $h^{-1}(0) = (-\infty, u - \varepsilon] \cup [v + \varepsilon, +\infty)$. Then $hp \in C_c^\infty(\mathbb{R}, V)$ is supported in $[u - \varepsilon, v + \varepsilon]$ and hence in $(-a, a)$. One checks easily that $\|f(x) - h(x)p(x)\| \leqslant e$ for all $x \in (a, b)$. $\qquad\square$

Remark 14.49. It should be noted that if I is an open interval, then $C_c^\infty(I, V)$ and $C_c(I, V)$ are naturally subspaces of $C_c(\mathbb{R}, V)$. This is not true when I is a closed or an half-open-half-closed interval. Therefore, Cor. 14.48 implies that any $f \in C(\mathbb{R}, V)$ supported in an open interval I can be uniformly approximated by smooth functions $\mathbb{R} \to V$ supported in I.

Remark 14.50. The above observation can be generalized: Let X be an LCH space. Assume that Ω is a nonempty *open* subset of X. Recall that Ω is LCH by Prop. 8.41. Then an element of $C_c(\Omega, V)$ is equivalently an element of $C_c(X, V)$ supported in U. The former gives the latter by "extension by zero"; the latter gives the former by restriction to Ω. We will say more about this in Sec. 15.4.

14.6 L^1-approximation; Riemann-Lebesgue lemma

A non-continuous integrable function cannot be approximated uniformly by smooth functions. However, it can be approximated by the latter under the L^1-norm. To avoid distraction, in this section we consider $\mathbb{C}$-valued functions. Fix $I \subset \mathbb{R}$ to be an interval containing at least two points. Then on $\mathscr{R}^1(I, \mathbb{C})$ one can define the L^1-**seminorm** to be

$$\|f\|_{L^1} \equiv \|f\|_{L^1(I,\mathbb{C})} = \int_I |f| \qquad (14.32)$$

It is easy to check that this is a **seminorm**, i.e., it satisfies the definition of a norm, except the assumption that $\|f\|_{L^1} = 0$ implies $f = 0$.

Proposition 14.51. *Let $f \in \mathscr{R}^1(I, \mathbb{C})$. Choose any $\varepsilon > 0$. Then there exists $g \in C_c^\infty(\mathbb{R}, \mathbb{C})$ supported in I such that $\|f - g\|_{L^1} < \varepsilon$. And there exists a **step function** h supported in I (i.e., a linear combination of functions of the form χ_E where E is a compact interval in I) such that $\|f - h\|_{L^1} < \varepsilon$.*

Note that if E is a bounded interval, then χ_E is clearly a linear combination of characteristic functions over compact intervals (including the single point sets). Thus, in the above definition of step functions, one can just assume that E is a bounded interval whose closure is in I.

The following proof shows that Prop. 14.51 can be easily generalized to the case that $f : I \to V$ is strongly integrable on compact subintervals of I and $\int_I |f| < +\infty$. (V is a Banach space.)

Proof. Let $a = \inf I, b = \sup I$. Since $\int_I |f| = \lim_{u \to a, v \to b} \int_a^b |f|$, there exist u, v such that $a < u < v < b$ and that $\int_a^u |f| + \int_v^b |f| < \varepsilon/2$. We claim that there exists a step function h supported in $J = [u, v]$ such that $\int_J |f - h| < \varepsilon/2$. Then $\int_I |f - h| < \varepsilon$, finishing the proof that f can be L^1-approximated by step functions supported in I.

Since $f|_J$ is strongly Riemann integrable, there exists $\sigma = \{a_0 < \cdots < a_n\} \in \mathcal{P}(J)$ such that $\omega(f, \sigma) < \frac{\varepsilon}{2}$. Let $J_i = [a_{i-1}, a_i]$. Pick any $\lambda_i \in f(J_i)$. Then $|f - \lambda_i \cdot \chi_{J_i}| \leqslant \operatorname{diam}(f(J_i))$ on J_i, and hence $\int_{J_i} |f - \lambda_i \cdot \chi_{J_i}| \leqslant \operatorname{diam}(f(J_i)) \cdot |J_i|$. Let $h = \sum_{i=1}^n \lambda_i \cdot \chi_{J_i}$. It follows that

$$\left| \int_J f - \int_J h \right| \leqslant \sum_{i=1}^n \int_{J_i} |f - \lambda_i \cdot \chi_{J_i}| \leqslant \sum_{i=1}^n \operatorname{diam}(f(J_i)) \cdot |J_i| = \omega(f, \sigma) < \frac{\varepsilon}{2}$$

finishing the proof.

