Qiuzhen Lectures on Analysis

BIN GUI

September 9, 2023

1 Basic set theory and numbers

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

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

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

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

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

Topics in the problem sets marked with $\heartsuit$ are important problems that will be used later in this course.

1.1 Basic operations and axioms

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

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

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

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

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:

$$\{x \in X : x \text{ satisfies property} \dots\} \tag{1.3}$$

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

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

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

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

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

$$\bigcup_{A \in \mathscr{A}} A = \{x \in X : \text{there exists } A \in \mathscr{A} \text{ such that } x \in A\} \tag{1.4a}$$

$$\bigcap_{A \in \mathscr{A}} A = \{x \in X : \text{for all } A \in \mathscr{A} \text{ we have } x \in A\} \tag{1.4b}$$

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 $\{x \in \mathbb{R} : x \geqslant 0\}$ could be simply written as $\{x : x \geqslant 0\}$ or even $\{x \geqslant 0\}$. By the same token, the phrase "$\in X$" in (1.4) could be omitted. So we can also write

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

which are special cases of (1.4).

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

Example 1.3. We have

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

When we write $f : A \to B$, we always mean that A, B are sets and f is a function from A to B. A and B are called respectively the **domain** and the **codomain**

---

[1] 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 $\{a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} : a, b, c, d \in \mathbb{R}\}$ is not a field because it is not commutative ($\mathbf{ij} = -\mathbf{ji} = \mathbf{k}$). The set of rational functions $P(x)/Q(x)$, where P, Q are polynomials with coefficients in $\mathbb{R}$ and $Q \neq 0$, is a field.

of f. (Sometimes people also use the words "source" and "target" to denote A and B.)

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

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

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

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

is called the **restriction** of f to E. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

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

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

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

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

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

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

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

Now we are ready to define indexed families of sets.

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

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

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

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

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

$$f \text{ is injective} \iff \text{there is } g : B \to A \text{ such that } g \circ f = \mathrm{id}_A \qquad (1.5a)$$
$$f \text{ is surjective} \iff \text{there is } g : B \to A \text{ such that } f \circ g = \mathrm{id}_B \qquad (1.5b)$$
$$f \text{ is bijective} \iff \text{there is } g : B \to A \text{ such that } g \circ f = \mathrm{id}_A \text{ and } f \circ g = \mathrm{id}_B \quad (1.5c)$$

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 $S_i \neq \varnothing$ for all $i \in I$, then there exists a function (the **choice function**)

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

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

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

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

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

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

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

According to the above remark, one does not need Axiom of Choice to prove (1.5a) 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**) $\leqslant$ on A is a relation on $A \times A$ satisfying for all $a, b, c \in A$ that:

- (Reflexivity) $a \leqslant a$.

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

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

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

Example 1.16. The following are examples of orders.

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

- Let X be a set. Then $(2^X, \subset)$ is a partially ordered set.

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

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

- (Reflexivity) $a \sim a$.

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

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

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

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

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

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

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

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

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

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

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

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 π are equivalent, and that this equivalence is implemented by the bijection Φ. The formal definition of commutative diagrams is the following:

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

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

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

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

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

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

---

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Let us prove that the subset

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

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

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

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

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

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

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

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

Definition 1.31. Let $(X, \leqslant)$ be a 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 $\mathscr{A} \subset 2^X$, i.e., $\mathscr{A}$ is a set of subsets of X. Then $\mathscr{A}$ is bounded from above by X, and is bounded from below by $\varnothing$. Moreover,

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

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

By taking negative, we see that $\mathbb{R}$ also satisfies the greatest-lower-bound property.

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

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

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

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

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

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

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

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

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

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

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

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 $\operatorname{card}(A) = \operatorname{card}(B)$ (or simply $A \approx B$), if there is a bijection $f : A \to B$. We write $\operatorname{card}(A) \leqslant \operatorname{card}(B)$ (or simply $A \precsim B$) if A and a subset of B have the same cardinality.

