

# Qiuzhen Lectures on Analysis

BIN GUI

September 13, 2023

<b>1</b>	<b>Basic set theory and numbers</b>	<b>2</b>
1.1	Basic operations and axioms . . . . .	2
1.2	Partial and total orders, equivalence relations . . . . .	7
1.3	$\mathbb{Q}$ , $\mathbb{R}$ , and $\overline{\mathbb{R}} = [-\infty, +\infty]$ . . . . .	9
1.4	Cardinalities, countable sets, and product spaces $Y^X$ . . . . .	13
<b>2</b>	<b>Metric spaces, convergence of sequences, and continuous functions</b>	<b>19</b>
2.1	Introduction . . . . .	19
2.2	Basic definitions and examples . . . . .	21
2.3	Convergence of sequences . . . . .	23
2.4	Continuous maps of metric spaces . . . . .	27
2.5	Homeomorphisms and isometric isomorphisms; convergence in $\overline{\mathbb{R}}$ . . . . .	32
2.6	Problems and supplementary material . . . . .	36
<b>3</b>	<b>Sequential compactness and completeness</b>	<b>39</b>
3.1	Sequential compactness . . . . .	39
3.2	Outlook: sequentially compact function spaces . . . . .	43
3.3	Complete metric spaces and Banach spaces . . . . .	44
3.4	The Banach spaces $l^\infty(X, V)$ and $C(X, V)$ . . . . .	48
3.5	Problems and supplementary material . . . . .	51
	<b>Index</b>	<b>53</b>
	<b>References</b>	<b>55</b>

# 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:

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

If  $\mathbb{F}$  is any field (e.g.  $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ ), we let  $\mathbb{F}^\times = \mathbb{F} \setminus \{0\}$ .

Topics marked with **\*\*** are technical and/or their methods are rarely used in later studies. You can skim or skip them on first reading. Topics marked with **\*** are interesting, but not necessarily technical. They are not essential for understanding the later part of the notes.

## 1.1 Basic operations and axioms

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

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

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

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

assuming that rational and real numbers exist and satisfy the properties we are familiar with in high school mathematics.

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

$$\{x : x \text{ satisfies property...}\} \tag{1.1}$$

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

$$S = \{A : A \text{ is a set and } A \notin A\} \tag{1.2}$$

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:

$$\{x \in X : x \text{ satisfies property } \dots\} \quad (1.3)$$

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

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

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

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

$$\bigcup_{A \in \mathcal{A}} A = \{x \in X : \text{there exists } A \in \mathcal{A} \text{ such that } x \in A\} \quad (1.4a)$$

$$\bigcap_{A \in \mathcal{A}} A = \{x \in X : \text{for all } A \in \mathcal{A} \text{ we have } x \in A\} \quad (1.4b)$$

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 (1.1). You should interpret such definition as (1.3), where the set  $X$  is omitted because it is clear from the context. For instance, if the context is clear, the set  $\{x \in \mathbb{R} : x \geq 0\}$  could be simply written as  $\{x : x \geq 0\}$  or even  $\{x \geq 0\}$ . By the same token, the phrase " $\in X$ " in (1.4) could be omitted. So we can also write

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

which are special cases of (1.4).

**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  $X \setminus A$  (where  $A \subset X$ ) as  $A^c$  and call it the **complement** of  $A$ .

**Example 1.3.** We have

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

The readers may notice that these examples are not exactly in the form (1.4). 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  $A_1, \dots, A_n$  are sets, their **Cartesian product** exists:

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

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

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

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  $\emptyset$  if some  $A_i$  is  $\emptyset$ .

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

**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  $a + bi$  where  $a, b \in \mathbb{R}$ . Define

$$\begin{aligned}(a + bi) + (c + di) &= (a + c) + (b + d)i \\ (a + bi) \cdot (c + di) &= (ac - bd) + (ad + bc)i\end{aligned}$$

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

If  $z = a + bi$ , 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  $\bar{z} = a - bi$ . Then  $|z|^2 = z\bar{z}$ .

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

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  $f : A \rightarrow B$ , 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**

---

<sup>1</sup>The readers can easily find the definition of fields online or through textbooks (e.g. [Rud-P, Def. 1.12]). We will not present the full definition of fields in the notes. Just 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  $\{a + bi + cj + dk : a, b, c, d \in \mathbb{R}\}$  is not a field because it is not commutative ( $ij = -ji = k$ ). 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.

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

$$f|_E : E \rightarrow B \quad x \mapsto f(x)$$

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,  $\emptyset$  (as a subset of  $\emptyset \times B$ ) is the only function from  $\emptyset$  to  $B$ . (A false assumption implies any statement.) If  $A \neq \emptyset$ , there are no functions  $A \rightarrow \emptyset$ .

**Definition 1.8.** A function  $x : \mathbb{Z}_+ \rightarrow A$  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  $(x_n)$ ).

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:  $\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}$ . 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  $S : I \rightarrow 2^X$  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

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

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  $S : I \rightarrow 2^X$  is defined to be  $S_i = S(i) = [0, i)$ .

**Exercise 1.12.** Let  $f : A \rightarrow B$  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  $f^{-1}(b) \neq \emptyset$ .  $f$  is called **bijective** if it is both surjective and injective. Define the **identity maps**  $\text{id}_A : A \rightarrow A, a \mapsto a$  and  $\text{id}_B$  in a similar way. Prove that

$$f \text{ is injective} \iff \text{there is } g : B \rightarrow A \text{ such that } g \circ f = \text{id}_A \quad (1.5a)$$

$$f \text{ is surjective} \iff \text{there is } g : B \rightarrow A \text{ such that } f \circ g = \text{id}_B \quad (1.5b)$$

$$f \text{ is bijective} \iff \text{there is } g : B \rightarrow A \text{ such that } g \circ f = \text{id}_A \text{ and } f \circ g = \text{id}_B \quad (1.5c)$$

Show that the  $g$  in (1.5a) (resp. (1.5b), (1.5c)) is surjective (resp. injective, bijective).

The equivalence (1.5b) 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  $S_i \neq \emptyset$  for all  $i \in I$ , then there exists a function (the **choice function**)

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

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  $f : A \rightarrow B$  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  $f \circ g = \text{id}_B$ .

**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 (1.5a) and (1.5c). 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**)  $\leq$  on  $A$  is a relation on  $A \times A$  satisfying for all  $a, b, c \in A$  that:

- (Reflexivity)  $a \leq a$ .
- (Antisymmetry) If  $a \leq b$  and  $b \leq a$  then  $a = b$ .
- (Transitivity) If  $a \leq b$  and  $b \leq c$  then  $a \leq c$ .

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

**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  $(2^X, \subset)$  is a partially ordered set.
- $\mathbb{R}^2$  is a partially ordered set, if we define  $(a, b) \leq (c, d)$  to be  $a \leq c$  and  $b \leq d$ .

**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  $f : A \rightarrow B$ , there is an equivalence relation  $\sim$  on  $A$  and a bijective map  $\Phi : A/\sim \rightarrow B$  such that the following diagram commutes:

$$\begin{array}{ccc} & A & \\ \pi \swarrow & & \searrow f \\ A/\sim & \xrightarrow{\Phi} & B \end{array} \quad (1.6)$$

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, (1.6) shows that the maps  $f$  and  $\pi$  are equivalent, and that this equivalence is implemented by the bijection  $\Phi$ . 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.<sup>2</sup>

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

$$\Phi(a_1, \dots, a_m) = a_1 e_1 + \dots + a_m e_m \quad \Psi(b_1, \dots, b_n) = b_1 \varepsilon_1 + \dots + b_n \varepsilon_n$$

Let  $T : V \rightarrow W$  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  $A \in \mathbb{F}^{n \times m}$  such that the following diagram commutes:

$$\begin{array}{ccc} \mathbb{F}^m & \xrightarrow[\cong]{\Phi} & V \\ A \downarrow & & \downarrow T \\ \mathbb{F}^n & \xrightarrow[\cong]{\Psi} & W \end{array} \quad (1.7)$$

---

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



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  $\Phi$  (on the sources) and  $\Psi$  (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  $\mathbb{Z}^\times = \mathbb{Z} \setminus \{0\}$ ) 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

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

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

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

- (Addition preserves  $\leq$ ) If  $a \leq b$  then  $a + c \leq b + c$ .
- (Multiplication by  $\mathbb{F}_{\geq 0}$  preserves  $\leq$ ) If  $a, b \geq 0$  then  $ab \geq 0$ .

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

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

$$\begin{aligned}
a \geq b \wedge c \geq d &\implies a + c \geq b + d \\
a \geq 0 &\iff -a \leq 0 \\
a \geq 0 \wedge b \geq c &\iff ab \geq ac \\
a \leq 0 \wedge b \geq c &\iff ab \leq a \\
&a^2 \geq 0 \\
0 < a \leq b &\implies 0 < b^{-1} \leq a^{-1}
\end{aligned}$$

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

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

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 Archimedean 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  $\leq$  such that  $\mathbb{F}$  is an ordered field, and if the order  $\leq$  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  $-x + x \leq -x + y$ ). By Archimedean property, there exists  $n \in \mathbb{Z}_+$  such that  $n(y - x) > 1$ . So  $y - x > \frac{1}{n}$  and hence  $x + \frac{1}{n} < y$ .