Finally, we show that f can be L^1-approximated by smooth functions supported in I. It suffices to prove that any step function supported in I is so. By linearity and triangle inequality, it suffices to prove that χ_E is so, if E is a nonempty compact interval in I. For each $\varepsilon > 0$, it is easy to construct a piecewise linear

(and hence continuous) function $g_0 : \mathbb{R} \to \mathbb{R}$ supported in $(a,b) = \mathrm{Int}(I)$ such that $\int_I |\chi_E - g_0| < \varepsilon$. Choose a bounded open interval $(c,d) \subset (a,b)$ containing $\mathrm{Supp}(g_0)$. By Cor. 14.48, there exists $g \in C_c^\infty(\mathbb{R}, \mathbb{R})$ supported in (c,d) such that $|g(x) - g_0(x)| < \varepsilon/(d-c)$ for all $x \in (c,d)$. It follows that

$$\int_I |g - g_0| = \int_c^d |g - g_0| \leqslant \varepsilon$$

Thus $\int_I |\chi_e - g| < 2\varepsilon$, finishing the proof. $\qquad\square$

We give an application of Prop. 14.51 by proving Riemann-Lebesgue lemma. The RL lemma plays a fundamental role in Fourier analysis, see for instance Pb. 14.6 and Cor. 14.56.

Theorem 14.52 (Riemann-Lebesgue lemma). *Let $f \in \mathscr{R}^1(\mathbb{R}, \mathbb{C})$. Then*

$$\lim_{t\to+\infty} \int_\mathbb{R} f(x)e^{\mathbf{i}tx}dx = \lim_{t\to-\infty} \int_\mathbb{R} f(x)e^{\mathbf{i}tx}dx = 0$$

The idea of the proof is the following: We L^1-approximate f by some compactly supported step function $g \in C_c(\mathbb{R}, \mathbb{C})$. Then it suffices to prove the RL lemma for g. By linearity, it suffices to assume that $g = \chi_{[a,b]}$. This special case can be proved easily.

Proof. By Prop. 14.51, for every $\varepsilon > 0$ there exists a compactly supported step function $g : \mathbb{R} \to \mathbb{C}$ such that $\int_\mathbb{R} |f - g| < \varepsilon$. So $|\int_\mathbb{R} f(x)e^{\mathbf{i}tx} - g(x)e^{\mathbf{i}tx}dx| < \varepsilon$ by Prop. 14.27. Thus

$$\limsup_{t\to\pm\infty} \left| \int_\mathbb{R} f(x)e^{\mathbf{i}tx}dx \right| \leqslant \limsup_{t\to\pm\infty} \left| \int_\mathbb{R} g(x)e^{\mathbf{i}tx}dx \right| + \varepsilon$$

Since ε is arbitrary, it suffices to prove that $\lim_{t\to\pm\infty} \int_\mathbb{R} g(x)e^{\mathbf{i}tx}dx = 0$. Since g is a linear combination of functions of the form $\chi_{[a,b]}$ where $-\infty < a < b < +\infty$, it suffices to prove the RL lemma for this function:

$$\int_\mathbb{R} \chi_{[a,b]} \cdot e^{\mathbf{i}tx}dx = \int_a^b e^{\mathbf{i}tx}dx = \frac{e^{\mathbf{i}nb} - e^{\mathbf{i}na}}{\mathbf{i}t}$$

converges to 0 as $t \to \pm\infty$. $\qquad\square$

14.7 Problems and supplementary material

Let V be a Banach space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$.

Problem 14.1. Solve Exe. 14.8.

Problem 14.2. Prove Thm. 14.34 using Thm. 14.35.