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

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

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

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

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

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

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

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

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

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

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

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

Proof. By the previous example, it suffices to prove $\mathbb{R} \approx (-2, 2)$. Define $f : \mathbb{R} \to \mathbb{R}$ as follows. We let $f(0) = 0$. Let $n \in \mathbb{N}$ and suppose that $f(n)$ has been defined. Then $f|_{(n,n+1]}$ is $f(x) = 2^{-n} \cdot (x - n) + f(n)$. This defines f on $[0, +\infty)$. On $(-\infty, 0)$, f is defined by $f(x) = -f(-x)$. Then f gives a bijection $\mathbb{R} \to (-2, 2)$, finishing the proof. (Checking the surjectivity of f is tedious. But it suffices to check that f is injective, and conclude the proof using Schröder-Bernstein.) $\qquad\square$

Alternatively, one may use the tangent function to give a bijection between $(-\pi/2, \pi/2)$ and $\mathbb{R}$. I have avoided this method, since piecewisely linear functions are 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

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

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

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

Clearly, a set A is finite iff either $A \approx \varnothing$ or $A \approx \{1, \dots, n\}$ 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 bounded of A, there is $x_{n+1} \in A$ larger than x_n. So $(x_n)_{n \in \mathbb{N}}$ can be constructed inductively.) This gives an injection $\mathbb{N} \to A$. Therefore $A \succsim \mathbb{N}$, and hence $A \approx \mathbb{N}$ by Schröder-Bernstein.

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

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

where $g(0), g(1), g(2), \ldots$ are $(0,0)$, $(1,0)$, $(0,1)$, $(2,0)$, $(1,1)$, $(0,2)$, $(3,0)$, $(2,1)$, $(1,2)$, $(0,3)$, etc., as shown by the figure

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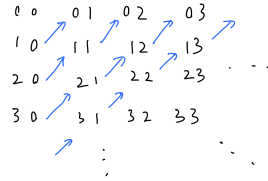
As an application, we prove the extremely important fact that $\mathbb{Q}$ is countable.

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

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

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

Definition 1.50. Let X, Y be sets. Then

$$Y^X = \{\text{functions } f : X \to Y\} \tag{1.9}$$

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

$$Y^\varnothing = \{\varnothing\} \tag{1.10}$$

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

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

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

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

$$Y^X = \underbrace{Y \times Y \times \cdots}_{\text{card}(X) \text{ pieces}} \tag{1.11}$$

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

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

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** $\chi_A : X \to \{0, 1\}$ to be

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

Then we have

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

since the following map is bijective:

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

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

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

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

Proof. We have a bijection

$$\Phi : Y^X \to Y^{A_1} \times \cdots \times Y^{A_n}$$
$$f \mapsto (f|_{A_1}, \ldots, f|_{A_n}) \tag{1.12}$$

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

Exercise 1.56. Assume that $A_1, \ldots, A_n$ are subsets of X. Define Φ by (1.12). Prove that Φ is injective iff $X = A_1 \cup \cdots \cup A_n$. Prove that Φ is surjective iff $A_1, \ldots, A_n$ are pairwise disjoint.

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

Proof. The special case that $m = 0$ (i.e. $X = \varnothing$, cf. (1.10)) and $m = 1$ is clear. When $m > 1$, assume WLOG that $X = \{1, \ldots, m\}$. Then $X = \{1\} \sqcup \cdots \sqcup \{m\}$. Apply Prop. 1.55 to this disjoint union. We see that $Y^X \simeq Y \times \cdots \times Y \simeq \{1, \ldots, n\}^m$ has n^m elements. $\qquad\square$

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

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

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

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

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

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

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

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

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

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

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

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

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

2 Metric spaces, 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 $f : [0, 1] \to \mathbb{R}$ minimizing or maximizing the expression

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

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

of functions)

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

where $K : [0,1]^2 \to \mathbb{R}$ is continuous and $f : [0,1] \to \mathbb{C}$ is subject to the condition $\int_0^1 |f(x)|^2 dx = 1$. 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), 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:

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

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

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

To carry out step (B), we need to <u>define suitable geometric structures for a function space X so that the convergence of sequences in X and the continuity of functions $S : X \to \mathbb{R}$ can be defined and studied in a similar pattern as that for $\mathbb{R}^n$</u>. **Metric** (of a metric space) and **topology** (of a topological space) are such geometric structures. As we shall learn, the topology of a metric space is uniquely determined by the convergence of sequences in this space. Step (C) can be carried out if every sequence in X has a convergent subsequent, i.e., if X is sequentially compact. Thus, we need a good criterion for sequential compactness of subsets of a function space. Arzelà-Ascoli theorem, the $C([a,b])$-version of Bolzano-Weierstrass 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 $d : X \times X \to \mathbb{R}_{\geq 0}$ 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 $[0, 1] \to \mathbb{R}$ and d is a metric on X. To show that $f, g \in X$ are equal, instead of checking that infinitely many values are equal (i.e. $f(t) = g(t)$ for all $t \in \mathbb{R}$), it suffices to check that one value (i.e. $d(f, g)$) is zero.