Let us prove that the subset

$$A = \left\{ k \in \mathbb{Z} : \frac{k}{n} \leq x \right\}$$

is nonempty and bounded from above in  $\mathbb{Z}$ . By Archimedean property, there is  $m \in \mathbb{Z}_+$  such that  $m > nx$ , i.e.  $\frac{m}{n} > x$ . So for each  $k \in \mathbb{Z}_+$  satisfying  $k \geq m$ , we have  $\frac{k}{n} = \frac{m}{n} + \frac{k-m}{n} > x$ . 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  $\frac{l}{n} > -x$ . So  $-\frac{l}{n} < x$ , 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  $x < \frac{k+1}{n}$ . Since  $\frac{k}{n} \leq x$ , we have

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

This proves  $x < p < y$  with  $p = \frac{k+1}{n}$ . □

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

**Definition 1.30.** Let  $(X, \leq)$  be a partially ordered set and  $E \subset X$ . An **upper bound of  $E$  in  $X$**  is an element  $x \in X$  satisfying  $e \leq x$  for all  $e \in E$ . An upper bound  $x \in X$  of  $E$  is called a **least upper bound** or a **supremum** if  $x \leq y$  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  $\inf E$ . □

**Definition 1.31.** Let  $(X, \leq)$  be a partially ordered set. 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  $\mathcal{A} \subset 2^X$ , i.e.,  $\mathcal{A}$  is a set of subsets of  $X$ . Then  $\mathcal{A}$  is bounded from above by  $X$ , and is bounded from below by  $\emptyset$ . Moreover,

$$\sup \mathcal{A} = \bigcup_{A \in \mathcal{A}} A \quad \inf \mathcal{A} = \bigcap_{A \in \mathcal{A}} A$$

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

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  $\tau : \mathbb{Q} \rightarrow \mathbb{F}$  and  $\iota : \mathbb{Q} \rightarrow \mathbb{R}$  preserving the algebraic operations and the order of  $\mathbb{Q}$ , and use a commutative diagram to indicate that  $\tau$  and  $\iota$  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  $\mathbb{R}$ .)  $\square$

**Definition 1.36.** Let  $-\infty, +\infty$  be two different symbols, and extend the total order  $\leq$  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

$$\begin{aligned} x \pm \infty &= \pm\infty + x = \pm\infty & +\infty - (-\infty) &= +\infty \\ x \cdot (\pm\infty) &= \pm\infty \cdot x = \begin{cases} \pm\infty & \text{if } x > 0 \\ \mp\infty & \text{if } x < 0 \end{cases} \\ \frac{x}{\pm\infty} &= 0 \\ \frac{\pm\infty}{x} &= x^{-1} \cdot (\pm\infty) \quad \text{if } x \neq 0 \end{aligned}$$

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

$$\begin{aligned} [a, b] &= \{x \in \overline{\mathbb{R}} : a \leq x \leq b\} & (a, b) &= \{x \in \overline{\mathbb{R}} : a < x < b\} \\ (a, b] &= \{x \in \overline{\mathbb{R}} : a < x \leq b\} & [a, b) &= \{x \in \overline{\mathbb{R}} : a \leq x < b\} \end{aligned} \tag{1.8}$$

So  $\mathbb{R} = (-\infty, +\infty)$  and  $\overline{\mathbb{R}} = [-\infty, +\infty]$ .

In this course, unless otherwise stated, an interval always means one of the four sets in (1.8). 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  $\text{card}(A) = \text{card}(B)$  (or simply  $A \approx B$ ), if there is a bijection  $f : A \rightarrow B$ . We write  $\text{card}(A) \leq \text{card}(B)$  (or simply  $A \lesssim B$ ) if  $A$  and a subset of  $B$  have the same cardinality.

**Exercise 1.39.** Show that  $\text{card}(A) \leq \text{card}(B)$  iff there is an injection  $f : A \rightarrow B$ , iff there is a surjection  $g : B \rightarrow A$ . (You need either Axiom of Choice or its consequence (1.5b) to prove the last equivalence.)

It is clear that  $\approx$  is an equivalence relation on the collection of sets. It is also true that  $\lesssim$  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  $A \approx B$ .*

★ *Proof.* Assume WLOG that  $A \subset B$ . Let  $f : B \rightarrow A$  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  $B_n \setminus A_n \rightarrow B_{n+1} \setminus A_{n+1}$  (since  $f$  gives bijections  $B_n \rightarrow B_{n+1}$  and  $A_n \rightarrow A_{n+1}$ ). Therefore, we have a bijection  $g : B \rightarrow A$  defined by

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

where “otherwise” means either  $x \in C$  or  $x \in A_n \setminus B_{n+1}$  for some  $n$ . □

Intuition about the above proof: View  $B$  as an onion. The layers of  $B$  are  $B_n \setminus A_n$  (the odd layers) and  $A_n \setminus B_{n+1}$  (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.*  $f : (0, 1) \rightarrow (a, b)$  sending  $x$  to  $(b - a)x + a$  is a bijection.  $\square$

**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

$$f : \mathbb{R} \rightarrow (-1, 1) \quad f(x) = \begin{cases} \frac{x}{1+x} & \text{if } x \geq 0 \\ -f(-x) & \text{if } x < 0 \end{cases} \quad (1.9)$$

is bijective.  $\square$

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 (1.9) 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,  $\mathbb{R} = (-\infty, +\infty) \approx \overline{\mathbb{R}}$ .  $\square$

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

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

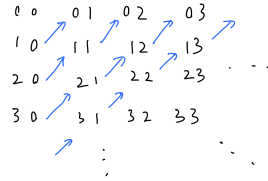
**Remark 1.45.** Let  $A \subset \mathbb{N}$ . If  $A$  is bounded above, then  $A \subset \{0, \dots, n\}$  and hence  $A$  is finite. If  $A$  is not bounded above, then we can construct a strictly increasing sequence  $(x_n)_{n \in \mathbb{N}}$  in  $A$ . (Pick any  $x_0 \in A$ . Suppose we have  $x_n \in A$ . Since  $x_n$  is not an upper bound 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  $\mathbb{N} \rightarrow A$ . Therefore  $A \gtrsim \mathbb{N}$ , and hence  $A \approx \mathbb{N}$  by Schröder-Bernstein.

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

**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  $A_1, A_2, \dots$  be countable sets. Since each  $A_i$  is countable, there is a surjection  $f_i : \mathbb{N} \rightarrow A_i$ . Thus, the map  $f : \mathbb{N} \times \mathbb{N} \rightarrow \bigcup_i A_i$  defined by  $f(i, j) = f_i(j)$  is surjective. Therefore, it suffices to show that there is a surjection  $\mathbb{N} \rightarrow \mathbb{N} \times \mathbb{N}$ . This is true, since we have a bijection  $g : \mathbb{N} \rightarrow \mathbb{N} \times \mathbb{N}$

where  $g(0), g(1), g(2), \dots$  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  $\mathbb{Q} \lesssim \mathbb{Z}$ . 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  $\mathbb{Q} \lesssim \mathbb{Z} \times \mathbb{Z} \approx \mathbb{Z}$ . □

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  $X \times \mathbb{N} \approx X$ .*

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

**Definition 1.50.** Let  $X, Y$  be sets. Then

$$Y^X = \{\text{functions } f : X \rightarrow Y\} \quad (1.10)$$

A more precise definition of  $Y^X$  (in the spirit of (1.3)) is  $\{f \in X \times Y \mid f : X \rightarrow Y \text{ is a function}\}$ . Note that by Rem. 1.7,

$$Y^\emptyset = \{\emptyset\} \quad (1.11)$$

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

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

$$Y^{\{1, \dots, n\}} \rightarrow Y^n \quad f \mapsto (f(1), \dots, f(n))$$

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

$$Y^X = \underbrace{Y \times Y \times \cdots}_{\text{card}(X) \text{ pieces}} \quad (1.12)$$

This generalizes the intuition in Def. 1.8 that a function  $f : \mathbb{Z}_+ \rightarrow Y$  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:  $\square$

**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 if 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**  $\chi_A : X \rightarrow \{0, 1\}$  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:

$$2^X \rightarrow \{0, 1\}^X \quad A \mapsto \chi_A$$

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

$$\begin{aligned}\Phi : Y^X &\rightarrow Y^{A_1} \times \cdots \times Y^{A_n} \\ f &\mapsto (f|_{A_1}, \dots, f|_{A_n})\end{aligned}\tag{1.13}$$

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

**Exercise 1.56.** Assume that  $A_1, \dots, A_n$  are subsets of  $X$ . Define  $\Phi$  by (1.13). Prove that  $\Phi$  is injective iff  $X = A_1 \cup \cdots \cup A_n$ . Prove that  $\Phi$  is surjective iff  $A_1, \dots, A_n$  are pairwise disjoint.