⋆ **Problem 14.3.** Define $f(x) = \left(\int_0^x e^{-t^2} dt \right)^2$ and $g(x) = \int_0^1 \frac{e^{-x^2(t^2+1)}}{t^2 + 1} dt$

1. Show that $f' + g' = 0$, and conclude that $f + g = \frac{\pi}{4}$.

2. Use part 1 to prove

$$\int_{-\infty}^{+\infty} e^{-t^2} dt = \sqrt{\pi} \tag{14.33}$$

Problem 14.4. Let $f : \mathbb{R} \to V$ be strongly integrable on compact intervals, and assume $\int_{\mathbb{R}} |f| < +\infty$. Let $g \in C^1(\mathbb{R}, \mathbb{R})$. Assume that $\|g\|_\infty, \|g'\|_\infty < +\infty$. (This is automatic when g has compact support.) Use Thm. 14.35 to prove that $f * g$ is well-defined, that $f * g \in C^1(\mathbb{R}, \mathbb{R})$, and that $(f * g)' = f * g'$.

Note. Since f is not assumed to be continuous, you cannot use Thm. 14.34 *directly* to compute $(f * g)'$ and to show $f * g \in C^1$. You have two options: (1) Use Thm. 14.35. (2) Use Prop. 14.51 to L^1-approximate f by a continuous compactly-supported function. Then apply Thm. 14.34 (together with Cor. 14.36) to that continuous function. Whichever method you use, I suggest you think about how you can use the other method to solve the problem. □

Remark 14.53. In particular, if $g \in C_c^\infty(\mathbb{R}, \mathbb{R})$, then the above problem shows that $f * g \in C^\infty(\mathbb{R}, \mathbb{R})$ and $(f * g)^{(n)} = f * g^{(n)}$ for all $n \in \mathbb{N}$.

⋆ **Problem 14.5.** The translation of each $f \in \mathscr{R}^1(\mathbb{R}, \mathbb{C})$ by $t \in \mathbb{R}$ is defined to be

$$f_t : \mathscr{R}^1(\mathbb{R}, \mathbb{C}) \to \mathscr{R}^1(\mathbb{R}, \mathbb{C}) \qquad f_t(x) = f(x - t)$$

Prove that the translation map is continuous under the L^1-seminorm, in the sense that $\lim_{t \to 0} \|f - f_t\|_{L^1} = 0$. Namely, prove that

$$\lim_{t \to 0} \int_{\mathbb{R}} |f - f_t| = 0$$

Hint. Use Prop. 14.51 to L^1-approximate f by compactly supported continuous functions or step functions. (Both types of functions will work.) □

Definition 14.54. Let $f \in \mathscr{R}([-\pi, \pi], \mathbb{C})$. For each $n \in \mathbb{Z}$, define the n-th **Fourier coefficient** to be

$$\widehat{f}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-inx} dx$$

We call $\sum_{n=-\infty}^{+\infty} \widehat{f}(n) e^{inx}$ the **Fourier series** of f.

Definition 14.55. For each $N \in \mathbb{N}$, define the **Dirichlet kernel**

$$D_N(x) = \sum_{n=-N}^{N} e^{inx} = \frac{\sin(N + \frac{1}{2})x}{\sin(x/2)}$$

(When $x \in 2\pi\mathbb{Z}$, the RHS above should be $2N + 1$.) From $\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{inx}dx = \delta_{n,0}$ we easily see $\frac{1}{2\pi}\int_{-\pi}^{\pi} D_N(x)dx = 1$.

Problem 14.6. Let $f \in \mathscr{R}(\mathbb{R}, \mathbb{C})$ have period 2π. Define $s_N(f; x) = \sum_{n=-N}^{N} \hat{f}(n)e^{inx}$.