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

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

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

To see this fact, note that $d(x, y) > 0$. Choose r, ρ 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. $\qquad\square$

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:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Example 2.12. In Exp. 2.10, we have $\delta \leqslant \rho \leqslant d \leqslant 2\delta$. So $\delta \approx \rho \approx d$.

Convention 2.13. Given finitely many metric spaces $X_1, \ldots, X_N$, the metric on their product space $X = X_1 \times \cdots \times X_N$ is chosen to be any one that is equivalent to the ones defined in Exp. 2.10. In the case that each X_i is a subset of $\mathbb{R}$ or $\mathbb{C}$, 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 denotes an object $(Y, d|_Y)$ where $Y \subset X$ and $d|_Y$ is the restriction of d to Y, namely, for all $y_1, y_2 \in Y$ we set

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

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

2.3 Convergence of sequences

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

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

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

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.)

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

$$\text{Property } P \text{ holds for } \textbf{sufficiently large } n \tag{2.4}$$

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

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

The following are some elementary examples and properties of convergent sequences.

Example 2.19. $\lim\limits_{n\to\infty} \dfrac{1}{n} = 0$.

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

Proposition 2.20. *Let $\mathbb{F} \in \{\mathbb{Q}, \mathbb{R}\}$ and $(x_n), (y_n)$ be sequences in $\mathbb{R}$ converging to $x, y \in \mathbb{R}$. If $x_n \leqslant y_n$ for all n, then $x \leqslant y$.*

Proof. If $y < x$, let $\varepsilon = x - y$. Then 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 $y_n < x_n$. $\square$

Example 2.21. Let $(x_n)_{n\in\mathbb{Z}_+}$ be a sequence in X, and let $x \in X$. It is easy to see that

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

Example 2.22. Suppose that (a_n) and (b_n) are sequences in $\mathbb{R}_{\geqslant 0}$, that $a_n \leqslant b_n$ for all n, and that $b_n \to 0$. Then $a_n \to 0$.

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

More generally, we have:

Proposition 2.23 (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 $\lim_{n \to \infty} a_n = 0$ and that $d(x_n, x) \leq a_n$ for all n. Then $\lim_{n \to \infty} x_n = x$.*

Proof. This follows immediately from Exp. 2.21 and 2.22. □

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

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

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

$$\mathbf{x}_n \to \mathbf{y} \iff d_X(\mathbf{x}_n, \mathbf{y}) \to 0 \iff \max_{1 \leq j \leq N} d_j(x_j, y_j) \to 0 \tag{2.5}$$

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 $x_{i,n} \to y_i$ (for all i).

Conversely, assume that for every i we have $x_{i,n} \to y_i$. Then for every $\varepsilon > 0$ there is $K_i \in \mathbb{Z}_+$ such that $d_i(x_{i,n}, y_i) < \varepsilon$ for every $n \geq K_i$. Let $K = \max\{K_1, \ldots, K_N\}$. 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. □

Proposition 2.25. *Let d, δ be two equivalent metrics on a set X. Let $(x_n)_{n \in \mathbb{Z}_+}$ and x be in X. Then $(x_n)_n$ converges to x under the metric d iff $(x_n)_n$ converges to x under δ. Namely, $d(x_n, x) \to 0$ iff $\delta(x_n, x) \to 0$.*

Proof. Prove it yourself. □

We now give some criteria for the divergence of a sequence.

Definition 2.26. 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 2.27. Note that if E is bounded, then for *any* $q \in X$ there exists $\widetilde{R} \in \mathbb{R}_{>0}$ such that $E \subset B_X(q, \widetilde{R})$. (Indeed, choose $\widetilde{R} = R + d(p, q)$, then by triangle inequality, $B(p, R) \subset B(q, \widetilde{R})$.)

From this Remark, we immediately see:

Example 2.28. 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 2.29. A finite union of bounded subsets is bounded.

Proposition 2.30. *Let $(x_n)_{n \in \mathbb{Z}_+}$ be a convergent sequence in a metric space X. Then $(x_n)_{n \in \mathbb{Z}_+}$ is bounded.*

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

Proof. Suppose that $x_n \to x$. Then for each $\varepsilon > 0$, say $\varepsilon = 1$, all but finitely many elements of x_n (say $x_1, \ldots, x_N$) are in $B(x, 1)$. So this whole sequence is in $A = \{x_1\} \cup \cdots \{x_N\} \cup B(x, 1)$. Since each $\{x_i\}$ is bounded, and since $B(x, 1)$ is bounded, A is bounded by Exp. 2.29. $\qquad \square$

Remark 2.31. Prop. 2.30 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. 2.35.) 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. 2.34)

Definition 2.32. 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 2.33. If A and B are partially ordered set, we say a function $f : A \to B$ is **increasing** (resp. **strictly increasing**), if for each $x, y \in A$ we have

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

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

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

Proposition 2.34. *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 $\{x_{n_k} : k \in \mathbb{Z}_+\}$. $\qquad\square$

Example 2.35. 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. 2.31 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.