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

*Proof.* The special case that  $m = 0$  (i.e.  $X = \emptyset$ , cf. (1.11)) and  $m = 1$  is clear. When  $m > 1$ , assume WLOG that  $X = \{1, \dots, m\}$ . Then  $X = \{1\} \sqcup \cdots \sqcup \{m\}$ . Apply Prop. 1.55 to this disjoint union. We see that  $Y^X \simeq Y \times \cdots \times Y \simeq \{1, \dots, n\}^m$  has  $n^m$  elements.  $\square$

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

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

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

Assume that  $X \approx 2^X$ . So there is a bijection  $\Phi : X \rightarrow 2^X$  sending each  $x \in X$  to a subset  $\Phi(x)$  of  $X$ . Motivated by Russell's Paradox (1.2), we define

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

Since  $\Phi$  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.  $\square$

**Remark 1.59.** Write  $\{1, \dots, n\}^X$  as  $n^X$  for short. Assuming that real numbers have decimal, binary, or (more generally) base- $n$  presentations where  $n \in \mathbb{Z}_{\geq 2}$ , then  $\mathbb{R} \approx n^{\mathbb{N}}$ . So by Prop. 1.58,  $\mathbb{N} \lesssim \mathbb{R}$ , i.e.  $\mathbb{R}$  is uncountable. The base- $n$  presentations of real numbers suggest that  $\text{card}(n^{\mathbb{N}})$  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  $n \in \mathbb{Z}_{\geq 2}$  we have  $2^X \lesssim n^X \lesssim \mathbb{N}^X$ . 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  $X$ .  $\square$

## 2 Metric spaces, convergence of sequences, and continuous functions

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

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, because they share a common theme: one regards the set of functions as a space  $X$ , and views each function as a point of that space. This viewpoint allows one to study the analytic properties of function spaces by using the intuitions from  $\mathbb{R}$  and  $\mathbb{R}^n$ . (Sequential) compactness, completeness, and separability are prominent examples of such properties. Their importance was already recognized by Fréchet by the time he defined metric spaces.

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  $f : [0, 1] \rightarrow \mathbb{R}$  minimizing or maximizing the expression

$$S(f) = \int_0^1 L(t, f(t), f'(t)) dt \quad (2.1)$$

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

of functions)

$$S(f) = \int_0^1 \int_0^1 f(x)K(x, y)\overline{f(y)}dxdy \quad (2.2)$$

where  $K : [0, 1]^2 \rightarrow \mathbb{R}$  is continuous and  $f : [0, 1] \rightarrow \mathbb{C}$  is subject to the condition  $\int_0^1 |f(x)|^2 dx = 1$ . 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 and points of a continuous function. 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  $S : X \rightarrow \mathbb{R}$  is continuous (i.e. if  $f_n$  converges to  $f$  in the way we define, then  $S(f_n) \rightarrow S(f)$ ).
- (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  $S : X \rightarrow \mathbb{R}$  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 subsequence, i.e., if  $X$  is sequentially compact. Thus, we need a good criterion for sequential compactness of 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 continuity of functions. 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  $d : X \times X \rightarrow \mathbb{R}_{\geq 0}$  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)  $d(x, z) \leq d(x, y) + d(y, z)$ .

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

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

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  $[0, 1] \rightarrow \mathbb{R}$  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”:

$$d(x_0, x_n) \leq \sum_{j=1}^n d(x_{j-1}, x_j) \tag{2.3}$$

**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  $B(x, r) \cap B(y, \rho) = \emptyset$ . This is called the **Hausdorff property**.

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

We will see (cf. Prop. 2.18) 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  $d(x, y) = |x - y|$

**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:

$$\begin{aligned} d_1(x, y) &= |x_1 - y_1| + \cdots + |x_n - y_n| \\ d_\infty(x, y) &= \max\{|x_1 - y_1|, \dots, |x_n - y_n|\} \end{aligned}$$

**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:

$$\begin{aligned} d_1(z, w) &= |z_1 - w_1| + \cdots + |z_n - w_n| \\ d_\infty(z, w) &= \max\{|z_1 - w_1|, \dots, |z_n - w_n|\} \end{aligned}$$

**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  $1 \leq p < +\infty$ , and  $d_\infty = \lim_{p \rightarrow +\infty} d_p$ . 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  $x = (x_1, \dots, x_N) \in X$  and  $y = (y_1, \dots, y_N) \in Y$ . Then the following are metrics on  $X$ :

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

With respect to the metric  $\delta$ , 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.  $d, \delta, \rho$  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

$$d_1(x, y) \leq \alpha d_2(x, y) \quad d_2(x, y) \leq \beta d_1(x, y)$$

This is an equivalence relation. More generally, we may write  $d_1 \lesssim d_2$  if  $d_1 \leq \alpha d_2$  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  $\delta \leq \rho \leq d \leq 2\delta$ . So  $\delta \approx \rho \approx d$ .

**Convention 2.13.** Given finitely many metric spaces  $X_1, \dots, X_N$ , the metric on their product space  $X = X_1 \times \dots \times X_N$  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  $\mathbb{C}$ , this convention is compatible with Convention 2.8, and we 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** is denoted 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  $\lim_{n \rightarrow \infty} x_n = x$  (or  $x_n \rightarrow x$ ), if:

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

Equivalently, this means that every (nonempty) open ball centered at  $x$  contains **all but finitely many**  $x_n$ .<sup>3</sup>

**Exercise 2.17.** Show that in the above definition of limits, it suffices to consider rational numbers  $\varepsilon > 0$ . (Note: You need to use Prop. 1.29.)

---

<sup>3</sup>“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...”.

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

$$x_n \text{ satisfies property } P \text{ for sufficiently large } n \quad (2.4)$$

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

**Proposition 2.18.** *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  $B(x, r) \cap B(y, \rho) = \emptyset$ . By the definition of  $x_n \rightarrow x$ , there exists  $N_1 \in \mathbb{Z}_+$  such that  $x_n \in B(x, r)$  for all  $n \geq N_1$ . Similarly,  $x_n \rightarrow y$  means that there is  $N_2 \in \mathbb{Z}_+$  such that  $x_n \in B(y, \rho)$  for all  $n \geq N_2$ . Choose any  $n \geq N_1, N_2$  (e.g.  $n = \max\{N_1, N_2\}$ ). Then  $x_n \in B(x, r) \cap B(y, \rho) = \emptyset$ , impossible.  $\square$

### 2.3.1 Methods for proving convergence and computing limits

**Example 2.19.**  $\lim_{n \rightarrow \infty} \frac{1}{n} = 0$ .

*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  $n \geq N$  we have  $1/n < \varepsilon$ .  $\square$

**Proposition 2.20.** *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  $x_n \leq y_n$  for all  $n$ , then  $x \leq y$ .*

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

**Proposition 2.21.** *Let  $(x_n)_{n \in \mathbb{Z}}$  be a sequence in  $[a, b] \subset \mathbb{R}$ . If  $(x_n)$  is increasing (resp. decreasing), then  $\lim_{n \rightarrow \infty} x_n$  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  $A - \varepsilon < x_n \leq A$  and so  $|x_n - A| < \varepsilon$ .  $\square$

**Example 2.22.** 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

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



**Example 2.23.** Suppose that  $(a_n)$  and  $(b_n)$  are sequences in  $\mathbb{R}_{\geq 0}$ , that  $a_n \leq b_n$  for all  $n$ , and that  $b_n \rightarrow 0$ . Then  $a_n \rightarrow 0$ .

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

More generally, we have:

**Proposition 2.24 (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  $\mathbb{R}_{\geq 0}$  such that  $\lim_{n \rightarrow \infty} a_n = 0$  and that  $d(x_n, x) \leq a_n$  for all  $n$ . Then  $\lim_{n \rightarrow \infty} x_n = x$ .

*Proof.* This follows immediately from Exp. 2.22 and 2.23.  $\square$

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  $(a_n) \in \mathbb{R}_{\geq 0}$  converging to 0 such that the inequality  $d(x_n, x) \leq a_n$  holds. And very often, a good (hard) analyst is one who knows how to find such good sequences!

**Proposition 2.25.** 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  $\mathbf{x}_n = (x_{1,n}, \dots, x_{N,n})$  be a sequence in  $X$ . Let  $\mathbf{y} = (y_1, \dots, y_N)$ . Then

$$\lim_{n \rightarrow \infty} \mathbf{x}_n = \mathbf{y} \iff \lim_{n \rightarrow \infty} x_{i,n} = y_i \quad (\forall 1 \leq i \leq N)$$

*Proof.* We let  $d$  be the metric  $\delta$  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

$$\mathbf{x}_n \rightarrow \mathbf{y} \iff d_X(\mathbf{x}_n, \mathbf{y}) \rightarrow 0 \iff \max_{1 \leq j \leq N} d_j(x_{j,n}, y_j) \rightarrow 0 \quad (2.5)$$

Suppose that the RHS of (2.5) is true. Fix any  $1 \leq i \leq N$ . Then for every  $\varepsilon > 0$ , there is  $K \in \mathbb{Z}_+$  such that for all  $n \geq K$  we have  $\max_j d_j(x_{j,n}, y_j) < \varepsilon$ , and hence  $d_i(x_{i,n}, y_i) < \varepsilon$ . This proves  $x_{i,n} \rightarrow y_i$  (for all  $i$ ).

Conversely, assume that for every  $i$  we have  $x_{i,n} \rightarrow y_i$ . Then for every  $\varepsilon > 0$  there is  $K_i \in \mathbb{Z}_+$  such that  $d_i(x_{i,n}, y_i) < \varepsilon$  for every  $n \geq K_i$ . Let  $K = \max\{K_1, \dots, K_N\}$ . Then for all  $n \geq K$  we have  $\max_j d_j(x_{j,n}, y_j) < \varepsilon$ . This proves the RHS of (2.5).

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

*Proof.* Prove it yourself. (Or see Prop. 2.61.)  $\square$

### 2.3.2 Criteria for divergence

**Definition 2.27.** A subset  $E$  of a metric space  $X$  is called **bounded** if there exists  $p \in X$  and  $R \in \mathbb{R}_{>0}$  such that  $E \subset B_X(p, R)$ .

**Remark 2.28.** Note that if  $E$  is bounded, then for *any*  $q \in X$  there exists  $\tilde{R} \in \mathbb{R}_{>0}$  such that  $E \subset B_X(q, \tilde{R})$ . (Indeed, choose  $\tilde{R} = R + d(p, q)$ , then by triangle inequality,  $B(p, R) \subset B(q, \tilde{R})$ .)