1. Prove that $s_N(f; x) = (f * D_N)(x)$ where the convolution is defined with respect to the integral $\frac{1}{2\pi}\int_I$ where I is any interval of length 2π. (Note that both f and D_N have period 2π. So translating I does not affect the result.) In other words, prove that

$$s_N(f; x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)D_N(x - t)dt = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x - t)D_N(t)dt \qquad (14.34)$$

2. Fix $x \in \mathbb{R}$. Assume that there exist $A, B \in \mathbb{C}$ such that

$$\limsup_{t \to 0^+} \left| \frac{f(x - t) - A}{t} \right| < +\infty \qquad \limsup_{t \to 0^+} \left| \frac{f(x + t) - B}{t} \right| < +\infty \qquad (14.35)$$

(In other words, assume that there exist $\delta, M > 0$ such that $|(f(x-t)-A)/t| \leqslant M$ and $|(f(x + t) - B)/t| \leqslant M$ for all $0 < t < \delta$.) Prove that

$$\lim_{N \to +\infty} s_N(f; x) = \frac{A + B}{2} \qquad (14.36)$$

Hint for part 2. Choose $g : [-\pi, \pi] \to \mathbb{C}$ such that $g(t) = B$ if $t < 0$, and $g(t) = A$ if $t > 0$. Prove that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} g(t)D_N(t)dt = \frac{A + B}{2} \qquad (14.37)$$

Let $\varphi : [-\pi, \pi] \to \mathbb{R}$ such that $\varphi(t) = \frac{f(x - t) - g(t)}{\sin(t/2)}$ if $t \neq 0$. Use Lebesgue's criterion to show that $\varphi \in \mathscr{R}([-\pi, \pi], \mathbb{C})$. Use Riemann-Lebesgue lemma to prove $\lim_{N \to +\infty} \int_{-\pi}^{\pi} \varphi(t) \sin \left(N + \frac{1}{2} \right)t \cdot dt = 0$. $\qquad \square$

Corollary 14.56. *Let $g \in C^1([a - \pi, a + \pi], \mathbb{C})$. Let $f : \mathbb{R} \to \mathbb{C}$ be a 2π-periodic function such that $f(x) = g(x)$ if $a - \pi < x < a + \pi$, and that $f(a + \pi) = \dfrac{g(a - \pi) + g(a + \pi)}{2}$. Clearly $f \in \mathscr{R}(\mathbb{R}, \mathbb{C})$. Then the Fourier series of f converges pointwise to f.*

Proof. Immediate from part 2 of Pb. 14.6. □

Example 14.57. If we let $a = 0$ and $g(x) = x$, then one can compute that the Fourier series of f is (13.10). It converges pointwise but not uniformly to f.

15 A topological proof of Stone-Weierstrass theorem

15.1 *-algebras and subalgebras

Fix $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$.

Definition 15.1. An **$\mathbb{F}$-algebra** is defined to be a ring $\mathscr{A}$ (not necessarily having 1) which is at the same time also an $\mathbb{F}$-vector space (where the vector addition is equal to the ring addition) such that the ring multiplication and the scalar multiplication satisfy the associativity: For every $\lambda \in \mathbb{F}$ and $x, y \in \mathscr{A}$, we have

$$\lambda(xy) = (\lambda x)y = x(\lambda y) \tag{15.1}$$

An $\mathbb{F}$-algebra is called **unital** if $\mathscr{A}$, as a ring, has the identity 1. In this case, we write $\lambda \cdot 1$ as λ if $\lambda \in \mathbb{F}$.

An $\mathbb{F}$-algebra is called **commutative** or **abelian** if $xy = yx$ for all $x, y \in \mathscr{A}$.

If $\mathscr{A}$ is an $\mathbb{F}$-algebra, then an **($\mathbb{F}$-)subalgebra** is a subset $\mathscr{B}$ which is invariant under the ring addition, ring multiplication, and scalar multiplication. (Namely, $\mathscr{B}$ is a subring and also a subspace of $\mathscr{A}$.) If $\mathscr{A}$ is unital, then a **unital ($\mathbb{F}$-)subalgebra** of $\mathscr{A}$ is an $\mathbb{F}$-subalgebra containing the identity of $\mathscr{A}$. □

Remark 15.2. A unital $\mathbb{F}$-algebra $\mathscr{A}$ is equivalently a ring with unit 1, together with a ring homomorphism $\mathbb{C} \to Z(\mathscr{A})$ where $Z(\mathscr{A})$ is the **center** of $\mathscr{A}$, i.e.

$$Z(\mathscr{A}) = \{x \in \mathscr{A} : xy = yx \text{ for every } y \in \mathscr{A}\}$$

We leave it to the readers to check the equivalence.