2.4   Continuous maps of metric spaces

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

Definition 2.36. Let $f : X \to Y$ be a map of metric spaces. Let $x \in X$. We say that f is **continuous at** x if one of the following equivalent conditions hold:

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

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

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

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

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

Proof of the equivalence. (2)$\Leftrightarrow$(2'): Obvious.

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

$\neg$(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 $d(f(p), f(x)) \geqslant \varepsilon$. Thus, for every $n \in \mathbb{Z}_+$, by taking $\delta = 1/n$, we see that there exists $x_n \in X$ such that $d(x_n, x) < 1/n$ and $d(f(x_n), f(x)) \geqslant \varepsilon$. By Squeeze Theorem (Prop. 2.23), $x_n \to x$. But $f(x_n) \nrightarrow f(x)$. So (1) is not true. $\qquad\square$

Remark 2.37. One can compare the definition of continuous maps to that of linear maps. A map is continuous if it *preserves the convergence of sequences*, i.e., if it maps convergent sequences to convergent ones. A map (between vector spaces) is

linear if it perserves the addition and the scalar multiplication of vectors. In general, a good map between two sets with "structures" is a map which preserves the structures. (Thus, linear combinations encode the linear structures of vector spaces. Similarly, the convergence of sequences remembers the "topological" structures of metric spaces.)

Lemma 2.38. *Let $f : X \to Y$ be a map of metric spaces. Let $(B_i)_{i \in I}$ be a collection of open balls in X such that $X = \bigcup_{i \in I} B_i$. Suppose that for each i, the restriction $f|_{B_i} : B_i \to Y$ is continuous. Then f is continuous.*

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

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

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

Definition 2.39. A map of metric spaces $f : X \to Y$ is called **Lipschitz continuous** if there exists $L \in \mathbb{R}_{>0}$ (called **Lipschitz constant**) such that for all $x_1, x_2 \in X$,

$$d_Y\big(f(x_1), f(x_2)\big) \leqslant L \cdot d_X(x_1, x_2) \tag{2.6}$$

Lemma 2.40. *Lipschitz continuous maps are continuous.*

Proof. Suppose that $f : X \to Y$ is Lipschitz continuous with Lipschitz constant L. Suppose $x_n \to x$ in X. Then $L \cdot d(x_n, x) \to 0$. By (2.6) and Squeeze Theorem (Prop. 2.23), $f(x_n) \to f(x)$. (You can also use Def. 2.36-(2) to prove this lemma.) $\qquad \square$

Example 2.41. Let $\mathbb{F} \in \{\mathbb{Q}, \mathbb{R}, \mathbb{C}\}$. The map $z \in \mathbb{F} \backslash \{0\} \mapsto z^{-1} \in \mathbb{F}$ is continuous.

Proof. Let this map be f. Since $\mathbb{F}$ is covered by open balls $B(z, \delta)$ where $z \in \mathbb{F} \backslash \{0\}$ and $0 < \delta < |z|$, by Lem. 2.38, it suffices to prove that f is continuous when restricted to every such $B(z, \delta)$. Let $\varepsilon = |z| - \delta > 0$. Choose $x, y \in B(z, \delta)$. Then $|x| = |x - z + z| \geqslant |z| - |z - x| > \varepsilon$ by triangle inequality. Similarly, $|y| > \varepsilon$. So

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

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

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

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

Proposition 2.42. *Let* $\mathbb{F} \in \{\mathbb{Q}, \mathbb{R}, \mathbb{C}\}$. *Then the following maps are continuous:*

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

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

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

Denote the multiplication map by μ. We choose the metric on $\mathbb{F}^2$ to be $d(\mathbf{x}, \mathbf{x}') = \max\{|x_1 - x_1'|, |x_2 - x_2'|\}$. Since $\mathbb{F} \times \mathbb{F}$ is covered by open balls of the form $B(0, r) = \{(x, y) \in \mathbb{F}^2 : |x| < r, |y| < r\}$ where $0 < r < +\infty$, by Lem. 2.38 and 2.40, 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

$$
|\mu(x, y) - \mu(x', y')| = |xy - x'y'| \leqslant |(x - x')y| + |x'(y - y')|
$$
$$
< 2r \cdot \max\{|x - x'|, |y - y'|\} = 2r \cdot d((x, y), (x', y'))
$$