From this Remark, we immediately see:

**Example 2.29.** 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 2.30.** A finite union of bounded subsets is bounded.

**Proposition 2.31.** 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  $x_n \rightarrow x$ . 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  $A = \{x_1\} \cup \dots \cup \{x_N\} \cup B(x, 1)$ . Since each  $\{x_i\}$  is bounded, and since  $B(x, 1)$  is bounded,  $A$  is bounded by Exp. 2.30.  $\square$

**Remark 2.32.** Prop. 2.31 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. 2.36.) 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. 2.35)

**Definition 2.33.** 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  $(x_n)_{n \in \mathbb{Z}_+}$ .

**Definition 2.34.** If  $A$  and  $B$  are partially ordered set, we say a function  $f : A \rightarrow B$  is **increasing** (resp. **strictly increasing**), if for each  $x, y \in A$  we have

$$\begin{aligned} x \leq y &\implies f(x) \leq f(y) \\ &\text{resp.} \\ x < y &\implies f(x) < f(y) \end{aligned}$$

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).

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

**Proposition 2.35.** *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 2.36.** 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. 2.32 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. 3.16.)

## 2.4 Continuous maps of metric spaces

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

**Definition 2.37.** Let  $f : X \rightarrow Y$  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

$$\lim_{n \rightarrow \infty} x_n = x \quad \implies \quad \lim_{n \rightarrow \infty} f(x_n) = f(x)$$

(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) $\Leftrightarrow$ (2'): Obvious.

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

$\neg$ (2) $\Rightarrow \neg$ (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  $d(f(p), f(x)) \geq \varepsilon$ . 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  $d(f(x_n), f(x)) \geq \varepsilon$ . By Squeeze theorem (Prop. 2.24),  $x_n \rightarrow x$ . But  $f(x_n) \not\rightarrow f(x)$ . So (1) is not true.  $\square$

**Remark 2.38.** One can compare Def. 2.37-(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 preserves 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. 2.62.

**Remark 2.39.** In this section, we mainly use Def. 2.37-(1) to study continuity. But in later sections we will also use Def. 2.37-(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  $\overline{B}_{\mathbb{C}}(0, 1) = \{w \in \mathbb{C} : |w| \leq 1\}$  iff  $|z| < 1$ .

Thus, Def. 2.37-(2') says that for any map of metric spaces  $f : X \rightarrow Y$  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  $f^{-1}(B_Y(f(x), \varepsilon))$ .

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  $f : X \rightarrow Y$  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. 2.37. Another prominent example is sequential compactness vs. compactness. These two approaches represent two (very) different intuitions. (So it is surprising that these two things, which seem so different, 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.  $\square$

## 2.4.1 Methods for proving continuity

**Lemma 2.40.** Let  $f : X \rightarrow Y$  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  $f|_{B_i} : B_i \rightarrow Y$  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  $f(x_n) \rightarrow f(x)$ . 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  $0 < \delta < r - d(x, y)$ . Then triangle inequality implies  $B(x, \delta) \subset B(y, r)$ .

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

**Definition 2.41.** A map of metric spaces  $f : X \rightarrow Y$  is called **Lipschitz continuous** if there exists  $L \in \mathbb{R}_{>0}$  (called **Lipschitz constant**) such that for all  $x_1, x_2 \in X$ ,

$$d_Y(f(x_1), f(x_2)) \leq L \cdot d_X(x_1, x_2) \quad (2.6)$$

**Lemma 2.42.** *Lipschitz continuous maps are continuous.*

*Proof.* Suppose that  $f : X \rightarrow Y$  is Lipschitz continuous with Lipschitz constant  $L$ . Suppose  $x_n \rightarrow x$  in  $X$ . Then  $L \cdot d(x_n, x) \rightarrow 0$ . By (2.6) and Squeeze theorem (Prop. 2.24),  $f(x_n) \rightarrow f(x)$ . (You can also use Def. 2.37-(2) to prove this lemma.)  $\square$

**Example 2.43.** Let  $\mathbb{F} \in \{\mathbb{Q}, \mathbb{R}, \mathbb{C}\}$ . The map  $z \in \mathbb{F} \setminus \{0\} \mapsto z^{-1} \in \mathbb{F}$  is continuous.

*Proof.* Let this map be  $f$ . Since  $\mathbb{F}$  is covered by open balls  $B(z, \delta)$  where  $z \in \mathbb{F} \setminus \{0\}$  and  $0 < \delta < |z|$ , by Lem. 2.40, 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  $|x| = |x - z + z| \geq |z| - |x - z| > \varepsilon$  by triangle inequality. Similarly,  $|y| > \varepsilon$ . So

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

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

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

We have given a fancy way of proving that if  $(z_n)$  is a sequence in  $\mathbb{F} \setminus \{0\}$  converging to  $z \in \mathbb{F} \setminus \{0\}$ , then  $z_n^{-1} \rightarrow z^{-1}$ . 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 2.44.** *Let  $\mathbb{F} \in \{\mathbb{Q}, \mathbb{R}, \mathbb{C}\}$ . Then the following maps are continuous:*

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

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  $\mu$ . 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. 2.40 and 2.42, 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

$$\begin{aligned} |\mu(x, y) - \mu(x', y')| &= |xy - x'y'| \leq |(x - x')y| + |x'(y - y')| \\ &< 2r \cdot \max\{|x - x'|, |y - y'|\} = 2r \cdot d((x, y), (x', y')) \end{aligned} \quad (2.7)$$

By Exp. 2.43 and Prop. 2.46, 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  $\mu$  is continuous, thanks to Prop. 2.45. So  $\div$  is continuous.  $\square$

**Proposition 2.45.** Suppose that  $f : X \rightarrow Y$  and  $g : Y \rightarrow Z$  are continuous maps of metric spaces. Then  $g \circ f : X \rightarrow Z$  is continuous.

**Proposition 2.46.** Suppose that  $f_i : X_i \rightarrow Y_i$  is a map of metric spaces, where  $1 \leq i \leq N$ . Then the product map

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

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

*Proof of Prop. 2.45 and 2.46.* Immediate from Def. 2.37-(1) and Prop. 2.25.  $\square$

**Corollary 2.47 (Squeeze theorem).** Let  $\mathbb{F} \in \{\mathbb{Q}, \mathbb{R}\}$  and  $(x_n), (y_n), (z_n)$  be sequences in  $\mathbb{R}$ . Assume that  $x_n \leq y_n \leq z_n$  for all  $n$ . Assume that  $x_n$  and  $z_n$  both converge to  $A \in \mathbb{R}$ . Then  $\lim_{n \rightarrow \infty} y_n = A$ .

*Proof.* Let  $a_n = y_n - x_n$  and  $b_n = z_n - x_n$ . Then  $0 \leq a_n \leq b_n$ , and  $\lim_n b_n = \lim_n z_n - \lim_n x_n = 0$  because the subtraction map is continuous (Prop. 2.44). By Exp. 2.23,  $a_n \rightarrow 0$ . Since  $x_n \rightarrow A$ ,  $y_n = x_n + a_n$  converges to  $A$ , since the addition map is continuous by Prop. 2.44.  $\square$

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 2.48.** By Prop. 2.44,  $f(x, y, z) = x^2y + 5y^4z^7 - 3xyz^2$  is a continuous function on  $\mathbb{C}^3$ . Clearly  $z \in \mathbb{C} \mapsto \bar{z} \in \mathbb{C}$  is continuous. So  $g(x, y, z) = f(\bar{x}, \bar{y}, z) + 2f(z, x^2, xy^{-9}) - 5xy^{-2}\bar{z}^{-3}$  is a continuous function on  $\mathbb{C} \times \mathbb{C}^\times \times \mathbb{C}^\times$ .

**Example 2.49.** Let  $f, g : X \rightarrow \mathbb{F}$  be continuous functions where  $\mathbb{F} \in \{\mathbb{Q}, \mathbb{R}, \mathbb{C}\}$ . Then by Prop. 2.44 and 2.45,  $f \pm g$  and  $fg$  are continuous, and  $f/g$  is continuous when  $0 \notin g(X)$ . Here

$$(f \pm g)(x) = f(x) \pm g(x) \quad (fg)(x) = f(x)g(x) \quad (f/g)(x) = f(x)/g(x) \quad (2.8)$$

**Example 2.50.** Let  $f : X \rightarrow Y$  be a map of metric spaces. Let  $E, F$  be subsets of  $X, Y$  respectively. (Recall that the metrics subspaces are chosen as in Def. 2.14.)

- The inclusion map  $\iota_E : E \rightarrow X, x \mapsto x$  is clearly continuous. Thus, if  $f$  is continuous, then  $f|_E : E \rightarrow Y$  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  $\tilde{f} : X \rightarrow F$  be  $\tilde{f}(x) = f(x)$ . Then it is clear that  $f$  is continuous iff  $\tilde{f}$  is continuous.

**Proposition 2.51.** Let  $X_1, \dots, X_N$  be metric spaces. The following projection is clearly continuous:

$$\pi_{X_i} : X_1 \times \dots \times X_N \rightarrow X_i \quad (x_1, \dots, x_N) \mapsto x_i$$

**Proposition 2.52.** Let  $f_i : X \rightarrow Y_i$  be maps where  $X, Y_1, \dots, Y_N$  are continuous. Then

$$f_1 \vee \dots \vee f_N : X \rightarrow Y_1 \times \dots \times Y_N \quad x \mapsto (f_1(x), \dots, f_N(x))$$

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

**Example 2.53.** Let  $X$  be a metric space. Then  $d : X \times X \rightarrow \mathbb{R}, (x, y) \mapsto d(x, y)$  is Lipschitz continuous with Lipschitz constant 1 (by triangle inequality). So  $d$  is continuous.