Example 15.3. If V is a $\mathbb{F}$-vector space, then $\mathrm{End}(V)$, the set of $\mathbb{F}$ linear maps $V \to V$, is naturally an $\mathbb{F}$-algebra. If V is a normed vector space, then $\mathfrak{L}(V)$ is an $\mathbb{F}$-algebra.

Definition 15.4. A **(complex) *-algebra** is defined to be a $\mathbb{C}$-algebra together with an **antilinear map** $* : \mathscr{A} \to \mathscr{A}$ sending x to x^* (where "antilinear" means that for every $a, b \in \mathbb{C}$ and $x, y \in \mathscr{A}$ we have $(ax + by)^* = \bar{a}x^* + \bar{b}y^*$) such that for every $x, y \in \mathscr{A}$, we have

$$(x^*)^* = x \qquad (xy)^* = y^* x^*$$

Note that $*$ must be bijective. We call $*$ an **involution**. A ***-subalgebra** $\mathscr{B}$ is defined to be a subalgebra satisfying $x \in \mathscr{B}$ iff $x^* \in \mathscr{B}$. If $\mathscr{A}$ is a unital algebra with unit **1**, we say that $\mathscr{A}$ is a **unital *-algebra** if $\mathscr{A}$ is equipped with an involution $* : \mathscr{A} \to \mathscr{A}$ such that $\mathscr{A}$ is a *-algebra, and that

$$\mathbf{1}^* = \mathbf{1}$$

A unital *-subalgebra is a unital subalegbra and also a *-subalgebra.

Example 15.5. The set of complex $n \times n$ matrices $\mathbb{C}^{n \times n}$ is naturally a unital $*$-algebra if for every $A \in \mathbb{C}^{n \times n}$ we define $A^* = \overline{A}^{\mathrm{t}}$, the complex conjugate of the transpose of A.

In this chapter, we are mainly interested in abelian algebras.

Example 15.6. Let X be a set. Then $\mathbb{F}^X$ is naturally a unital $\mathbb{F}$-algebra, and $l^\infty(X, \mathbb{C})$ is its unital $\mathbb{F}$-subalgebra. If X is a topological space, then $C(X, \mathbb{F})$ is a unital $\mathbb{F}$-subalgebra of $\mathbb{F}^X$. If X is compact, then $C(X, \mathbb{F})$ is a unital $\mathbb{F}$-subalgebra of $l^\infty(X, \mathbb{F})$.

The following is our main example of this chapter.

Example 15.7. Let X be a set. Then $\mathbb{C}^X$ is a unital *-algebra if for every $f \in \mathbb{C}^X$ we define

$$f^* : X \to \mathbb{C} \qquad f^*(x) = \overline{f(x)} \tag{15.2}$$

Then $l^\infty(X, \mathbb{C})$ is a unital *-subalgebra of $\mathbb{C}^X$. Assume that X is a compact topological space. Then $C(X, \mathbb{C})$ is a unital *-subalgebra of $l^\infty(X, \mathbb{C})$. If $f_1, \ldots, f_n \in C(X, \mathbb{C})$, then $\mathbb{C}[f_1, \ldots, f_n]$, the set of polynomials of $f_1, \ldots, f_n$ with coefficients in $\mathbb{C}$, is a unital subalgebra of $C(X, \mathbb{C})$. And $\mathbb{C}[f_1, f_1^*, \ldots, f_n, f_n^*]$ is a unital *-subalgebra of $C(X, \mathbb{C})$.