By Exp. 2.41 and Prop. 2.44, the map $(x, y) \in \mathbb{F} \times \mathbb{F}^\times \mapsto (x, y^{-1}) \in \mathbb{F} \times \mathbb{F}$ is continuous. So its composition with the continuous map μ is continuous, thanks to Prop. 2.43. So $\div$ is continuous. $\qquad\square$

Proposition 2.43. *Suppose that* $f : X \to Y$ *and* $g : Y \to Z$ *are continuous maps of metric spaces. Then* $g \circ f : X \to Z$ *is continuous.*

Proposition 2.44. *Suppose that* $f_i : X_i \to Y_i$ *is a map of metric spaces, where* $1 \leqslant i \leqslant N$. *Then the product map*

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

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

Proof of Prop. 2.43 and 2.44. Immediate from Def. 2.36-(1) and Prop. 2.24. $\qquad\square$

Corollary 2.45 (Squeeze Theorem). *Let* $\mathbb{F} \in \{\mathbb{Q}, \mathbb{R}\}$ *and* $(x_n), (y_n), (z_n)$ *be sequences in* $\mathbb{R}$. *Assume that* $x_n \leqslant y_n \leqslant z_n$ *for all* n. *Assume that* x_n *and* z_n *both converge to* $A \in \mathbb{R}$. *Then* $\lim_{n \to \infty} y_n = A$.

Proof. Let $a_n = y_n - x_n$ and $b_n = z_n - x_n$. Then $0 \leqslant a_n \leqslant b_n$, and $\lim_n b_n = \lim_n z_n - \lim_n x_n = 0$ because the subtraction map is continuous (Prop. 1.7). By Exp. 2.22, $a_n \to 0$. Since $x_n \to A$, $y_n = x_n + a_n$ converges to A, since the addition map is continuous by Prop. 1.7. $\qquad\square$

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 2.46. Let $f, g : X \to \mathbb{F}$ be continuous functions where $\mathbb{F} \in \{\mathbb{Q}, \mathbb{R}, \mathbb{C}\}$. Then by Prop. 2.42 and 2.43, $f \pm g$ and fg are continuous, and f/g is continuous when $0 \notin g(X)$. Here

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

Example 2.47. Let $f : X \to Y$ 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 $\iota_E : E \to X, x \mapsto x$ is clearly continuous. Thus, if f is continuous, then $f|_E : E \to Y$ is continuous, since $f|_E = f \circ \iota_E$.

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

Proposition 2.48. *Let $X_1, \ldots, X_N$ be metric spaces. The following projection is clearly continuous:*

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

Proposition 2.49. *Let $f_i : X \to Y_i$ be maps where $X, Y_1, \ldots, Y_N$ are continuous. Then*

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

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

Example 2.50. Let X be a metric space. Then $d : X \times X \to \mathbb{R}, (x, y) \mapsto d(x, y)$ is Lipschiz continuous with Lipschitz constant 1 (by triangle inequality). So d is continuous.

Proposition 2.51. *Let X be a metric space and $E \subset X$. Define **distance function***

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

Then d_E is has Lipschitz constant 1. So d_E is continuous.

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

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

This proves that d_E has Lipschitz constant 1. $\qquad\square$

Definition 2.52. More generally, if E, F are subsets of a metric space E, we can define

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

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

Exercise 2.53. Let $E, F \subset X$. Prove that

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

Example 2.54. If X is a metric space and $p \in X$, then by Prop. 2.51 (or simply by triangle inequality), the function $d_p : x \in X \mapsto d(x, p) \in \mathbb{R}$ has Lipschitz constant 1 and hence is continuous. In particular, if $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, the function $z \in \mathbb{F} \mapsto |z|$ is continuous (since $|z| = d_{\mathbb{F}}(z, 0)$). Thus, if $f : X \to \mathbb{F}$ is continuous, then $|f| : X \to \mathbb{R}_{\geqslant 0}$ is continuous where

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

Example 2.55. Let $N \in \mathbb{Z}_+$. Then then the following function is continuous:

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

Similarly, the minimum function is continuous.

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

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

So $\max_2$ is continuous by Exp. 2.46 and 2.54.

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

Index

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