**Proposition 2.54.** Let  $X$  be a metric space and  $E \subset X$ . Define *distance function*

$$d(\cdot, E) : X \rightarrow \mathbb{R}_{\geq 0} \quad d(x, E) = \inf_{e \in E} d(x, e)$$

Then  $d(\cdot, E)$  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  $d(x, e) \leq d(x, y) + d(y, e)$ . Since  $d(x, E) \leq d(x, e)$ , we get  $d(x, E) \leq d(x, y) + d(y, e)$ . Applying  $\inf_{e \in E}$  to the RHS gives  $d(x, E) \leq d(x, y) + d(y, E)$ . Hence  $d(x, E) - d(y, E) \leq d(x, y)$ . Exchanging  $x$  and  $y$  gives

$$|d(x, E) - d(y, E)| \leq d(x, y) \quad (2.9)$$

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

**Definition 2.55.** More generally, if  $E, F$  are subsets of a metric space  $E$ , we can define

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

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

**Exercise 2.56.** Let  $E, F \subset X$ . Prove that

$$d(E, F) = \inf_{f \in F} d(E, f) \quad (2.11)$$

**Example 2.57.** If  $X$  is a metric space and  $p \in X$ , then by Prop. 2.54 (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  $f : X \rightarrow \mathbb{F}$  is continuous, then  $|f| : X \rightarrow \mathbb{R}_{\geq 0}$  is continuous where

$$|f|(x) = |f(x)| \quad (2.12)$$

**Example 2.58.** Let  $N \in \mathbb{Z}_+$ . Then the following function is continuous:

$$\max : \mathbb{R}^N \rightarrow \mathbb{R} \quad (x_1, \dots, x_N) \mapsto \max\{x_1, \dots, x_N\} \in \mathbb{R}$$

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

$$\max(x_1, x_2) = \frac{x_1 + x_2 + |x_1 - x_2|}{2} \quad (2.13)$$

So  $\max_2$  is continuous by Exp. 2.49 and 2.57.

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

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

### 2.5.1 General theory

**Definition 2.59.** A bijection of metric spaces  $f : X \rightarrow Y$  is called a **homeomorphism** if one of the following equivalent (cf. Def. 2.37) statements holds:

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



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

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

A special case of the above definition is:

**Definition 2.60.** Let  $X$  be a set with metrics  $d, \delta$ . We say that  $d$  and  $\delta$  induce the **same topology** on  $X$  (or that  $d, \delta$  are **topologically equivalent**) if one of the following clearly equivalent statements holds:

- (1) The map  $(X, d) \rightarrow (X, \delta), x \mapsto x$  is a homeomorphism.<sup>4</sup>
- (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  $\delta$ .

**Proposition 2.61.** Suppose that  $d, \delta$  are equivalent metrics on a set  $X$ . Then  $d, \delta$  are topologically equivalent.

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

**Exercise 2.62.** Let  $f : X \rightarrow Y$  be a map of metric spaces. We say that  $f : X \rightarrow Y$  is an **isometry** (or is **isometric**) if for all  $x_1, x_2 \in X$  we have

$$d_Y(f(x_1), f(x_2)) = d_X(x_1, x_2) \quad (2.14)$$

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.  $\square$

**Remark 2.63.** Isometric isomorphisms are important examples of homeomorphisms. That  $f : X \rightarrow Y$  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  $f : X \rightarrow Y$  is a bijection of sets. Suppose that  $Y$  is a metric space. Then there is a unique metric  $d_X$  on  $X$  such that  $f$  is an isometric isomorphism: one defines  $d_X$  using (2.14). 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  $f$ .  $\square$

---

<sup>4</sup>We prefer not to call this map the identity map, because the metrics on the source and on the target are different.

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

**Exercise 2.64.** Two metrics inducing the same topology are not necessarily equivalent metrics. For example, let  $f : [0, 1] \rightarrow [0, 1]$  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  $f : (X, d_X) \rightarrow (X, d_X)$  is a homeomorphism. So  $d_X$  and  $f^*d_X$  give the same topology on  $[0, 1]$  (cf. Exe. 2.65). Show that  $f^*d_X$  and  $d_X$  are not equivalent metrics.

**Exercise 2.65.** Let  $f : X \rightarrow Y$  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  $f : (X, d_X) \rightarrow (Y, d_Y)$  is a homeomorphism.

In particular, if  $f : X \rightarrow X$  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  $f : (X, d_X) \rightarrow (X, d_X)$  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 2.66.** We say that a sequence  $(x_n)$  in  $\overline{\mathbb{R}}$  **converges to**  $+\infty$  (resp.  $-\infty$ ), if for every  $A \in \mathbb{R}_{>0}$  there is  $N \in \mathbb{Z}_+$  such that for all  $n \geq N$  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  $n \geq N$  (i.e.  $x_n \in \mathbb{R}$  for sufficiently large  $n$ ), and that the subsequence  $(x_{k+N})_{k \in \mathbb{Z}_+}$  converges in  $\mathbb{R}$  to  $x$ .  $\square$

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. 2.66. We shall now find such a metric.

**Lemma 2.67.** Let  $-\infty \leq a < b \leq +\infty$  and  $-\infty \leq c < d \leq +\infty$ . Then there is a strictly increasing bijective map  $[a, b] \rightarrow [c, d]$ .

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

*Proof.* We have a strictly increasing bijection  $f : \mathbb{R} \rightarrow (-1, 1)$  defined by (1.9).  $f$  can be extended to a strictly increasing bijective map  $\overline{\mathbb{R}} \rightarrow [-1, 1]$  if we set  $f(\pm\infty) = \pm 1$ . Thus,  $f$  restricts to a strictly increasing bijection  $[a, b] \rightarrow [c', d']$  where  $c' = f(a)$ ,  $d' = f(b)$ . Choose a linear function  $g(x) = \alpha x + \beta$  (where  $\alpha > 0$ ) giving an increasing bijection  $[c', d'] \rightarrow [c, d]$ . Then  $g \circ f$  is the desired map.  $\square$

**Theorem 2.68.** Let  $\varphi : \overline{\mathbb{R}} \rightarrow [a, b]$  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. 2.66 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  $x_n \rightarrow x$  (in the sense of Def. 2.66) iff  $y_n \rightarrow y$  (under the Euclidean metric). Write  $\psi = \varphi^{-1}$ , which is a strictly increasing map  $[a, b] \rightarrow \overline{\mathbb{R}}$ . 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  $x_n \rightarrow x$ , 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  $y_n \rightarrow y$ . That  $y_n \rightarrow y$  implies  $x_n \rightarrow x$  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  $x_n \rightarrow +\infty$ , 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  $y_n \rightarrow b$ . Conversely, if  $y_n \rightarrow b$ , then for each  $A > 0$ , all but finitely many  $y_n$  are inside  $(\varphi(A), b]$  and hence  $> \varphi(A)$ . So all but finitely many  $x_n$  are  $> A$ .  $\square$

**Convention 2.69.** Unless otherwise stated, a metric on  $\overline{\mathbb{R}}$  is one that makes Def. 2.66 true, for instance  $\varphi^* d_{[a,b]}$  in Thm. 2.68. 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  $-\infty < a < b < +\infty$ .

**Remark 2.70.** By Thm. 2.68, 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  $x_n \leq y_n$  for all  $n$ , then  $A \leq B$ .
2. **Squeeze theorem:** Suppose that  $(x_n), (y_n), (z_n)$  are sequences in  $\overline{\mathbb{R}}$ ,  $x_n \leq y_n \leq z_n$  for all  $n$ , and  $x_n$  and  $z_n$  both converge to  $A \in \overline{\mathbb{R}}$ . Then  $y_n \rightarrow A$ .
3. Prop. 2.21 also holds for  $[a, b]$  in  $\overline{\mathbb{R}}$ : if  $(x_n)$  is an increasing resp. decreasing sequence in  $[a, b]$ , then  $\lim_n x_n$  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. 2.66. 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  $\varphi, \psi : \overline{\mathbb{R}} \rightarrow [0, 1]$  be a strictly increasing bijections where  $\psi \circ \varphi^{-1} : [0, 1] \rightarrow [0, 1]$  is  $x \mapsto x^2$ . Then by Exe. 2.64,  $\varphi^*d_{[0,1]}$  and  $\psi^*d_{[0,1]}$  are non-equivalent metrics giving the same topology 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. 2.68.

**Theorem 2.71.** *Let  $\varphi$  be a strictly increasing bijection in one of the following forms*

$$\begin{array}{ll} [a, b] \rightarrow [c, d] & (a, b) \rightarrow (c, d) \\ (a, b) \rightarrow (c, d] & [a, b) \rightarrow [c, d) \end{array}$$

where  $-\infty \leq a \leq b \leq +\infty$  and  $-\infty \leq c \leq d \leq +\infty$ . Then  $\varphi$  is a homeomorphism, i.e., if  $(x_n)$  and  $x$  are in the domain, then  $x_n \rightarrow x$  iff  $\varphi(x_n) \rightarrow \varphi(x)$  (in the sense of Def. 2.66).

*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  $(-\infty, +\infty] \rightarrow (c, d]$ , then it can also be proved  $(-\infty, +\infty] \rightarrow (a, b]$ . By composing the inverse of the second map with the first map, we see that the theorem holds for  $[a, b] \rightarrow [c, d]$ .