More generally, we have:

Example 15.8. Let $\mathscr{A}$ be an abelian unital $\mathbb{F}$-algebra. Let $\mathfrak{S} \subset \mathscr{A}$. Then

$$\mathbb{F}[\mathfrak{S}] = \mathrm{Span}_{\mathbb{F}}\{x_1^{n_1} \cdots x_k^{n_k} : k \in \mathbb{Z}_+, x_i \in \mathfrak{S}, n_i \in \mathbb{N}\} \tag{15.3}$$

the set of polynomials of elements in $\mathfrak{S}$, is the smallest unital $\mathbb{F}$-subalgebra containing $\mathfrak{S}$, called the **unital $\mathbb{F}$-subalgebra generated by** $\mathfrak{S}$. (Here, we understand $x^0 = 1$ if $x \in \mathscr{A}$.) Thus, if $\mathscr{A}$ is an abelian unital *-algebra, then $\mathbb{C}[\mathfrak{S} \cup \mathfrak{S}^*]$ (where $\mathfrak{S}^* = \{x^* : x \in \mathfrak{S}\}$) is the smallest unital *-algebra containing $\mathfrak{S}$, called the **unital *-subalgebra generated by** $\mathfrak{S}$.

15.2 Stone-Weierstrass (SW) theorems

The main goal of this chapter is to prove the following theorem.

Theorem 15.9 (**SW theorem, compact real version**)**.** *Let X be a compact Hausdorff space. Let $\mathscr{A}$ be unital subalgebra of $C(X, \mathbb{R})$ separating points of X. Then $\mathscr{A}$ is dense in $C(X, \mathbb{R})$ (under the l^∞-norm).*

Example 15.10. Let $X = [a, b]$ be a compact interval in $\mathbb{R}$. Then $\mathbb{R}[x]$ is a unital subalgebra of $C(X, \mathbb{R})$. It separates points of $[a, b]$ because it contains id. Thus, by SW theorem, $\mathbb{R}[a, b]$ is dense in $C([a, b], \mathbb{R})$. This special case was proved in Thm. 14.45.

More generally, let $X = I_1 \times \cdots \times I_N$ where each $I_i \subset \mathbb{R}$ is a compact interval. Let $\pi_i : X \to I_i$ be the projection onto the i-th component, i.e., the i-th coordinate function. Then $\pi_1, \ldots, \pi_N$ separate points of X, since $\pi_1 \times \cdots \times \pi_N$ is the identity map of X. Thus, by SW theorem, $\mathbb{R}[\pi_1, \ldots, \pi_N]$ is dense in $C(X, \mathbb{R})$. In fact, we will first prove this special case before we prove the general SW theorem. □

From the above version of SW theorem it is easy to prove:

Theorem 15.11 (SW theorem, compact complex version). *Let X be a compact Hausdorff space. Let $\mathscr{A}$ be a unital *-subalgebra of $C(X, \mathbb{C})$ separating points of X. Then $\mathscr{A}$ is dense in $C(X, \mathbb{R})$ (under the l^∞-norm).*

Proof. Let $\mathrm{Re}\mathscr{A} = \{\mathrm{Re}f = (f + f^*)/2 : f \in \mathscr{A}\}$. Since for each $f \in \mathscr{A}$ we have $f^* \in \mathscr{A}$, we know that $\mathrm{Re}f \in \mathscr{A}$. This proves that $\mathrm{Re}\mathscr{A} \subset \mathscr{A}$. Since $f = \mathrm{Re}(f) - \mathbf{i}\mathrm{Re}(\mathbf{i}f)$, we conclude

$$\mathscr{A} = \mathrm{Re}\mathscr{A} + \mathbf{i}\mathrm{Re}\mathscr{A} \tag{15.4}$$

In other words, elements of $\mathscr{A}$ are precisely of the form $\alpha + \mathbf{i}\beta$ where $\alpha, \beta \in \mathrm{Re}\mathscr{A}$. From this, it is clear that $\mathrm{Re}\mathscr{A}$ is a unital subalgebra of $C(X, \mathbb{R})$ separating points of X. Thus, by Thm. 15.9, $\mathrm{Re}\mathscr{A}$ is dense in $C(X, \mathbb{R})$. It is clear from (15.4) that $\mathscr{A}$ is dense in $C(X, \mathbb{C})$. □

Example 15.12. Let $S^1 = \{z \in \mathbb{C} : |z| = 1\}$. For each $n \in \mathbb{Z}$, let $e_n : S^1 \to \mathbb{C}$ be defined by $e_n(e^{\mathbf{i}x}) = e^{\mathbf{i}nx}$. In other words, $e_n(z) = z^n$. Then $\mathbb{C}[e_1, e_{-1}] = \mathrm{Span}_{\mathbb{C}}\{e_n : n \in \mathbb{Z}\}$ is a unital *-subalgebra of $C(S^1, \mathbb{C})$. It separates points of S^1 since e_1 does. Therefore, by SW theorem, $\mathbb{C}[e_1, e_{-1}]$ is dense in $\mathbb{C}(S^1, \mathbb{C})$.