Let us consider  $\varphi : (-\infty, +\infty] \rightarrow (c, d]$ .  $\varphi$  can be extended to a strictly increasing bijection  $\varphi : \overline{\mathbb{R}} \rightarrow [c, d]$  by letting  $\varphi(-\infty) = c$ . It suffices to prove that this  $\varphi$  is a homeomorphism. When  $-\infty < c < d < +\infty$ , then the theorem holds by Thm. 2.68. If one of  $c, d$  is  $\pm\infty$ , the same argument as in the proof of Thm. 2.68 proves that  $\varphi$  is a homeomorphism. We leave it to the readers to fill in the details.  $\square$

## 2.6 Problems and supplementary material

**Definition 2.72.** 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  $f : A \rightarrow \mathbb{R}$  is **subadditive** if for every  $x, y \in A$  we have  $f(x + y) \leq f(x) + f(y)$ .

**Problem 2.1.** Consider the following increasing functions:

$$\begin{array}{ll} f_1 : \mathbb{R}_{\geq 0} \rightarrow [0, 1] & f_1(x) = \frac{x}{1+x} \\ f_2 : \mathbb{R}_{\geq 0} \rightarrow [0, 1] & f_2(x) = \min\{x, 1\} \end{array}$$

Prove that they are subadditive functions.

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

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

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

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

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

Prove that  $\delta$  is a metric, and that  $\delta$  and  $d$  are topologically equivalent.

**Proposition 2.73.** *Let  $(X, d)$  be a metric space. Then there is a **bounded metric**  $\delta$  on  $X$ , where “bounded” means  $\sup d(X \times X) < +\infty$ , such that  $d$  and  $\delta$  induce the same topology.*

*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:

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

□

**Problem 2.3.** Let  $(X_i, d_i)_{i \in \mathbb{Z}_+}$  be an sequence of metric spaces. Assume that  $d_i \leq 1$  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

$$d(f, g) = \sup_{i \in \mathbb{Z}_+} \frac{d_i(f(i), g(i))}{i} \quad (2.15)$$

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)  $\lim_{n \rightarrow \infty} f_n = g$  under the metric  $d$ .
- (b)  $f_n$  **converges pointwisely** to  $g$ , namely,  $\lim_{n \rightarrow \infty} f_n(i) = g(i)$  for every  $i \in \mathbb{Z}_+$ .

**Remark 2.74.** 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

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

also defines a metric, and that (a) (with  $d$  replaced by  $\delta$ ) and (b) equivalent. So  $\delta$  and  $d$  (defined by (2.15)) 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 a later problem.

★ **Problem 2.4.** Let  $X = \bigsqcup_{i \in I} A_i$  be a disjoint union of metric spaces  $(A_i, d_i)$ . Assume that  $d_i \leq 1$  for all  $i$ . For each  $x, y \in X$ , define

$$d(x, y) = \begin{cases} d_i(x, y) & \text{if } x, y \in A_i \text{ for some } i \in I \\ \frac{1}{2} & \text{otherwise} \end{cases}$$

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  $\lim_{n \rightarrow \infty} x_n = x$  mean in terms of the convergence in each  $A_i$ ?

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  $n \rightarrow \infty$ . 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  $f : X \rightarrow \mathbb{R}$  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 2.69.)

*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  $A - 1/n < f(x_n) \leq A$  (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  $f(x_n) \rightarrow A$  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  $f : X \rightarrow \overline{\mathbb{R}}$ , which is continuous (cf. Exp. 2.50). Since  $f(x_n) \rightarrow A$ , its subsequence  $f(x_{n_k})$  also converges to  $A$ . But since  $x_{n_k} \rightarrow x$  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  $A \in \mathbb{R}$ .  $\square$

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  $X_1, \dots, X_N$  be sequentially compact metric spaces. Then  $X = X_1 \times \dots \times X_N$  is sequentially compact.

*Proof.* Since  $X = (X_1 \times \dots \times X_{N-1}) \times X_N$ , 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,  $(b_n)$  has a convergent subsequence  $(b_{n_{k_l}})$ . So  $(a_{n_{k_l}}, b_{n_{k_l}})$  is a convergent subsequence of  $(a_n, b_n)$ .  $\square$

**Proposition 3.5.** Let  $f : X \rightarrow Y$  be a continuous surjective 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  $f(x)$ .  $\square$

**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  $d$ , equivalently,  $\sup d(X \times X) < +\infty$ .

*Proof.* Fix  $p \in X$ . The function  $d_p : x \in X \mapsto d(x, p) \in \mathbb{R}_{\geq 0}$  is continuous by Exp. 2.57. So, by Lem. 3.2,  $d_p$  is bounded by some  $0 < R < +\infty$ . So  $X = \overline{B}_X(p, R)$ .  $\square$

### 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  $x$ .

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

$$\alpha_n = \inf\{x_k : k \geq n\} \quad \beta_n = \sup\{x_k : k \geq n\} \quad (3.1)$$

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

$$\liminf_{n \rightarrow \infty} x_n = \sup\{\alpha_n : n \in \mathbb{Z}_+\} = \lim_{n \rightarrow \infty} \alpha_n \quad (3.2a)$$

$$\limsup_{n \rightarrow \infty} x_n = \inf\{\beta_n : n \in \mathbb{Z}_+\} = \lim_{n \rightarrow \infty} \beta_n \quad (3.2b)$$

(cf. Rem. 2.70), called respectively the **limit inferior** and the **limit superior** of  $(x_n)$ .



**Theorem 3.10.** Let  $(x_n)$  be a sequence in  $\overline{\mathbb{R}}$ , and let  $S$  be the set of cluster points of  $(x_n)$ . Then  $\liminf_{n \rightarrow \infty} x_n$  and  $\limsup_{n \rightarrow \infty} x_n$  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  $\alpha_{n_k} \leq x_{n_k} \leq \beta_{n_k}$ , we have  $A \leq x \leq B$  by Rem. 2.70. 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  $n_1 < \dots < n_k$  has been constructed. By the definition of  $\beta_{1+n_k}$ , there is  $n_{k+1} \geq 1 + n_k$  such that  $x_{n_{k+1}}$  is close to  $\beta_{1+n_k}$ , say

$$\beta_{1+n_k} - \frac{1}{k} < x_{n_{k+1}} \leq \beta_{1+n_k} \quad (3.3)$$

Since the left most and the right most of (3.3) both converge to  $B$  as  $k \rightarrow \infty$ , by Squeeze theorem (Cor. 2.47) we conclude  $\lim_k x_{n_k} = B$ .

In general, by Lem. 2.67 and Thm. 2.71, there is an increasing (i.e. order-preserving) homeomorphism (i.e. topology-preserving map)  $\varphi : \overline{\mathbb{R}} \rightarrow [0, 1]$ . Then  $\varphi(\beta_n) = \sup\{\varphi(x_k) : k \geq n\}$  (cf. Exe. 3.12) 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  $B$ .  $\square$

**Remark 3.11.** 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)  $\varphi : X \rightarrow Y$  where  $Y$  is in the convenient case. Then the result on  $Y$  can be translated via  $\varphi^{-1}$  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.  $\square$

**Exercise 3.12.** Let  $X, Y$  be partially ordered sets. Let  $\varphi : X \rightarrow Y$  be an increasing bijection whose inverse is also increasing. (Namely,  $\varphi$  induces an equivalence of partially ordered sets). 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 3.13** (Bolzano-Weierstrass). *Let  $[a_1, b_1], \dots, [a_N, b_N]$  be closed intervals in  $\overline{\mathbb{R}}$ . Then  $[a_1, b_1] \times \dots \times [a_N, b_N]$  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. 3.10,  $(x_n)$  has a subsequence  $(x_{n_k})$  converging to some  $x \in \overline{\mathbb{R}}$ . (E.g.  $x = \limsup_n x_n$ .) Since  $a \leq x_{n_k} \leq b$ , we have  $a \leq x \leq b$  by Rem. 2.70.  $\square$

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  $\mathbb{R}$ .

### 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 3.14.** *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. 2.35.

(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  $\lim_{n \rightarrow \infty} x_n = x$ . Suppose not. Then there exists  $\varepsilon > 0$  such that for every  $N \in \mathbb{Z}_+$  there is  $n \geq N$  such that  $d(x_n, x) \geq \varepsilon$ . Thus, one can inductively construct a subsequence  $(x_{n_k})$  of  $(x_n)$  such that  $d(x_{n_k}, x) \geq \varepsilon$  for all  $k$ . Since  $X$  is sequentially compact,  $(x_{n_k})$  has a subsequence  $x'_n$  converging to  $x' \in X$ . So  $d(x'_n, x) \geq \varepsilon$  for all  $n$ . Since the function  $y \in X \mapsto d(y, x)$  is continuous (Exp. 2.57), we have  $\lim_{n \rightarrow \infty} d(x'_n, x) = d(x', x)$ . This proves that  $d(x', x) \geq \varepsilon > 0$ . However,  $x', x$  are both cluster points of  $(x_n)$ , and so  $x = x'$ . This gives a contradiction.  $\square$

**Remark 3.15.** Thm. 3.14 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.

**Corollary 3.16.** *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. 2.31 and 2.35.

(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  $\mathbb{R}$ . By Bolzano-Weierstrass,  $X$  is sequentially compact. Thus, by Thm. 3.14,  $(x_n)$  has a subsequence converging in  $X$  and hence in  $\mathbb{R}^N$ .  $\square$

**Corollary 3.17.** *The following are true.*

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

*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. 3.10,  $A = \min S$ ,  $B = \max S$ . 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. 3.14 (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. 2.31. Conversely, if  $A = B$  and if  $(x_n)$  is bounded, say  $\alpha \leq x_n \leq \beta$  for all  $n$  where  $-\infty < \alpha < \beta < +\infty$ . Then  $\alpha \leq A \leq B \leq \beta$ . So  $A, B \in \mathbb{R}$ . By part 1,  $(x_n)$  converges to  $A \in \mathbb{R}$ .  $\square$

## 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 using the intuition from  $\mathbb{R}^N$ . 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. 3.50.) 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  $[0, 1]^{\mathbb{Z}_+}$ . 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

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

If  $\mathcal{X}$  satisfies a condition called “**equicontinuous**”, then a sequence  $f_n$  in  $\mathcal{X}$  converges *uniformly* to  $f \in \mathcal{X}$  iff  $f_n|_A$  converges *pointwisely* to  $f|_A$ . Thus, the sequential compactness of  $[-1, 1]^A$  under pointwise convergence topology implies the sequential compactness of  $\mathcal{X}$  under uniform convergence topology. This remarkable sequential compactness result on  $\mathcal{X}$  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  $f : [-\pi, \pi] \rightarrow \mathbb{C}$  satisfying  $f(-\pi) = f(\pi)$  has Fourier series expansion  $f(x) = \sum_{n \in \mathbb{Z}} a_n e^{inx}$  where  $a_n \in \mathbb{C}$ . However, for the sake of studying differential equations, one needs to consider series  $\sum_{n \in \mathbb{Z}} a_n e^{inx}$  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  $f(x) = \sum_{n \in \mathbb{Z}} a_n e^{inx}$  satisfying  $\sum_n |a_n|^2 \leq 1$ . Therefore, one lets  $\overline{B} = \{z \in \mathbb{C} : |z| \leq 1\}$  and considers  $\hat{f} : n \in \mathbb{Z} \mapsto a_n \in \mathbb{C}$  as an element of  $\overline{B}^{\mathbb{Z}}$ . The sequential compactness of  $\overline{B}^{\mathbb{Z}}$  helps one find the  $\hat{f}$  such that the corresponding  $f(x) = \sum_n \hat{f}(n) \cdot e^{inx}$  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 3.18.** 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  $m, n \geq N$  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 3.19.** 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  $n \geq N$  we have  $d(x_n, x_N) < \varepsilon$ .

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

**Proposition 3.20.** 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  $n \geq N$ . Since this is true for every  $m \geq N$ , we have  $d(x_m, x_n) \leq d(x, x_n) + d(x, x_m) < \varepsilon/2 + \varepsilon/2 = \varepsilon$ .  $\square$

**Definition 3.21.** A metric space  $X$  is called **complete** if every Cauchy sequence in  $X$  converges.

We have many examples of complete metric spaces:

**Theorem 3.22.** 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  $x_n \rightarrow x$ .

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  $m, n \geq \mathbb{Z}_+$ . Since  $x_{n_k} \rightarrow x$ , there is  $k \geq N$  such that  $d(x_{n_k}, x) < \varepsilon/2$ . Since  $n_k$  is strictly increasing over  $k$ , we have  $n_k \geq k$ . So  $n_k \geq N$ . So we can let  $m = n_k$ . This gives  $d(x_n, x_{n_k}) < \varepsilon/2$ . Therefore  $d(x_n, x) \leq d(x_n, x_{n_k}) + d(x_{n_k}, x) < \varepsilon$  for all  $n \geq N$ .  $\square$

**Example 3.23.** 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. 3.22.  $X$  is complete.

**Corollary 3.24.**  $\mathbb{R}^N$  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. 3.23,  $X$  is complete. So  $(x_n)$  converges to some  $x \in X$ .  $\square$

**Definition 3.25.** We say that 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 3.26.** Let  $A$  be a metric subspace of a complete metric space  $Y$ . Recall that the metric of  $A$  inherits from that of  $Y$  (cf. Def. 2.14). Then  $A$  is a closed subset of  $Y$  iff  $A$  is complete.

*Proof.* First, assume that  $A$  is closed. Let  $(x_n)$  be a Cauchy sequence in  $A$ . Then it is a Cauchy sequence in  $Y$ . So  $x_n \rightarrow x \in X$  because  $Y$  is complete. So  $x \in A$  by the definition of closedness.

Next, we assume that  $A$  is complete. Choose a sequence  $(x_n)$  in  $A$  converging to a point  $x \in X$ . By Prop. 3.20,  $(x_n)$  is a Cauchy sequence in  $X$ , and hence a Cauchy sequence in  $A$ . Since  $A$  is complete, there is  $a \in A$  such that  $x_n \rightarrow a$ . So we must have  $x = a$  because any sequence has at most one limit in a metric space. This proves  $x \in A$ . So  $A$  is closed.  $\square$

**Example 3.27.** Let  $-\infty < a < b < +\infty$ . Then  $(a, b)$  is not complete (under the Euclidean metric), because  $(a, b)$  is not closed in the complete metric space  $\mathbb{R}$ . (For sufficiently large  $n$ ,  $b - 1/n$  is in  $(a, b)$ , but  $\lim_{n \rightarrow \infty} (b - 1/n) = b$  is not in  $(a, b)$ .)

**Example 3.28.** By Prop. 1.29, for each  $x \in \mathbb{R} \setminus \mathbb{Q}$ , 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.

**Exercise 3.29.** Let  $d, \delta$  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  $\delta$ .

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

**Exercise 3.30.** Find a non-complete metric  $\delta$  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 functions 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 3.31.** Let  $V$  be a vector space over  $\mathbb{F}$  with zero vector  $0_V$ . A function  $\|\cdot\|: V \rightarrow \mathbb{R}_{\geq 0}$  is called a **norm** if for every  $u, v \in V$  and  $\lambda \in \mathbb{F}$ , the following hold:

- (Subadditivity)  $\|u + v\| \leq \|u\| + \|v\|$ .
- (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 3.32.** Let  $V$  be a vector space. If  $V$  is a normed vector space, then

$$d(u, v) = \|u - v\| \quad (3.4)$$

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 3.33.** 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).

**Example 3.34.** We always assume that the norm on  $\mathbb{F}^N$  is the **Euclidean norm**

$$\|(a_1, \dots, a_N)\| = \sqrt{|a_1|^2 + \dots + |a_N|^2} \quad (3.5)$$

The canonical metric it gives is the Euclidean metric. Thus, by Cor. 3.24,  $\mathbb{F}^N$  is a Banach space.

If  $(\lambda_n)$  is a sequence in  $\mathbb{F}$  converging to  $\lambda$ , 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  $\lambda x$  by checking that each component of  $\lambda_n x_n$  converges to the corresponding component of  $\lambda x$ . This is due to Prop. 2.25. However, if  $(x_n)$  is in general a sequence in a normed vector space, this method fails. So we need a different argument:

**Proposition 3.35.** Let  $V$  be a normed vector space. The following maps are continuous

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

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. 2.43.

*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 (2.7), 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. 2.40 and Lem. 2.42.  $\square$

### 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 uniformly 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 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 3.36.** Let  $X$  be a vector space 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  $\lambda \in \mathbb{F}$ :

$$\begin{aligned} f + g : X &\rightarrow V & (f + g)(x) &= f(x) + g(x) \\ \lambda f : X &\rightarrow V & (\lambda f)(x) &= \lambda f(x) \end{aligned}$$

We also define the **absolute value function**

$$|f| : X \rightarrow \mathbb{R}_{\geq 0} \quad x \in X \mapsto \|f(x)\| \quad (3.6)$$

**Definition 3.37.** Let  $X$  be a set and let  $V$  be a normed vector space. For each  $f \in V^X$ , define the  **$l^\infty$ -norm**

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

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

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

which is a vector subspace of  $V^X$ . Then  $l^\infty(X, V)$  is a normed vector space under the  $l^\infty$ -norm.



**Exercise 3.38.** Prove that for every  $f, g \in V^X$  and  $\lambda \in \mathbb{F}$  we have

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

(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  $l^\infty(X, V)$ .  $\square$

**Definition 3.39.** 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  $\lim_{n \rightarrow \infty} \|f - f_n\|_\infty = 0$ . In this case, we write  $f_n \rightrightarrows f$ .

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

In more detail, the uniform convergence of  $f_n$  to  $f$  means that “for every  $\varepsilon > 0$  there is  $N \in \mathbb{Z}_+$  such that for all  $n \geq N$  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, because if  $\|f - f_n\|_\infty \rightarrow 0$ , then clearly  $\|f_n(x) - f(x)\| \rightarrow 0$  for each  $x \in X$ .

**Example 3.40.** Let  $f_n : (0, 1) \rightarrow \mathbb{R}$  be  $f_n(x) = x^n$ . Then  $f_n$  converges pointwisely to 0. 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 3.41.** The uniform convergence of sequences in  $l^\infty(X, V)$  is induced by the  $l^\infty$ -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. 3.3. When  $X$  is countable, we have seen in Pb. 2.3 that the pointwise convergence in  $V^X$  is also given by a metric.  $\square$

**Theorem 3.42.** Let  $X$  be a set, and let  $V$  be an  $\mathbb{F}$ -Banach space. Then  $l^\infty(X, V)$  is an  $\mathbb{F}$ -Banach space.