One often views

$$C(S^1, \mathbb{C}) = \{g \in C([-\pi, \pi], \mathbb{C}) : g(-\pi) = g(\pi)\} \tag{15.5}$$

since any g in the RHS corrsponds bijectively to f in the LHS by setting $f(e^{\mathbf{i}t}) = g(t)$. Therefore, the above conclusion is that *any $g \in C([-\pi, \pi], \mathbb{C})$ satisfying $g(-\pi) = g(\pi)$ can be approximated uniformly by functions of the form $\sum_{n=-N}^{N} a_n \cdot e^{\mathbf{i}nx}$ where $a_n \in \mathbb{C}$.* This property is fundamental to the theory of Fourier series. □

Example 15.13. We continue the above discussion. $\mathbb{C}[e_1] = \mathrm{Span}_{\mathbb{C}}\{e_n : n \in \mathbb{N}\}$ is a unital subalgebra (though not a *-subalgebra) of $C(S^1, \mathbb{C})$ separating points of S^1. Let us prove that it is not dense in $C(S^1, \mathbb{C})$.

Proof. We view functions on S^1 as those on $[-\pi, \pi]$ having the same values at π and at $-\pi$. We claim that e_{-1} is not in the closure of $\mathbb{C}[e_1]$. Indeed, it can be checked that $\int_{-\pi}^{\pi} e_{-1}e_1 = 2\pi$, and that $\int_{-\pi}^{\pi} p \cdot e_1 = 0$ for every $p \in \mathbb{C}[e_1]$ (since this is true when $p = e_n$ and $n \geqslant 0$). If e_{-1} is in the closure of $\mathbb{C}[e_1]$, then there is a sequence $(p_k)_{k \in \mathbb{Z}_+}$ in $\mathbb{C}[e_1]$ converging uniformly on $[-\pi, \pi]$ to e_{-1}. By Cor. 13.21, $0 = \int_{-\pi}^{\pi} p_k e_1$ converges to $\int_{-\pi}^{\pi} e_{-1}e_1 = 2\pi$ as $k \to \infty$, impossible. $\qquad\square$

Example 15.14. Let X, Y be compact metric spaces spaces. Let $\mathscr{A}$ be the subspace of $C(X \times Y, \mathbb{R})$ spanned by elements of the form fg where $f \in C(X, \mathbb{R})$ and $g \in C(Y, \mathbb{R})$. (More precisely, one should understand fg as $(f \circ \pi_X) \cdot (g \circ \pi_Y)$ where π_X, π_Y are the projections of $X \times Y$ onto X and Y respectively.) By the Urysohn functions (cf. Rem. 7.118), we see that $C(X, \mathbb{R})$ resp. $C(Y, \mathbb{R})$ separates points of X resp. Y. (This is in fact also true when X, Y are compact Hausdorff spaces.) Thus $\mathscr{A}$ is a unital subalgebra of $C(X \times Y, \mathbb{R})$ separating points of $X \times Y$. By SW theorem, $\mathscr{A}$ is dense in $C(X \times Y, \mathbb{R})$. Nevertheless, we will prove SW theorem by first proving this special case: see Cor. 15.33.[1]

15.3 Proof of SW, I: polynomial approximation on $[0, 1]^{\mathscr{I}}$

Starting from this section, we begin our proof of SW Thm. 15.9. We first explain our strategy of the proof. Let X be a compact Hausdorff space.