*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  $m, n \geq N$  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  $n \geq N$  and all  $x$ ,

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

for every  $m \geq N$ . Let  $m \rightarrow \infty$ . Then by the continuity of subtraction and taking norm (cf. Prop. 3.35.), we obtain  $\|f_n(x) - f(x)\| \leq \varepsilon$  for all  $n \geq N$  and  $x \in X$ . In other words,  $\|f_n - f\|_\infty \leq \varepsilon$  for all  $n \geq N$ . In particular,  $\|f\|_\infty \leq \|f_N\|_\infty + \|f_N - f\|_\infty < +\infty$  by (3.9). This proves  $f \in l^\infty(X, V)$  and  $f_n \rightrightarrows f$ .  $\square$

Mathematicians used to believe that “if a sequence of continuous functions  $f_n : [0, 1] \rightarrow \mathbb{R}$  converges pointwisely to a function  $f : [0, 1] \rightarrow \mathbb{R}$ , 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  $f : \mathbb{R} \rightarrow \mathbb{R}$  be a function with period  $2\pi$  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 pointwisely 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. This was the first time the importance of uniform convergence was realized.

**Definition 3.43.** Let  $X, Y$  be metric spaces (or more generally, topological spaces to be defined later). Then  $C(X, Y)$  denotes the set of continuous functions from  $X$  to  $Y$ .

**Theorem 3.44.** Let  $X$  be a sequentially compact metric space. Let  $V$  be a Banach space. Then  $C(X, V)$  is a closed linear subspace of  $l^\infty(X, V)$ .

*Proof.* For any  $f \in C(X, V)$ , the absolute value function  $|f| : x \in X \mapsto \|f(x)\|$  is continuous. Thus by the sequential compactness of  $X$  and Lem. 3.2,  $|f|$  is bounded on  $X$ . This proves that  $\|f\|_\infty < +\infty$ . Thus  $C(X, V)$  is a subset of  $l^\infty(X, V)$ . Using Prop. 3.35, one checks easily that  $C(X, V)$  is a linear subspace of  $l^\infty(X, V)$ .

It remains to prove that  $C(X, V)$  is closed in  $l^\infty(X, V)$ . Choose a sequence  $(f_n)$  in  $C(X, V)$  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. 2.37-(2'). (One can also use Def. 2.37-(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  $n \geq N$  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

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

This finishes the proof.  $\square$

We emphasize that in the above proof, the sequential compactness of  $X$  is used only to show that  $C(X, V)$  is a subset of  $l^\infty(X, V)$ .

**Convention 3.45.** Unless otherwise stated, if  $X$  is sequentially compact metric space (or more generally, a compact Hausdorff space to be defined latter), and if  $V$  is a normed vector space, the norm on  $C(X, V)$  is chosen to be the  $l^\infty$ -norm.

**Corollary 3.46.** *Let  $X$  be a sequentially compact metric space. Let  $V$  be a Banach space. Then  $C(X, V)$  is a Banach space (under the  $l^\infty$ -norm).*

*Proof.* This follows immediately from Prop. 3.26, Thm. 3.44, and the fact that  $l^\infty(X, V)$  is complete (Thm. 3.42).  $\square$

## 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  $n \geq N$  such that  $d(x_n, x) < \varepsilon$ .

**Remark 3.47.** 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  $n \geq N$  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  $n \geq N$ ,  $x_n$  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  $\neg P$ ”.  $\square$

**Remark 3.48.** 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).

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

**Remark 3.49.** You will notice that your proof of Pb. 3.2 is slightly simpler than the proof we gave for Thm. 3.10. 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.

**Theorem 3.50 (Tychonoff theorem, countable version).** *Let  $(X_n)_{n \in \mathbb{Z}_+}$  be a sequence of metric spaces where each  $X_n$  is sequentially compact. Then the product space  $S = \prod_{n \in \mathbb{Z}_+} X_n$  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**.

*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  $x_{1,1}, x_{1,2}, x_{1,3} \dots$  whose value at  $n = 1$  converges in  $X_1$ . Since  $X_2$  is sequentially compact, we can choose a subsequence  $x_{2,1}, x_{2,2}, x_{2,3}, \dots$  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  $(x_{i,j})_{i,j \in \mathbb{Z}_+}$ :

$$\begin{array}{cccc} x_{1,1} & x_{1,2} & x_{1,3} & \cdots \\ x_{2,1} & x_{2,2} & x_{2,3} & \cdots \\ x_{3,1} & x_{3,2} & x_{3,3} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{array} \quad (3.10)$$

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  $n$ ,  $\lim_{j \rightarrow \infty} x_{n,j}(n)$  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  $n$ ,  $(x_{i,i})_{i \geq n}$  is a subsequence of the  $n$ -th line, whose value at  $n$  therefore converges in  $X_n$ . Thus  $\lim_{i \rightarrow \infty} x_{i,i}(n)$  converges in  $X_n$ . Thus, by Pb. 2.3,  $(x_{i,i})_{i \in \mathbb{Z}_+}$  converges under any metric inducing the product topology.  $\square$

**Problem 3.3.** Let  $V$  be a normed vector space. For every  $f, g \in V^X$  define

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

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

# Index

- Absolute function  $|f|$ , 48
- All but finitely many  $x_n$ , 23
- Archimedean property, 10
- Banach space, 47
- Bolzano-Weierstrass theorem, 42
- Bounded metric, 37
- Bounded sequence, 26
- Bounded subset, 26
- Cardinality  $\text{card}(A)$ , 13
- Cauchy sequence, 45
- Characteristic function, 16
- Closed subsets, 46
- Cluster point of a sequence, 40
- Commutative diagram, 8
- Complete metric space, 45
- Convergence in  $\mathbb{R}$ , 34
- Countable, 14
- Countably infinite, 14
- Diagonal method, 51
- Disjoint union, 15
- Distance function  $d(\cdot, E)$ , 31
- Endpoints of an interval, 12
- Equivalent metrics, 23
- Euclidean metric, 22
- Euclidean norm, 47
- Eventually, 51
- Field, 4
- Field extension, 10
- Frequently, 51
- Increasing and decreasing, 26
- Indexed family of sets, 5
- Infinimum, 11
- Interior point, 28
- Interval, 12
- Isometric metric spaces, 33
- Isometry and isometric isomorphism, 33
- Isomorphism, 12
- Limit inferior and superior, 40
- Linear maps, 8
- Lipschitz constant, 29
- Lipschitz continuous, 29
- Matrix representation, 9
- Metric space, 21
- Metric subspace, 23
- Monotonic, 26
- Norm, 47
- Normed vector space, 47
- Open set, 28
- Ordered field, 9
- Ordered field extension=ordered subfield, 10
- Pointwise convergence, 49
- Power set  $2^X$ , 5
- Product map, 30
- Product space, 16
- Pullback metrics, 33
- Quotient sets, 8
- Real number, 12
- Schröder-Bernstein theorem, 13
- Sequentially compact, 39
- Squeeze theorem, 25, 30, 35
- Strictly increasing and strictly decreasing, 26
- Strictly monotonic, 26
- Subadditivity, 36, 47
- Subfield, 10
- Sufficiently large, 24
- Supremum, 11
- Topologically equivalent metrics, 33
- Triangle inequality, 21
- Tychonoff theorem, countable version, 51
- Uniform convergence, 49
- Upper bound, 11
- Vector spaces, 8
- Well defined, 9
- $\bigsqcup_{\alpha \in \mathcal{A}} A_\alpha$ , the disjoint union, 15
- $A \setminus B$ , 3
- $A^c$ , the complement of  $A$ , 3

$B_X(x, r) = B(x, r)$  and  $\overline{B}_X(x, r) = \overline{B}(x, r)$ , 21

$\mathbb{C}$ , the set of complex numbers, 4

$C(X, Y)$ , the set of continuous functions  $X \rightarrow Y$ . See Conv. 3.42 also, 50

$d(x, E)$ , 31

$d(E, F)$ , 32

$\mathbb{F}^\times = \mathbb{F} \setminus \{0\}$ , 2

$|f|$ , 48

$f|_E$ , the restriction of  $f$  to  $E$ , 5

$f^*d_Y$ : pullback metric, 33

$f_n \rightrightarrows f$ , 49

$\mathbb{R}^{n \times m}$ , the set of  $n \times m$  matrices, 8

$\text{id}_A$ , 6

$\inf E$ , 11

$\|\cdot\|_{l^\infty} = \|\cdot\|_\infty$ , 48

$l^\infty(X, V)$ , 48

$\liminf, \limsup$ , 40

$\mathbb{N} = \{0, 1, 2, \dots\}$ , 2

$n^X = \{1, \dots, n\}^X$ , 17

$\mathbb{Q}$ , the field of rational numbers, 9

$\overline{\mathbb{R}} = [-\infty, +\infty] = \mathbb{R} \cup \{-\infty, +\infty\}$ , 12

$\sup E$ , 11

$V^X$  as a vector space, 48

$\prod_{i \in I} X_i$ , 16

$Y^X$ , the set of functions  $X \rightarrow Y$ , 15

$\mathbb{Z}_+ = \{1, 2, \dots\}$ , 2

$\chi_A$ , the characteristic function of  $A$ , 16

## References

- [Axl] Axler, S. (2015). Linear algebra done right. 3rd ed.
- [Mun] Munkres, J. (2000). Topology. Second Edition.
- [Rud-P] Rudin, W. (1976). Principles of Mathematical Analysis. 3rd ed.

YAU MATHEMATICAL SCIENCES CENTER, TSINGHUA UNIVERSITY, BEIJING, CHINA.  
*E-mail:* binguimath@gmail.com      bingui@tsinghua.edu.cn