We first consider the special case that X is metrizable. Then by Thm. 8.45, X is homeomorphic to a closed (and hence compact) subset of $[0, 1]^{\mathbb{Z}_+}$. Therefore, we may assume that X is a closed subset of $[0, 1]^{\mathbb{Z}_+}$. As we will see, any continuous real-valued function on X can be extended to a continuous function on $[0, 1]^{\mathbb{Z}_+}$. (This is due to **Tietze extension theorem**, which will be proved in the next section.) Thus, it suffices to prove SW theorem for $[0, 1]^{\mathbb{Z}_+}$. In fact, as we will see, it suffices to prove that any element of $C([0, 1]^{\mathbb{Z}_+}, \mathbb{R})$ can approximated uniformly by polynomials, i.e., by elements of $\mathbb{R}[\pi_1, \pi_2, \dots]$ where $\pi_n : [0, 1]^{\mathbb{Z}_+} \to [0, 1]$ is the projection onto the n-th component. This task will be achieved in this section.

The above method can be easily genearlized to the case that X is not necessarily metrizable. In fact, since $\mathscr{A} \subset C(X, \mathbb{R})$ separates points of X, as in the proof of Thm. 8.45, $\mathscr{A}$ enables us to embed X onto a compact subset of $[0, 1]^{\mathscr{I}}$ where $\mathscr{I}$ is an index set. Therefore, we need to show that $[0, 1]^{\mathscr{I}}$ is compact. This is indeed due to:

Theorem 15.15 (Tychonoff theorem). *Let $(X_\alpha)_{\alpha \in \mathscr{I}}$ be a family of compact topological spaces. Then the product space $S = \prod_{\alpha \in \mathscr{I}} X_\alpha$ (equipped with the product topology) is compact.*

---

[1]More precisely, we don't need Exp. 15.14 to prove SW. But our method of proving SW immediately implies Exp. 15.14. To put it differently, we don't need the full power of the SW theorem to prove Exp. 15.14.

Index

References

[Axl]   Axler, S. (2015). Linear algebra done right. 3rd ed.

[Che92]   Chernoff, P. R. (1992). A simple proof of Tychonoff's theorem via nets. The American mathematical monthly, 99(10), 932-934.

[Die-H]   Dieudonné, J. (1983). History of Functional Analyais. North Holland, 1st edition.

[Fol]   Folland, G. B. (1999). Real analysis: modern techniques and their applications (Vol. 40). John Wiley & Sons.

[Hau14]   Hausdorff, F. (1914) Grundzüge der Mengenlehre

[Jah]   Jahnke, H. N. (2003). A history of analysis (No. 24). American Mathematical Soc..

[Kel]   Kelley, J. L., General topology.

[Kli]   Kline, M. (1990). Mathematical Thought from Ancient to Modern Times.

[Lee]   Lee, J. M., (2013). Introduction to Smooth manifolds, 2nd ed. Springer.

[MS22]   Moore, E. H., & Smith, H. L. (1922). A general theory of limits. American journal of Mathematics, 44(2), 102-121.

[Mun]   Munkres, J. (2000). Topology. Second Edition.

[NB97]   Narici, L., & Beckenstein, E. (1997). The Hahn-Banach theorem: the life and times. Topology and its Applications, 77(2), 193-211.

[RS]   Reed, M., & Simon, B. (1972). Methods of Modern Mathematical Physics I: Functional analysis, Academic Press, New York.

[Rud-P]   Rudin, W. (1976). Principles of Mathematical Analysis. 3rd ed.

[Rud-R]   Rudin, W. (1987). Real and complex analysis. 3rd ed.

[Sim-R]   Simon, B. (2015). Real analysis. A comprehensive course in analysis, part 1. Providence, RI: American Mathematical Society (AMS)

[Sto37]   Stone, M. H. (1937). Applications of the theory of Boolean rings to general topology. Transactions of the American Mathematical Society, 41(3), 375-481.

[Sto48]   Stone, M. H. (1948). The generalized Weierstrass approximation theorem. Mathematics Magazine, 21(5), 167-184, 237-254.

[Wil]   Willard, S. (1970). General topology.

[Zor-2]   Zorich, V. A.(2016). Mathematical analysis II, 2nd ed.

YAU MATHEMATICAL SCIENCES CENTER, TSINGHUA UNIVERSITY, BEIJING, CHINA.

E-mail: binguimath@gmail.com   bingui@tsinghua.edu.cn